a set of probability measure one. Thus, we have proved that $\{y_{n_k} : \geq 1\}$ converges to y_o in L^2.

To complete this proof, we must show that $\{y_n : n \geq 1\}$ converges y_o in L^2. Given $\delta > 0$, there is k such that

$$E[(y_m - y_{n_k})^2]^{\frac{1}{2}} < \delta/2 \tag{1.16}$$

r all $m \geq n_k$ and

$$E[(y_{n_k} - y_o)^2]^{\frac{1}{2}} < \delta/2 \,. \tag{1.17}$$

the Triangle Inequality,

$$E[(y_m - y_o)^2]^{\frac{1}{2}} < \delta \text{ for } m \geq n_k \,. \quad \blacksquare \tag{1.18}$$

Notes

. We follow the usual convention of viewing the equivalence class of random variables that are equal almost surely as a unique random variable.

This notion of convergence is induced by the norm, $||y|| = E(y^2)^{\frac{1}{2}}$ on the linear space L^2.

The definition of continuity given here only considers the behavior of the function π in the neighborhood of zero. However, since c is linear, continuity at a single point immediately translates into continuity at all points.

This lemma is a special case of the Riesz Representation Theorem.

We take the notation $E(z|x)$ to mean the expectation conditioned on the sigma algebra generated by x.

The unique extension of measures from algebras to sigma algebras follows from the Carathéodory Theorem. See Halmos (1950) or Royden (1968) for a discussion of this result.

Recall that a projection is defined only up to an equivalence class of random variables that are almost surely equal. Some (but not necessarily all) members of this equivalence class will be measurable with respect to B_o.

By $P[x(t)|H(t-1)]$ we mean the vector containing the projections of components $x(t)$ onto $H(t-1)$.

3

Exact Linear Rational Expectations Models: Specification and Estimation

by Lars Peter HANSEN and Thomas J. SARGENT

Introduction

A distinguishing characteristic of econometric models that incorporate rational expectations is the presence of restrictions across the parameters of different equations. These restrictions emerge because people's decisions are supposed to depend on the stochastic environment which they confront. Consequently, equations describing variables affected by people's decisions inherit parameters from the equations that describe the environment. As it turns out, even for models that are linear in the variables, these cross-equation restrictions on the parameters can be complicated and often highly nonlinear.

This paper proposes a method for conveniently characterizing cross-equation restrictions in a class of linear rational expectations models, and also indicates how to estimate statistical representations satisfying these restrictions. For most of the paper, we restrict ourselves to models in which there is an *exact* linear relation across forecasts of future values of one set of variables and current and past values of some other set of variables. The key requirement is that all of the variables entering this relation must be observed by the econometrician. While probably only a minority of rational expectations models belong to this class, it does contain interesting models that have been advanced to study the term structure of interest rates, stock prices, consumption and permanent income, the dynamic demand for factors of production, and many other subjects.

It is useful to compare the class of exact models with the class studied by Hansen and Sargent (1980a). The differences lie entirely in the interpretations of the *error terms* in the equations that are permitted. In Hansen and Sargent (1980a), random processes which the econometrician treats as disturbances in decision rules can have a variety of sources. Disturbance terms can be interpreted as reflecting shocks to technologies or preferences observed by private agents but not by the econometrician. Disturbances can also be interpreted as reflecting interactions with *hidden* decision variables which are simultaneously chosen by private agents but unobserved by the econometrician. Finally, disturbances can be interpreted, along the lines of Shiller (1972), as reflecting the phenomenon that in forecasting the future, private agents use larger information sets than the econometrician can consider because of data limitations. Of these alternative interpretations of error terms, only the last one can be accommodated within the class of exact models of the present paper. While this limitation on the permissible interpretations of error terms excludes many rational expectations models, a variety of interesting examples still remains within the general class of exact linear rational expectations models.

In linear rational expectations models, the cross-equation restrictions can be characterized very conveniently by working in terms of a vector moving-average representation for the variables being modeled. By straightforward applications of the Wiener-Kolmogorov least squares prediction formulas, these restrictions can readily be deduced. Once the restrictions are deduced, the parameters of the model can be estimated by maximizing one of various approximations to the likelihood function. The vector moving-average representation that incorporates the rational expectations restrictions is nested within less constrained vector moving-average representations, and a likelihood ratio statistic can be computed to test the model.

The ease of characterizing the restrictions and calculating estimates is a great virtue of specifying the model in vector moving-average form. However, an identification question must be addressed before this strategy can be implemented. Without *a priori* restrictions on their parameters, many vector moving-average representations are consistent with a given set of second moments. A natural and practically important question is whether the cross-equation rational expectations restrictions provide enough prior information to identify a unique moving-average representation. For the case of exact linear rational expectations models, Section 3 provides results that characterize identification. Insofar as the identification question is concerned, there are substantial differences between exact rational expectations models and models that admit one or more of the additional interpretations of the error terms described above. It is the special nature of the identification problem in these exact models, and not anything special about the appropriate

methods either of representing the models or of estimating them, that causes us to restrict this paper mainly to analyzing exact linear rational expectations models. In Section 6, we briefly indicate how both our methods for model specification and estimation carry over to *inexact* linear rational expectations models.

As Shiller (1979) and Hansen and Hodrick (1980) have indicated for several special examples of our general model, it is possible to devise powerful tests of such models without estimating the complete vector process subject to the model's restrictions. However, for many applications, the analyst wants more than just a test of the model, and desires a complete representation of the vector process. Indeed, our interest in the identification and estimation of constrained moving-average models is not entirely motivated by the exact linear rational expectations models that occupy most of our attention in the present paper. As we indicate in Section 6, the restrictions that emerge in the present models strongly resemble those that characterize rational expectations models which can accommodate additional interpretations of disturbance terms (e.g. Hansen and Sargent 1980a). This makes constrained moving-average estimation a more generally useful method for estimating the parameters needed to overcome Lucas's (1976) critique of econometric policy evaluation procedures.

1. General Model

In this section we specify a general time series econometric model and consider representations of solutions to that model. We begin by specifying in turn the information of economic agents, the information set of the econometrician, and the economic model.

Economic Agents' Information Set

Let J^+ denote the set of nonnegative integers, J the set of all integers, and $x \equiv \{x_t : t \in J\}$ be a p-dimensional, vector stochastic process that is covariance stationary and has finite second moments. For simplicity, we assume that $Ex_t = 0$. The time t common information set of economic agents, denoted Ω_t, is the set of all random variables with finite second moments that are (Borel measurable) functions of current and past values of x. In making decisions at time t, economic agents are assumed to forecast optimally conditioned on Ω_t. Let $\mathrm{P}(\cdot|\Omega_t)$ denote the projection operator that maps random variables with finite second moments into optimal forecasts based on Ω_t. This operator can be interpreted as an expectation operator conditioned on current and past values of x.

Econometrician's Information Set

The econometrician is assumed to observe only the first q components of the x process. We denote the resulting process $y \equiv \{y_t : t \in J\}$. Let Σ_t denote the closed (in mean-square) information set generated by taking linear combinations of current and past values of y. Instead of forecasting using conditional expectations, the econometrician uses the linear least squares projections onto Σ_t. In general, forecast errors using $P(\cdot|\Omega_t)$ will have smaller second moments than forecast errors using $P(\cdot|\Sigma_t)$.

We assume that y is linearly regular and has full rank. These assumptions are sufficient to guarantee that elements of Σ_t can be represented as

$$\sum_{j=0}^{\infty} \alpha_j \cdot v_{t-j} \tag{1.1}$$

where $v \equiv \{v_t : t \in J\}$ is a serially uncorrelated q-dimensional process satisfying $Ev_t = 0$, $Ev_t v_t' = I$ where the entries of v_t are in Σ_t. Also, $\alpha \equiv \{\alpha_j : j \in J\}$ is a sequence in $\mathbf{R}^q$ satisfying

$$\sum_{j=0}^{\infty} \alpha_j \cdot \alpha_j < \infty . \tag{1.2}$$

The v process is said to be *fundamental* for the y process and v_t is the new uncorrelated information that is added to Σ_{t-1} in forming Σ_t. Since $P(v_{t+1}|\Sigma_t) = 0$, it is more convenient to represent Σ_t in terms of current and past values of v instead of current and past values of y.

We assume that the econometrician uses a subset of the information observed by economic agents and that he calculates linear least squares projections in making these forecasts. These are convenient assumptions for studying many dynamic economic models using time series methods.

Economic Model

Our general model is of the form:

$$P\Big\{[A(L)\ ;\ B(L^{-1})]y_t|\Omega_{t-\ell}\Big\} = 0 \tag{1.3}$$

where

$$A(z) = A_n(z)/A_d(z)\ , \tag{1.4}$$

$$B(z) = B_n(z)/B_d(z)\ , \tag{1.5}$$

$A_n(z)$ is an $(r \times r)$ matrix polynomial with a determinant that has zeros outside the unit circle of the complex plane, $B_n(z)$ is an $[r \times (q-r)]$ matrix polynomial, and $A_d(z)$ and $B_d(z)$ are scalar polynomials with zeros outside the unit circle of the complex plane.

