

Exact Linear Rational Expectations Models:  
Specification and Estimation

Lars Peter Hansen

Carnegie-Mellon University

Thomas J. Sargent

University of Minnesota and  
Federal Reserve Bank of Minneapolis

Research Department Staff Report 71  
Federal Reserve Bank of Minneapolis

September 1981

ABSTRACT

This paper describes how to specify and estimate rational expectations models in which there are exact linear relationships among variables and expectations of variables that the econometrician observes.

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

Kenneth Singleton, Lawrence Christiano, Martin Eichenbaum, Katherine Schipper, and Robert Townsend provided useful comments. Martin Eichenbaum and Judy Sargent performed the numerical calculations that are reported in Section 3. This research was supported in part by NSF Grant SES-8007016.

## 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 are 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 restriction across forecasts of future values of one set of variables and current and past values of some other set of variables. While probably only a minority of rational expectations models belong to this class, it does contain interesting models that have been advanced to study forward markets, 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 (1980). The differences lie entirely in the interpretations of the "error terms" in the

equations that are permitted. In Hansen and Sargent (1980), 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 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 maximizing one of various approximations to the likelihood function. In

this paper we focus on the frequency domain approximation proposed by Hannan (1970).

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 2 provides three lemmas 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 5, we briefly indicate how both our methods for model specification and estimation carry over to inexact linear rational expectations models.

While it is not their only purpose, the methods described here lead to straightforward tests of both the model under

consideration and the rational expectations hypothesis. The vector moving average representation that incorporates the rational expectations restrictions is nested within less constrained vector moving average representations. A likelihood ratio statistic can be computed to test the model. In the context of applications of exact linear rational expectations models to stock prices and the term structure of interest rates, the likelihood ratio test is in a relevant sense more powerful than are the variance bounds tests proposed and used by LeRoy and Porter (1979), Shiller (1979a), and Singleton (1980a). In Section 4, we briefly indicate the relation between the variance bounds and likelihood ratio tests.

The main goal of this paper is to describe procedures for estimating complete linear vector stochastic processes subject to rational expectations restrictions. As Geweke (1979), Shiller (1979a), and Hansen (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 5, 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 (1980)). 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.

For our empirical examples, in Section 3 we return once more to that serviceable laboratory for students of expectations, the term structure of interest rates. Our main intention is to illustrate our procedures with the help of a convenient body of data. As the various examples of Section 1 are intended to emphasize, the procedures are also applicable to many other examples.

### 1. General Model

We begin by specifying a general model and by giving some examples. Let us assume that  $y \equiv \{y_t : t \in J\}$  is a discrete time vector stochastic process where  $J$  is the set of integers.

We partition  $y$  into  $(y_1', y_2')'$ . Let  $A(L^{-1}) \equiv \sum_{j=0}^{\infty} a_j L^{-j}$  where  $L$  is the lag operator and hence  $L^{-j}$  shifts the time index of a variable forward  $j$  periods. We assume that  $A(z^{-1})$  can be represented as

$$(1.1) \quad A(z^{-1}) = \frac{A_1(z)}{A_2(z)}$$

where  $A_1(z)$  is a square matrix polynomial conformable with  $y_1$  and  $A_2(z)$  is a scalar polynomial with zeroes inside the unit circle. Furthermore, we assume that the maximum order of the polynomial elements of  $A_1(z)$  is less than the order of  $A_2(z)$ . These assumptions guarantee that the Laurent series expansions about zero in a region containing the unit circle of the elements of

$$(1.2) \quad \frac{A_1(z)}{A_2(z)}$$

are indeed one-sided in nonpositive powers of  $z$ . Denote this expansion  $\sum_{j=0}^{+\infty} a_j z^{-j}$ . Let  $B(L) \equiv \sum_{j=0}^{\infty} b_j L^j$  where  $b_j$  has the same row dimension as  $y_1$  and column dimension as  $y_2$ . We assume that  $B(z)$  can be represented as

$$(1.3) \quad B(z) = \frac{B_1(z)}{B_2(z)}$$

where  $B_1(z)$  is a matrix polynomial and  $B_2(z)$  is a scalar polynomial with zeroes all outside the unit circle. This latter assumption guarantees that  $B(z)$  has a power series representation appropriate for a region containing the unit circle. Let this power series representation be given by  $\sum_{j=0}^{\infty} b_j z^j$ .

