PACS File #2700

Methods for Estimating Continuous Time Rational Expectations Models From Discrete Time Data

Lars Peter Hansen
and
Thomas J. Sargent

April 1980

Staff Report #59

Carnegie-Mellon University

University of Minnesota
and
Federal Reserve Bank of Minneapolis

This paper describes methods for estimating the parameters of continuous time linear stochastic rational expectations models from discrete time observations. The economic models that we study are continuous time, multiple variable, stochastic, linear-quadratic rational expectations models. The paper shows how such continuous time models can properly be used to place restrictions on discrete time data. Various heuristic procedures for deducing the implications for discrete time data of these models, such as replacing derivatives with first differences, can sometimes give rise to very misleading conclusions about parameters. The idea is to express the restrictions imposed by the rational expectations model on the continuous time process of the observable variables. Then the likelihood function of a discrete-time sample of observations from this process is obtained. Estimators are obtained by maximizing the likelihood function with respect to the free parameters of the continuous time model.

Our thanks go to Judy Sargent, who calculated the example in Section 5.

1. Introduction

This paper describes methods for estimating the parameters of continuous time linear stochastic rational expectations models from discrete time observations. The economic models that we study are continuous time, multiple variable, stochastic, linear-quadratic versions of the costly-adjustment models of Lucas [13, 14, 15], Mortensen [19], Treadway [29, 30] and Gould [5]. Those authors all specified their models with agents acting in continuous time. The present paper in effect shows how such continuous time models can properly be used to place restrictions on discrete time data. As Sims [27] and Geweke [4] have emphasized, various heuristic procedures for deducing the implications for discrete time data of these models, such as replacing derivatives with first differences, can sometimes give rise to very misleading conclusions about parameters.

The main idea underlying our estimation procedures can be described quite simply. The idea is first conveniently to express the restrictions imposed by the rational expectations model on the continuous time process of the observable variables. Then the likelihood function of a discrete-time sample of observations from this process is obtained. The proposal is to maximize this likelihood function with respect to the free parameters of the continuous time model.

The present paper describes solutions of several technical problems that must be solved in order to implement this estimation strategy. First, we describe convenient methods for calculating the optimal decision rules which comprise the central feature of the model. The formulas that we present are about as close to being in an analytic closed form as is technically possible. In particular, by using linear prediction theory, we obtain a closed form formula for the cross-equation restrictions implied by rational expectations in continuous time. We also describe how the method of Vaughan [31] can be used to factor the spectral-density like matrix that appears in the Euler equations for our problem. Taken together, these methods provide computationally rapid and analytically convenient methods of deducing the continuous time stochastic process from the parameters of the model.

Second, we describe convenient methods for deducing approximations to the likelihood function of a discrete time sample, to be viewed as a function of the parameters of the continuous time model.

The formulas that we describe are general in the sense of accommodating many interrelated decision variables and very rich specifications of the continuous time stochastic processes for the driving variables.

The present paper is a sequel to Hansen and Sargent [8], in which we argued that the cross-equation restrictions delivered by rational expectations can solve the aliasing problem by uniquely identifying the parameters of the continuous time model from statistics of the discrete time observations. The enterprise of the present paper relies on those earlier results on identification. Those earlier results were stated in the context of a vector first-order differential equation system. Like P. C. B. Phillips [2], we adopted that setup because it facilitated discussion of the identification problem. For purposes of actually implementing the estimators, there are great advantages to having the formulas of the present paper, which do not require stacking the system into a first-order vector differential equation. Even though substitute formulas could be derived in the context of Phillips [2] or Hansen and Sargent [8] by suitably stacking the system into a first-order equation, those alternative formulas would probably be less useful computationally because of the wasteful presence of many zeroes in the system matrices.

A variety of examples fit into the general class of models studied in this paper. Several examples can be obtained by adapting the discrete-time examples given by Hansen and Sargent [10], or by reading the papers of Lucas [13, 14, 15], Mortensen [19], Treadway [29, 30], or Gould [5].

This paper is organized as follows. Section 2 describes and solves the optimum problem faced by the agent in the model. Section 3 discusses practical means of computing the decision rule. Section 4 then advances an interpretation of the "error term" in the equation to be fit by the econometrician. Section 5 describes how the free parameters of the continuous time model can be estimated. Section 6 describes modifications required when some of the discrete time data are unit averages (i.e., integrals of flows such as GNP or sales) while others are point-in-time observations. Various technical matters are described in three appendices.

2. The Model

This section describes and solves the problem that is faced by the "agent" in the model.[1]

We define the following objects. Let

y_t be an $(n \times 1)$ vector of variables that enter the agent's objective function and which the agent can control at time t.

x be a $(p \times 1)$ vector of uncontrollable stochastic processes. At time t the agent observes $\{x_u : u \le t\}$. The vector x_t can be partitioned as

$$x_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{bmatrix}$$

where x_{1t} and x_{2t} are both $(n \times 1)$ vectors.

h be an $(n \times 1)$ vector of constants.

H_1 and H_2 be $(n \times n)$ symmetric matrices, where H_1 is positive semidefinite and H_2 is positive definite.

$G(D) = G_0 + G_1 D + \cdots + G_{m_1} D^{m_1}$, where D is the derivative operator with respect to time $(D = \frac{d}{dt})$, $G_0, G_1, \ldots, G_{m_1}$ are each $(n \times n)$, and G_{m_1} is nonsingular.

r be a fixed positive discount rate.

At time t, the agent has the instantaneous return function

$$(2.1) \qquad F(x_t, y_t, Dy_t, \ldots, D^{m_1} y_t, t) = e^{-rt} \{(h + x_{1,t} + x_{2,t})' y_t - \tfrac{1}{2} y_t' H_1 y_t - \tfrac{1}{2} [G(D) y_t]' H_2 [G(D) y_t]\} \quad .$$

The agent faces an x process which affects his instantaneous return through (2.1), but which he cannot control. The vector $x_t = [x_{1t}', x_{2t}', x_{3t}']'$ is described by the stochastic differential equation

$$(2.2) \qquad \theta(D)\, x_t = \psi(D)\, w_t \quad .$$

The operator $\theta(D)$ can be represented as

$$\theta(D) = \theta_0 + \theta_1 D + \cdots + \theta_{m_2} D^{m_2}$$

where θ_j is a $(p \times p)$ matrix, while the operator $\psi(D)$ can be represented as

$$\psi(D) = \psi_0 + \psi_1 D + \cdots + \psi_{m_3} D^{m_3}$$

where ψ_j is a $(p \times p)$ matrix. We assume that the zeroes of $\det \theta(s)$ are less than $r/2$ in real part, and that the zeroes of $\det \psi(s)$ are less than or equal to zero in real part. The $(p \times 1)$ vector w is a continuous time white noise with generalized autocovariance function

$$E\, w_t w_{t-u}' = \delta(t-u)\, V e^{rt}$$

where V is a positive definite matrix, and where δ is the Dirac delta generalized function. Further, define $\varphi(s) = \theta(s)^{-1}\psi(s)$ where s is a complex variable.[a] We assume that the order of the denominator polynomial of $\varphi_{ij}(s)$ exceeds the order of the numerator polynomial for all $i = 1, \ldots, m_2$ and $j = 1, \ldots, m_3$. We note as an implication of the assumption that the zeroes of $\det \psi(s)$ are less than or equal to zero in real part, it follows that the error in forecasting x_{t+u} from a linear function of $\{x_\nu : \nu \le t\}$ can be expressed as an integral of $\{w_{t+\nu} : 0 \le \nu \le u\}$. We also note that while w is not a physically realizable stochastic process, the restrictions placed on $\varphi_{ij}(s)$ imply that x is physically realizable.[3]

The assumption that the zeroes of $\det\theta(s)$ are less than $r/2$ in real part implies that x_t is of mean exponential order less than $r/2$. Further, the assumptions that the zeroes of $\det\theta(s)$ are less than zero in real part, and that $E w_t w'_{t-u} = \delta(t-u) V e^{rt}$ imply that the variance of x is of exponential order less than r. Alternative assumptions about $\theta(s)$ and $E w_t w'_{t-u}$ would also be workable econometrically. In particular, we could assume that the zeroes of $\det\theta(s)$ are less than zero in real part, and that $E w_t w'_{t-u} = \delta(t-u) V$. With minor qualifications and modifications, the procedures that we describe below would apply.

At time zero, the agent is assumed to maximize the objective function [4]

$$\text{(2.3)} \qquad \lim_{T\to\infty} E_0 \int_0^T F(x_t, y_t, D y_t, \ldots, D^{m_1} y_t, t)\, dt$$

subject to (2.1), (2.2) and initial conditions $y_0 = y_0^o$, $D y_0 = (D y_0)^o, \ldots, D^{m_1-1} y_0 = (D^{m_1-1} y_0)^o$. Here E_t denotes the mathematical expectation operator conditioned on information Ω_t that the agent possesses at time t. The information set Ω_t includes at least y_{t-u} for $u > 0$ and x_{t-u} for $u \ge 0$.

The agent is assumed to maximize (2.3) subject to (2.2) and the initial conditions by choosing a time invariant linear contingency plan expressing $D^{m_1} y_t$ as a function of the information that is possessed at t. This linear decision rule or policy function can be expressed as

$$D^{m_1} y_t = P(y_t, Dy_t, \ldots, D^{m_1 - 1} y_t, \{x_{t-u}; u \geq 0\}, c)$$

where c is a vector of constants and P is a linear function. That a time invariant policy maximizes (2.3) follows from the specification of the return function (2.1) and from the infinite horizon in (2.3). The linearity of the optimum policy function can be rationalized either by assuming that x is Gaussian, or else by simply restricting the optimization to the class of linear policy functions.

We proceed to derive an expression for the optimum decision rule that is econometrically tractable. We emphasize here that the econometric tractability of our results is what causes us to exploit some of the special features of the problem and to solve it by classical procedures. As in the corresponding discrete time problem (see Hansen and Sargent [10]), it is possible to formulate the problem (2.1) - (2.3) as a continuous time stochastic optimal linear regulator problem, and to solve it by methods described, for example, by Kwakernaak and Sivan [12]. However, the classical method of solution described here has the advantage of delivering expressions for the optimum decision rules that are as close to being in closed form as is possible. For the purposes of implementing the nonlinear estimators that we describe below, this feature of our solution is a great practical advantage.

Our solution procedure mimics that used by Hansen and Sargent [10] for discrete time analogues of problem (2.1) - (2.3). We exploit what Theil and Simon called the certainty equivalence property, and first solve the certain version of our problem that emerges upon setting w_t identically to zero for all $t \geq 0$. By using the variational methods described by Luenberger [18], the Euler equation for the certainty version of problem (2.1) - (2.3) can be shown to be

$$(2.4) \qquad \frac{\partial F}{\partial y_t} + (-1) D \frac{\partial F}{\partial D y_t} + \cdots + (-1)^{m_1} D^{m_1} \frac{\partial F}{\partial D^{m_1} y_t} = 0 \quad .$$

The Euler equation (2.4) is a set of necessary conditions that the path for y_t implied by the optimal rule must satisfy. From (2.1) we have that

$$(2.5) \quad \begin{aligned} &\text{(a)} \quad \frac{\partial F}{\partial y_t} = \{h + x_{1,t} + x_{2,t} - H_1 y_t - G_0' H_2 [G(D) y_t]\} e^{-rt} \\ &\text{(b)} \quad \frac{\partial F}{\partial D^j y_t} = - G_j' H_2 [G(D) y_t] e^{-rt} \quad . \end{aligned}$$

Differentiating (2.5b) j times with respect to t gives

$$(2.6) \qquad D^j \frac{\partial F}{\partial D^j y_t} = - (D - r)^j G_j' H_2 [G(D) y_t] e^{-rt} \quad \text{for } j = 1, 2, \ldots, m \quad .$$

Substituting (2.5) and (2.6) into the Euler equation (2.4) gives

$$h + x_{1,t} + x_{2,t} - H_1 y_t - [G(r - D)' H_2 G(D)] y_t = 0$$

or

$$(2.7) \qquad [H_1 + G(r - D)' H_2 G(D)] y_t = h + x_{1,t} + x_{2,t} \quad .$$

The Euler equation (2.7) gives a set of necessary conditions for an optimum. However, they are not sufficient conditions because a class of nonoptimal paths or decision rules for y also satisfy (2.7). In order to justify the particular solution to the Euler equation that we choose, some additional technical conditions must be imposed on our problem. The general conditions on our problem that are sufficient to validate our solution procedure are given in Appendix A. Briefly, the conditions require that the associated optimal linear regulator problem is "stabilizable" and "detectable". Alternative sets of restrictions can be imposed on our problem that are sufficient to satisfy these requirements. For example, the following two alternative sets of conditions are sufficient:

> Assumption 1: In addition to the assumptions already made, it is assumed that H_1 is of full rank and so is positive definite.
>
> Assumption 2: In addition to the assumptions already made, it is assumed that the zeroes of $\det G(s)$ are less than $r/2$ in real part.

It is to be emphasized that each of these sets of assumptions is sufficient, but not necessary to validate our solution procedure. Oftentimes, for a given applied problem, it is a routine matter to verify the weaker stabilizability and detectability sufficient conditions given in Appendix A.

For the remainder of this paper we will assume that H_1 is positive definite (Assumption 1). Under this condition, an additional necessary condition for optimality, namely,

$$(2.8) \qquad \int_0^{\infty} e^{-rt}\, y_t' H_1 y_t \, dt < +\infty$$

is sufficient to determine the unique optimum path or decision rule for y that satisfies the Euler equation (2.7).