Solution

Let Λ_t be generated by current and past values of a serially uncorrelated, q-dimensional process $w \equiv \{w_t : t \in J\}$ where $Ew_t = 0$ and $Ew_t w_t' = I$. Clearly, $P(w_{t+1}|\Lambda_t) = 0$ and the entries of w_t are in Λ_t. We assume that solutions to the model can be represented as time invariant linear functions of current and past values of w:

$$y_t = C(L)w_t = \begin{bmatrix} C_1(L) \\ C_2(L) \end{bmatrix} w_t \tag{1.6}$$

where

$$C(z) \equiv \sum_{j=0}^{\infty} c_j z^j \quad \text{and} \quad \sum_{j=0}^{\infty} \operatorname{trace}(c_j c_j') < \infty . \tag{1.7}$$

Partition C_1 is $(r \times q)$ and partition C_2 is $[(q-r) \times q]$.

One possible candidate for w is $w = v$, in which case $\Sigma_t = \Lambda_t$ for all t. We do not restrict ourselves to this specification and instead allow Λ_t to be strictly larger than Σ_t. For the purposes of solving the model, we take Λ_t to be a subset of Ω_t. Since some of the components of Ω_t may not be observed by the econometrician, it is possible for Λ_t to be strictly larger than Σ_t.

Applying the Law of Iterated Projections to (1.3) gives

$$P\Big\{[A(L)\ ;\ B(L^{-1})]y_t|\Lambda_{t-\ell}\Big\} = 0 \tag{1.8}$$

or equivalently

$$P\Big\{[A(L)\ ;\ B(L^{-1})]y_t|\Lambda_t\Big\} = D_1(L)w_t \tag{1.9}$$

where $D_1(z)$ is an $(r \times q)$ matrix polynomial with degree $\ell - 1$. When $\ell = 0$, $D_1(z) = 0$. Also, (1.6) implies that

$$[0\ ;\ I]\, y_t = D_2(L)w_t \tag{1.10}$$

where $D_2(z) = C_2(z)$. Our goal is to solve for $C(z)$ in terms of the triple $[A(z),\, B(z),\, D(z)]$ where

$$D(z) = \begin{bmatrix} D_1(z) \\ D_2(z) \end{bmatrix} . \tag{1.11}$$

Relation (1.9) implies that

$$\left\{ [A(z) \; ; \; B(z^{-1})] \begin{bmatrix} C_1(z) \\ C_2(z) \end{bmatrix} \right\}_+ = D_1(z) , \tag{1.12}$$

or equivalently

$$A(z)C_1(z) + [B(z^{-1})C_2(z)]_+ = D_1(z) \tag{1.13}$$

where $[\cdot]_+$ is the annihilation operator that is defined as follows. The matrix $B(z^{-1})C_2(z)$ is well defined for $\rho < |z| < 1$ for some $\rho < 1$. In particular, ρ can be chosen so that ρ^{-1} is the modulus of the smallest zero of $B_d(z)$. The matrix function $B(z^{-1})C_2(z)$ has a two-sided Laurent series expansion in the region $\{z : \rho < |z| < 1\}$. Then $[B(z^{-1})C_2(z)]_+$ is the function defined by the power series expansion where the negative powers of z are ignored. Negative powers of z correspond to future values of w_t. Hence, the projection of these terms onto Σ_t is zero. Solving for $C_1(z)$ gives

$$\begin{aligned} C_1(z) &= A(z)^{-1}\{D_1(z) - [B(z^{-1})D_2(z)]_+\} \\ C_2(z) &= D_2(z) . \end{aligned} \tag{1.14}$$

Equation (1.14) can be viewed as linear transformations mapping the matrix function $D(z)$ into the matrix function $C(z)$. Throughout this paper we will think of alternative solutions to (1.8) as being indexed by alternative choices of $D(z)$. The only restriction we have placed on $D(z)$ is that $D_1(z)$ be a polynomial with degree $\ell - 1$, and that $D_2(z)$ have a power series expansion with square summable coefficients.

To obtain an alternative convenient representation of $C(z)$, we follow Hansen and Sargent (1980a) by characterizing the annihilation operator as applied to the function $B(z^{-1})D_2(z)$ in terms of properties of this function inside the unit disk of the complex plane. The functions $D_1(z)$ and $D_2(z)$ are both analytic inside the unit disk, and the function $B(z^{-1})$ is analytic everywhere in this same region except at the zeros of $B_d(z^{-1})$ and possibly at $z = 0$. The zeros of $B_d(z)$ are all outside the unit circle of the complex plane, implying that the zeros

of $B_d(z^{-1})$ are all inside the unit circle. Since $B_d(z)$ is a finite-order polynomial, $B_d(z^{-1})$ has a finite number of zeros. Hence $B(z^{-1})D_2(z)$ is analytic everywhere inside the unit disk except at a finite number of points. At each of the points for which it is not analytic, $B(z^{-1})D_2(z)$ has a finite-order pole.

To compute $[B(z^{-1})D_2(z)]_+$, we take a *partial fractions decomposition* whereby we form the principal parts of the Laurent series expansion about each of the poles and subtract the sum of these principal parts from $B(z^{-1})D_2(z)$. Let z_k denote the k^{th} pole and $G_k(z^{-1})$ denote the principal part of the Laurent series expansion of $B(z^{-1})D_2(z)$ about z_k. Recall that the principal part, $G_k(z^{-1})$, of a Laurent series expansion about z_k consists only of the terms $(z - z_k)^j$ for strictly negative powers of j. Since z_k is a finite-order pole, $G_k(z^{-1})$ is the sum of only a finite number of such terms. Then $B(z^{-1})D_2(z) - G(z^{-1})$ is analytic inside the unit circle of the complex plane where

$$G(z^{-1}) \equiv \sum_k G_k(z^{-1}) . \tag{1.15}$$

In other words, $B(z^{-1})D_2(z) - G(z^{-1})$ has a power series expansion that is convergent in the unit disk of the complex plane.

By construction, $G(z)$ satisfies the following restriction:

Restriction R1: $G(z)$ is an $(r \times q)$ function that is analytic in an open set containing $\{z : |z| \leq 1\}$ and $G(0) = 0$.

For any function G that satisfies *R1*, G has a power series expansion with a leading coefficient zero and a radius of convergence that exceeds one. Consequently, in a domain containing $\{z : |z| \geq 1\}$, $G(z^{-1})$ can be represented as

$$G(z^{-1}) = \sum_{j=1}^{\infty} g_j z^{-j} . \tag{1.16}$$

Since $B(z^{-1})D_2(z) - G(z^{-1})$ is analytic in the unit disk and $G(z)$ satisfies *R1*,

$$[B(z^{-1})D_2(z)]_+ = B(z^{-1})D_2(z) - G(z^{-1}) . \tag{1.17}$$

In summary, one convenient way to compute the left side of (1.14) is first to compute $G(z^{-1})$ and then subtract it from $D_1(z) - B(z^{-1})D_2(z)$. Substituting into (1.14), we have that

$$\begin{aligned} C_1(z) &= A(z)^{-1}[D_1(z) - B(z^{-1})D_2(z) + G(z^{-1})] \\ C_2(z) &= D_2(z) . \end{aligned} \tag{1.18}$$

Equation (1.18) implies that

$$[A(z);\ B(z^{-1})]\,C(z) = z^{\ell} H(z^{-1})\ , \tag{1.19}$$

where

$$H(z^{-1}) \equiv z^{-\ell}\,[D_1(z) + G(z^{-1})]\ . \tag{1.20}$$

Then $H(z)$ satisfies *R1* by construction. Furthermore, for any $C(z)$ satisfying (1.19) for some $H(z)$ that satisfies *R1*, there exists an *admissible* $D(z)$ such that (1.14) is satisfied where an admissible $D(z)$ is one with square summable coefficients and an upper $(r \times q)$ partition that is a matrix polynomial with degree $\ell - 1$. In accordance with relation (1.20),

$$D_1(z) = \sum_{j=1}^{\ell} d_j\, z^j\ , \tag{1.21}$$

and in accordance with (1.14),

$$D_2(z) = C_2(z)\ . \tag{1.22}$$

Therefore, an equivalent characterization to $C(z)$ satisfying (1.14) for some admissible $D(z)$ is that $C(z)$ satisfies (1.19) for some $H(z)$ satisfying *R1*. Characterization (1.14) has the advantage that the solution is explicitly parameterized by $D(z)$. However, characterization (1.19) also will turn out to also be useful in Section 3 when we investigate issues pertaining to identification of $D(z)$ and hence $C(z)$.

In checking (1.19), it suffices to verify that $H(z)$ is analytic in the region $\{z : |z| < 1\}$ rather than in an open set containing $\{z : |z| \leq 1\}$ because the left side of (1.19) is analytic in the region $\{z : \rho < |z| < 1\}$. Consequently, when $H(z)$ is analytic inside the unit circle of the complex plane, it is also analytic in the larger set $\{z : |z| < (1/\rho)\}$.

2. Examples

In this section we consider five illustrations of the general model specified and solved in section one.