Suppose that our theorizing informs us that

$$(1.4) \quad E[A(L^{-1})y_{1t} | \Omega_t] = B(L)y_{2t}$$

where  $\Omega_t$  is an information set that includes at least  $\{y_t, y_{t-1}, \dots\}$  and  $E[\cdot | \Omega_t]$  is the expectations operator conditional on  $\Omega_t$ . We add the assumption that  $y$  is a linearly indeterministic, mean zero, covariance stationary stochastic process.<sup>1</sup> This allows us to exploit convenient results from linear, least squares prediction theory. Letting  $\hat{E}[\cdot | \Phi_t]$  denote the linear least squares projection on  $\Phi_t$ , where  $\Phi_t$  is the closed (under the root mean square norm) linear space generated by  $\{y_t, y_{t-1}, \dots\}$ , equation (1.4) implies that

$$(1.5) \quad \hat{E}[A(L^{-1})y_{1t} | \Phi_t] = B(L)y_{2t}.$$
<sup>2</sup>

Equation (1.5) can be derived from (1.4) by employing a simple iterated projections argument. From Wold's Decomposition Theorem we know that  $y$  can be represented as

$$(1.6) \quad \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} C_1(L) \\ C_2(L) \end{bmatrix} u_t$$

or

$$y_t = C(L)u_t$$

where  $u_t \in \phi_t$ ,  $Eu_t u_t' = I$ ,  $u$  is serially uncorrelated,  $C(L)$  is one-sided in nonnegative powers of  $L$ , and  $u$  has the same dimension as  $y$ .<sup>3</sup>

Equation (1.5) imposes some restrictions on the matrix polynomial  $C(L)$ . First, write

$$(1.7) \quad y_{1t} = C_1(L)u_t$$

and note that

$$(1.8) \quad A(L^{-1})y_{1t} = A(L^{-1})C_1(L)u_t.$$

Using the Wiener-Kolmogorov prediction formula,<sup>4</sup> we know that

$$(1.9) \quad \hat{E}[A(L^{-1})y_{1t} | \phi_t] = [A(L^{-1})C_1(L)]_+ u_t$$

where  $[ ]_+$  is the annihilation operator<sup>5</sup> that instructs us to ignore nonnegative powers of  $L$ . Second, write

$$(1.10) \quad y_{2t} = C_2(L)u_t$$

and note that

$$(1.11) \quad B(L)y_{2t} = B(L)C_2(L)u_t.$$

Using relation (1.5) and equating coefficients in equations (1.9) and (1.11) we obtain

$$(1.12) \quad B(L)C_2(L) = [A(L^{-1})C_1(L)]_+$$

Equation system (1.12) summarizes the restrictions that the rational expectations model imposes on the vector moving average representation.

It is now of interest for us to consider some examples.

Example (i): Relationship Between a k Period Forward Price and a k-Step-Ahead Spot Price.

Let  $y_{1t}$  be the spot price at time period  $t$  and let the first element of  $y_{2t}$  be the  $k$ -step-ahead forward price. The remainder of the  $y_{2t}$  vector can contain other information available to economic agents. Rational expectations theorizing can lead one to the following model:<sup>6</sup>

$$(1.13) \quad E[y_{1t+k} | \Omega_t] = b_0 y_{2t}$$

where  $b_0 = [1, 0]$  implying that

$$(1.14) \quad \hat{E}[A(L^{-1})y_{1t} | \Phi_t] = B(L)y_{2t}$$

where  $A(L^{-1}) = L^{-k}$  and  $B(L) = b_0$ . In this case

$$(1.15) \quad [A(L^{-1})C(L)]_+ = \sum_{j=0}^{\infty} c_{k+j} L^j$$

where  $C(L) = \sum_{j=0}^{\infty} c_j L^j$ . This provides us with a simple, convenient expression for restrictions (1.12).

Example (ii): Relationship Between Long- and Short-Term Interest Rates.

Let  $y_{1t}$  be the one-period interest rate at time  $t$  and let the first element of  $y_{2t}$  be the  $k$ -period interest rate. Sargent (1979a) has employed the following model of the term structure of interest rates:

$$(1.16) \quad E\left[\frac{1}{k}(y_{1t} + y_{1t+1} + \dots + y_{1t+k-1}) \mid \Omega_t\right] = b_0 y_{2t}$$

where  $b_0 = [1, 0]$ . This implies that

$$(1.17) \quad \hat{E}[A(L^{-1})y_{1t} \mid \Phi_t] = B(L)y_{2t}$$

where  $A(L^{-1}) = \frac{1}{k} + \frac{1}{k}L^{-1} + \dots + \frac{1}{k}L^{-k+1}$  and  $B(L) = b_0$ .