To construct the rule for y that satisfies (2.7) and (2.8) we argue as follows. Let the Laplace transform of $H_1 + G(r-D)' H_2 G(D)$ be defined as

$$L(s) = H_1 + G(r-s)' H_2 G(s) \quad .$$

Let $s^* = s - r/2$. Then we have

$$L^*(s^*) \equiv L(s^* + r/2) = H_1 + G(r/2 - s^*)' H_2 \, G(r/2 + s^*) \quad .$$

The matrix function L^* is the Laplace transform of the autocovariance function of a certain generalized stochastic process. It is an implication of the spectral factorization theorem described by Rozanov [25] that $L^*(s^*)$ can be represented as

$$\text{(2.9)} \qquad L^*(s^*) = C(-s^*)' C(s^*)$$

where $C(s^*) = C_0 + C_1 s^* + \cdots + C_{m_1} s^{*m_1}$, where C_j is an $(n \times n)$ matrix, and the zeroes of $\det C(s^*)$ are less than zero in real part. The representation (2.9) is unique up to premultiplication of $C(s^*)$ by a unitary matrix. From (2.9) and the definition of L^* in terms of L, it follows that

$$L(s) = L^*(s - r/2) = C(-s + r/2)' C(s - r/2) \quad .$$

Using the above representation for $L(s) = H_1 + G(r-s)' H_2 G(s)$, we can write the Euler equation as

$$C(-D + r/2)' C(D - r/2) y_t = \Gamma h + x_{1,t} + x_{2,t} \rceil \quad .$$

It can be shown, as a consequence of the condition that H_1, H_2, and G_{m_1} are of full rank, that the unique solution of the Euler equation that also satisfies (2.8) can be represented as

$$C(D - \tfrac{r}{2})\, y_t = C(-D + \tfrac{r}{2})'^{-1}(h + x_{1,t} + x_{2,t}) \quad . \tag{2.10}$$

In order to convert (2.10) to an econometrically useful representation, we first obtain a matrix partial fractions decomposition of $C(-s^*)'^{-1}$. Assume that the zeroes of $\det C(s^*)'$ are distinct, so that

$$\det C(s^*)' = \rho_0^* (s^* - \rho_1^*) \cdots (s^* - \rho_k^*)$$

where $k = n m_1$. Then we have

$$C(s^*)'^{-1} = \frac{\operatorname{adj} C(s^*)'}{\rho_0^*(s^* - \rho_1^*)\cdots(s^* - \rho_k^*)}$$

or

$$C(s^*)'^{-1} = \frac{B_1}{s^* - \rho_1^*} + \cdots + \frac{B_k}{s^* - \rho_k^*} \tag{2.11}$$

where

$$B_j = \frac{\rho_0^{*-1} \operatorname{adj} C(\rho_j^*)'}{\prod_{\substack{h=1 \\ h\neq j}}^{k} (\rho_j^* - \rho_h^*)} \quad .$$

We therefore have

$$C(-s + \tfrac{r}{2})'^{-1} = \frac{B_1}{-s - \rho_1^* + \frac{r}{2}} + \cdots + \frac{B_k}{-s - \rho_k^* + \frac{r}{2}}$$

$$= \frac{B_1}{-s - \rho_1} + \cdots + \frac{B_k}{-s - \rho_k}$$

where $\rho_j = \rho_j^* - \frac{r}{2}$ for $j = 1, \ldots, k$. Notice that since the zeroes ρ_j^* of $\det C(s^*)'$ are less than zero in real part by construction, it follows that ρ_j is less than $-\frac{r}{2}$ in real part for $j = 1, \ldots, k$.

Next, we recall from the property of Laplace transforms that

$$(2.12)\qquad \frac{B_j}{-D-\rho_j}\,[h + x_{1,t} + x_{2,t}] = B_j \int_0^{\infty} e^{\rho_j u}\,(h + x_{1,t+u} + x_{2,t+u})\,du \quad .$$

Using (2.12) in conjunction with (2.11) and (2.10) shows that the decision rule at time t can be expressed as

$$(2.13)\qquad C(D - r/2)\,y_t = \sum_{j=1}^{k} B_j \int_0^{\infty} e^{\rho_j u}(h + x_{1,t+u} + x_{2,t+u})\,du \quad .$$

Now represent $C(D - r/2)$ as

$$(2.14)\qquad C(D - r/2) = A_0 + A_1 D + \cdots + A_{m_1} D^{m_1} = A(D)$$

where

$$A_j = \binom{m_1}{j}(-\,r/2)^{m_1-j}\,C_{m_1} + \binom{m_1-1}{j}(-\,r/2)^{m_1-j-1}\,C_{m_1-1} + \cdots + \binom{j}{j}C_j$$

for $j = 0, 1, \ldots, m_1$. Substituting (2.14) into (2.13), pre-multiplying by $A_{m_1}^{-1}$, and rearranging gives

$$(2.15)\qquad D^{m_1} y_t = -A_{m_1}^{-1}\left(A_0 + A_1 D + \cdots + A_{m_1-1} D^{m_1-1}\right) y_t$$

$$+ A_{m_1}^{-1} \sum_{j=1}^{k} B_j \int_0^{\infty} e^{\rho_j s}(h + x_{1,t+u} + x_{2,t+u})\,du \quad .$$

Equation (2.15) expresses the decision rule for $D^{m_1} y_t$ in terms of y_t, $D y_t$, ..., $D^{m_1-1} y_t$ and the actual entire future values of the time functions x_1 and x_2. Our assumption about (2.2) that the zeroes of $\det\theta(s)$ are less than $r/2$ in real part, and our finding that the real part of ρ_j is less than $-r/2$ imply that the integrals on the right side of (2.15) converge.

With or without uncertainty,[5/] it can be verified that the decision rule (2.15) satisfies both the Euler equation (2.7) and the additional necessary condition (2.8). However, where there is uncertainty (i.e., where w_t or $\psi(D)$ in (2.2) is not identically zero), the rule (2.15) requires the agent to use more information than we have assumed that he possesses. That is, with uncertainty the rule (2.15) is "anticipative" or "nonrealizable" because it requires the agent to know future values of x_1 and x_2. To derive the appropriate "realizable" or "nonanticipative" linear decision rule under uncertainty that satisfies the first order necessary conditions for the problem, one simply replaces future $x_{1,t+u}$ and $x_{2,t+u}$ in (2.15) with the corresponding linear least squares forecasts, conditioned on information that the agent possesses at time t. Then under uncertainty the optimal decisions satisfy

$$(2.16)\qquad D^{m_1} y_t = -A_{m_1}^{-1}\left(A_0 + A_1 D + \cdots + A_{m_1-1} D^{m_1-1}\right) y_t + A_{m_1}^{-1} \sum_{j=1}^{k} B_j E_t \int_0^{\infty} e^{\rho_j u} (h + x_{1,t+u} + x_{2,t+u})\, du \quad .$$

To make (2.16) econometrically operational, it remains to derive a closed form expression for the terms

$$E_t \int_0^{\infty} e^{\rho_j u} x_{t+u}\, du \qquad \text{for } j = 1, \ldots, k \quad .$$

Let

$$(2.17)\qquad \mu_1(\rho, D) = \theta(-\rho)^{-1} \sum_{j=0}^{m_2-1} D^j \sum_{k=j}^{m_2-1} \theta_{k+1} (-\rho)^{k-j}$$

$$(2.18)\qquad \mu_2(\rho, D) = -\theta(-\rho)^{-1} \sum_{j=0}^{m_3-1} D^j \sum_{k=j}^{m_3-1} \psi_{k+1} (-\rho)^{k-j}$$

In Appendix B, it is proved that

$$\text{(2.19)} \qquad E_t \int_0^{\infty} e^{\rho_j u} x_{t+u}\, du = \mu_1(\rho_j, D)\, x_t + \mu_2(\rho_j, D)\, w_t \quad .$$

Substituting (2.19) into (2.16) gives our final expression for the decision rule

$$\text{(2.20)} \qquad D^{m_1} y_t = -A_{m_1}^{-1} \sum_{j=0}^{m_1-1} A_j D^j y_t$$

$$+ A_{m_1}^{-1} \sum_{j=1}^{k} B_j M [\mu_1(\rho_j, D)\, x_t + \mu_2(\rho_j, D)\, w_t] + C(r/2)^{\prime -1} h$$

where M is an $(n \times p)$ matrix such that $Mx_t = x_{1t} + x_{2t}$. We also have

$$\text{(2.2)} \qquad \theta(D)\, x_{t+s} = \psi(D)\, w_{t+s} \quad .$$

Equations (2.20) and (2.2) form a statistical model of the joint (y_t, x_t) stochastic process. The model is subject to an extensive set of cross-equation restrictions which are summarized in the following equations:

$$C(-D + r/2)' C(D - r/2) = H_1 + G(r - D)' H_2 G(D)$$

$$\text{(2.11)} \qquad C(s^*)^{\prime -1} = \frac{B_1}{s^* - \rho_1^*} + \cdots + \frac{B_k}{s^* - \rho_k^*}$$

where

$$B_j = \frac{\rho_0^{*-1} \operatorname{adj} C(\rho^*)'}{\prod_{\substack{h \neq j \\ h=1}}^{k} (\rho_j^* - \rho_h^*)}$$

$$\text{(2.14)} \qquad C(D - r/2) = A_0 + A_1 D + \cdots + A_{m_1} D^{m_1}$$

(2.17) $$\mu_1(\rho, D) = \theta(-\rho)^{-1} \sum_{j=0}^{m_2-1} D^j \sum_{k=j}^{m_2-1} \theta_{k+1} (-\rho)^{k-j}$$

(2.18) $$\mu_2(\rho, D) = -\theta(-\rho)^{-1} \sum_{j=0}^{m_3-1} \sum_{k=j}^{m_3-1} \psi_{k+1} (-\rho)^{k-j} .$$

These cross-equation restrictions play a key role in determining the identification and optimal estimation of the parameters $\{H_1, H_2, G(D), \theta(D), \psi(D), V\}$ of the continuous time model from a discrete time data record.

Before we proceed to issues of identification and estimation, several preliminary issues must be dealt with. In the next section, we describe feasible computational methods for computing the optimal decision rule (2.20). While this section can be skipped on first reading, it is important practically since procedures such as are described are needed in order to compute and to maximize the likelihood function of the discrete time data record. The succeeding section then specializes (2.20) and (2.2) somewhat in order to build a tractable model of the "error term" facing the econometrician. In section 5, we will finally return to describing procedures for estimating the free parameters of the model.

3. Computing the Decision Rule

There are two nontrivial computational tasks involved in implementing our formula (2.20) for the decision rule. First, there is the task of factoring the matrix $L^*(s^*)$ in the manner called for in (2.9). Second, there is the task of conveniently computing the matrix partial fractions decomposition of $C(s^*)'^{-1}$ indicated in (2.11). In this section, we describe how each of these jobs can be done.

First, consider the job of achieving the factorization

$$H_1 + G(r - s)' H_2 G(s) = C(-s + r/2)' C(s - r/2)$$

where the zeroes of $\det C(s - r/2)$ are less than $r/2$ in real part. Here we describe a procedure analogous to one used by Hansen and Sargent [9] for a discrete time model. The idea is to use the solution of the nonstochastic optimal linear regulator problem for an auxiliary problem to compute the "feedback part" of the control law, and to deduce the factorization of $H_1 + G(r - s)' H_2 G(s)$ from the feedback part.

The relevant auxiliary problem is simply the version of the certainty problem in which $x_t = 0$ and $w_t = 0$ for all t and $h = 0$. From (2.15) it is immediate that the optimal decision rule for the auxiliary problem is

$$D^{m_1} y_t = - A_{m_1}^{-1} (A_0 + A_1 D + \cdots + A_{m_1 - 1} D^{m_1 - 1}) y_t \quad . \tag{3.1}$$

Our first goal is to solve for $A_{m_1}^{-1} A_0$, ..., $A_{m_1}^{-1} A_{m_1 - 1}$ from the objective function parameters G_0, ..., G_{m_1}, H_1 and H_2. To do this, we formulate the auxiliary problem as a continuous time optimal linear regulator problem (see Kwakernaak and Sivan [12]). Let

$$U_{t+s} = G(D) y_{t+s}$$

$$X_{t+s} = \left(D^{m_1 - 1} y'_{t+s}, D^{m_1 - 2} y'_{t+s}, \ldots, y'_{t+s}\right)'$$

$\mathcal{D}$ be an $n \times m_1 n$ matrix given by $[0 \ I]$

$$Z_t = \mathcal{D} X_t = y_t$$

$$\mathcal{R}_1(t) = + \frac{1}{2} e^{-rt} H_1$$

$$\mathcal{R}_2(t) = + \frac{1}{2} e^{-rt} H_2 \quad .$$

The law of motion for the state vector X_t is

$$\begin{bmatrix} D^{m_1} y_t \\ D^{m_1-1} y_t \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ D^2 y_t \\ D y_t \end{bmatrix} = \begin{bmatrix} -G_{m_1}^{-1} G_{m_1-1} & -G_{m_1}^{-1} G_{m_1-2} & \cdots & -G_{m_1}^{-1} G_1 & -G_{m_1}^{-1} G_0 \\ I & 0 & & 0 & 0 \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 0 & & & 0 & 0 \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} D^{m_1-1} y_{t+s} \\ D^{m_1-2} y_{t+s} \\ \cdot \\ \cdot \\ \cdot \\ D y_{t+s} \\ y_{t+s} \end{bmatrix} + \begin{bmatrix} G_{m_1}^{-1} \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 0 \end{bmatrix} U_t$$

which we write

$$(3.2) \qquad DX_t = \mathcal{A} X_t + \mathcal{B} U_t \quad .$$

The objective for the auxiliary problem can be represented as minimizing

$$(3.3) \qquad \lim_{T \to \infty} E_0 \int_0^T [Z_t' \mathcal{R}_1(t) Z_t + U_t' \mathcal{R}_2(t) U_t] \, dt$$

subject to the law of motion (3.2). This establishes that the auxiliary problem is an example of the optimal linear regulator problem.

The parameters of the integrand of the criterion function (3.3) are dependent on t via $\mathcal{R}_1$ and $\mathcal{R}_2$. It is convenient for us to transform the variables so that this dependence disappears. This will allow us to use results for the "time invariant" optimal linear regulator problem. For a given t, let

$$X_t^* = e^{-r\frac{t}{2}} X_t$$

$$U_t^* = e^{-r\frac{t}{2}} U_t$$

$$Z_t^* = e^{-r\frac{t}{2}} Z_t \quad .$$

Note that

$$DX_t^* = - r/2\, X_t^* + e^{-r\frac{s}{2}} DX_t \quad .$$

Substituting from (3.2), we have

(3.4) $$DX_t^* = \mathcal{A}^* X_t^* + \mathcal{B} \dot{U}_t^*$$

where

$$\mathcal{A}^* = \mathcal{A} - r/2\, I \quad .$$

Criterion function (3.3) can be written

(3.5) $$\lim_{T \to \infty} E_0 \int_0^T [Z_t^{*\prime} H_1 Z_t^* + U_t^{*\prime} H_2 \dot{U}_t^*]\, dt$$

where

$$Z_t^* = \mathcal{D} X_t^* \quad .$$

Minimization of (3.5) subject to (3.4) is a time invariant version of the optimal linear regulator problem.