Example 1: Lognormal Model of Bond Pricing

An implication of the intertemporal asset pricing model as studied by LeRoy (1973), Rubinstein (1976), Lucas (1978), Breeden (1979), and Cox, Ingersoll and Ross (1985), among others, is that the price of a pure discount bond that pays one dollar n time periods into the future satisfies:

$$\exp(p_t^n) = \mathrm{P}\,[\exp(m_{t+n} - m_t)|\Omega_t] \tag{2.1}$$

where $\exp\,(p_t^n)$ is the bond price, $\exp\,(m_t)$ is the indirect marginal utility of money, and Ω_t is the information set available to economic agents, all at date t. Following Hansen and Singleton (1983), Hansen and Hodrick (1983), Campbell (1986), Harvey (1988), and Hansen and Singleton (1990), suppose that $m_t - m_{t-1}$ is one component of x_t where x is a stationary Gaussian process.[1] Then (2.1) implies that

$$\begin{aligned} p_t^n &= \mathrm{P}\,[(m_{t+n} - m_t)|\Omega_t] + \mathbf{c}_n \\ &= \mathrm{P}\,[(m_{t+1} - m_t) + (m_{t+2} - m_{t+1}) + \ldots \\ &\quad + (m_{t+n} - m_{t+n-1})|\Omega_t] + \mathbf{c}_n \end{aligned} \tag{2.2}$$

where $\mathbf{c}_n$ is equal to one-half the conditional variance of $m_{t+n} - m_t$. Abstracting from the constant term, relation (2.2) is a special case of the general model described in section one with $y_{1t} = p_t^n$; the first entry of y_{2t} equal to $m_t - m_{t-1}$ and the remaining entries being variables that are observed by the econometrician and are potentially useful in forecasting future values of $m_t - m_{t-1}$; $A(z) = 1$, $B(z) = (z + z^2 + \ldots + z^n)\,[1;\ 0]$, and $\ell = 0$. Therefore, $C(z)$ satisfies

$$\begin{aligned} C_1(z) &= -[1;\ 0]\ [(z^{-1} + z^{-2} + \ldots + z^{-n})\, D_2(z)]_+ \\ C_2(z) &= D_2(z)\ . \end{aligned} \tag{2.3}$$

Next suppose that the indirect marginal utility for money is not observable to an econometrician. Instead observations are available on the price of a one-period discount bond. Setting $n = 1$ in (2.2) results in

$$p_t^1 = \mathrm{P}\,[(m_{t+1} - m_t)|\Omega_t] + \mathbf{c}_1\ . \tag{2.4}$$

Applying the Law of Iterated Projections (see Lemma 3.5 in Chapter 2) to (2.2) and substituting from (2.4), it follows that

$$p_t^n = \mathrm{P}\,[(p_t^1 + p_{t+1}^1 + \ldots + p_{t+n-1}^1)|\Omega_t] + \mathbf{c}_n - n\mathbf{c}_1\ . \tag{2.5}$$

Again abstracting from the constant term, relation (2.5) is also in the form of the general model presented in section one where now the first entry of y_{2t} is p_t^1 and $B(z) = (1 + z + \ldots + z^{n-1})\,[1;\ 0]$.

Finally, replacing $-p_t^n/n$ in (2.5) by the one-period rate of return on an n-period discount bond and p_t^1 by the one-period rate of return on a one-period bond, one obtains Sargent's (1979) version of a rational expectations model of the term structure of interest rates.

Example 2: Present-Value Model

Let d_t denote the time t dividend or payout, and p_t the time t value to owning a claim to this dividend process from time t forward. Suppose these two variables are related via a present-value formula

$$p_t = \mathrm{P}\,(\sum_{j=0}^{\infty} \lambda^\tau\, d_{t+\tau} \mid \Omega_t)\ , \tag{2.6}$$

where λ is a constant discount factor.[2] In LeRoy and Porter (1981) and Shiller (1981), d_t is the dividend paid to share-holders of a stock and p_t is the price of the stock. In Hamilton and Flavin (1986) and Roberds (Chapter 6), d_t is the net surplus of a government (taxes minus expenditures) and p_t is government debt.

As noted in Campbell and Shiller (1987), a present-value model of the form (2.6) is a special case of an exact linear rational expectations model as defined in section one. To see this, let $y_{1t} = p_t$, d_t be the first entry of y_{2t}, $A(z) = 1$, $B(z) = -1/(1-\lambda z)\,[1;\,0]$ and $\ell = 0$. The solution to the model is

$$\begin{aligned} C_1(z) &= [1;\,0]\,[zD_2(z) - \lambda D_2(\lambda)]/(z-\lambda) \\ C_2(z) &= D_2(z)\ . \end{aligned} \tag{2.7}$$

Example 3: Permanent Income Model of Consumption

Following Hall (1978), Flavin (1981), Hansen (1987) and Sargent (1987b), we consider a rational expectations version of the permanent income model of the form:

$$c_t - \rho a_t = (1-\lambda)\ \mathrm{P}\,(\sum_{j=0}^{\infty} \lambda^j\, e_{t+j} \mid \Omega_t) \tag{2.8}$$

where a_t is the asset-holding, c_t is consumption and e_t is the endowment or labor income of a consumer at time t. The parameters ρ and λ are related via $\lambda = 1/(1+\rho)$. This model has a built-in stochastic singularity because of the resource constraint:

$$c_t + a_{t+1} - a_t = \rho a_t + e_t\ . \tag{2.9}$$

For this reason, instead of treating a, c and e as distinct processes, we investigate implications for $c_t - \rho a_t$ and e_t. Expressed in terms of these transformed process, the model in equation (2.8) is essentially the same as the present-value model.

The permanent income model is known to have the property that $c_{t+1} - c_t$ is uncorrelated with y_{t-j} for $j \in J^+$. To verify this property, note that

$$\begin{aligned} c_{t+1} - c_t &= (c_{t+1} - \rho a_{t+1}) - (c_t - \rho a_t) + \rho(a_{t+1} - a_t) \\ &= (c_{t+1} - \rho a_{t+1}) - (c_t - \rho a_t) - \rho(c_t - \rho a_t) + \rho e_t\ . \\ &= (L^{-1} - \lambda^{-1})\,(c_t - \rho a_t) + \rho e_t\ . \\ &= \left[\lambda^{-1}(\lambda - L)\,C_1(L) + \rho[1;\ 0]\,LC_2(L)\right] w_{t+1} \\ &= -(1-\lambda)\ [1;\ 0]\,D_2(\lambda) w_{t+1}\ . \end{aligned} \tag{2.10}$$

The implications of (2.10) are investigated further in Chapter 5.

Example 4: Martingale Model with Temporal Aggregation

As noted in Hansen and Singleton (1983, 1990), Grossman, Melino and Shiller (1987) and Hall (1988), several intertemporal asset pricing models have the implication that some linear combination of a vector of variables observed by an econometrician is a martingale difference sequence. Denote this linear combination by $\gamma \cdot y_t$ where $\gamma' = [\gamma_1\, \gamma_2']$ is a q-dimensional vector and γ_1 is different from zero. Hence

$$\mathrm{P}\,(\gamma \cdot y_t | \Omega_{t-1}) = 0\ . \tag{2.11}$$

This model becomes a special case of the general model presented in section one by letting $A(z)$ be a scalar given by the first entry of γ, $B(z)$ be a $(q-1)$-dimensional row vector given by the remaining $(q-1)$-dimensional subvector of γ', and ℓ be one.

Instead of observing y_t directly, suppose that an econometrician observes a temporally-aggregated version of y_t. As in Grossman, Melino and Shiller (1987), Hall (1988), and Hansen and Singleton (1990), suppose that

$$\gamma \cdot y_t = \int_0^1 \alpha(\tau) \cdot dW(t-\tau) \tag{2.12}$$

where W is a continuous-time, vector martingale with stationary increments, and α is vector-valued, continuous function on $[0,\,1]$. Relation (2.12) clearly implies that the discrete time process $\gamma \cdot y_t$ is a martingale difference sequence. Let y_t^a denote the temporally aggregated version of y_t. Then

$$\gamma \cdot y_t^a = \int_0^1 \left[\int_0^1 \alpha(\tau) dW(t-\tau-s) \right] ds \, . \tag{2.13}$$

As emphasized by Hall (1988), it follows from (2.13) that

$$\mathrm{P}\,(\gamma \cdot y_t^a \mid \Omega_{t-2}) = 0 \, . \tag{2.14}$$

In other words, temporal aggregation has the effect of shifting the information set back one additional time period so that ℓ is set to two. Thus $C(z)$ for the model with temporal aggregation is given by

$$\begin{aligned} C_1(z) &= (\gamma_1)^{-1} [D_1(z) - \gamma_2' D_2(z)] \\ C_2(z) &= D_2(z) \end{aligned} \tag{2.15}$$

where $D_1(z)$ is a first-order polynomial.