Shiller (1979a) has employed a similar model in which

$$(1.18) \quad E\left[\frac{1-\gamma}{1-\gamma} \frac{1}{k}(y_{1t} + \gamma y_{1t+1} + \dots + \gamma^{k-1} y_{1t+k-1}) \mid \Omega_t\right] = b_0 y_{2t}$$

where  $0 < \gamma < 1$  and  $b_0 = [1, 0]$ . Thus

$$(1.19) \quad \hat{E}[A(L^{-1})y_{1t} | \Phi_t] = B(L)y_{2t}$$

where  $A(L^{-1}) = \frac{1-\gamma}{1-\gamma} (1 + \gamma L^{-1} + \dots + \gamma^{k-1} L^{-k+1})$  and  $B(L) = b_0$ .  
 Explicit expressions for restrictions (1.12) can be obtained in a straightforward manner.

Example (iii): Stock Prices and Dividends

Let  $y_{1t}$  be the rate of dividends on a stock at time  $t$  and let the first element of  $y_{2t}$  be the price of the stock. Shiller (1979b) has studied the following model of stock prices:

$$(1.20) \quad E\left[\sum_{j=0}^{\infty} \gamma^j y_{1t+j} | \Omega_t\right] = b_0 y_{2t} \quad 0 < \gamma < 1,$$

where  $b_0 = [1, 0]$ . This model implies that

$$(1.21) \quad \hat{E}[A(L^{-1})y_{1t} | \Phi_t] = B(L)y_{2t}$$

where  $A(L^{-1}) = \frac{1}{(1 - \gamma L^{-1})}$  and  $B(L) = b_0$ . In this model,  $\gamma$  is the discount factor applied to dividends.

Example (iv): Consumption and Permanent Income

The following example derives from Hall's (1978) version of the permanent income theory of consumption. Let  $w_t$  be labor income at  $t$ , let  $c_t$  be consumption, and let  $A_t$  be nonhuman assets. Then consider the consumption function

$$(1.22) \quad c_t = \frac{\beta\rho}{1+\rho} A_t + \beta\rho E\left[\sum_{j=0}^{\infty} (1+\rho)^{-j} w_{t+j} \mid \Omega_t\right],$$

where  $\rho$  is the discount rate used to define permanent labor income, and  $\beta$  is the marginal propensity to consume out of permanent income. This model falls within our framework, upon making the following identifications: let the first two components of  $y_{2t}$  be  $(c_t, A_t)$ , and set  $y_{1t} = w_t$ ,

$$(1.23) \quad A(L^{-1}) = \beta\rho / [(1 - (1 + \rho)^{-1} L^{-1})],$$

$$(1.24) \quad B(L) = [1, -\beta\rho/(1 + \rho), 0].$$

The applicability of both Hall's testing procedures and the statistical model of the present paper depend critically on the consumption function being an exact equation, or equivalently, on "transitory consumption" being identically zero.

Example (v): Demand Functions for Factors of Production

Sargent (1978b) and Kennan (1979) have estimated linear demand functions for labor that are derived from optimizing a quadratic objective function subject to linear constraints. We focus on Sargent's version, though by reinterpreting  $w_t$  below as output, Kennan's example fits our framework also. Assuming no shocks to technology and a single factor of production, the demand function turns out to be

$$n_t = \lambda n_{t-1} - \frac{\lambda}{\delta} E\left[\sum_{j=0}^{\infty} (\lambda\gamma)^j w_{t+j} \mid \Omega_t\right],$$

where  $0 < \lambda < 1$ ,  $0 < \gamma < 1$ ,  $\delta > 0$ ,  $n_t$  is the stock of employment at  $t$ , and  $w_t$  is the real wage. To put this model within our framework, let the first element of  $y_{2t}$  be  $n_t$ , set  $w_t = y_{1t}$ ,  $A(L^{-1}) = -\frac{\lambda}{\delta}[1 - \lambda\gamma L^{-1}]^{-1}$ , and  $B(L) = [1 - \lambda L, 0]$ .

Multiple factor versions of this example along the lines of Hansen and Sargent (1981a) can easily be constructed. This list of examples could readily be extended to incorporate versions of linear rational expectations models that have been used to analyze a wide variety of macroeconomic and microeconomic phenomena.<sup>7</sup> The preceding examples are sufficient to illustrate the variety of models that reside within the class we are studying.