We assume that the pair $(\mathcal{A}^*, \mathcal{B})$ is stabilizable and the pair $(\mathcal{A}^*, \mathcal{D})$ is detectable (see Kwakernaak and Sivan [12] and Appendix A). It follows that our time invariant auxiliary problem satisfies sufficient conditions for solution via the eigenvalue decomposition of the state - co-state transition matrix that is described by Vaughan [31], let

$$\mathcal{M} = \begin{bmatrix} -\mathcal{A}^* & \mathcal{B} H_2^{-1} \mathcal{B}' \\ \mathcal{D}' H_1 \mathcal{D} & \mathcal{A}^{*\prime} \end{bmatrix} \quad .$$

The matrix $\mathcal{M}$ is known to have eigenvalues that are symmetric with respect to the imaginary axis. Assuming the eigenvalues are distinct, we let Λ

denote the matrix with eigenvalues on the diagonal such that

$$(3.6) \qquad \Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & -\Lambda_1 \end{bmatrix}$$

where the real parts of the eigenvalues in Λ_1 are greater than zero in real part. We define $\mathfrak{W}$ to be a nonsingular, normalized matrix whose columns are eigenvectors corresponding to the eigenvalues Λ, so that

$$\mathfrak{W}^{-1} \mathfrak{M} \mathfrak{W} = \Lambda \quad .$$

Partition $\mathfrak{W}$ conformably with Λ in (3.6) to obtain

$$\mathfrak{W} = \begin{bmatrix} \mathfrak{W}_{11} & \mathfrak{W}_{12} \\ \mathfrak{W}_{21} & \mathfrak{W}_{22} \end{bmatrix} \quad .$$

The optimal control for the auxiliary problem is given by

$$(3.7) \qquad U_t = - H_2^{-1} \mathfrak{B}' \mathfrak{W}_{21} \mathfrak{W}_{11}^{-1} X_t = F X_t \quad .$$

Partition F conformably with X_t to obtain

$$F = [F_{m_1-1} \cdots F_0] \quad .$$

Using the definitions of U_t and X_t, we have

$$G_{m_1} D^{m_1} y_t + \cdots + G_0 y_t = F_{m_1-1} D^{m_1-1} y_t + \cdots + F_0 y_t$$

or

$$(3.8) \qquad D^{m_1} y_t = G_{m_1}^{-1} (F_{m_1-1} - G_{m_1-1}) D^{m_1-1} y_t + \cdots + G_{m_1}^{-1} (F_0 - G_0) y_t \quad .$$

Comparing (3.8) to (3.1) we conclude that

$$(3.9)\qquad \begin{aligned} A_{m_1}^{-1} A_0 &= -G_{m_1}^{-1}(F_0 - G_0) \\ &\vdots \\ A_{m_1}^{-1} A_{m_1-1} &= -G_{m_1}^{-1}(F_{m_1-1} - G_{m_1-1}) \end{aligned} \quad .$$

Thus we have a numerical algorithm for determining $A_{m_1}^{-1} A_0, \ldots A_{m_1}^{-1} A_{m_1-1}$ given the criterion function parameters H_1 and $G(D)$.

To complete our discussion of the factorization of $H_1 + G(r-s)' H_2 G(s)$, recall that

$$(3.10)\qquad H_1 + G(r-s)' H_2 G(s) = C(-s + r/2)' C(s - r/2) \quad ,$$

$$C(s - r/2) = A(s)$$

and

$$C_{m_1} = A_{m_1} \quad .$$

From (3.9) and (3.10), we know how to compute

$$A_{m_1}^{-1}(A_0 + A_{m_1}^{-1} A_1 s + \ldots + I s^{m_1}) = C_{m_1}^{-1} C(s - r/2) \quad .$$

We can also compute

$$(3.11)\qquad C_{m_1}^{-1} C(-s + r/2) = A_{m_1}^{-1} A_0 + A_{m_1}^{-1} A_1 (r-s) + A_{m_1}^{-1} A_2 (r-s)^2 + \ldots + I(r-s)^{m_1} \quad .$$

From

$$H_1 + G(r-s)' H_2 G(s) = C(-s + r/2)' C(s - r/2)$$

we can verify that

$$G_{m_1}' H_2 G_{m_1} = C_{m_1}' C_{m_1} \quad .$$

Now

$$H_1 + G(r - s)'H_2 G(s) = C(-s+r/2)'C(s-r/2)$$

$$= C(-s+r/2)'C_{m_1}^{-1\prime} C_{m_1}' C_{m_1} C_{m_1}^{-1} C(s-r/2)$$

$$= C(-s+r/2)'C_{m_1}^{-1\prime} G_{m_1}' H_2 G_{m_1} C_{m_1}^{-1} C(s-r/2) \quad .$$

The feedforward polynomial is simply

$$C(-s+r/2)'C_{m_1}^{-1\prime} G_{m_1}' H_2 G_{m_1} = C(-s+r/2)'C_{m_1}$$

which can be computed from (3.11) given knowledge of $G_{m_1}' H_2$ and the feedback polynomial

$$C_{m_1}^{-1} C(s - r/2) \quad .$$

The feedback polynomial is determined from numerically solving the auxiliary problem. So we have described an algorithm capable of factoring $H_1 + G(r - s)'H_2 G(s)$ in the desired way.

The second computational issue that we wish to mention is the partial fractions decomposition (2.11), which is equivalent with attaining a decomposition of

$$A_{m_1}^{-1} C(-s+r/2)^{-1\prime} = [C(-s+r/2)'C_{m_1}]^{-1} \quad .$$

While the formula given under (2.11) is correct, a numerically efficient algorithm for calculating the ρ^{*}'s and B_j's is described in some detail in Hansen and Sargent [10]. There the authors suggest an adaptation of an algorithm proposed by Emre and Hüseyin [3]. It is straightforward to adapt Hansen and Sargent's computation to the present problem, essentially by replacing "z" in Hansen and Sargent's formulas with "s".

4. A Model of the Disturbance Term

Our agent has been assumed to maximize over contingency plans for y_t the criterion

$$(4.1) \qquad \lim_{T\to\infty} E_0 \int_0^T e^{-rt} \{(h_0 + x_{1,t} + x_{2,t})' y_t - \tfrac{1}{2} y_t' H_1 y_t - \tfrac{1}{2} [G(D) y_t]' H_2 [G(D) y_t]\} \, dt$$

subject to initial values of y_0, D_{y_0}, ..., $D^{m_1 - 1} y_0$ and also subject to the law of motion

$$(4.2) \qquad \theta(D) \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \psi(D) \, w_t \quad .$$

The maximizing choice of y_t was shown to have the representation

$$(2.20) \qquad D^{m_1} y_t = - A_{m_1} \sum_{j=0}^{m_1 - 1} A_j D^j y_t + A_{m_1}^{-1} \sum_{j=1}^{k} B_j M[\mu_1 (\rho_j, D) x_t + \mu_2 (\rho_j, D) w_t] + C(r/2)'^{-1} h \quad ,$$

where A_j, μ_1, μ_2, ρ_j, B_j, and $C(\cdot)$ are all defined in section 2. We have assumed that w is "fundamental for the x process" in the sense that linear least squared errors in forecasting x_{t+v} on the basis of $\{x_{t-u} : u \geq 0\}$ can be expressed as an integral of $\{w_{t+u} : v \geq u \geq 0\}$. This assumption implies that w_t can be expressed as a function of current and lagged x's . Equivalently, since we have assumed that the zeroes of $\det \psi(s)$ in (4.2) are less than zero in real part, upon operating on both sides of (4.2) with $\psi(D)^{-1}$ we obtain

$$\psi(D)^{-1} \theta(D) x_t = w_t \quad .$$

Since the zeroes of det $\psi(D)$ are less than zero in real part, the poles of $\psi(D)^{-1}\theta(D)$ are less than zero in real part. Thus, the above equation expresses w_t as a convolution integral of current and past x's. This observation means that equation (2.20) expresses the law of motion for $D^{m_1} y_t$ as an exact function of $D^{m_1-1} y_t, \ldots, y_t$ and the observations on current and past x's which the agent possesses. From the agent's point of view, there is no random term on the right side of the decision rule (2.20). This reflects the fact that the agent is playing a "game against nature" and so finds it optimal to employ a nonrandom strategy.

For econometric purposes, it is necessary to have a specification that includes a random error, at least from the econometrician's perspective. In the interests of specifying a tractable and plausible model of the error term, we shall assume that the econometrician possesses less information than does the economic agent. In particular, we assume that the econometrician never observes x_1 while the private agent does. As noted previously, the economic agent sees x_t at all real t. Since this data record is continuous, the agent, in effect, observes derivatives of x_t also. On the other hand, the econometrician only has data on the levels of $[y_t', x_{2,t}', x_{3,t}']$ at discrete points in time, t belonging to the integers, and does not observe their time derivatives.

In the interests of constructing a tractable model of the error term, we further restrict the specification of the stochastic process governing x_t. We define

$$z_t = \begin{bmatrix} x_{2,t} \\ x_{3,t} \end{bmatrix}, \qquad x_t = \begin{pmatrix} x_{1,t} \\ z_t \end{pmatrix} .$$

We partition $\theta(D)$, $\psi(D)$, and w_t conformably with the partitioning

$x_t' = x_{1,t}', z_t'$, and assume that

$$w_t = \begin{pmatrix} w_{1,t} \\ w_{z,t} \end{pmatrix}$$

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

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

where $\theta^1(D)$ and $\psi^1(D)$ are each $(n \times n)$ operators and $\theta^2(D)$ and $\psi^2(D)$ are each $(p-n) \times (p-n)$ operators. Substituting the special assumptions about θ and ψ into (4.2) and premultiplying both sides by $\psi(D)^{-1}$ gives

(4.4) $$\psi^1(D)^{-1}\, \theta^1(D)\, x_{1,t} = w_{1t} \quad ,$$
$$\psi^2(D)^{-1}\, \theta^2(D)\, z_t = w_{z,t} \quad ,$$

where the poles of $\psi^1(s)^{-1}$ and $\psi^2(s)^{-1}$ are each less than zero in real part by virtue of our assumption that the zeroes of $\det \psi(s)$ are less than zero in real part. Equation (4.4) together with these conditions on the location of the poles of $\psi^1(s)^{-1}$ and $\psi^2(s)^{-1}$ imply that $w_{1,t}$ is contained in the space spanned by $\{x_{1,u} : u \le t\}$, and that $w_{z,t}$ is contained in the space spanned by $\{z_u : u \le t\}$. Therefore, w_{1t} is fundamental for x_{1t} and $w_{z,t}$ is fundamental for z_t. It then follows from (4.4) that for all t and u,

(4.5) $$E[x_{1,t} \mid x_{1,v}, z_v \text{ for } v \le u] = E[x_{1,t} \mid x_{1,v} \text{ for } v \le u]$$

and

(4.6) $$E[z_t \mid x_{1,v},\ z_v \text{ for } v \le u] = E[z_t \mid z_v \text{ for } v \le u] \quad .$$

Equation (4.5) asserts that z fails to Granger cause x_1, while equation (4.6) asserts that x_1 fails to Granger cause z. Recall, however, that where $E w_t w'_{t-u} = \delta(t-u) V e^{rt}$, we have permitted V to have nonzero off-diagonal elements, so that x_{1t} and z_t are permitted to be correlated.

We now derive a special version of the decision rule (2.20) that incorporates the special assumption (4.3) about $\theta(D)$ and $\psi(D)$. Define μ_1^2 and μ_2^2 analogously to μ_1 and μ_2 in (2.17) and (2.18) to be

$$\mu_1^2(\rho, D) = \theta^2(-\rho)^{-1} \sum_{j=0}^{m_2-1} D^j \sum_{k=j}^{m_2-1} \theta^2_{k+1}(-\rho)^{k-j} \tag{4.7}$$

$$\mu_2^2(\rho, D) = -\theta^2(-\rho)^{-1} \sum_{j=0}^{m_3-1} D^j \sum_{k=j}^{m_3-1} \psi^2_{k+1}(-\rho)^{k-j} \tag{4.8}$$

Employing (4.4) we can write the decision rule for $D^{m_1} y_t$ as

$$A_{m_1}^{-1} A(D)\, y_t = A_{m_1}^{-1} \sum_{j=1}^{k} B_j J_2[\mu_1^2(\rho_j, D) z_t + \mu_2^2(\rho_j, D)\, w_{z,t}] + C(r/2)'^{-1} h + \varepsilon_t \tag{4.9}$$

where J_2 is an $n \times (p-n)$ matrix such that $J_2 z = x_2$ and

$$\varepsilon_t = A_{m_1}^{-1} \sum_{j=1}^{k} B_j E_t \int_0^{+\infty} e^{\rho_j s} x_{1,t+s}\, ds \quad . \tag{4.10}$$

Since ε is not observed by the econometrician, we view it as the disturbance vector in the decision rule. In Appendix B it is shown that

$$E_t \int_0^{\infty} e^{\rho_j u} x_{1,t+u}\, du = -\left[\frac{\theta^1(D)^{-1} \psi^1(D) - \theta^1(-\rho_j)^{-1} \psi^1(-\rho_j)}{D + \rho_j}\right] w_{1,t} \quad .$$

Hence

$$\varepsilon_t = -A_{m_1}^{-1} \sum_{j=1}^{k} B_j \left[\frac{\theta^1(D)^{-1} \psi^1(D) - \theta^1(-\rho_j)^{-1} \psi^1(-\rho_j)}{D + \rho_j}\right] w_{1,t} \,. \tag{4.11}$$

Let $\theta^1(s)^{-1} = \frac{\theta^a(s)}{\theta^d(s)}$ where θ^a is the adjoint of θ^1 and θ^d is the determinant of θ^1. Applying $\theta^d(D)$ to both sides of (4.11) we obtain

$$(4.12) \qquad \theta^d(D)\,\varepsilon_t = -A_{m_1}^{-1} \sum_{j=1}^{k} B_j \frac{\theta^a(D)\,\psi^1(D) - \theta^d(D)\,\theta^1(-\rho_j)\psi^1(-\rho_j)}{D+\rho_j} w_{1,t} \quad .$$

Applying $\theta^d(D)$ to both sides of (4.9) and using (4.12) then gives

$$(4.13) \qquad \theta^d(D)\,A_{m_1}^{-1}A(D)\,y_t = A_{m_1}^{-1}\sum_{j=1}^{k} B_j J_2[\theta^d(D)\mu_1^2(\rho_j,D)z_t + \theta^d(D)\mu_2^2(\rho_j,D)w_{z,t}]$$

$$- A_{m_1}^{-1}\sum_{j=1}^{k} B_j\left[\frac{\theta^a(D)\psi^1(D) - \theta^d(D)\theta^1(-\rho_j)^{-1}\psi^1(-\rho_j)}{D+\rho_j}\right] w_{1,t}$$

$$+ \theta^d(0)\;C(\tfrac{r}{2})'^{-1}\,h \quad .$$

The stochastic differential equation (4.13) and the equation

$$(4.14) \qquad \theta^2(D)\,z_t = \psi^2(D)\,w_{z,t}$$

form a system of linear stochastic differential equations in (y_t', z_t') that is driven by the vector white noise (w_{1t}', w_{zt}'). Ignoring the constant in (4.13), the system can be represented compactly as

$$(4.15) \qquad \begin{bmatrix} K_{11}(D) & K_{12}(D) \\ 0 & K_{22}(D) \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} Q_{11}(D) & Q_{12}(D) \\ 0 & Q_{22}(D) \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{z,t} \end{bmatrix}$$

or

$$K(D)\begin{bmatrix} y_t \\ z_t \end{bmatrix} = Q(D)\,w_t$$

where

(4.16)

$$K_{11}(D) = \theta^{d}(D)\, A_{m_1}^{-1} A(D)$$

$$K_{12}(D) = - A_{m_1}^{-1} \sum_{j=1}^{k} B_j\, J_2\, \theta^{d}(D)\, \mu_1^2(\rho_j, D)$$

$$K_{22}(D) = \theta^{2}(D)$$

$$Q_{11}(D) = - A_{m_1}^{-1} \sum_{j=1}^{k} B_j \left[\frac{\theta^{a}(D)\psi^{1}(D) - \theta^{d}(D)\theta^{1}(-\rho_j)^{-1}\psi^{1}(-\rho_j)}{D + \rho_j}\right]$$

$$Q_{12}(D) = A_{m_1}^{-1} \sum_{j=1}^{k} B_j\, J_2\, \theta^{d}(D)\, \mu_2^2(\rho_j, D)$$

$$Q_{22}(D) = \psi^{2}(D) \quad .$$

The model (4.15) is subject to an extensive set of cross-equation restrictions, which we collect here for convenience:

(4.17)

$$C(-D+\tfrac{r}{2})'\, C(D-\tfrac{r}{2}) = H_1 + G(r-D)'\, H_2\, G(D)$$

$$C(D-\tfrac{r}{2}) = A_0 + A_1 D + \ldots + A_{m_1} D^{m_1}$$

$$B_j = \frac{\rho_0^{*-1}\ \mathrm{adj}\ C(\rho_j^*)}{\prod_{h\neq j} (\rho_j^* - \rho_h^*)} \qquad j = 1, \ldots, k \qquad k = n \cdot m_1$$

$$\det C(s^*)' = \rho_0^*\, (s^* - \rho_1^*) \cdot \cdot \cdot (s^* - \rho_k^*)$$

$$\rho_j = \rho_j^* - \frac{r}{2}$$

$$\mu_1^2(\rho, D) = \theta^{2}(-\rho)^{-1} \sum_{j=0}^{m_2-1} D^j \sum_{k=j}^{m_2-1} \theta_{k+1}^2\, (-\rho)^{k-j}$$

$$\mu_2^2(\rho, D) = -\theta^{2}(-\rho)^{-1} \sum_{j=0}^{m_3-1} D^j \sum_{k=j}^{m_3-1} \psi_{k+1}^2\, (-\rho)^{k-j}$$

$$\theta^{1}(D)\, x_{1,t} = \psi^{1}(D)\, w_{1,t}\,, \quad \theta^{a}(s) = \mathrm{adj}\, \theta^{1}(s)\,, \quad \theta^{d}(s) = \det \theta^{1}(s)$$

$$\theta^{2}(D)\, z_t = \psi^{2}(D)\, w_{z,t}$$

$$x_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{bmatrix} \equiv \begin{bmatrix} x_{1,t} \\ z_t \end{bmatrix} \quad .$$

The free parameters of the model are the parameters H_1, H_2, G_0, ..., G_{m_1} of the agent's criterion function, and the parameters of $\theta(D)$ and $\psi(D)$ which appear in his "constraints", that is, in the law of motion for x. 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_1, H_2, $G(D)$, $\theta(D)$ and $\psi(D)$.

The restrictions that we have imposed in this section on the relationship of the information possessed by the econometrician, who only sees (y_t, z_t), to that possessed by the agent, who sees (y_t, x_{1t}, z_t), delivers the continuous time model (4.15). In this model, the white noise process w_1 gives rise to an econometrically tractable error term in the projection of y_t on the continuous time process $\{z_u : u \leq t, u \text{ real}\}$. The model (4.15) is now in a form that we can consider the problem of estimating the parameters of the continuous time model from discrete time data on $\{y_t, z_t, t = 1, 2, ..., T\}$.

5. Estimation and Identification

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 (4.15) involves derivatives and convolution integrals of (y, z), the econometrician faces a massive problem of systematically missing data. The approach that we take to the estimation problem that the econometrician faces is conceptually straightforward. We seek to obtain an expression for the likelihood function of a discrete record of data on (y_t, z_t) for $t = 1, ..., T$. The estimator we recommend is obtained by maximizing the likelihood function of this discrete time data record with respect to the free parameters of the continuous time model H_1, H_2, $G(D)$, $\theta(D)$ and $\psi(D)$. The parameters of the continuous time model are in general identified from the discrete time likelihood function because of the extensive cross-equation rational expectations restrictions described in (4.17). That such cross-equation restrictions could break the aliasing identification problem was the message of Hansen and Sargent [7].

In this section, we show how the likelihood function of a discrete time record of data can be calculated from (4.15) - (4.16). Our first task is to convert (4.15) into a slightly altered form. Recall (4.15),

$$K(D) \begin{bmatrix} y_t \\ z_t \end{bmatrix} = Q(D)\, w_t \quad , \tag{4.15}$$

and our specification that

$$E\, w_t w'_{t-u} = \delta(t - u)\, e^{rt}\, V$$

where V is a positive definite symmetric matrix. We shall introduce the following transformations of variables

$$\begin{aligned} \tilde{y}_t &= e^{-r/2\, t}\, y_t \\ \tilde{z}_t &= e^{-r/2\, t}\, z_t \\ \tilde{w}_t &= e^{-r/2\, t}\, w_t \quad . \end{aligned} \tag{5.1}$$

Then it is readily verified that (4.15) is equivalent with

$$K(D + \tfrac{r}{2}) \begin{bmatrix} \tilde{y}_t \\ \tilde{z}_t \end{bmatrix} = Q(D + \tfrac{r}{2})\, \tilde{w}_t \quad . \tag{5.2}$$

Notice that

$$E\, \tilde{w}_t \tilde{w}'_{t-u} = \delta(t - u)\, V \quad .$$

Further recall from section 2 that the zeroes of $\det C(s - r/2) = \det A(s)$ were shown to be less than $r/2$ in real part, and that the zeroes of $\det\theta(s)$ were assumed to be less than $r/2$ in real part. Therefore, since from (4.16) we have

$$K_{11}(s) = \theta^{d}(s)\, A_{m_1}^{-1}\, A(s)$$

$$K_{22}(s) = \theta^{2}(s) \qquad ,$$

it follows from (4.15) that the zeroes of $\det K(s)$ are less than $r/2$ in real part. It immediately follows that the zeroes of $\det K(s + r/2)$ are less than zero in real part. This condition on the zeroes of $\det K(s + r/2)$ together with the assumption that $E\,\tilde{v}_t\,\tilde{v}_t' = \delta(t - s)\,V$ implies that if we regard the system as having started up in the infinite past, then (5.2) describes a continuous time covariance stationary stochastic process. Let us rewrite (5.2) as

$$(5.3) \qquad \tilde{K}(D) \begin{bmatrix} \tilde{y}_t \\ \tilde{z}_t \end{bmatrix} = \tilde{Q}(D)\,\tilde{v}_t$$

where

$$(5.4) \qquad \begin{aligned} \tilde{K}(D) &= K(D + r/2) \\ \tilde{Q}(D) &= Q(D + r/2) \end{aligned} \qquad .$$

From the preceding observations, the zeroes of $\det\tilde{K}(s)$ are less than zero in real part.

The covariogram of the $(\tilde{y}, \tilde{z})$ process is defined as the $(p \times p)$ matrix function

$$R(\tau) = E \begin{bmatrix} \tilde{y}_t \\ \tilde{z}_t \end{bmatrix} \begin{bmatrix} \tilde{y}_{t-\tau} \\ \tilde{z}_{t-\tau} \end{bmatrix}' \qquad .$$

The spectral density matrix of the $(\tilde{y}, \tilde{z})$ process is defined as the Fourier transform of $R(\tau)$, namely,

$$(5.6) \qquad S(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau)\, d\tau \quad .$$

It can be shown (see Phillips [21] or Kwakernaak and Sivan [12]) that for our model (5.2), $S(\omega)$ is given by

$$(5.7) \qquad S(\omega) = K(i\omega+\tfrac{r}{2})^{-1} Q(i\omega+\tfrac{r}{2})\, V\, Q(-i\omega+\tfrac{r}{2})' K(-i\omega+\tfrac{r}{2})^{-1'}$$

or

$$S(\omega) = \tilde{K}(i\omega)^{-1} \tilde{Q}(i\omega)\, V\, \tilde{Q}(-i\omega)' \tilde{K}(-i\omega)^{-1'} \quad .$$

Equation (5.7) provides a convenient expression for the spectral density of the continuous time process $(\tilde{y}_t, \tilde{z}_t)$. For our estimation problem, to obtain the likelihood function we require an expression for the spectral density of the discrete time data. The discrete time spectrum $S^d(\omega)$ is related to the continuous time spectrum via the "folding" relationship

$$(5.8) \qquad S^d(\omega) = \sum_{k=-\infty}^{\infty} S(\omega + 2\pi k) \quad , \quad -\pi \le \omega \le \pi \quad .$$

(For example, see Koopmans [11].)

For the econometric applications that we have in mind, creating $S^d(\omega)$ numerically from (5.8) is feasible, but much more expensive than the following procedure, which utilizes results of A. W. Phillips [21].

Assume that the zeroes of $\det \tilde{K}(s)$ are distinct and let

$$\det \tilde{K}(s) = \lambda_0 (s - \lambda_1) \cdots (s - \lambda_{m_4})$$

where m_4 is the number of zeroes of $\det \tilde{K}(s)$.

Define

$$(5.9) \qquad W_j = \frac{-[\operatorname{adj} \tilde{K}(\lambda_j)]\tilde{Q}(\lambda_j) V \tilde{Q}(-\lambda_j)'[\operatorname{adj} \tilde{K}(-\lambda_j)]'}{2\lambda_0^2 \lambda_j \prod_{\substack{k=1 \\ k\neq j}}^{m_4} (\lambda_j - \lambda_k)(-\lambda_j - \lambda_k)} .$$

Expanding (5.7) in partial fractions gives

$$(5.10) \qquad S(\omega) = \sum_{j=1}^{m_4} \frac{W_j}{(i\omega - \lambda_j)} - \sum_{j=1}^{m_4} \frac{W_j'}{(i\omega + \lambda_j)} .$$

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

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

Therefore, taking inverse Fourier transforms of each side of (5.10) gives

$$(5.11) \qquad R(s) = \begin{cases} \sum_{j=1}^{m_4} W_j e^{\lambda_j s} & \text{for} \quad s \geq 0 \\ \sum_{j=1}^{m_4} W_j' e^{-\lambda_j s} & \text{for} \quad s < 0 \end{cases}$$

By sampling (5.11) at integer s, we obtain the covariogram of the discrete time process $(\tilde{y}_t, \tilde{z}_t : t = 0, \pm 1, \pm 2, \dots)$. The essential element in writing down the likelihood function of the discrete time data as a function of the model's free parameters[5] H_1, H_2, $G(s)$, $\theta(s)$, and $\psi(s)$ is the ability to represent $R(\tau)$ sampled at the integers as a function of those free parameters. Expression (5.11) and the steps leading up to it accomplish this task.

It will also prove useful to have another expression for the spectrum of the sampled $(\tilde{y}_t, \tilde{z}_t)$ process. To derive it, let $\alpha_j = e^{\lambda_j}$ and write

(5.11) sampled at the integers as

$$(5.12) \qquad R(\tau) = \begin{cases} \sum_{j=1}^{m_4} W_j \, \alpha_j^{\tau} & \tau \geq 0 \\ \sum_{j=1}^{m_4} W_j' \, (\alpha_j^{-1})^{\tau} & \tau < 0 \end{cases} .$$

Define the covariance generating function

$$(5.13) \qquad g(\zeta) = \sum_{\tau=-\infty}^{\infty} R(\tau) \, \zeta^{\tau} .$$

Using (5.12) we readily obtain

$$(5.14) \qquad g(\zeta) = \sum_{j=1}^{m_4} W_j \left(\frac{1}{1 - \alpha_j \zeta} \right) + \sum_{j=1}^{m_4} W_j' \left(\frac{\zeta^{-1} \alpha_j}{1 - \zeta^{-1} \alpha_j} \right) .$$

Evaluating (5.14) at $\zeta = e^{-i\omega}$ gives the spectral density of the integer-sampled process $(\tilde{y}_t, \tilde{z}_t)$:

$$(5.15) \qquad S^d(\omega) = g(e^{-i\omega}) = \sum_{j=1}^{m_4} W_j \left(\frac{1}{1 - \alpha_j e^{-i\omega}} \right) + \sum_{j=1}^{m_4} W_j' \left(\frac{\alpha_j e^{+i\omega}}{1 - \alpha_j e^{i\omega}} \right) .$$

With these results in hand, we can now indicate how to construct the likelihood function for a set of observations on $(\tilde{y}_t, \tilde{z}_t)$, $t = 1, \ldots, T$, assuming that w_t is a Gaussian process. Define the stacked matrices of observations on $\tilde{y}_t$ and $\tilde{z}_t$, $t = 1, \ldots, T$ as

$$(5.15) \qquad \bar{y}_T = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{y}_T \end{bmatrix} , \qquad \bar{z}_T = \begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{z}_T \end{bmatrix} .$$

Define the covariance matrix of $(\bar{y}_T', \bar{z}_T')$

$$(5.16) \qquad \Gamma_T = E \begin{bmatrix} \bar{y}_T \\ \bar{z}_T \end{bmatrix} \begin{bmatrix} \bar{y}_T \\ \bar{z}_T \end{bmatrix}' .$$

The $pT \times pT$ covariance matrix Γ_T can be computed as a function of the free parameters of the continuous time model H_1, H_2, $G(s)$, $\theta(s)$, and $\psi(s)$ using (5.15), (5.16) and the results leading up to (5.12). The normal log likelihood for $(\bar{y}_T', \bar{z}_T)$ is given by

$$(5.17) \qquad \mathcal{L}_T^* = -\frac{1}{2} T p \log 2\pi - \frac{1}{2} \log \det \Gamma_T - \frac{1}{2} [\bar{y}_T' \, \bar{z}_T'] \, \Gamma_T^{-1} \begin{bmatrix} \bar{y}_T \\ \bar{z}_T \end{bmatrix} .$$

The log likelihood function (5.16) is to be maximized with respect to the free parameters H_1, H_2, $G(s)$, $\theta(s)$, and $\psi(s)$ of the continuous time model. These parameters make their appearance in (5.17) through the covariance matrix Γ_T.

The maximization of (5.17) 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, (5.17) is a formidable function because the $pT \times pT$ matrix Γ_T must be inverted each time (5.17) is evaluated for different points in the space of free parameters of $\{H_1, H_2, G(s), \theta(s), \psi(s)\}$. Since the matrix Γ_T is liable to be very large, this difficulty has led researchers such as Hannan [7] and Phadke and Kadem [20] to propose frequency domain approximations to the normal likelihood function that economize on computations. Let the periodogram of the $(\tilde{y}_t, \tilde{z}_t)$ process at frequency $\omega_j = \frac{2\pi j}{T}$ $j = 1, \ldots, T$ be the $p \times p$ matrix $I(\omega_j)$. The approximations used are then

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

$$\log \det \Gamma_T \approx \sum_{j=1}^{T} \log \det S^d(\omega_j) .$$

Substituting the above approximations into (5.17) gives the approximate log likelihood function

(5.18) $$\mathcal{L}_T^{**} = -\frac{1}{2}Tp\log 2\pi - \frac{1}{2}\sum_{j=1}^{T}\log\det S^d(\omega_j) - \frac{1}{2}\sum_{j=1}^{T}\text{trace}\,[S^d(\omega_j)^{-1}I(\omega_j)] .$$

Equation (5.15) and the results leading up to it express $S^d(\omega_j)$ as a function of the free parameters of the continuous time model $\{H_1, H_2, G(s), \theta(s), \psi(s)\}$. By maximizing (5.18) instead of (5.17), 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 (5.18) is a good approximation to (5.17) in the sense that maximizing it delivers estimators asymptotically equivalent to those obtained by maximizing (5.17).