Example 5: Dynamic Demand Function for a Factor of Production

Sargent (1978) and Kennan (1979) have studied linear demand functions for factors of production that are derived from optimizing a quadratic objective function subject to linear constraints. We focus on Sargent's version, though a reinterpretation of the variables will yield Kennan's model as well. Assuming that there are no shocks to the technology and a single factor of production, the demand function turns out to be

$$n_t = \delta n_{t-1} - \theta\,\mathrm{P}\left[\sum_{j=0}^{\infty} (\beta\delta)^r p_{t+j} \mid \Omega_t\right] \tag{2.16}$$

where $0 < \beta < 1$, $0 < \delta < 1$, $\theta > 0$, n_t is the amount of the factor demanded at date t and p_t is the rental rate of the factor at date t. This model is of the form described in section one with $y_{1t} = n_t$, the first entry of y_{2t} equal to p_t, $A(z) = (1 - \delta z)$, $B(z) = 1/(1 - \beta\delta z)\,[\theta;\ 0]$ and $\ell = 0$. Then $C(z)$ satisfies:

$$\begin{aligned} C_1(z) &= [-\theta;\ 0]\ [zD_2(z) - \beta\delta D_2(\beta\delta)]/[(1-\delta z)\,(z - \beta\delta)]; \\ C_2(z) &= D_2(z) \, . \end{aligned} \tag{2.17}$$

Multiple factor versions of this example can be constructed easily along the lines of Hansen and Sargent (1981a) and Kollintzas (1985).

The list of examples could be extended to incorporate linear rational expectations models which have been used to study a wide variety of macroeconomic and microeconomic phenomena. The preceding examples are sufficient to illustrate the variety of *exact* linear rational expectations models.

3. Identification

As in section one, we study models with vector moving-average representations

$$y_t = C(L) w_t \tag{3.1}$$

where y is a q-dimensional covariance stationary process and $C(z)$ satisfies (1.19) for some $H(z)$ satisfying *R1*. In other words, $C(z)$ satisfies the cross-equation restrictions implied by the underlying economic model. Since the coefficients of the power series expansion of $C(z)$ are square summable, $C(z)$ can be extended from the interior of the unit disk to the unit circle by letting

$$C[\exp(-i\theta)] = \sum_{j=0}^{\infty} c_j \exp(-i\theta j) \tag{3.2}$$

be defined almost everywhere on $[-\pi, \pi]$ as a matrix of mean-square limits.[3] The spectral density of y is given by

$$S(\theta) \equiv C[\exp(-i\theta)] C[\exp(i\theta)]' \tag{3.3}$$

where the prime denotes transposition (but not complex conjugation). The spectral density generates the autocovariances of y via the formula:

$$E(y_t y_{t-\tau}') = (1/2\pi) \int_{-\pi}^{\pi} \exp(i\theta\tau) S(\theta) d\theta \, . \tag{3.4}$$

Taken together, formulas (3.3) and (3.4) give the autocovariances of the process y as a function of $C(z)$ used in representing y as a moving average of a vector white noise.

Without constraining C, it is known that there are multiple choices of C that imply the same spectral density and hence the same sequence of autocovariances. Loosely put, there is an equivalence class of matrix functions C which can be generated one from another by post-multiplying an original C by functional counterparts to orthogonal matrices. It is not possible to distinguish members of this equivalence class on the basis of the implied autocovariances of the time series. In other words, there is an identification problem in representing y as a one-sided moving-average of a vector white noise of disturbances.

In many circumstances, especially in problems involving prediction, this identification problem is resolved in part by constructing a moving-average representation in terms of a process v that is fundamental for y:

$$y_t = F(L)v_t \ . \tag{3.5}$$

We refer to (3.5) as a *fundamental moving-average representation.* Given one fundamental moving-average representation characterized by a function $F(z)$, all other fundamental representations are obtained by post-multiplying F by orthogonal matrices U, i.e. by matrices of real numbers that satisfy $UU' = I$.

Once we relax the restriction that the moving-average representation be fundamental, there is a much larger class of observationally equivalent moving-average representations. Again these moving-average representations are obtained by post-multiplying $F(z)$ by $U(z)$ except that now $U(z)$ can be an orthogonal matrix function, an object that we now define.

Definition: $U(z)$ is said to be a q-dimensional ***orthogonal matrix function*** if $U(z)$ is a $(q \times q)$ matrix function of a complex variable with a power series expansion $U(z) = \sum_{j=0}^{\infty} u_j z^j$ satisfying

(i) the entries of u_j are real numbers for $j \in J^+$;
(ii) $\sum_{j=0}^{\infty} \operatorname{trace}(u_j u_j') < \infty$;
(iii) $U[\exp(-i\theta)]\, U[\exp(i\theta)]' = I$ almost everywhere.

Given a matrix function $F(z)$ determining a fundamental moving-average representation and a q-dimensional orthogonal matrix function $U(z)$, we can construct other moving-average representations via the formula:

$$C(z) = F(z)U(z) \tag{3.6}$$

[e.g. see Rozanov (1967), page 62]. In light of (i), the coefficient matrices of the power series expansion of C have entries that are real numbers. In light of (ii) and (iii), the boundary values of C on the unit circle satisfy:

$$\begin{aligned} &C[\exp(-i\theta)]\, C[\exp(+i\theta)]' = \\ &\qquad F[\exp(-i\theta)]\, U[\exp(-i\theta)]\, U[\exp(i\theta)]'\, F[\exp(i\theta)]' \\ &\qquad = F[\exp(-i\theta)]\, F[\exp(i\theta)]' \ . \end{aligned} \tag{3.7}$$

Consequently, it follows from (3.3) and (3.4) that the spectral density function and autocovariance sequence implied by $C(z)$ is identical to that implied by $F(z)$. As a result, for a given fundamental representation (3.5), we can construct alternative observationally equivalent moving-average representations of the form (3.1) by selecting a q-dimensional orthogonal matrix function $U(z)$, forming $C(z)$ as in (3.6), and forming w_t via

$$w_t = U(L^{-1})' v_t \tag{3.8}$$

where

$$U(L^{-1})' = \sum_{j=0}^{\infty} u_j' L^{-j} \ . \tag{3.9}$$

Note that (iii) implies that $U(L^{-1})' = U(L)^{-1}$. From (3.8) it follows that $v_t = U(L)w_t$, so that v_t depends only on current and past values of w_t. Therefore, the linear space Λ_t generated by current and past values of w_t is at least as large as the linear space Σ_t generated by current and past values of v_t. Whenever u_j is different from zero for some $j \geq 1$, Λ_t is strictly larger than Σ_t.

Consider the class of $C(z)$'s constructed via (3.6) and indexed by alternative q-dimensional orthogonal functions $U(z)$. Suppose that at least one member of this class, say $\hat{C}(z)$, satisfies (1.19) for some $\hat{H}(z)$ satisfying *R1*. Not all of the remaining members of this class of $C(z)$'s will necessarily satisfy the cross-equation restrictions. Hence the cross-equation restrictions go at least part of the way in reducing the class of observationally equivalent moving-average representations. The remainder of this section studies the extent of this reduction.

Put somewhat differently, we are interested in characterizing the class of observationally equivalent $D(z)$'s. Formula (1.14) describes the mapping from *admissible* $D(z)$'s, i.e. $D(z)$'s for which $D_1(z)$ is a polynomial with degree $\ell - 1$, to the class of $C(z)$'s that satisfy the restrictions. Formulas (3.3) and (3.4) delineate the mapping from the class of $C(z)$'s used in moving-average representations for y to the family of spectral density functions and autocovariance sequences for y. Taken together, these formulas give a mapping from the space of admissible $D(z)$'s to the space of admissible autocovariances for y. The identification question we analyze pertains to the inverse of this mapping. What is the class of admissible $D(z)$'s associated with a given admissible autocovariance sequence?

Since we focus only on admissible autocovariance sequences, there exists at least one admissible $D(z)$ that generates this sequence, say $\hat{D}(z)$. Corresponding to this $\hat{D}(z)$ is a $\hat{C}(z)$ and an $\hat{H}(z)$ given by formula (1.20). Consider any other $C(z) = F(z)U(z)$ for some q-dimensional orthogonal matrix function $U(z)$. The matrix functions $C(z)$ and $\hat{C}(z)$ are related via:

$$C(z) = \hat{C}(z)\,\hat{U}(z^{-1})'\,U(z) \,. \tag{3.10}$$

Since $\hat{C}(z)$ and $\hat{H}(z)$ satisfy equation (1.19),

$$[A(z);\ B(z^{-1})]\,\hat{C}(z) = z^{\ell}\,\hat{H}(z^{-1}) \,. \tag{3.11}$$

Post-multiply both sides of (3.11) by $\hat{U}(z^{-1})'\,U(z)$ to obtain

$$[A(z);\ B(z^{-1})]\,C(z) = z^{\ell}\hat{H}(z^{-1})\,\hat{U}(z^{-1})'\,U(z) \,. \tag{3.12}$$

Therefore, $C(z)$ satisfies the restrictions if and only if:

Restriction R2: $\hat{H}(z)\,\hat{U}(z)'\,U(z^{-1})$ is analytic on the domain $\{z : |z| < 1\}$.

We have established:

Lemma 1: Suppose that $C(z) = F(z)\,U(z)$ for some q-dimensional orthogonal matrix function $U(z)$. Then $C(z)$ satisfies (1.14) for some admissible $D(z)$ if, and only if $U(z)$ satisfies *R2*.

Since $\hat{H}(z)$ is analytic inside the unit circle of the complex plane, one convenient sufficient condition for *R2* is:

Restriction R3: $\hat{U}(z)'\,U(z^{-1})$ is analytic on $\{z : |z| < 1\}$.