2a. Identification

We study models with vector moving average representations

$$(2.1) \quad y_t = C(L)u_t$$

where  $y$  is an  $(n \times 1)$  covariance stationary vector stochastic process with mean zero,  $C(L) = \sum_{j=0}^{\infty} c_j L^j$  where  $c_j$  is an  $(n \times n)$  matrix and  $L$  is the lag operator, and where  $u$  is an  $(n \times 1)$  vector white noise with  $Eu_t u_s' = 0$  for  $t \neq s$ ,  $Eu_t u_t' = I$  for all  $t$  and  $Eu_t = 0$  for all  $t$ . The theoretical spectral density matrix of  $y$  is given by

$$(2.2) \quad S(\omega) = C(e^{-i\omega})C(e^{-i\omega})'$$

where the prime denotes transposition and complex conjugation, and where the spectral density matrix is defined as the Fourier transform of the cross-covariogram of  $y$ ,

$$(2.3) \quad S(\omega) = \sum_{\tau=-\infty}^{\infty} E y_t y_{t-\tau}' e^{-i\omega\tau}.$$

Without imposing constraints on the  $c_j$ 's it is well known that there are multiple choices of  $C(L)$  that will satisfy (2.2). Loosely put, there is an equivalence class of matrix lag operators that satisfy (2.2), members of which can be generated from another by post multiplying  $C(L)$  by matrices of "Blaschke factors." Since the spectral density matrix  $S$  summarizes all of

the population covariance properties of the  $y$  time series, alternative choices of matrices  $C(L)$  in the equivalence class characterized by (2.2) are observationally equivalent. Thus there is an identification problem in representing  $y$  as a one-sided moving average of white noise disturbances. In many circumstances, especially in problems involving prediction, this identification problem is partially resolved by choosing a "fundamental" representation, that is, a  $C(L)$  for which the associated contemporaneous white noise  $u_t$  lies in the space spanned by  $\{y_t, y_{t-1}, \dots\}$ .

In this paper, we are interested in estimating vector moving average representations in which rational expectations and economic theory impose a set of cross-equation restrictions on  $C(L)$ . As a prolegomenon to estimation, we study the question of whether these cross-equation restrictions eliminate the multiplicity of moving average representations. Put somewhat differently, the question is whether, assuming that the rational expectations restrictions are correct, there is among the equivalence class of  $C(L)$ 's that satisfy (2.2), a unique  $C(L)$  that satisfies the rational expectations restrictions. The answer to this question is somewhat ambiguous. However, we shall show that for one class of parameterizations of  $C(L)$  that might seem interesting for applied work, the cross-equation rational expectations restrictions do not eliminate the identification problem associated with the multiplicity of moving average representations.

As mentioned in Section 1, it is known from linear prediction theory that linearly indeterministic, covariance stationary stochastic processes have multiple moving average representations. The problem is that as given in (2.1)  $C(L)$  and  $u$  are not unique. There is both a relatively trivial and a nontrivial sense in which this is true. We describe the trivial sense first. Let  $D$  be an orthogonal matrix and form

$$(2.4) \quad u_t^* = Du_t.$$

Note that

$$(2.5) \quad \begin{aligned} Eu_t^* u_t^{*'} &= EDu_t u_t' D' \\ &= DD' \\ &= I. \end{aligned}$$

Let  $C^*(L) = C(L)D'$  and we obtain the equivalent representation

$$(2.6) \quad y_t = C^*(L)u_t^* = C(L)D'Du_t = C(L)u_t.$$

Thus, the relatively trivial identification problem emerges because each of the white noise vectors generated by multiplication by an orthogonal matrix spans the same linear space, i.e., each one gives rise to the same information space. To proceed with estimation, sufficient normalizations must be imposed to eliminate this nonuniqueness, trivial though it may be.

The nontrivial sense in which multiple moving average representations emerge is that  $y$  can be represented in terms of different vector white noises that span different linear spaces. In particular, Wold's Decomposition Theorem informs us that there exists a representation such that  $u_t \in \Phi_t$ , i.e.,  $u_t$  can be recovered from current and past observations of  $y$ . This is referred to as a fundamental representation. However, there also exist moving average representations that are not fundamental. In these nonfundamental representations the current and past  $u$ 's span a space that is strictly larger than  $\Phi_t$  and in particular  $u_t \notin \Phi_t$ .

The restrictions we derived in Section 1 are appropriate for a fundamental representation. This raises a question as to whether those restrictions are invariant under the multiplicity of representations. If we conclude that the restrictions hold for one of the representations, can we expect them to hold for another one of the moving average representations? There are two approaches to answering this question. The first approach is to employ transform methods, since shifting from one representation to another can be accomplished via multiplication by orthogonal matrices and Blaschke factors.<sup>8</sup> The second approach involves iterated projection arguments.<sup>9</sup> We will study this question first using projection arguments and then using  $z$  transform machinery. It turns out that the  $z$  transform approach provides us with a more complete answer.