Identification of the free parameters $\beta = \{H_1, H_2, G(s), \theta(s), \psi(s)\}$ of the continuous time model from the discrete time data record $\{\bar{y}_T, \bar{z}_T\}$ can be addressed as follows. Let the true values of the parameters be β^o, and the true spectral density matrix be $S^d(\omega;\beta^o)$. Let the value of the spectral density matrix implied by the parameter vector β be $S^d(\omega;\beta)$. Then consider the function

(5.19) $$\mathcal{L}_T^{+} = -\frac{1}{2}Tp\log 2\pi - \frac{1}{2}\sum_{j=1}^{T}\log\det S^d(\omega_j;\beta) - \frac{1}{2}\sum_{j=1}^{T}\text{trace}\,[S^d(\omega_j;\beta)^{-1}S_d(\omega_j;\beta^o)]$$

which is formed by setting $S_d(\omega_j;\beta^o) = I(\omega_j)$ in (5.18). Setting $S_d(\omega_j;\beta^o) = I(\omega_j)$ amounts to assuming that the sample moments equal population moments. The function (5.19) of β achieves a maximum at

$\beta = \beta^o$. If the parameter value $\beta = \beta^o$ is the unique maximizer of (5.19), the model is said to be identified.

Recall from (4.15) and (5.3) that the model being estimated is

$$(5.20)\qquad \begin{bmatrix} \tilde{K}_{11}(D) & \tilde{K}_{12}(D) \\ 0 & \tilde{K}_{22}(D) \end{bmatrix} \begin{bmatrix} \tilde{y}_t \\ \tilde{z}_t \end{bmatrix} = \begin{bmatrix} \tilde{Q}_{11}(D) & \tilde{Q}_{12}(D) \\ 0 & \tilde{Q}_{22}(D) \end{bmatrix} \begin{bmatrix} \tilde{w}_{1,t} \\ \tilde{w}_{z,t} \end{bmatrix}$$

where

$$\tilde{K}_{11}(D) = \theta^d(D + \tfrac{r}{2}) A_{m_1}^{-1} A(D + \tfrac{r}{2})$$

$$\tilde{K}_{12}(D) = -A_{m_1}^{-1} \sum_{j=1}^{k} B_j J_2 \theta^d(D+\tfrac{r}{2}) \mu_1(\rho_j, D+\tfrac{r}{2})$$

$$\tilde{K}_{22}(D) = \theta^2(D+\tfrac{r}{2})$$

$$(5.21)\qquad \tilde{Q}_{11}(D) = -A_{m_1}^{-1} \sum_{j=1}^{k} B_j \left[\frac{\theta^a(D+\frac{r}{2}) \psi^1(D+\frac{r}{2}) - \theta^d(D+\frac{r}{2}) \theta^1(-\rho_j) \psi^1(-\rho_j)}{D - \frac{r}{2} + \rho_j} \right]$$

$$\tilde{Q}_{12}(D) = A_{m_1}^{-1} \sum_{j=1}^{k} B_j J_2 \theta^d(D + \tfrac{r}{2}) \mu_2(\rho_j, D + \tfrac{r}{2})$$

$$\tilde{Q}_{22} = \psi^2(D + \tfrac{r}{2}) \quad .$$

The parameters on the right side of (5.21) are written as functions of the free parameters of the model $\{H_1, H_2, G(s), \theta(s), \psi(s), V\}$ in equation (4.17). A distinguishing feature of the model (5.20) - (5.21) is the presence of an extensive set of restrictions across the parameters of the $\tilde{z}$ process and the $\tilde{y}$ process. In particular, the parameters of the operators $\tilde{K}_{12}(D)$ and $\tilde{Q}_{12}(D)$ are themselves nonlinear functions of the parameters of $\tilde{K}_{22}(D) = \theta^2(D + \frac{r}{2})$ and $\tilde{Q}_{22}(D) = \psi^2(D + \frac{r}{2})$. Such cross-equation restrictions are a hallmark of rational expectations models. Even more so than in models in which the agent and the econometrician use data at the same level of time aggregation, in the present setting these cross-equation restrictions play a crucial role in permitting identification of the parameters of the continuous-time model from the discrete-time data record. In particular suppose we consider the $\tilde{z}$ process from (5.20) in isolation from the $\tilde{y}$ process, namely,

(5.22) $$\tilde{K}_{22}(D)\ \tilde{z}_t = \tilde{Q}_{22}(D)\ \tilde{w}_{z,t}$$

or

$$\theta^2(D + \frac{r}{2})\ \tilde{z}_t = \psi^2(D + \frac{r}{2})\ \tilde{w}_{z,t}$$

with

$$E\,\tilde{w}_{zt}\,\tilde{w}'_{zt-u} = \delta(t - u)\ V_2 \quad .$$

The parameters of $\theta^2(D + \frac{r}{2})$, $\psi^2(D + \frac{r}{2})$ and V_2 are among the free parameters of the continuous-time model. One can imagine proceeding as above to write down the likelihood function of a discrete-time record of $\bar{z}_T$ given the model (5.22). However, it is known that the parameters of (5.22) are not in general identified from the discrete-time likelihood functions. That is, the counterpart of (5.19) in general has a multiplicity of maximizers. This is the classical aliasing identification problem for linear stochastic systems. We summarize this identification problem for (5.22) in Appendix C, where it is shown that in general alternative settings for the free parameters of $\tilde{K}_{22}(D)$ and $\tilde{Q}_{22}(D)$ give rise to the <u>same</u> discrete-time autocovariance function $R(\tau)$ sampled at the integers, and so to the same discrete-time spectrum $S^d(\omega)$. So in the absence of the cross-equation restrictions (5.21) and (4.17), even the free parameters of the driving process $\tilde{z}_t$ are not identifiable.[6/] Thus, if even the parameters of $\tilde{K}_{22}$ and $\tilde{Q}_{22}$ of the driving process $\tilde{z}$ are to be identifiable, the cross-equation restrictions (5.21) and (4.17) must be exploited and the $(\tilde{y}, \tilde{z})$ must be estimated jointly.

In many cases, the cross-equation restrictions (5.21) and (4.17) are sufficiently stringent that the parameters of the continuous-time model are identified from the likelihood function of the discrete-time data record $[\bar{y}_T, \bar{z}_T]$. The question of identification is studied in some detail by Hansen and Sargent [8]. In the present context, it is useful to illustrate how the cross-equation rational expectations restrictions serve to identify the continuous time parameters in the following example. We let x_{1t} in (4.1) be a scalar stochastic process. We let x_{2t} equal z_t, another scalar stochastic process. The process x_{1t} will generate the error term in the econometrician's model. We let the agent choose the scalar stochastic process y_t, to maximize

$$\lim_{T\to\infty} E\frac{1}{T}\int_0^T \{(x_{1t} + z_t)\,y_t - \frac{1}{2}H_1 y_t^2 - \frac{1}{2}H_2\,[G(D)\,y_t]^2\}\,dt$$

subject to y_t given at t and the laws of motion

$$x_{1t} = DW_{1t}$$

$$(\theta_0^2 + \theta_1^2 D + \theta_2^2 D^2)\,z_t = (\psi_0^2 + \psi_1^2 D)\,w_{zt} \quad .$$

We have set $r = 0$. We set $H_1 = \frac{1}{2}$, $H_2 = 1$, $G(D) = G_0 + G_1 D$, $G_0 = \sqrt{\frac{1}{2}}$, $G_1 = 1$. We further assumed that the $E\,w_t w'_{t-s} = I\,\delta(t-s)$, where I is the (2×2) identity matrix.

We computed the bivariate (y_t, z_t) process implied by this optimization problem for two separate settings for $\theta^2(D)$ and $\psi^2(D)$. For the first example, we chose

$$\theta^2(D) = (2 + 2D + D^2)\,, \quad \psi^2(D) = (2\sqrt{2} + 2D) \quad .$$

For the second example, we chose

$$\theta^2(D) = (54.04479 + 2\,D + D^2)\ , \quad \psi^2(D) = (14.70303 + 2.0\,D)\ .$$

It can be verified using the methods of Appendix C that these two processes for z_t have been constructed to give rise to autocovariances for z_t that agree at the integers. In particular, the first example is chosen so that

$$E\,z_t\,z_{t-\tau} = R_{22}(\tau) = e^{-\tau} \cos \tau \quad \text{for} \quad \tau > 0 \quad ,$$

while the second example was chosen so that

$$R_{22}(\tau) = e^{-\tau} \cos\,((1 + 2\,\pi)\,\tau) \quad \text{for} \quad \tau > 0 \quad .$$

For examples 1 and 2, Tables 1 and 2, respectively, report $K(D)$ and $Q(D)$, calculated according to (4.16) and (4.17). The tables also report parts of the covariogram $R(\tau)$ and the spectral density $S^d(\omega)$ of the sampled process, calculated according to the formulas described above. Notice that the $R_{22}(\tau)$'s in the two examples agree at the integers, but not in between the integers. Notice also that the discrete spectral densities of the z process, $S^d_{22}(\omega)$, are equal in the two examples. These features illustrate the presence of the aliasing identification problem for the z_t process considered in isolation from the y_t process. Next, notice that even at the integers, $R_{11}(\tau)$, $R_{12}(\tau)$ and $R_{21}(\tau)$ disagree between the two examples. Similarly, $S^d_{11}(\omega)$, $S^d_{12}(\omega)$, and $S^d_{21}(\omega)$ disagree between the two examples. The cross-equation rational expectation restrictions are responsible for these differences. On the basis of such differences, the analyst can discriminate among the continuous time models from discrete time observations on y_t and z_t.

Table 1

EXAMPLE 1

$$\left[\begin{pmatrix}1 & -.6\\ 0 & 2\end{pmatrix}+\begin{pmatrix}1 & -.2\\ 0 & 2\end{pmatrix}D+\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}D^2\right]\begin{bmatrix}y_t\\ z_t\end{bmatrix}=\left\{\begin{bmatrix}-1 & -.4\\ 0 & 2\sqrt{2}\end{bmatrix}+\begin{bmatrix}0 & 0\\ 0 & 2\end{bmatrix}D\right\}\begin{pmatrix}w_{1t}\\ w_{zt}\end{pmatrix}.$$

roots of det $K(s) = 0$: $\rho_1 = -1 + 1i$
$\rho_2 = -1 - i$
$\rho_3 = -1$

τ	$R_{11}(\tau)$	$R_{12}(\tau)$	$R_{21}(\tau)$	$R_{22}(\tau)$
0	.606	.333	.333	2.000
.25	.490	.404	.226	1.509
.50	.390	.405	.139	1.064
.75	.305	.364	.072	.691
1.00	.236	.302	.025	.397
1.25	.179	.234	- .005	.180
1.50	.134	.169	- .024	.031
1.75	.099	.112	- .033	- .061
2.00	.072	.066	- .035	- .112
2.25	.052	.032	- .032	- .132
2.50	.037	.007	- .028	- .131
2.75	.026	- .008	- .022	- .118
3.00	.019	- .017	- .017	- .098
3.25	.013	- .021	- .012	- .077
3.50	.010	- .021	- .008	- .056
3.75	.007	- .020	- .004	- .038
4.00	.006	- .017	- .002	- .023
4.25	.004	- .013	- .000	- .012

		$S^d_{12}(\omega)$		$S^d_{21}(\omega)$		
ω	$S^d_{11}(\omega)$	re	im	re	im	$S^d_{22}(\omega)$
.000	1.284	.638	.000	.638	- .000	2.343
.049	1.282	.638	- .019	.638	.019	2.346
.098	1.275	.639	- .038	.639	.038	2.353
.147	1.263	.639	- .057	.639	.057	2.365
.196	1.248	.641	- .077	.641	.077	2.382
.245	1.229	.642	- .097	.642	.097	2.403
.294	1.207	.643	- .117	.643	.117	2.428
.343	1.181	.644	- .137	.644	.137	2.456
.392	1.154	.645	- .157	.645	.157	2.488
.441	1.125	.645	- .177	.645	.177	2.522

Table 2

EXAMPLE 2

$$\left\{\begin{pmatrix}1.0 & -.05259\\ 0 & 54.04479\end{pmatrix}+\begin{pmatrix}1.0 & -.01753\\ 0 & 2.0\end{pmatrix}D+\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}D^2\right\}\begin{pmatrix}y_t\\ z_t\end{pmatrix}=\left\{\begin{pmatrix}-1.0 & -.03506\\ 0 & 14.70303\end{pmatrix}+\begin{pmatrix}0 & 0\\ 0 & 2.0\end{pmatrix}D\right\}\begin{pmatrix}w_{1t}\\ w_{2t}\end{pmatrix}$$

roots of det $K(s) = 0$: $\rho_1 = -1 + (1+2\pi)i$

$\rho_2 = -1 - (1+2\pi)i$

$\rho_3 = -1$

τ	$R_{11}(\tau)$	$R_{12}(\tau)$	$R_{21}(\tau)$	$R_{22}(\tau)$
0	.500	.027	.027	2.000
.25	.389	- .005	- .006	- .385
.50	.303	- .026	- .013	-1.064
.75	.236	.004	.009	.643
1.00	.184	.007	.004	.397
1.25	.143	- .010	- .007	- .543
1.50	.111	- .004	.000	- .031
1.75	.087	.004	.004	.341
2.00	.067	- .001	- .001	- .112
2.25	.052	- .004	- .002	- .164
2.50	.041	.001	.001	.131
2.75	.031	.000	.000	.048
3.00	.024	- .002	- .001	- .098
3.25	.019	- .000	.000	.008
3.50	.015	.000	.000	.056
3.75	.011	- .000	- .000	- .026
4.00	.009	- .000	- .000	- .023

		$S^d_{12}(\omega)$		$S^d_{21}(\omega)$		
ω	$S^d_{11}(\omega)$	re	im	re	im	$S^d_{22}(\omega)$
.000	1.083	.031	.00000	.031	.00000	2.343
.049	1.080	.031	.00009	.031	- .00009	2.346
.098	1.073	.031	.00019	.031	- .00019	2.353
.147	1.061	.031	.00027	.031	- .00027	2.365
.196	1.046	.032	.00033	.032	- .00033	2.382
.245	1.026	.032	.00036	.032	- .00036	2.403
.294	1.003	.032	.00037	.032	- .00037	2.428
.343	.977	.033	.00034	.033	- .00034	2.456
.392	.950	.034	.00029	.034	- .00029	2.488
.441	.920	.034	.00019	.348	- .00019	2.522

6. Time Averaged Data

The procedures of the preceding section 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 quarter, for example. Observations on GNP, sales, and man-hours are usually recorded in this way. When some or all of the data are recorded in this way, the procedures of the preceding section must be modified. Fortunately, these modifications are straightforward and utilize virtually the same technical apparatus employed in Section 5. Basically, the idea is simply to take into account the implications of time averaging of the continuous time data for the resulting discrete time spectral density and covariogram.