Restriction *R3* is always satisfied when U is a constant orthogonal matrix independent of z. Hence the restrictions are always satisfied for fundamental moving-average representations. More generally, note that

$$\begin{aligned} y_t &= \hat{C}(L)\hat{w}_t \\ &= C(L)\,U(L^{-1})'\,\hat{U}(L)\hat{w}_t \\ &= C(L)w_t \,. \end{aligned} \tag{3.13}$$

Consequently, the w process must satisfy:

$$w_t = U(L^{-1})'\,\hat{U}(L)\,\hat{w}_t \,. \tag{3.14}$$

When restriction *R3* is satisfied, the entries of w_t are in the closed linear space $\hat{\Lambda}_t$ generated by current and past values of $\hat{w}_t$. Hence the closed linear space Λ_t generated by current and past values of w_t is no larger than $\hat{\Lambda}_t$. It follows from Lemma 1 that the restrictions are preserved for any moving-average representation associated with an information reduction. Alternatively, this special case of Lemma 1 could be proved by assuming that (1.8) is satisfied for the information set $\hat{\Lambda}_{t-\ell}$ and applying the Law of Iterated Projections (see Chapter 2, Lemma 3.5) to show that (1.8) is also satisfied for the smaller information set $\Lambda_{t-\ell}$.

Restriction *R2* turns out to be weaker than restriction *R3*. Thus Lemma 1 covers cases other than those associated with an information reduction. To see this, consider the following example.

Example 1: Let λ be any real number such that $|\lambda| < 1$. Form the spectral decomposition of the symmetric positive semidefinite matrix

$$\hat{H}(\lambda)'\,\hat{H}(\lambda) = U_1 V U_1' \tag{3.15}$$

where $U_1U_1' = U_1'U_1 = I$ and V is a diagonal matrix with strictly positive real numbers in the diagonal entries of its first block and zeros in the second block. Then the second column block of $\hat{H}(\lambda)U_1$ contains only zeros. As in Rozanov (1967, page 47), we construct an orthogonal matrix function $U_2(z)$ of the form

$$U_2(z) = \begin{bmatrix} I & 0 \\ 0 & \beta(z)I \end{bmatrix} \tag{3.16}$$

where $\beta(z)$ is the *Blaschke* factor

$$\beta(z) \equiv (z-\lambda)/(1-\lambda z) \,. \tag{3.17}$$

Notice that

$$\begin{aligned} \beta(z)\beta(z^{-1}) &= (z-\lambda)\,(z^{-1}-\lambda)/[(1-\lambda z)\,(1-\lambda z^{-1}] \,. \\ &= 1 \,. \end{aligned} \tag{3.18}$$

Consequently, $U_2(z)$ is a q-dimensional orthogonal matrix function. Let

$$U(z) = \hat{U}(z)\,U_1U_2(z) \,. \tag{3.19}$$

Then $U(z)$ also is a q-dimensional orthogonal matrix function. Note that

$$\hat{H}(z)\hat{U}(z)'U(z^{-1}) = \hat{H}(z)U_1U_2(z^{-1}) \,. \tag{3.20}$$

Although $U_2(z^{-1})$ has a pole at $z = \lambda$, $\hat{H}(z)U_1U_2(z^{-1})$ is analytic on $\{z : |z| < 1\}$ because the singularity at $z = \lambda$ is removable by construction. This follows from the fact that $\hat{H}(\lambda)U_1$ has all zeros in its second column block whereas the pole of $U_2(z^{-1})$ at $z = \lambda$ occurs in its second row block. Consequently, $U(z^{-1})$ given by (3.19) satisfies *R2*. Furthermore,

$$\hat{U}(z)'\,U(z^{-1}) = U_1U_2(z^{-1})\ , \tag{3.21}$$

which has a pole at $z = \lambda$ that is not removable. Therefore, $U(z)$ fails to satisfy *R3*.

Following the construction in Example 1, a rich family of $U(z)$'s can be generated, each member of which satisfies *R2* but fails to satisfy *R3*. For instance, any real value of λ for which $|\lambda| < 1$, can be used. In addition, given any finite number of $U(z^{-1})$'s that satisfy *R2* but fail to satisfy *R3*, the product of these $U(z^{-1})$'s possesses this same property. Finally, a construction similar to that given in Example 1 will work for complex values of λ with some minor modifications. First, the matrix $\hat{H}(\lambda)'$ on the left side of (3.15) should be replaced by its conjugate in constructing a (conjugate) symmetric matrix that is positive semidefinite. Second, the coefficient matrices of the power series expansion of the resulting $U(z)$ may not have all real entries. However, by repeating the construction using the complex conjugate of λ in the second stage and post multiplying by an appropriately chosen unitary matrix, this complication can be avoided. In summary, there exists a rich collection of q-dimensional orthogonal matrix functions that satisfy *R2* but not *R3*.

In practice, one typically adopts a finite-dimensional parameterization of $\hat{D}_2(z)$. It is of interest to see what impact this specification has on the analysis. Suppose that $\hat{D}_2(z)$ is given by the ratio of polynomials:

$$\hat{D}_2(z) = \hat{D}_n(z)/\hat{D}_d(z) \tag{3.22}$$

where $\hat{D}_n(z)$ and $\hat{D}_d(z)$ are finite-order polynomials and the zeros of $\hat{D}_n(z)$ are all outside the unit circle. Let N denote the order of $\hat{D}_n(z)$. Suppose we admit only $C(z)$'s constructed from $D_2(z)$'s of the form:

$$D_2(z) = D_n(z)/\hat{D}_d(z) \tag{3.23}$$

where the order of $D_n(z)$ does not exceed N. Our candidate for $D_n(z)$ is given by

$$D_n(z) = \hat{D}_n(z)\,\hat{U}(z^{-1})'\,U(z)\ . \tag{3.24}$$

By assumption, $D_2(z)$ and hence the right side of (3.24) is analytic inside the unit circle of the complex plane. However, the right side of (3.24) will not necessarily be a polynomial. Hence the restriction that $D_n(z)$ be a finite order polynomial with the same degree as $\hat{D}_n(z)$ limits further the class of admissible $C(z)$'s.

Consider first the case in which the more stringent restriction *R3* is satisfied. In this case the right side of (3.24) is analytic everywhere in the complex plane. Furthermore, since $\hat{U}(z^{-1})'U(z)$ has a one-sided Laurent series expansion in terms of negative powers of z, it can be extended to be analytic at the point ∞. Therefore

$$\lim_{z\to\infty} z^{-N+1}\,\hat{D}_n(z)\,\hat{U}(z^{-1})'\,U(z) = 0\ . \tag{3.25}$$

It follows that $\hat{D}_n(z)\,\hat{U}(z^{-1})'\,U(z)$ has a finite order pole at ∞ and that the order of this pole does not exceed N. Therefore, $\hat{D}_n(z)\,\hat{U}(z^{-1})'\,U(z)$ is a finite order polynomial and its order does not exceed N. Hence requiring that $D_n(z)$ be an N^{th} order polynomial does not alter the analysis of *R3*.

We consider next the more general restriction *R2*, and ask whether *R2* can be satisfied even though *R3* is violated.

Example 2: We show how to modify the construction in Example 1 to accommodate the order restriction on $D_n(z)$. Since $A(z)$, $B(z)$ and $D(z)$ are matrices of rational functions, so is $\hat{H}(z)$. In Example 1 we exploited the fact that the matrix $\hat{H}(\lambda)'\,\hat{H}(\lambda)$ is always singular. However, in this example we must confine our attention to only a finite number of values of λ. In particular, let

$$\det\begin{bmatrix}\hat{H}(z^{-1})\\ \hat{D}_n(z)\end{bmatrix} = \hat{\delta}_n(z)/\hat{\delta}_d(z) \tag{3.26}$$

where $\hat{\delta}_n(z)$ and $\hat{\delta}_d(z)$ do not have any zeros in common. Suppose that there exists a real number λ such that $|\lambda| < 1$ and λ^{-1} is a zero of $\hat{\delta}_n(z)$ but not a zero of $\hat{D}_d(z)$ [i.e., λ is not a pole of $\hat{D}_2(z)$]. Such a λ does not always exist, although it often does.

As in Example 1, we use λ to construct $U(z)$. In this case we begin by forming the spectral decomposition of the symmetric, positive semi-definite matrix:

$$\hat{H}(\lambda)'\,\hat{H}(\lambda) + \hat{D}_n(\lambda^{-1})'\,\hat{D}_n(\lambda^{-1}) = U_1VU_1'\ . \tag{3.27}$$

At least one entry of V, the bottom right one, must be zero because λ^{-1} is a zero of $\hat{\delta}_n(z)$. Note that by construction, the second column blocks of $\hat{H}(\lambda)U_1$ and $\hat{D}_n(\lambda^{-1})U_1$ are zero. Use the same orthogonal matrix functions $U_2(z)$ and $U(z)$ that were used in Example 1:

$$\hat{D}_n(z)\,\hat{U}(z^{-1})'\,U(z) = \hat{D}_n(z)U_1U_2(z)\ . \tag{3.28}$$

As in Example 1, $H(z)$ is analytic inside the unit circle. However, in this case we also preserve the requirement that $D_n(z)$ be a polynomial with an order no greater than N because the second column block of $\hat{D}_n(\lambda^{-1})U_1$ is zero and the first order pole of $U_2(z)$ occurs in the second row block. Consequently, the singularity of $\hat{D}_n(z)U_1U_2(z)$ at λ^{-1} is removable.