In Section 1 we derived the following restrictions:

$$(2.7) \quad B(L)C_2(L) = [A(L^{-1})C_1(L)]_+.$$

Suppose that these restrictions hold for a particular moving average representation, not necessarily a fundamental one. Then it is true that these restrictions must hold for any other representation obtained by multiplying the vector white noise by an orthogonal matrix. The argument to prove this proposition can proceed in two ways. One strategy is to note that any white noise generated by an orthogonal transformation of the initial white noise must span the same linear space. Restrictions (2.6) imply that

$$(2.8) \quad \hat{E}[A(L^{-1})y_{1t} | \Sigma_t] = B(L)y_{2t}$$

where  $\Sigma_t$  is the linear space spanned by current and past values of the white noise  $u_t$  and where  $u_t \in \Omega_t$ . Multiplying  $u$  by an orthogonal matrix will give rise to a new noise  $u^*$ , but the new information set  $\Sigma_t^*$  spanned by  $(u_t^*, u_{t-1}^*, \dots)$  will be identical to  $\Sigma_t$ . That restrictions (2.6) continue to hold is just a simple application of the Wiener-Kolmogorov prediction formulas. Alternatively, the proposition can be verified mechanically by carrying out the multiplication of  $C(L)$  by the orthogonal matrix. We state the conclusion in Lemma 1.

Lemma 1: Suppose  $y_t = C(L)u_t$  where  $C(L)$  satisfies (2.7) and  $u$  is a white noise. Let  $D$  be an orthogonal matrix and let  $u_t^* = Du_t$  and  $C^*(L) = C(L)D'$ . Then  $C^*(L)$  satisfies (2.7).

The second lemma we discuss deals with the nontrivial sense in which there are multiple representations. It states the following:

Lemma 2: Suppose  $y_t = C(L)u_t$  where  $C(L)$  satisfies (2.7) and  $u$  is a white noise. Let  $u^*$  be fundamental for  $y$  and let  $C^*(L)$  be an operator  $C^*(L) = \sum_{j=0}^{\infty} c_j^* L^j$  such that  $y_t = C^*(L)u_t^*$ . Then  $C^*(L)$  satisfies (2.7).

The proof of this lemma is provided in Appendix A. It turns out to be a simple application of an iterated projections argument. After noting that  $\Sigma_t^* = \phi_t \subset \Sigma_t$ , then the proof is straightforward. The message that emerges from Lemma 2 is that once we know that the restrictions hold for a particular moving average representation, we know they must hold for any fundamental representation. A stronger result, that if the restrictions hold for one moving average representation they must hold for any other one, is apparently not true. However, it turns out that if we confine the class of moving average representations in an interesting way, the restrictions hold for all such representations. Before formally considering this result, it is fruitful to focus on some examples.

Example (i):            A case in which restrictions holding for a nonfundamental representation implies that they hold for a fundamental representation.

For our first example, suppose that we have rational expectations restrictions of the form

$$(2.9) \quad E[y_{1t+1} | \Omega_t] = y_{2t}.$$

Also assume that

$$(2.10) \quad \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 + 5L + 6L^2 & 1 \\ 5 + 6L & 0 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

where  $Eu_{1t}^2 = Eu_{2t}^2 = 1$ ,  $Eu_{1t}u_{2t} = 0$ , and  $(u_1, u_2)$  is a vector white noise. We assume that  $u_{1t}$  and  $u_{2t}$  are functions of elements in agents' information set  $\Omega_t$ . Then equations (2.9) and (2.10) imply that

$$(2.11) \quad \hat{E}[y_{1t+1} | \Sigma_t] = y_{2t}$$

where  $\Sigma_t$  is the closed linear space generated by  $\{u_{1t}, u_{1t-1}, \dots, u_{2t}, u_{2t-1}, \dots\}$ . Using (2.10), we have

$$(2.12) \quad y_{1t+1} = u_{1t+1} + 5u_{1t} + 6u_{1t-1} + u_{2t+1}.$$

Taking the linear least squares projection onto  $\Sigma_t$  gives

$$(2.13) \quad \hat{E}[y_{1t+1} | \Sigma_t] = 5u_{1t} + 6u_{1t-1} = \left[ \frac{1 + 5L + 6L^2}{L} \right]_+ u_{1t}.$$

Using representation (2.10) we also verify that

$$(2.14) \quad 5u_{1t} + 6u_{1t-1} = y_{2t}.$$

Thus, restrictions (2.11) are satisfied by representation (2.10).