The preceding section deduced the continuous time autocovariance function

$$R(s) = \sum_{j=1}^{m_4} W_j e^{\lambda_j s} \qquad \text{for} \quad s \geq 0 \quad , \tag{6.1}$$

and supplied forumlas for the W_j's in terms of the parameters of the economic model. Recall that

$$R(s) \equiv E\begin{pmatrix} \tilde{y}_t \\ \tilde{z}_t \end{pmatrix} \begin{pmatrix} \tilde{y}_{t-s} \\ \tilde{z}_{t-s} \end{pmatrix}' \qquad , \qquad s \geq 0 \quad . \tag{6.2}$$

We can indicate completely the effects of time averaging by supposing that $\tilde{y}_t$ and $\tilde{z}_t$ are both scalar processes, so that $R(s)$ is a (2×2) matrix function of s, with $R(s) = \begin{bmatrix} r_{11}(s) & r_{12}(s) \\ r_{21}(s) & r_{22}(s) \end{bmatrix}$. We suppose that the discrete data on $\tilde{z}_t$ are point in time, while those on $\tilde{y}_t$ are unit averaged. In particular, we consider the unit averaged process

$$\bar{y}_t = \int_0^1 \tilde{y}_{t-s}\, ds \quad .$$

We define the cross-covariogram of the joint $(\bar{y}_t, \tilde{z}_t)$ process as

$$\bar{R}(\tau) = E \begin{bmatrix} \bar{y}_t \\ \tilde{z}_t \end{bmatrix} \begin{bmatrix} \bar{y}_t \\ \tilde{z}_t \end{bmatrix}' \qquad \text{for integer } \tau$$

$$\equiv \begin{bmatrix} \bar{r}_{11}(\tau) & \bar{r}_{12}(\tau) \\ \bar{r}_{21}(\tau) & \bar{r}_{22}(\tau) \end{bmatrix} \quad .$$

Evidently, $\bar{r}_{22}(\tau) = r_{22}(\tau)$, so that we have only to compute the terms $\bar{r}_{11}(\tau)$ and $\bar{r}_{12}(\tau)$. Evaluating these terms 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 observations.

We first compute $r_{11}(\tau)$. We shall find it convenient to handle the terms for $\tau \geq 1$ separately from $r_{11}(0)$. We first wish to compute $\bar{r}_{11}(\tau)$ for $\tau \geq 1$. We have

$$\bar{r}_{11}(\tau) = E \int_0^1 x_{1t-s}\, ds \int_0^1 x_{1t-\tau-u}\, du$$

$$= \int_0^1 \int_0^1 E\, [x_{1t-s}\, x_{1t-\tau-u}]\, ds\, du$$

$$= \int_0^1 \int_0^1 r_{11}(\tau + u - s)\, ds\, du \quad .$$

Let us write $W_j = \begin{bmatrix} w_j^{11} & w_j^{12} \\ w_j^{21} & w_j^{22} \end{bmatrix}$. Substituting this into the above line gives

$$\text{(6.3)} \qquad \bar{r}_{11}(\tau) = \sum_{j=1}^{m_4} w_j^{11} \int_0^1 \int_0^1 e^{\lambda_j(\tau+u-s)}\, ds\, du \qquad \text{for } \tau \geq 1 \quad .$$

We thus have to evaluate the double integral

$$\int_0^1 \int_0^1 e^{\lambda_j(\tau+u-s)}\, ds\, du = e^{\lambda_j \tau} \int_0^1 e^{\lambda_j u}\, du \int_0^1 e^{-\lambda_j s}\, ds$$

$$= e^{\lambda_j \tau} \left[\frac{e^{\lambda_j u}}{\lambda_j}\right]_0^1 \left[\frac{e^{-\lambda_j s}}{-\lambda_j}\right]_0^1$$

$$= e^{\lambda_j \tau} \left[\frac{e^{\lambda_j} - 1}{\lambda_j}\right] \left[\frac{-e^{-\lambda_j} + 1}{\lambda_j}\right] \quad .$$

Substituting back into (6.3) we obtain

$$(6.4) \qquad \bar{r}_{11}(\tau) = \sum_{j=1}^{m_4} w_j^{11} \left[\frac{e^{\lambda_j} - 1}{\lambda_j}\right] \left[\frac{-e^{-\lambda_j} + 1}{\lambda_j}\right] e^{\lambda_j \tau} \qquad \text{for} \quad \tau \geq 1 \quad .$$

Next we compute

$$\bar{r}_{11}(0) = \int_0^1 \int_0^1 r_{11}(u - s)\, du\, ds$$

$$(6.5) \qquad = \int_0^1 \int_0^s r_{11}(u - s)\, du\, ds + \int_0^1 \int_s^1 r_{11}\ (u - s)\, du\, ds \quad .$$

From (6.1) we need to evaluate the integral

$$\int_0^1 \int_0^s e^{-\lambda_j(u - s)}\, du\, ds = \int_0^1 e^{\lambda_j s} \int_0^s e^{-\lambda_j u}\, du\, ds$$

$$= \int_0^1 e^{\lambda_j s} \left[\frac{e^{-\lambda_j u}}{-\lambda_j}\right]_0^s ds$$

$$= \int_0^1 e^{\lambda_j s} \left[\frac{1 - e^{-\lambda_j s}}{\lambda_j}\right] ds$$

$$= \frac{1}{\lambda_j}\left[\frac{e^{\lambda_j} - 1}{\lambda_j} - 1\right] \quad .$$

We also need the integral

$$\int_0^1 \int_s^1 e^{\lambda_j(u - s)}\, du\, ds = \int_0^1 e^{-\lambda_j s} \left[\frac{e^{\lambda_j u}}{\lambda_j}\right]_s^1 ds$$

$$= \int_0^1 e^{-\lambda_j s} \left[\frac{e^{\lambda_j} - e^{\lambda_j s}}{\lambda_j}\right] ds$$

$$= \frac{1}{\lambda_j}\left[\frac{e^{\lambda_j}(1 - e^{-\lambda_j})}{\lambda_j} - 1\right]$$

$$= \frac{1}{\lambda_j}\left[\frac{e^{\lambda_j} - 1}{\lambda_j} - 1\right] \quad .$$

Using the results from these integrations in (6.1) and (6.5) gives

$$(6.6) \qquad \bar{r}_{11}(0) = \sum_{j=1}^{m_4} \frac{2}{\lambda_j}\left[\frac{e^{\lambda_j} - 1}{\lambda_j} - 1\right] w_j^{11} \quad .$$

Collecting our results, we have

$$(6.6)\qquad \bar{r}_{11}(0) = \sum_{j=1}^{m_4} \frac{2}{\lambda_j}\left[\frac{e^{\lambda_j}-1}{\lambda_j} - 1\right] w_j^{11}$$

$$(6.4)\qquad \bar{r}_{11}(\tau) = \sum_{j=1}^{m_4} w_j^{11}\left[\frac{e^{\lambda_j}-1}{\lambda_j}\right]\left[\frac{-e^{-\lambda_j}+1}{\lambda_j}\right] e^{\lambda_j \tau} \quad .$$

Using (6.4) and (6.6), we want to derive the spectral density of the discrete time unit-sampled process for $\bar{y}_t$. Let

$$v_j^{11} = w_j^{11}\left[\frac{e^{\lambda_j}-1}{\lambda_j}\right]\left[\frac{-e^{-\lambda_j}+1}{\lambda_j}\right]$$

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

Then from (6.4) it follows that

$$(6.7)\qquad \begin{aligned} \bar{r}_{11}(\tau) &= \sum_{j=1}^{m_4} v_j^{11}\, \alpha_j^{\tau} && \text{if } \tau > 1 \\ &= \sum_{j=1}^{m_4} \frac{2}{\lambda_j}\left[\frac{e^{\lambda_j}-1}{\lambda_j}\right] w_j^{11} && \text{if } \tau = 0 \quad . \end{aligned}$$

To obtain the z-transform of $\bar{r}_{11}(\tau)$ we first calculate

$$\sum_{\tau=1}^{\infty} v_j^{11}\, \alpha_j^{\tau}\, z^{\tau} = \frac{v_j^{11}\, \alpha_j\, z}{1 - \alpha_j z} \quad .$$

Hence the z-transform of $\bar{r}_{11}$ is

$$(6.8)\qquad \sum_{j=1}^{m_4} \frac{v_j^{11}\,\alpha_j z}{(1-\alpha_j z)} + \sum_{j=1}^{m_4} \frac{v_j^{11}\,\alpha_j z^{-1}}{(1-\alpha_j z^{-1})} + \sum_{j=1}^{m_4} \frac{2}{\lambda_j}\left[\frac{e^{\lambda_j}-1}{\lambda_j} - 1\right] w_j^{11} \quad .$$

Substituting $e^{-i\omega}$ for z in (6.8) gives the discrete time spectral density for the unit averaged process $\bar{y}_t$.

Next we compute for $\tau \geq 0$

$$E\,\tilde{z}_t\,\bar{y}_{t-\tau} = \int_0^1 r_{21}\,(\tau + u)\,du \quad .$$

From (6.1) we need to evaluate

$$\int_0^1 e^{\lambda_j(\tau+u)}\, du = e^{\lambda_j \tau}\left[\frac{e^{\lambda_j u}}{\lambda_j}\right]_0^1 = \frac{e^{\lambda_j \tau}}{\rho_j}\,[e^{\lambda_j} - 1] \quad .$$

Using this integral in (6.1) gives

$$\bar{r}_{21}(\tau) = \sum_{j=1}^{m_4} w_j^{21}\, \alpha_j^{\tau}\, \frac{(e^{\lambda_j} - 1)}{\lambda_j} \qquad \tau \geq 0 \quad .$$

The z-transform of $\bar{r}_{21}$ is then

$$\text{(6.9)} \qquad \sum_{j=1}^{m_4} \frac{w_j^{21}\, \alpha_j\, z}{(1-\alpha_j z)}\, \frac{(e^{\lambda_j} - 1)}{\lambda_j} + \sum_{j=1}^{m_4} \frac{w_j^{12}\, \alpha_j\, z^{-1}\, (e^{\lambda_j} - 1)}{(1 - \alpha_j\, z^{-1})\, \lambda_j} + \sum_{j=1}^{m_4} w_j^{21}\, \frac{(e^{\lambda_j} - 1)}{\lambda_j} \quad .$$

Substituting $e^{-i\omega}$ for z in (6.9) gives the discrete time cross-spectrum between the unit averaged process $\bar{y}_t$ and the point-in-time sampled process $\tilde{z}_t$.

Finally, the z-transform of $\bar{r}_{22}$ is

$$\text{(6.10)} \qquad \sum_{j=1}^{m_4} \frac{w_j^{22}\, \alpha_j\, z}{1 - \alpha_j\, z} + \sum_{j=1}^{m_4} \frac{w_j^{22}\, \alpha_j\, z^{-1}}{1 - \alpha_j\, z^{-1}} + \sum_{j=1}^{m_4} w_j^{22} \quad .$$

The construction leading to (6.8) can be used to calculate the discrete time cross-spectrum or cross-covariogram between two processes each of which has been unit averaged. Similarly, the calculations leading to (6.9) construct the cross-spectrum and cross-covariogram between a unit averaged and a point-in-time sampled process. Thus, (6.8) and (6.9) could be used to construct the spectral density and matrix covariogram of an $(N \times 1)$ vector process, some of whose members are unit averaged, while the others are point-in-time observations.

With (6.8) and (6.9) in hand, the estimation strategy advocated in Section 5 can be implemented in an appropriate way. In particular, an approximate likelihood function can be constructed for a discrete-time data set which is an arbitrary mixture of series, some of which are unit averaged while others are point-in-time observations. As in Section 5, the approximate likelihood function is to be viewed as a function of the free parameters of the underlying continuous time model, and to be maximized with respect to them.

7. Conclusion

This paper has shown how to estimate members of a class of continuous time rational expectations models from discrete time observations. As Sims [27], Geweke [4], and others have documented in somewhat related contexts, serious errors in inference about parameters can be made if the analyst ignores the temporal aggregation problem that exists when economic activity is proceeding at a finer level of time than are the analyst's observations. The results in this paper provide a set of methods for resolving this time aggregation problem in essentially an ideal fashion.

The tools developed in this paper could be used to study the nature of the approximation errors that would be committed by an analyst who erroneously assumed that economic activity is occurring in discrete time when it is in fact going on continuously. This problem has a variety of interesting aspects in the context of rational expectations models, some of which we propose to study in subsequent work. We also plan to implement the estimators described in this paper in several applications.

Appendix A

In section 2, we considered the Euler equation

$$\text{(A.1)} \qquad C(-D+\tfrac{r}{2})'C(D-\tfrac{r}{2})\,y_t = [h+x_{1,t}+x_{2,t}] \quad ,$$

where $H_1 + G(r-s)'H_2G(s) = C(-s+\frac{r}{2})'C(s-\frac{r}{2})$ and the zeroes of $\det C(-s+\frac{r}{2})'$ exceed zero in real part and the zeroes of $\det C(s-\frac{r}{2})$ are less than zero in real part. We asserted that under some side conditions on our problem, the appropriate solution of the Euler equation that satisfies $\lim_{T\to\infty} E_0 \frac{1}{T}\int_0^T e^{-rt}\, y_t' H_1 y_t\, dt < +\infty$ is found from (A.1) by solving the "stable roots backwards and the unstable roots forwards." In this appendix we indicate the nature of these side conditions.