Like Example 1, Example 2 assumes that λ is real. However, $\hat{\delta}_n(z)$ may have complex zeros that exceed one in modulus and are not simultaneously zeros of $\hat{D}_d(z)$. Such zeros come in complex conjugate pairs. These conjugate pairs of zeros can also be used in an analogous (but slightly more complicated) fashion to construct orthogonal matrix functions that respect the order restriction on $D_n(z)$. The order restriction does limit us to a finite number of choices of λ because $\hat{\delta}_n(z)$ has a finite number of zeros. Finite products of such orthogonal matrix functions can be formed so long as any given λ is not used more times than its multiplicity as a zero of $\hat{\delta}_n(z)$.

Example 3: We show by illustration that it is possible for $\hat{\delta}_n(z)$ to have a zero outside the unit circle of the complex plane that is not simultaneously a zero of $\hat{D}_d(z)$. Set $A(z) = 1$, $B(z^{-1}) = z^{-1}$, $\ell = 0$, $\hat{D}_n(z) = [1 + 2z + z^2,\ 1]$ and $\hat{D}_d(z) = 1$. Hence $n = 2$, $\hat{D}_1(z) = 0$, and $\hat{H}(z^{-1}) = [1 \quad 1]$. Note that

$$\det \begin{bmatrix} \hat{H}(z^{-1}) \\ D_n(z) \end{bmatrix} = -2z - z^2 = \hat{\delta}_n(z)\ . \tag{3.29}$$

The polynomial $\hat{\delta}_n(z)$ has one zero outside the unit circle, namely $z = -2$. Therefore, $\lambda = -1/2$ can be used in the construction described in Example 2.

In summary, in this section we have investigated the extent to which the cross-equation restrictions implied by members of the class of exact rational expectations models can be used to narrow the family of observationally equivalent moving-average representations. We have found

that the function $D_2(z)$ that can be used to index alternative solutions is not identified. Even when entries of $D_2(z)$ are restricted to be ratios of polynomials with prespecified orders, $D_2(z)$ can fail to be identified.

The standard approach in reduced-form time series analysis is to focus exclusively on fundamental moving-average representations. As we have indicated, fundamental representations will satisfy the restrictions implied by the model. Requiring that $C(z)$ be fundamental is equivalent to the restriction that

$$C(z) = F(z)U \tag{3.30}$$

for some orthogonal matrix U that is independent of z. However, it may be computationally tedious to restrict the family of $D(z)$'s so that resulting $C(z)$'s satisfy (3.30). Furthermore, this extra restriction is *ad hoc*, and the underlying information structure faced by economic agents is still left unidentified.[4]

While these observations are discouraging from the standpoint of identifying $D_2(z)$, they do not overturn the validity of likelihood-based inferences. The restricted and unrestricted likelihood functions will typically have multiple peaks, but the peaks of interest will all have the same value.

The analysis in this section has taken $A(z)$ and $B(z^{-1})$ as given and has focused on the identification of $D(z)$ and hence $C(z)$. More generally, $A(z)$ and $B(z^{-1})$ may only be known up to a finite-dimensional parameter vector. It is often possible to identify the parameters governing $A(z)$ and $B(z^{-1})$ even though identification of $D(z)$ is problematic. When $A(z)$ and $B(z)$ are not identified, the family of observationally equivalent $D(z)$'s is likely to be expanded.

4. Restrictions Implied for First Differences

In the analysis considered thus far, we have assumed that the y process is covariance stationary. Although we could view this assumption as being appropriate for deviations about a linear time trend, an alternative strategy is to assume that the first difference of y_2 is covariance stationary where the process $y' \equiv [y_1' \; y_2']$ is partitioned in a manner compatible with $C(z)$. We maintain the model restrictions (1.3) which for convenience are written below:

$$P\Big\{[A(L)\ ;\ B(L^{-1})]\, y_t|\Omega_{t-\ell}\Big\} = 0\ . \tag{4.1}$$

Recall that $B(z) = \sum_{j=0}^{\infty} b_j z^j$. Let

$$b_j^* \equiv \sum_{k=j}^{\infty} b_j \quad \text{for } j = 0, 1, \dots ,$$

(4.2)

$$y_{2t}^* \equiv y_{2t} - y_{2,t-1}$$

and

$$B^*(z) \equiv \sum_{j=1}^{\infty} b_j^*(z) .$$

Now $b_j = b_j^* - b_{j+1}^*$, implying that[5]

$$B(L^{-1})y_{2t} = B^*(L^{-1})y_{2t}^* + b_0^* y_{2t} . \tag{4.3}$$

Substituting (4.3) into (4.1), we obtain

$$P\Big[A(L)y_{1t} + b_0^* y_{2t} + B^*(L^{-1})y_{2t}^* | \Omega_{t-\ell}\Big] = 0 \tag{4.4}$$

Define

$$y_{1t}^* \equiv A(L)y_{1t} + b_0^* y_{2t} . \tag{4.5}$$

Then we can write the first-difference model as

$$P\Big\{[I \ ; \ B^*(L^{-1})] y_t^* | \Omega_{t-\ell}\Big\} = 0 \tag{4.6}$$

where $y^{*\prime} \equiv [y_1^{*\prime} \ , \ y_2^{*\prime}]$ is assumed to be covariance stationary. This is just a special case of the general model presented in Section 1.

The first-difference model derived here is usefully compared to that employed by Sargent (1979). In particular, Sargent first differenced (4.1) and projected both sides onto $\Omega_{t-\ell-1}$ to obtain

$$P\Big\{[A(L) \ ; \ B(L^{-1})] \ (y_t - y_{t-1}) | \Omega_{t-\ell-1}\Big\} = 0 . \tag{4.7}$$

Although restrictions (4.7) can be tested using procedures discussed in this paper, some implications of (4.1) are lost by projecting onto $\Omega_{t-\ell-1}$ rather than $\Omega_{t-\ell}$. On the other hand, (4.6) involves a projection onto $\Omega_{t-\ell}$ rather than $\Omega_{t-\ell-1}$ and imposes more restrictions than (4.7). Therefore, it is quite possible that the procedures proposed in this section can detect empirical contradictions of the hypothesis (4.1) that Sargent's procedure could not.

5. Likelihood Estimation and Inference

In this section we describe briefly how to conduct estimation and inference using the method of maximum likelihood with a Gaussian likelihood function. First we illustrate how to impose the restrictions implied by the model on the moving-average coefficients. Then we describe two alternative approaches to evaluating the likelihood function.

Imposing the Restrictions

For pedagogical purposes, we focus on the following special case of the model described in section one. Suppose $A(z) = I$ and $B(z) = b_0 + b_1 z + \dots + b_k z^k$ and $\ell = 0$. Furthermore, we consider rational parameterizations of $D_2(z)$ of the form described in Section 3:

$$D_2(z) = D_n(z)/D_d(z) \tag{5.1}$$

where $D_n(z) = d_{n0} + d_{n1}z + \dots + d_{nk}z^k$ and $D_d(z) = d_{d0} + d_{d1}z + \dots + d_{dk}z^k$. The restrictions that the polynomials $B(z)$, $D_n(z)$ and $D_d(z)$ all have the same order is made only for notational convenience. Additional zero restrictions can be imposed on these polynomials in a straightforward fashion. To guarantee that $D_d(z)$ has a power series expansion with square summable coefficients, we restrict the zeros of $D_d(z)$ to be outside the unit circle of the complex plane. A convenient way to accomplish this is to use a parameterization suggested by Monahan (1984).

For the special case of the model considered here, (1.18) simplifies to

$$\begin{aligned} C_1(z) &= -B(z^{-1})\, D_2(z) + G(z^{-1}) \\ C_2(z) &= D_2(z) \end{aligned} \tag{5.2}$$

where $G(z)$ satisfies *R1*. Substituting (5.1) into (5.2) yields

$$C_1(z)D_d(z) = -B(z^{-1})\, D_n(z) + D_d(z)G(z^{-1}) . \tag{5.3}$$

There are two important implications of (5.3). First, $C_1(z)D_d(z)$ is a finite-order polynomial with maximum order k. This follows from the fact that the largest positive power of z that occurs on the right side of (5.3) is k. Hence

$$C_1(z) = C_n(z)/D_d(z) \tag{5.4}$$

for some k^{th}-order polynomial $C_n(z)$. Second, $G(z)$ is a finite order polynomial with maximum order k. This follows from the fact that the largest (in absolute value) negative power of z that occurs in $B(z^{-1})$ is $-k$.