Representation (2.10) implies that current and past  $(y_1, y_2)$ 's can be expressed in terms of current and past  $(u_1, u_2)$ 's. This yields the implication that  $\Sigma_t \supset \Phi_t$ , i.e., that at least as much information is contained in  $\Sigma_t$  as in  $\Phi_t$ . Notice that a fundamental representation for  $(y_1, y_2)$  is given by

$$(2.15) \quad \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1+5L+6L^2 & 1 \\ 5+6L & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{6+5L}{5+6L} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_{1t}^* \\ u_{2t}^* \end{bmatrix} \\ = \begin{bmatrix} \frac{1}{\sqrt{2}}(6L + 5L^2) & \frac{1}{\sqrt{2}}(2 + 5L + 6L) \\ \frac{1}{\sqrt{2}}(6 + 5L) & \frac{1}{\sqrt{2}}(5 + 6L) \end{bmatrix} \begin{bmatrix} u_{1t}^* \\ u_{2t}^* \end{bmatrix} \\ = C^*(L)u_t^*$$

where

$$(2.16) \quad \begin{bmatrix} u_{1t}^* \\ u_{2t}^* \end{bmatrix} = \begin{bmatrix} \frac{5 + 6L}{6 + 5L} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

In (2.16)  $u^*$  is obtained from  $u$  using matrix lag polynomials formed with "Blaschke factors." The theoretical spectral density for  $u^*$  is given by

$$(2.17) \quad \begin{bmatrix} \frac{5+6e^{-i\omega}}{6+5e^{-i\omega}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{5+6e^{i\omega}}{6+5e^{i\omega}} & 0 \\ 0 & 1 \end{bmatrix} = I$$

verifying that  $u^*$  is a vector white noise. The determinant of  $C^*(z)$  is given by

$$(2.18) \quad \det C^*(z) = -(6 + 5z).$$

Since  $\det C^*(z)$  has its zero at  $z = -6/5$  which is outside the unit circle, the matrix polynomial  $C^*(L)$  has a one-sided inverse. This verifies that  $u^*$  is fundamental for  $y^*$ . Finally,

$$(2.19) \quad \begin{bmatrix} \frac{1(6L + 5L^2)}{\sqrt{2}L} \end{bmatrix}_+ = \frac{1}{\sqrt{2}}(6 + 5L)$$

$$\begin{bmatrix} \frac{1(2 + 5L + 6L^2)}{\sqrt{2}L} \end{bmatrix}_+ = \frac{1}{\sqrt{2}}(5 + 6L)$$

verifying that the restrictions (2.7) hold for  $C^*(L)$ . This example, therefore, illustrates the implications of Lemma 2: the restrictions hold for a nonfundamental representation, and therefore also for a fundamental representation.

Example (ii):            An example where restrictions hold for a fundamental representation, but not for a nonfundamental representation.

Suppose that

$$(2.20) \quad \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \frac{1 - 5L^2 + 2L^3}{2 - L} & 1 \\ 1 - 2L & 0 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

where  $Eu_{1t}^2 = Eu_{2t}^2 = 1$ ,  $Eu_{1t}u_{2t} = 0$  and  $(u_1, u_2)$  is a vector white noise. Consider the restrictions

$$(2.21) \quad \hat{E}[y_{1t+1} | \Sigma_t] = y_{2t}.$$

where  $\Sigma_t$  is the closed linear space generated by  $\{u_{1t}, u_{1t-1}, \dots, u_{2t}, u_{2t-1}, \dots\}$ . These restrictions are not satisfied by (2.20) since we have

$$(2.22) \quad \left[ \frac{1 - 5L^2 + 2L^3}{(2 - L)L} \right]_+ \neq 1 - 2L$$

Now let

$$(2.23) \quad \begin{bmatrix} u_{1t}^* \\ u_{2t}^* \end{bmatrix} = \begin{bmatrix} \frac{1 - 2L}{2 - L} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} .$$

The theoretical spectral density function for  $u^*$  is given by

$$(2.24) \quad \begin{bmatrix} \frac{1 - 2e^{-i\omega}}{2 - e^{-i\omega}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1 - 2e^{i\omega}}{2 - e^{i\omega}} & 0 \\ 0 & 1 \end{bmatrix} = I$$

verifying that  $u^*$  is a vector white noise. From (2.20) and (2.23) we see that

$$(2.25) \quad \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 + 2L - L^2 & 1 \\ 2 - L & 0 \end{bmatrix} \begin{bmatrix} u_{1t}^* \\ u_{2t}^* \end{bmatrix}$$

$$= C^*(L)u_t^*.$$

Now

$$(2.26) \quad \det C^*(z) = -(2 - z)$$

which has a zero at  $z = 2$  that is outside the unit circle. Thus  $u^*$  is fundamental for  $y$ . Restrictions (2.21) imply that

$$(2.27) \quad \hat{E}[y_{1t+1} | \phi_t] = y_{2t}.$$

Since

$$(2.28) \quad \left[ \frac{1 + 2L - L^2}{L} \right]_+ = 2 - L \quad \text{and}$$

$$\left[ \frac{1}{L} \right]_+ = 0,$$

$C^*(L)$  satisfies restrictions (2.27). This example illustrates that the converse of Lemma 2 is not true in general. The cross-equation restrictions hold for the fundamental representation (2.25) but not for the nonfundamental representation (2.20). In this example  $u_t^* \in \Phi_t$ , however,  $u_t \notin \Phi_t$ .