Our method of solving (A.1) chooses the unique solution that delivers a closed loop system governing the joint (y_t, x_t) process that is of exponential order less than $r/2$. Under the appropriate regularity conditions, the optimal closed loop system matrix for the discounted infinite time optimal linear regulator problem is known to be of exponential order less than $r/2$. Consider the discounted optimal linear regulator problem, to minimize

$$\lim_{T\to\infty} \int_0^T e^{-rt}\,[Y_t'Y_t + U_t'\,Q\,U_t]\,dt$$

subject to

$$DX_t = AX_t + BU_t$$

$$Y_t = CX_t$$

where X is $(N\times 1)$, A is $(N\times N)$, B is $(K\times N)$, C is $(L\times N)$, and Q is a $(K\times K)$ positive definite matrix. Define the transformed variables $\tilde{X}_t = e^{-r/2\,t}X_t$ and $\tilde{U}_t = e^{-r/2\,t}\,U_t$. Then transform the problem to the undiscounted problem, to minimize

$$\lim_{T\to\infty} \int_0^T [\tilde{Y}_t' \tilde{Y}_t + \tilde{U}_t Q \tilde{U}_t] \, dt$$

subject to

$$D\tilde{X}_t = [A - \frac{r}{2} I] \tilde{X}_t + B \tilde{U}_t$$

$$\tilde{Y}_t = C \tilde{X}_t \quad .$$

With suitable definitions of X, Y, U, A, B, and C, the certainty version of our problem is a version of this deterministic linear regulator problem. For example, we would define $X(t)' = [y(t)', x(t)']$. It is known that if the pair $[A - \frac{r}{2} I, B]$ is <u>stabilizable</u>, and the pair $(A - \frac{r}{2} I, C)$ is <u>detectable</u>, then the optimum closed loop system for $\tilde{X}_t$ is stable, and the associated optimum closed loop system governing X_t is of exponential order less than $r/2$. For the problem analyzed in the text, sufficient conditions for $[A - \frac{r}{2} I, B]$ to be stabilizable and $(A - \frac{r}{2} I, C)$ to be detectable will be satisfied by virtue of the conditions that we have imposed on the zeroes of $\det \theta(s)$, provided that some side conditions are imposed on H_1 and the G_j's. For example, imposing full rank on H_1 will imply stabilizability of $(A - \frac{r}{2} I, B)$ and detectability of $(A - \frac{r}{2} I, C)$, as can be checked using the conditions for stabilizability and detectability described by Kwakernaak and Sivan [12]. The optimal closed loop system describing (y_t, x_t) of the text can be of exponential order less than $r/2$ if and only if the Euler equation (A.1) is solved in the manner in the text, namely by premultiplying both sides of (A.1) by $C(-D + \frac{r}{2})'^{-1}$. This procedure picks out the unique optimal solution of the Euler equations.

It should be noted that the full rank condition on H_1 is sufficient to imply that the optimal closed loop system is of exponential order less than $r/2$, but is not necessary. For example, an alternative set of sufficient conditions would drop the full rank condition on H_1 and replace it with the condition that the zeroes of $\det G(s)$ are less than $r/2$ in real part (the condition that the zeroes of $\det \theta(s)$ are less than $r/2$ must also be retained). With this alternative condition, the stabilizability and detectability conditions for the linear regulator problem are met.

The full rank condition on H_1 is appropriate for many problems. We conclude this appendix by indicating in terms of classical methods how the nonsingularity of H_1 is sufficient to make our solution to the Euler equation (2.7) the unique one that satisfies condition (2.8).

First, we rewrite (2.7) as

$$\text{(A.2)} \qquad [H_1 + G(r-D)'H_2 G(D)]y_t = x_t^* \qquad \text{for} \quad t \geq 0$$

where

$$x_t^* = h + x_{1,t} + x_{2,t} \qquad \text{and} \qquad \lim_{t \to -\infty} e^{-rt/2} x_t^* = 0 \; .$$

Second, we obtain a particular solution to (A.2). To accomplish this, we take a partial fractions decomposition of $[H_1 + G(r-s)'H_2 G(s)]^{-1}$ of the form

$$[H_1 + G(r-s)'H_2 G(s)]^{-1} = \frac{G_1^*}{s+\rho_1} + \cdots + \frac{G_k^*}{s+\rho_k} + \frac{H_1^*}{s-\rho_1-r} + \cdots + \frac{H_k^*}{s-\rho_k-r} \; .$$

A particular solution to (A.2) is given by

$$y_t^p = -\int_0^{+\infty} (G_1^* e^{\rho_1 u} + \cdots + G_k^* e^{\rho_k u})\, x_{t+u}^*\, du + \int_0^s (H_1^* e^{\rho_1 u} + \cdots + H_k^* e^{\rho_k u}) e^{ru} x_{t-u}^*\, du \; .$$

Third, we obtain a general solution to the homogeneous differential equation in which x_t^* is set to zero for $s \geq 0$. Let $A_1^*, \ldots, A_k^*, B_1^*, \ldots, B_k^*$ be n dimensional vectors with unit norm such that

$$[H_1 + G(r+\rho_j)' H_2 G(-\rho_j)] A_j^* = 0$$

$$[H_1 + G(-\rho_j)' H_2 G(r+\rho_j)] B_j^* = 0 \qquad \text{for} \quad j = 1, \ldots, k \quad .$$

The general solution to the homogeneous equation is

$$y_t^h = c_1 e^{-\rho_1 t} A_1^* + \cdots + c_k e^{-\rho_k t} A_k^* + f_1 e^{(\rho_1 + r)t} B_1^* + \cdots + f_k e^{(\rho_k + r)t} B_k^*$$

where $c_1, \ldots, c_k, f_1, \ldots, f_k$ are arbitrary scalar constants. Finally, we obtain a representation for the solutions to the Euler equation by adding y_t^p and y_t^h.

Since there are $2k$ arbitrary constants and only m_1 initial condition vectors $\overset{\circ}{y}_0$, $D\overset{\circ}{y}_0, \ldots, D^{m_1 - 1}\overset{\circ}{y}_0$ corresponding to k initial conditions, we have a whole family of solutions to the Euler equation (A.2). However, we know that (2.8) has to be satisfied for the optimal decision path. We rewrite the requirement below

$$\text{(A.3)} \qquad \int_0^{+\infty} e^{-rs} y_t' H_1 y_t < +\infty$$

where the matrix H_1 is positive definite. Recall that $-\rho_j > r/2$ for $j = 1, \ldots, k$. For a nonzero c_j

$$\int_0^{+\infty} e^{-rs} e^{-2\rho_j s} c_j^2 A_j^{*\prime} H_1 A_j^* \, ds = +\infty \quad .$$

Thus in order for (A.3) to be satisfied, it must be the case that $c_j = 0$ for $j = 1, \ldots, k$. The initial condition vectors $\overset{\circ}{y}_0$, $D\overset{\circ}{y}_0, \ldots, D^{m_1 - 1}\overset{\circ}{y}_0$ uniquely determine the correct values for $f_1, \ldots, f_k$. The solution for $D^{m_1} y_t$ provided in the paper corresponds to the solution described above.

Appendix B

In this appendix we discuss the derivation of the optimal prediction formulas used in the decision rule deviation in section 2. We accomplish this task by employing Laplace transforms and appealing to some results from elementary complex analysis. The strategy we employ is analogous to one used for discrete-time optimal prediction analyzed in Appendix A of Hansen and Sargent [9].

Let us begin with a convolution operator Ψ defined by a square integrable function ψ. More specifically, for a scalar stochastic process "a" we define

$$\Psi[a]_t = \int_{-\infty}^{+\infty} \psi(s)\, a_{t-s}\, ds$$

where

$$\int_{-\infty}^{+\infty} \psi(s)^2\, ds < +\infty \quad .$$

The Laplace transform of this operator is given by

$$\tilde{\Psi}(\zeta) = \int_{-\infty}^{+\infty} e^{-\zeta s}\psi(s)\; ds = \tilde{\Psi}^{+}(\zeta) + \tilde{\Psi}^{-}(\zeta)$$

$$\tilde{\Psi}^{+}(\zeta) = \int_{0}^{+\infty} e^{-\zeta s}\psi(s)\; ds$$

$$\tilde{\Psi}^{-}(\zeta) = \int_{-\infty}^{0} e^{-\zeta s}\psi(s)\, ds \quad .$$

Writing $\zeta = \zeta_1 + \zeta_2 i$ where ζ_1 and ζ_2 are real variables, we note that $\tilde{\Psi}^{+}(\zeta)$ defines an analytic function for $\zeta_1 > 0$ and $\tilde{\Psi}^{-}(\zeta)$ defines an analytic function for $\zeta_1 < 0$. For $\zeta_1 = 0$ we define

$$\tilde{\Psi}^{+}(i\,\zeta_2) = \lim_{\zeta_1 \downarrow 0} \Psi^{+}(\zeta_1 + \zeta_2\, i)$$

$$\tilde{\Psi}^{-}(i\,\zeta_2) = \lim_{\zeta_1 \uparrow 0} \Psi^{-}(\zeta_1 + \zeta_2\, i)$$

where the above limits are taking with respect to the L^2 metric defined by Lebesque measure on the real line. Thus Ψ is at least well defined almost everywhere for $\zeta_1 = 0$. It turns out that $\tilde{\psi}$ given by $\tilde{\psi}(\zeta_2) = \tilde{\Psi}^+(i\,\zeta_2) + \tilde{\Psi}^-(i\,\zeta_2)$ is the Fourier transform of ψ.

The annihilation operator applied to Laplace transforms of convolution operators is defined by

$$[\tilde{\Psi}(\zeta)]_+ = \tilde{\Psi}^+(\zeta) \quad .$$

Thus the annihilation operator instructs us to "ignore" $\psi(s)$ for $s < 0$. We now restrict ourselves to cases in which $\tilde{\Psi}^-(\zeta)$ is analytic for $\zeta_1 < R$ for some $R > 0$. Under this additional assumption $\tilde{\Psi}(\zeta)$ defines an analytic function in the region $0 < \zeta_1 < R$. This prepares us for considering a lemma analogous to the lemma in Appendix A of Hansen and Sargent [9].

<u>Lemma</u> Suppose A is a meromorphic function in $0 \leq \zeta$ such that

(i) $A(\zeta) = \tilde{\Psi}(\zeta)$ for $0 < \zeta_1 < R$.

(ii) The function A has at most a finite number of singularities $\zeta^1, \zeta^2, \ldots, \zeta^k$ in $0 < \zeta_1$, with $P_1, P_2, \ldots, P_k$ denoting the corresponding principal parts of the Laurent series expansion of A at these points.

(iii) $\lim_{\zeta \to \infty} A(\zeta) = 0$.

Then $$[\tilde{\Psi}(\zeta)]_+ = A(\zeta) - \sum_{j=1}^{k} P_j(\zeta) \quad .$$

<u>Proof</u>: Let $B(\zeta) = A(\zeta) - \sum_{j=1}^{k} P_j(\zeta)$. A standard result from analytic function theory assures us that B is analytic in $0 \le \zeta_1$. The function P_j is analytic for $|\zeta| \ne \zeta_j$ and $\lim_{\zeta \to \infty} P_j(\zeta) = 0$ for $j = 1, 2, \ldots, k$. Thus $\lim_{\zeta \to \infty} B(\zeta) = 0$. Using a result in Beltami and Wohlers [2] we know that there exists a generalized function that has its support $[0, \infty)$ and has B as its Laplace transform. Since

$$\frac{1}{(\zeta - \zeta^j)^m} = - \int_{-\infty}^{0} s^m e^{-\zeta s} e^{\zeta^j s} \, ds \qquad \text{for } \zeta_1 \le \zeta_1^j$$

we are guaranteed that there is a square integrable function that has B as its Laplace transform. We are also assured that there is a square integrable function that has $\sum_{j=1}^{k} P_j$ as its Laplace transform and has $(-\infty, 0]$ as its support set. This is sufficient to deliver the desired conclusion that $\Psi^{+}(\zeta) = B(\zeta)$.

We now wish to show how to use this lemma to solve the prediction problem

$$E_t \int_0^{+\infty} e^{\rho s} x_{t+s} \, ds$$

where

$$\theta(D)\, x_t = \psi(D)\, w_t$$

as in equation (2.2) and the real part of ρ is negative. We begin by assuming that x is a covariance stationary linearly indeterministic process. Embedded in this assumption is the requirement that the zeroes of $\theta(\zeta)$ are less than zero in real part. Now

$$x_t = \theta(D)^{-1} \psi(D)\, w_t = \varphi(D)\, w_t \quad .$$

Since the order of the numerator polynomial of $\varphi_{ij}(\zeta)$ is less than the order of the denominator and since the zeroes of the denominator polynomial are less than zero in real part, we know that $\varphi_{ij}(\zeta)$ is the Laplace transform of a one-sided convolution operator. In other words

$$\varphi_{ij}(D) v_{jt} = \int_0^{+\infty} \varphi^*_{ij}(s)\, w_{jt-s}\, ds$$

for some square summable function φ^*_{ij} where w_j is the j^{th} element of w. Evaluating the integral,

$$\int_0^{+\infty} e^{\rho s} e^{\zeta s}\, ds = -\frac{1}{\rho + \zeta}$$

for ζ_1 less than real part of $-\rho$. Therefore

$$\int_0^{+\infty} e^{\rho s} x_{t+s} ds = -\frac{1}{\rho + D}\; x_t = -\frac{1}{\rho + D}\, \varphi(D)\, w_t \qquad .$$

Remembering that the forecasting errors in forecasting x_{t+s} given Ω_t can be expressed as an integral of $\{w_{t+u} : 0 \le u \le \zeta\}$ we use the continuous-time version of the Weiner-Kolmogorov formulas to ascertain that

$$E_t \int_0^{+\infty} e^{\rho s} x_{t+s} ds = \left[\frac{-1}{\rho + D}\, \varphi(D)\right]_+ w_t \qquad .$$

Our goal is to evaluate

$$\left[\frac{1}{\rho + \zeta}\, \varphi(\zeta)\right]_+$$

using the lemma in this appendix. The meromorphic function $\frac{\varphi(\zeta)}{\rho + \zeta}$ has as its only singularity in $\zeta_1 \ge 0$ a first order pole at $-\rho$. The principal part of the Laurent series expansion of $\frac{\varphi(\zeta)}{\rho + \zeta}$ at ρ is $\varphi(-\rho)$.