Substituting (5.4) into (5.3) gives

$$C_n(z) = -B(z^{-1})D_n(z) + D_d(z)G(z^{-1}) \tag{5.5}$$

which can be viewed as a system of linear equations in the coefficients of $C_n(z)$ and $G(z)$. Furthermore, there is a recursive structure to these equations as evident by the fact that the equations determining the coefficients of $G(z)$ do not involve the coefficients of $C_n(z)$. Written in matrix notation, the first set of equations is

$$\begin{aligned}\begin{bmatrix}0\\0\\\vdots\\0\end{bmatrix} = &-\begin{bmatrix}b_k & 0 & \dots & 0\\ b_{k-1} & b_k & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ b_1 & b_2 & \dots & b_k\end{bmatrix}\begin{bmatrix}d_{n0}\\ d_{n1}\\ \vdots\\ d_{nk-1}\end{bmatrix}\\ &+\left\{\begin{bmatrix}d_{d0} & 0 & \dots & 0\\ d_{d1} & d_{d0} & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ d_{d,k-1} & d_{d,k-2} & \dots & d_{d0}\end{bmatrix}\otimes I\right\}\begin{bmatrix}g_k\\ g_{k-1}\\ \vdots\\ g_1\end{bmatrix}\end{aligned} \tag{5.6}$$

where $G(z) = g_1 z + g_2 z^2 + \ldots + g_k z^k$. Since the matrix in $\{\cdot\}$ in (5.6) is nonsingular, this system of equations can be solved for the coefficients of G. Given these coefficients, one can then compute the coefficients of $C_n(z)$ by

$$\begin{aligned}\begin{bmatrix}c_{n0}\\ c_{n1}\\ \vdots\\ c_{nk}\end{bmatrix} = &-\begin{bmatrix}b_0 & b_1 & \dots & b_k\\ 0 & b_0 & \dots & b_{k-1}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & b_0\end{bmatrix}\begin{bmatrix}d_{n0}\\ d_{n1}\\ \vdots\\ d_{nk}\end{bmatrix}\\ &+\left\{\begin{bmatrix}d_{d1} & d_{d2} & \dots & d_{dk}\\ d_{d2} & d_{d3} & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & 0\end{bmatrix}\otimes I\right\}\begin{bmatrix}g_1\\ g_2\\ \vdots\\ g_k\end{bmatrix}\end{aligned} \tag{5.7}$$

Therefore, the coefficients of $C_n(z)$ can be computed using some simple matrix manipulations.[6]

Likelihood Evaluation

A Gaussian likelihood function can be expressed as a function of the mean vector and covariance matrix of the data, (say $y_1, y_2, \ldots, y_T$), and of the unknown parameters. In this paper we have abstracted from restrictions on the mean vector, and have instead investigated the covariance restrictions. These covariance restrictions can be viewed equivalently as restrictions on the spectral density of y. In sections one, three and the first part of this section, we described the mapping from $D(z)$ to the spectral density matrix of y. This representation of the restriction makes it convenient to follow the suggestions of Hannan (1970), Robinson (1977), Dunsmuir and Hannan (1976), and Dunsmiur (1978) by using a frequency domain approximation to the likelihood function.

This approximation is formed as follows. First, form the finite Fourier transform of the y sequence:

$$Y(\theta_j) = \sum_{t=1}^{T} y_t e^{-i\theta_j t} \tag{5.8}$$

where $\omega_j = \frac{2\pi j}{T}$, $j = 1, 2, \ldots, T-1$. We omit frequency zero from consideration since sample means are subtracted from our time series. The periodogram is defined as

$$I(\theta_j) = \frac{1}{T} Y(\theta_j) Y(-\theta_j)' . \tag{5.9}$$

Then the log likelihood of the sample $\{y_t : t = 1, \ldots, T\}$ is approximated by

$$\begin{aligned} L = \frac{nT}{2} \log 2\pi - \frac{1}{2} \sum_{j=1}^{T-1} \log \det S(\theta_j)\\ - \frac{1}{2} \sum_{j=1}^{T-1} \operatorname{trace} S(\theta_j)^{-1} I(\theta_j)\end{aligned} \tag{5.10}$$

where $S(\theta_j)$ is the theoretical spectral density defined in (3.3) and (2.3). Note that $S(\theta_j)$ can be expressed in terms of the free parameters of $D(z)$ using formulas (1.14) and (3.3). To emphasize the dependence on $C(z)$ and hence on $D(z)$, it is worthwhile substituting (3.3) into (5.10) to obtain

$$(5.11)\quad \begin{aligned} L = & -\frac{nT}{2}\log 2\pi - \frac{1}{2}\sum_{j=1}^{T-1}\log\det\,[C(e^{-i\theta_j})\,C(e^{i\theta_j})'] \\ & -\frac{1}{2}\sum_{j=1}^{T-1}\operatorname{trace}\,[C(e^{-i\theta_j})C(e^{i\theta_j})']I(\theta_j)\,. \end{aligned}$$

In computing (5.11), it is useful to exploit the fact that

$$\log\det[C(e^{-i\theta})C(e^{i\theta})'] = \log\det\{C[e^{-i(2\pi-\theta)}]C[e^{i(2\pi-\theta)}]'\}\,,$$

and

$$(5.12)\quad \begin{aligned} & \operatorname{trace}\ [C(e^{-i\theta})\,C(e^{i\theta})']^{-1}I(\theta) \\ & \quad = \operatorname{trace}\ \{C[e^{-i(2\pi-\theta)}]\,C[e^{i(2\pi-\theta)}]'\}^{-1}I(2\pi-\theta)\,. \end{aligned}$$

These formulas permit (5.10) to be rewritten in terms of sums over only $T/2$ frequencies. The free parameters of $D(z)$ are estimated by maximizing (5.11) subject to the restrictions (1.14).

Alternatively, by restricting $D(z)$ appropriately, one can often avoid making approximations to the likelihood function and instead use filtering methods to evaluate it. For instance, consider the model and the parameterization of $D(z)$ given in the first subsection. In this case, $C(z)$ can be represented as

$$(5.13)\quad \begin{aligned} C_1(z) &= C_n(z)/D_d(z) \\ C_2(z) &= D_n(z)/D_d(z) \end{aligned}$$

where $C_n(z)$, $D_n(z)$ and $D_d(z)$ are finite-order polynomials. In this case, the y process is a vector autoregressive moving-average (ARMA) process. It is known that such a process has a state-space representation [e.g. see Anderson and Moore (1979), page 236].

Recall that the joint density of $y_1, y_2, \ldots, y_T$ can be expressed as a product of conditional densities, say the product of the density of y_T conditioned on $y_1, y_2, \ldots, y_{T-1}$, the density of y_{T-1} conditioned on $y_1, y_2, \ldots, y_{T-2}, \ldots$, and the marginal density of y_1. Each of these densities is Gaussian and hence can be constructed given knowledge of the conditional means and covariance matrices. Given the state-space representation, one can calculate recursively the required conditional expectations and conditional covariance matrices using one of several possible filtering algorithms described in Anderson and Moore (1979).

The initial covariance matrix required for these algorithms can be computed using a doubling algorithm [see Anderson and Moore (1979), page 67].[7]

The analysis in this section took $A(z)$ and $B(z^{-1})$ as given and focused on the identification of $D(z)$ and hence $C(z)$. If $A(z)$ and $B(z^{-1})$ are only known up to a finite-dimensional parameter vector, it will only be more difficult to identify of $D(z)$. However, even if $D(z)$ is not fully identified, it is often possible to identify the parameters governing $A(z)$ and $B(z^{-1})$.

6. Inexact Models with Hidden Variable Interpretations of Disturbances

So far, this paper has been confined to analyzing exact linear rational expectations models. In this section, we briefly indicate which aspects of the analysis readily carry over to more general linear rational expectations models and which aspects require modification. It turns out that the methods of representing the cross-equation restrictions and estimating the parameters both carry over with minimal modification. However, the treatment of identification must be modified substantially.

Following Sargent (1978) and Hansen and Sargent (1980a), we assume that a subvector of y_{2t} is not observed by the econometrician. For this reason, we partition y_{2t} as $y_{2t}' = [y_{2t}^{o\prime}, y_{2t}^{u\prime}]$ where the process y_2^o is observed by the econometrician while the process y_2^u is not. We partition $C(z)$ accordingly, so that the moving-average representation for y is:

$$(6.1)\quad \begin{bmatrix} y_{1t} \\ y_{2t}^o \\ y_{2t}^u \end{bmatrix} = \begin{bmatrix} C_1(L) \\ C_2^o(L) \\ C_3^u(L) \end{bmatrix} w_t\,.$$

Similarly, we partition $B(z) = [B^o(z);\ B^u(z)]$. Hence we write equation (1.3) as

$$(6.2)\quad A(L)y_{1t} + \mathrm{P}\,[B^o(L^{-1})y_{2t}^o|\Omega_t] = -\,\mathrm{P}\,[B^u(L^{-1})y_{2t}^u|\Omega_t]$$

where we have assumed that $\ell = 0$. We refer to such a model as an *inexact* rational expectations model because of the term $\mathrm{P}\,[B^u(L^{-1})y_{2t}^u|\Omega_t]$. The introduction of this term means that the relationship observed by the econometrician would not be exact even if economic agents could forecast perfectly.