We now turn to Lemma 3, which indicates that for a wide class of cases, the rational expectations restrictions do remain intact under alternative moving average representations. In such cases, the rational expectations restrictions fail uniquely to identify one from the equivalence class of moving average representations that is obtained by "flipping roots" via Blaschke factors. This result is of substantial use in interpreting econometric results.

Lemma 3: Suppose

- (i)  $y_t = C(L)u_t$  is a fundamental representation;
- (ii)  $C(z)$  is a matrix of rational functions with  $\mu(z)$  as the lowest common denominator polynomial of the elements of  $C(z)$ ;
- (iii)  $\det C(z) = \frac{\rho_1(z)\rho_2(z)}{\gamma(z)}$   
 where  $\rho_1(z)$ ,  $\rho_2(z)$ , and  $\gamma(z)$  are finite order polynomials,  $\rho_2(z)$  does not have zeroes in common with

$b_2(z)$  and  $\mu(z)$ , and  $\gamma(z)$  does not have zeroes in common with  $\rho_1(z)$  and  $\rho_2(z)$ ;

(iv)  $C(L)$  satisfies

$$B(L)C_2(L) = [A(L^{-1})C_1(L)]_+.$$

Consider any other representation  $y_t = C^*(L)u_t^*$ , not necessarily fundamental, that satisfies

(ii')  $C^*(z)$  is a matrix of rational function with  $\mu(z)$  as a common denominator polynomial of the elements of  $C^*(z)$ ;

$$(iii') \det C^*(z) = \frac{\rho_1(z)\rho_2^*(z)}{\gamma(z)}$$

where  $\rho_2^*(z)$  is a finite order polynomial that does not have a zero at  $z = 0$ ;

Then  $C^*(L)$  also satisfies the restrictions in (iv).

This lemma, which is proved in the Appendix A, implies that in an important class of models, the rational expectations restrictions do not uniquely identify the moving average representation. Thus, suppose we begin with a fundamental representation  $y_t = C(L)u_t$ . We can think of finding a lowest common denominator for the rational lag operator in each row of  $C(L)$ . This will allow us to write

$$(2.29) \quad C(z) = H(z)J(z)$$

where  $H(z)$  is a diagonal matrix consisting of the reciprocals of the common denominator of each row on each diagonal and  $J(z)$  is a finite order matrix polynomial. The zeroes of the denominators

of the diagonal elements of  $H(z)$  are zeroes of  $\mu(z)$ . We know that

$$(2.30) \quad \det C(z) = \det H(z) \det J(z)$$

and thus if  $\theta$  is a zero of  $\rho_2(z)$ , then

$$(2.31) \quad \det J(\theta) = 0.$$

We can always find an orthogonal matrix  $D$  such that  $J(\theta)D$  has zeroes in the first column. Let  $C^*(z)$  be given by

$$(2.32) \quad C^*(z) = H(z)J(z)DG(z)$$

where

$$(2.33) \quad G(z) = \begin{bmatrix} \frac{1 - \theta z}{z - \theta} & 0 \\ 0 & I \end{bmatrix} .$$

Dividing out common factors we can represent  $J(z)DG(z)$  as a finite order matrix polynomial.<sup>10</sup> Furthermore, the order of the polynomial elements of  $J(z)DG(z)$  are, in general, the same as those of  $J(z)D$  as long as the polynomial elements in any row of  $J(z)$  have the same order. Associated with  $C^*(L)$  we can define a white noise vector  $u^*$  such that  $y_t = C^*(L)u_t^*$ , where

$$(2.34) \quad u_t = D \begin{bmatrix} \frac{1 - \theta L}{L - \theta} & 0 \\ 0 & I \end{bmatrix} u_t^*$$