Since $\frac{\varphi_{ij}(\zeta)}{\rho+\zeta}$ is a rational function whose denominator polynomial order exceeds its numerator polynomial order,

$$\lim_{\zeta \to +\infty} \frac{\varphi(\zeta)}{\rho+\zeta} = 0 \quad .$$

Thus the lemma is appropriate and

$$\left[\frac{\varphi(\zeta)}{\rho+\zeta}\right]_+ = \frac{\varphi(\zeta) - \varphi(-\rho)}{\rho+\zeta} \quad .$$

We conclude that

$$\text{(B.1)} \qquad E_t \int_0^{+\infty} e^{\rho s} x_{t+s} ds = -\left[\frac{\varphi(D) - \varphi(-\rho)}{D+\rho}\right] w_t \quad .$$

Formula (B.1) provides a solution to the prediction problem discussed in section 2. An alternative representation of this solution is also valuable. First, write

$$\frac{\varphi(\zeta)-\varphi(-\rho)}{\zeta+\varphi} = \frac{\theta(\zeta)^{-1}\psi(\zeta) - \theta(-\rho)^{-1}\psi(\zeta) + \theta(-\rho)^{-1}\psi(\zeta) - \theta(-\rho)^{-1}\psi(-\rho)}{\rho+\zeta}$$

$$= \frac{[\theta(\zeta)^{-1} - \theta(-\rho)^{-1}]\psi(\zeta) + \theta(-\rho)^{-1}[\psi(\zeta) - \psi(-\rho)]}{\rho+\zeta} \quad .$$

Recalling that $\psi(D) w_t = \theta(D) x_t$, we have

$$\left[\frac{\varphi(D)-\varphi(-\rho)}{D+\rho}\right] w_t = \left[\frac{\theta(D)^{-1} - \theta(-\rho)^{-1}}{D+\rho}\right] \psi(D) w_t + \theta(-\rho)^{-1}\left[\frac{\psi(D)-\psi(-\rho)}{D+\rho}\right] w_t$$

$$= \theta(-\rho)^{-1}\left[\frac{\theta(-\rho) - \theta(D)}{D+\rho}\right] x_t + \theta(-\rho)^{-1}\left[\frac{\psi(D)-\psi(-\rho)}{D+\rho}\right] w_t \quad .$$

Now

$$\frac{\theta(\zeta)-\theta(-\rho)}{\zeta+\rho} = \theta_1 + \theta_2(-\rho) + \cdots + \theta_{m_2}(-\rho)^{m_2-1}$$

$$+ [\theta_2 + \theta_3(-\rho) + \cdots + \theta_{m_2}(-\rho)^{m_2-2}]\zeta + \cdots + \theta_{m_2}\zeta^{m_2-1} \quad .$$

An analogous formula can be obtained for

$$\frac{[\psi(D) - \psi(-\rho)]}{D + \rho} \quad .$$

Thus

$$\text{(B.2)} \qquad E_t \int_0^{+\infty} e^{\rho s} x_{t+s} = \theta(-\rho)^{-1} \sum_{j=0}^{m_2-1} D^j \sum_{k=j}^{m_2-1} \theta_{k+1} (-\rho)^{k-j} x_t$$

$$- \theta(-\rho)^{-1} \sum_{j=0}^{m_3-1} D^j \sum_{k=j}^{m_3-1} \psi_{k+1} (-\rho)^{k-j} w_t \quad .$$

This provides us with an expression for the solution of the forecasting problem in terms of x_t, v_t, m_2-1 derivatives of x_t and m_3-1 derivatives of v_t .

Although x is not assumed to be covariance stationary in the text, it is assumed that x can be transformed into a covariance stationary process. It is easily verified by using the transformation suggested in the text that formula (B.2) remains valid.

Appendix C

In this appendix we discuss the aliasing phenomenon in the context of a continuous time vector mixed autoregressive moving average process. The notation in this appendix does not quite match up with that in the text, but the analysis can be thought of as applying to the $\tilde{z}$ process discussed in the text. For sake of simplicity, tildes and subscripts have been suppressed. Also we adopt a slightly different normalization. More specifically.

(C.1) $$\theta(D)\, z_t = \psi(D)\, w_{zt}$$

where

$$\theta(D) = \theta_0 + \theta_1 D + \cdots + I\, D^{m_2}$$

$$\psi(D) = \psi_0 + \psi_1 D + \cdots + \psi_{m_3} D^{m_3}$$

$$E\, w_{zt} w'_{zt-u} = I\, \delta(t-u) \quad .$$

Both ψ and θ are $(p-n) \times (p-n)$ and the elements of w_z are linear combinations of the elements of the instantaneous forecast errors in forecasting z from its past. We impose the normalization that ψ_{m_3} is lower triangular. For the purposes of this discussion we make the additional assumption that $m_3 = m_2 - 1$. Let

$$\mathcal{R}(s) = E[z_t\, z'_{t-s}]$$

and

$$S(\zeta) = \int_{-\infty}^{+\infty} e^{-\zeta s}\, \mathcal{R}(s)\, ds \quad .$$

As noted in the text we know that

(C.2) $$S(\zeta) = \theta(\zeta)^{-1}\, \psi(\zeta)\, \psi(-\zeta)'\, \theta(-\zeta)^{-1\prime} \quad .$$

Let $\lambda_1, \ldots \lambda_{m_5}$ denote the zeroes of $\det \theta(\zeta)$ where $m_5 = (n-p)m_2$. Assuming these zeroes are distinct we let

$$W_j = - \frac{\operatorname{Adj} \theta(\lambda_j) \psi(\lambda_j) \psi(-\lambda_j)' \operatorname{Adj} \theta(-\lambda_j)'}{2\lambda_j \prod_{\substack{k=1 \\ k \neq j}}^{m_4} (\lambda_j - \lambda_k)(-\lambda_j - \lambda_k)} .$$

Expanding $S(\zeta)$ in matrix partial fractions we obtain

$$S(\zeta) = \sum_{j=1}^{m_5} \frac{W_j}{\zeta - \lambda_j} - \sum_{j=1}^{m_5} \frac{W_j'}{\zeta + \lambda_j} .$$

Following the reasoning in section 5 we have

$$\mathcal{R}(s) = \sum_{j=1}^{m_5} W_j e^{\lambda_j s} \qquad \text{for } s \geq 0 .$$

We are now in a position to address the aliasing phenomenon. Recall that $e^{2\pi i k} = 1$ for all integer k . Suppose that λ_1 is a complex root. Without loss of generality we can let λ_2 denote its complex conjugate. Now

$$e^{\lambda_1 s} = e^{(\lambda_1 + 2\pi i k)s}$$

and

$$e^{\lambda_2 s} = e^{(\lambda_2 - 2\pi i k)s}$$

for s sampled at the integers and for any integer k . Thus

$$\mathcal{R}^+(s) = W_1 e^{(\lambda_1 + 2\pi i k)s} + W_2 e^{(\lambda_2 - 2\pi i k)s} + \sum_{j=3}^{m_5} W_j e^{\lambda_j s}$$

is equal to $\mathcal{R}(s)$ for s sampled at the nonnegative integers. If we can construct a continuous time stochastic process z^+ that has $\mathcal{R}^+$ as its autocovariance function, then when sampled at the integers this stochastic process will have the same covariance properties as z sampled at the integers. It will not be possible to distinguish the parameters governing z^+ from the parameters governing z from discrete-time data. This is what is meant by the aliasing phenomenon.

We proceed to illustrate how to construct such a z^+ process using results from Phillips [21]. We are interested in constructing a $(p - n) \times (p - n)$ matrix polynomial

$$\theta^+(\zeta) = \theta_0^+ + \theta_1^+ \zeta + \cdots + I \zeta^{m_2} \quad .$$

This is accomplished by solving the following equations

$$\theta^+(\lambda_1 + 2\pi ik) W_1 = 0$$

$$\theta^+(\lambda_2 - 2\pi ik) W_2 = 0$$

$$\theta^+(\lambda_j) W_j = 0 \qquad \text{for } j = 3, \ldots, m_5 \quad .$$

Phillips [21] shows that the W_j's are of rank one. Thus these equations provide $(p - n)^2 m_2$ linearly independent equations in the $(p - n)^2 m_2$ elements of $\theta_0^+, \theta_1^+, \ldots, \theta_{m_2-1}^+$. Solving these equations yields the matrix polynomial $\theta^+(\zeta)$. Define

$$\psi^+(\zeta)\, \psi^+(-\zeta)' = \theta^+(\zeta) \left[\frac{W_1}{\zeta - \lambda_1 - 2\pi ik} - \frac{W_1'}{\zeta + \lambda_1 + 2\pi ik} + \frac{W_2}{\zeta - \lambda_2 + 2\pi ik} - \frac{W_2'}{\zeta + \lambda_2 - 2\pi ik} + \sum_{j=3}^{m_5} \frac{W_j}{\zeta - \lambda_j} - \sum_{j=3}^{m_5} \frac{W_j'}{\zeta + \lambda_j} \right] \theta^+(-\zeta)' \quad .$$

It can be verified that ψ^+ is a $(m_2 - 1)$ order polynomial and can be chosen so that $\psi^+_{m_2-1}$ is lower triangular. However, $\psi^+(\zeta)$ is not necessarily constrained to have real coefficients in its polynomial representation. For choices of k in which $\psi^{+}(\zeta)$ is a real polynomial we can generate a z^{+} process so that

$$\theta^{+}(D)\ z_t^{+} = \psi^{+}(D)\ w_t^{+} \quad .$$

This process has $\Re^{+}(s)$ as its autocovariance function. By choosing different values of k we can generate a family of parameters that are observationally equivalent from the standpoint of discrete-time data. The constraint that $\psi^{+}(\zeta)$ have real coefficients in its polynomial representation provides additional identifying information. This is analogous to results obtained in Hansen and Sargent [8]. There they show that restricting the continuous time innovation intensity matrix to be positive semidefinite is potentially important in identifying the parameters of a first-order stochastic differential equation from discrete-time data.

Footnotes

1. In what follows, the term "agent" should be interpreted liberally enough to include the fictitious social planner who computes the equilibrium of the model, as in Lucas [14].

2. Under uncertainty, i.e., with a positive definite V matrix, the criterion function (2.3) becomes unbounded as $T \to \infty$, under our assumptions. The decision rule that we compute is optimal for criterion function (2.3) under uncertainty, in the sense that it maximizes

$$\lim_{T\to\infty} \frac{1}{T} E_0 \int_0^T F(x_t, y_t, Dy_t, \ldots, D^{m_1} y_t, t)\, dt \quad .$$

3. The assumption that $E w_t w'_{t-u} = \delta(t-u) V e^{rt}$ is made with a view toward guaranteeing that the random process $(e^{-r/2t} y_t, e^{-r/2t} x_t)$ will be covariance stationary. Large parts of our results would go through under a weaker condition on $E w_t w'_{t-s}$. In particular, all of our results on the equilibrium decision rule would obtain under the very general assumption that $E w_t w'_{t-u} = \delta(t-u) V_t$ where V_t is any sequence of positive definite matrices. Also for many applications, the assumption that $E w_t w'_{t-u} = \delta(t-u) V$ would imply that the (y_t, x_t) process itself is covariance stationary. However, making (y_t, x_t) covariance stationary in addition requires stricter conditions on H_1 and $G(D)$ than we have imposed. Covariance stationarity is desirable because it facilitates estimation and underlies the procedures advocated in section 5.

4. As it stands now, (2.2) is overparameterized, in the sense that normalizations must be imposed on θ_0, ψ_0, and V to make the parameters of (2.2) identifiable from the continuous time spectral density of x. For example, $\theta_0 = I$, $\psi_0 = I$ is one workable normalization. At this point, we impose no particular normalization. All of our results up to section 4 will hold for any acceptable normalization, at which point we shall specialize the setup somewhat.

5. The matrix V also belongs in this list in general. However, the parameters of V can always be normalized, that is set entirely *a priori*, provided enough parameters of $\theta(D)$ and $\psi(D)$ are left free.

6. In the model where agents and the econometricians both operate on the basis of data sampled at the same equispaced discrete points in time, the parameters of the $\tilde{z}$ process are identified without taking into account the restrictions across the $\tilde{y}$ and $\tilde{z}$ processes. See Hansen and Sargent [9], [10].

References

[1] Bard, Y., 1974, Nonlinear parameter estimation (Academic Press, New York).

[2] Beltrami, E. J., and M. R. Wohlers, 1966, Distributions and the boundary values of analytic function (Academic Press, New York).

[3] Emre, E., and Ö. Hüseyin, 1975, Generalization of Leverrier's algorithm to polynomial matrices of arbitrary degree, IEEE Transactions on Automatic Control, 136.

[4] Geweke, J. B., 1978, Temporal aggregation in the multivariate regression model, Econometrica, 46, 643 - 662.

[5] Gould, J. P., 1968, Adjustment costs in the theory of investment of the firm, Review of Economic Studies, 35, 47 - 56.

[6] Granger, C. W. J., 1969, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37, 424 - 438.

[7] Hannan, E. J., 1970, Multiple time series (Wiley, New York).

[8] Hansen, L. P., and T. J. Sargent, 1979, Rational expectations and the aliasing phenomenon, manuscript.

[9] Hansen, L. P., and T. J. Sargent, 1980, Formulating and estimating dynamic linear rational expectations models, Journal of Economic Dynamics and Control, forthcoming.

[10] Hansen, L. P., and T. J. Sargent, 1980, Linear rational expectations models for dynamically interrelated variables, in R. E. Lucas, Jr., and T. J. Sargent, eds., Rational expectations and econometric practice (University of Minnesota Press, Minneapolis).

[11] Koopmans, L. H., 1974, The spectral analysis of time series (Academic Press, New York).

[12] Kwakernaak, H., and R. Sivan, 1972, Linear optimal control systems (Wiley, New York).

[13] Lucas, R. E., Jr., 1965, Distributed lags and optimal investment policy, manuscript.

[14] Lucas, R. E., Jr., 1966, Optimal investment with rational expectations, manuscript.

[15] Lucas, R. E., Jr., 1967, Adjustment costs and the theory of supply, Journal of Political Economy, 75, 321 - 334.

[16] Lucas, R. E., Jr., 1976, 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 (North Holland, Amsterdam).

[17] Lucas, R. E., Jr., and E. Prescott, 1971, Investment under uncertainty, Econometrica 39, 659 - 681.

[18] Luenberger, D. G., 1969, Optimization by vector space methods, (Wiley, New York).

[19] Mortensen, D. T., 1973, Generalized costs of adjustment and dynamic factor demand theory, Econometrica 41, 657 - 665.

[20] Phadke, M. S., and G. Kedem, 1978, Computation of the exact likelihood function of multivariate moving average models, Biometrika 65, 511 - 519.

[21] Phillips, A. W., 1959, The estimation of parameters in systems of stochastic differential equations, Biometrika 59, 67 - 76.

[22] Phillips, P. C. B., 1972, The structural estimation of a stochastic differential equation system, Econometrica 40, 1021 - 1041.

[23] Phillips, P. C. B., 1973, The problem of identification in finite parameter continuous time models, Journal of Econometrics 1, 351 - 362.

[24] Phillips, P. C. B., 1974, The estimation of some continuous time models, Econometrica 42, 803 - 824.

[25] Rozanov, Y. A., 1967, Stationary random processes (Holden-Day, San Francisco).

[26] Simon, H. A., 1956, Dynamic programming under uncertainty with a quadratic criterion function, Econometrica 24, 74 - 81.

[27] Sims, C. A., 1971, Discrete approximations to continuous time lag distributions in econometrics, Econometrica 39, 545 - 564.

[28] Theil, H., 1964, Optimal decision rules for government and industry (North Holland, Amsterdam).

[29] Treadway, A. B., 1969, On rational entrepreneur behavior and the demand for investment, Review of Economic Studies 36, 227 - 240.

[30] Treadway, A. B., 1970, Adjustment costs and variable inputs in the theory of the competitive firm, Journal of Economic Theory 2, 329 - 347.

[31] Vaughan, D. R., 1969, A negative-exponential solution to the matrix Riccati equation, IEEE Transactions on Automatic Control 14; 72 - 75.

[32] Wymer, C. R., 1972, Econometric estimation of stochastic differential equation systems, Econometrica 40, 565 - 577.