From the vantage of model specification and solution, an inexact model as given here is identical with an exact model. However, from the vantage point of identification and testable restrictions, an inexact model can be quite different. For instance, suppose that $B^u(z)$ is nonsingular inside and on the unit circle of the complex plane, implying that there are r entries of y_2^u. Partition $D_2(z)$ as

$$D_2(z) = \begin{bmatrix} D_2^o(z) \\ D_2^u(z) \end{bmatrix} . \tag{6.3}$$

Then

$$\mathrm{P}\,[B^u(L^{-1})y_{2t}^u|\Lambda_t] = K^u(L)w_t \tag{6.4}$$

where

$$K^u(z) \equiv [B^u(z^{-1})D_2^u(z)]_+ . \tag{6.5}$$

Inverting relationship (6.5), we have that

$$D_2^u(z) = [B^u(z^{-1})^{-1}\,K^u(z)]_+ . \tag{6.6}$$

Therefore, without restricting $D_2^u(z)$, there are no restrictions on $K^u(z)$, and hence relation (6.2) imposes no restrictions on the process $[y_1',\, y_2^{o\prime}]'$ observed by the econometrician. Consequently, the empirical content of model (6.2) is tied to the *a priori* restrictions which are imposed on $D_2(z)$.

One way to restrict $D_2(z)$ is to require that it be block diagonal:

$$D_2(z) = \begin{bmatrix} D_2^o(z) \\ D_2^u(z) \end{bmatrix} = \begin{bmatrix} D_2^{oo}(z) & 0 \\ 0 & D_2^{uu}(z) \end{bmatrix} \tag{6.7}$$

where $D_2^{uu}(z)$ is a square matrix. In this case,

$$K^u(z) = [0 \quad [B^u(z^{-1})\,D_2^{uu}(z)]_+] , \tag{6.8}$$

and

(6.9)
$$\begin{bmatrix} C_1(z) \\ C_2^o(z) \end{bmatrix} = \begin{bmatrix} -A(z)^{-1}[B^o(z^{-1})D_2^{oo}(z)]_+ & -A(z)^{-1}[B^u(z^{-1})D_2^{uu}(z)]_+ \\ D_2^{oo}(z) & 0 \end{bmatrix}$$

The first column partition of the right side of (6.9) is

$$\begin{bmatrix} -A(z)^{-1}[B^o(z^{-1})D_2^{oo}(z)]_+ \\ D_2^{oo}(z) \end{bmatrix} \tag{6.10}$$

which is a square matrix and has the same structure as $C(z)$ given in (1.14). Consequently, the analysis in Section 3 applies to this component. Notice that the unobserved process y_2^u contaminates only the upper $(q \times q)$ portion of the spectral density of $[y_1',\, y_2^{o\prime}]'$.

Recall that the number of columns of $D_2(z)$ is equal to the number of entries of the stochastic process y. This restriction was imposed to ensure that y is stochastically nonsingular. Since $D_2^{uu}(z)$ is assumed to be a square matrix, $D_2^{oo}(z)$ has $q - r$ more columns than rows (where r is the dimension of y_1). A second way to restrict $D_2(z)$ is to treat y_2^u and y_2^o symmetrically by requiring also that $D_2^{oo}(z)$ be a square matrix. In this case, the y process is stochastically singular but the process $[y_1',\, y_2^{o\prime}]'$ observed by the econometrician is, in general, stochastically nonsingular. Now the matrix function in (6.10) is not square and the analysis in Section 3 will no longer apply.

It is of interest to investigate further this second restriction on $D_2(z)$. Given the structure of the matrix $\begin{bmatrix} C_1(z) \\ C_2^o(z) \end{bmatrix}$,

(6.11)
$$\det \begin{bmatrix} C_1(z) \\ C_2^o(z) \end{bmatrix} = -\ \det[D_2^{oo}(z)]\ \det\left\{[-A(z)^{-1}]\,[B^u(z^{-1})D_2^{uu}(z)]_+\right\} .$$

Therefore, a zero of $\det \begin{bmatrix} C_1(z) \\ C_2^o(z) \end{bmatrix}$ must be either a zero of $\det[D_2^{oo}(z)]$ or a zero of $\det\{[-A(z)^{-1}[B^u(z^{-1})D_2^{uu}(z)]_2\}$. As argued by Sargent (1978) and Hansen and Sargent (1980a), if $D_2^{oo}(z)$ is restricted to be nonsingular inside the unit circle of the complex plane, then y_1 should not cause y_2^o in the sense of Granger (1969). This is one of the testable implications of the model.

Even when $D_2^{oo}(z)$ is nonsingular, moving-average representation (6.9) may fail to be fundamental because $\det\{-A(z)^{-1}\,[B^u(z^{-1})$ $D_2^{uu}(z)]_+\}$ may have zeros inside the unit circle of the complex plane. As emphasized by Hansen and Sargent (1980a), such zeros can exist even when $D_2^{uu}(z)$ is nonsingular at all points inside the unit circle. However, Hansen and Sargent (1980a) also note that one can construct a fundamental representation for the observable components of y that satisfies the restrictions. To see this, define

$$K^{uu}(z) \equiv [B^u(z^{-1})D_2^{uu}(z)]_+ . \tag{6.12}$$

When $K^{uu}(z)$ is singular at isolated points inside the unit circle, it follows from the Wold Decomposition Theorem that there exists a matrix $K^{uu*}(z)$ that is nonsingular inside the unit circle and satisfies:

(6.13)

$$K^{uu}[\exp(-i\theta)]\,K^{uu}[\exp(i\theta)]' \;=\; K^{uu^*}[\exp(-i\theta)]\,K^{uu^*}[\exp(i\theta)]'$$

almost everywhere (see also Section 3). Define

$$D_2^{uu^*}(z) \equiv [B^u(z^{-1})^{-1} K^{uu^*}(z)]_+ \,. \tag{6.14}$$

Then, D_2^* is observationally equivalent to D_2, where D_2^* is constructed by replacing $D_2^{uu}(z)$ by $D_2^{uu^*}(z)$. However, it is not necessarily true that

$$D_2^{uu}[\exp(-i\theta)]\,D_2^{uu}[\exp(i\theta)]' = D_2^{uu^*}[\exp(-i\theta)]\,D_2^{uu^*}[\exp(i\theta)]' \,, \tag{6.15}$$

so that the implied serial correlation properties for the unobservable component of y are different. Therefore, without additional restrictions, the spectral density function for the unobserved process y_2^u is not identified. Furthermore, $D^{uu^*}(z)$ may well have zeros inside the unit circle of the complex plane. In other words, it may require a moving-average representation for the unobserved process y_2^u which is *not* fundamental to construct a $D_2^*(z)$ associated with a fundamental moving-average representation for the observed process $[y_1', y_2^{o\prime}]'$. [8]

Conclusions

As indicated by the examples given in section two, the procedures described in this paper are applicable to a variety of linear rational expectations models. Therefore, the solution procedure, the characterization of identification, and the methods of estimation and inference for exact rational expectations models are useful tools for guiding a wide range of empirical applications. Chapter 6 of this volume contains an application.[9]

Notes

1. Actually, the assumption of stationarity of $m_t - m_{t-1}$ is not imposed in all of these papers. In Section 4 we see how to transform the model to accommodate unit roots in the process y.

2. This formulation does not include any explicit adjustment for the riskiness of the payout process. Such an adjustment, when appropriate, must be incorporated into a constant term or asset payouts and values must be scaled appropriately by a process of used price contingent claims. Alternatively, Campbell and Shiller (1988) propose log-linear approximations as a means of incorporating risk-adjustments.

3. The probability space underlying this mean-square convergence has $[-\pi, \pi]$ as the collection of sample points, the Borel measurable subsets of $[-\pi, \pi]$ as a sigma algebra, and Lebesgue measure scaled by $1/2\pi$ as a probability measure.

4. The matrix of U in (3.30) is not identified even when (3.30) is imposed; however, the information structure associated with these $C(z)$'s is identified. As shown by Campbell and Shiller (1988), the present-value model (Example 2 in Section 2) implies a linear restriction on a finite-order vector autoregressive representation for y. This provides a convenient parameterization which enforces (3.30).

5. The rearranging of terms in the infinite sum in order to obtain (4.3) is justified given our assumptions about $B(z)$ and y.

6. This representation of the restrictions is superior to that used by Sargent (1979). Although the present restrictions are nonlinear in the coefficients of $D_d(z)$, they are considerably easier to compute than the restrictions that Sargent (1979) imposed directly on the vector autoregressive representation.

7. Time domain approximations to the likelihood function that assume $C(z)$ is nonsingular inside and on the unit circle of the complex plane may be inconvenient for many parameterizations of $D(z)$ because it may be difficult to restrict the parameterization of $D(z)$ in such a way as to ensure that $C(z)$ is nonsingular.

8. The working paper form of this chapter (Hansen and Sargent 1981e) contained an application to the term structure of interest rates.

9. A very similar phenomenon occurs in Futia (1981). In Futia's analysis two alternative rational expectations equilibria are compared. One endows economic agents directly with observations on an exogenous forcing process and the other presumes that information about this process can only be extracted from observations on endogenously determined prices. Futia showed that when endogenous information is sufficient to reveal the exogenous information, the two equilibria coincide. Otherwise the second equilibrium fails to exist.