It can be verified that current and past values of  $u^*$  will span a larger space than current past values of  $y$ . Thus  $u^*$  is not fundamental for  $y$ . Calculating the  $\det C^*(z)$ , we determine that

$$(2.35) \quad \det C^*(z) = \det C(z) \frac{(1 - \theta z)}{(z - \theta)}$$

$$= \frac{\rho_1(z) \rho_2(z)}{\gamma(z)} \frac{(z - \theta)}{(1 - \theta z)}$$

$$= \frac{\rho_1(z) \rho_2^*(z)}{\gamma(z)}$$

where

$$\rho_2^*(z) = \rho_2(z) \frac{(1 - \theta z)}{(z - \theta)}$$

Since  $\theta$  is a zero of  $\rho_2(z)$ , we can divide out the common factors and claim that  $\rho_2^*(z)$  can be expressed as a polynomial. Lemma 3 allows us to assert that the rational expectations restrictions hold for  $C^*(L)$  given that they hold for  $C(L)$ . This is true even though  $u^*$  generates a larger information set than  $u$ .

Using Lemma 3 and the procedure described above, we can generate alternative nonfundamental one-sided moving average representations that satisfy the restriction. This is true as long as  $\rho_2(z)$  is not a constant and the polynomial elements in

any row of  $J(z)$  have the same order. Our choice of a particular zero of  $\rho_2(z)$  was arbitrary. Any element in the set of zeroes of  $\rho_2(z)$  could have been employed. Furthermore, this procedure could have been repeated using different zeroes of  $\rho_2(z)$ . All of this indicates that the rational expectations restrictions cannot, in general, be expected to pin down a unique moving average representation, but rather these restrictions will, in general, hold over a multiplicity of such representations. From the standpoint of estimation, this implies that multiple peaks in the likelihood function may exist. This is true even if we normalize to avoid the relatively trivial identification problem.

On the other hand, these problems in identification are not necessarily all that damaging. In cases in which the polynomial orders for elements in given rows of  $J(z)$  are assumed to be different, both the relatively trivial identification problem of Lemma 1 and the deeper identification problem of Lemma 3 are often less severe and sometimes nonexistent. In cases in which  $\rho_2(z)$  is constant, then Lemma 3 cannot be applied to generate alternative moving average representations that satisfy the restrictions. For example, if the  $y$  process is assumed to have a finite order vector autoregressive representation, the polynomial  $\rho_2(z)$  is constant. Even if there is a multiplicity of moving average representations, the possibility of testing the rational expectations restrictions is not destroyed, nor are the chances of estimating parameters of economic agents' assumed objective functions in dynamic optimization models ruined. Also, it is straightforward but sometimes computationally tedious to infer an

estimate of a fundamental representation from estimates of a possibly nonfundamental representation.

Lemma 3 provides us with a justification for ignoring the location of numerator zeroes for rational forms of  $C(z)$ . The same cannot be said of denominator zeroes of  $C(z)$ . If we let  $\mu(z)$  be the lowest common denominator polynomial of the elements of  $C(z)$ , then the zeroes of  $\mu(z)$  must be constrained to lie outside the unit circle. The prediction formulas used for deriving the rational expectations restrictions rely on the elements of  $C(L)$  being one-sided in nonnegative powers of  $L$ . When  $\mu(z)$  has zeroes inside the unit circle, this one-sided constraint is violated.

2b. Estimation

In estimating linear time series models, full scale maximum likelihood estimation with Gaussian density function is computationally impractical for most applications. This has lead to the proposal of several alternative procedures which use approximations to the likelihood function that ease the computational burden. These approximations are constructed so that the resulting estimators will be asymptotically equivalent to the maximum likelihood estimator. Hansen and Sargent (1980) describe a couple of these approximations and discuss how to impose the rational expectations restrictions for a closely related class of models.

For the purpose of this paper we follow the suggestion of Hannan (1970), Robinson (1976), Phadke and Kedem (1978), and Kohn (1979) and approximate the log likelihood function of a sample  $\{y_t, t = 1, 2, \dots, T\}$  as follows.<sup>11</sup> First we define the Fourier transform of the  $y$  sequence as

$$(2.36) \quad Y(\omega_j) = \sum_{t=1}^T y_t e^{-i\omega_j t}$$

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

$$(2.37) \quad I(\omega_j) = \frac{1}{T} Y(\omega_j)Y(\omega_j)'$$

where the prime denotes complex conjugation and transposition.

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

$$(2.38) \quad L = \frac{nT}{2} \log 2\pi - \frac{1}{2} \sum_{j=1}^{T-1} \log \det S(\omega_j) \\ - \frac{1}{2} \sum_{j=1}^{T-1} \text{trace } S(\omega_j)^{-1} I(\omega_j)$$

where  $S(\omega_j)$  is the theoretical spectral density defined in (2.2) and (2.3). Note that  $S(\omega_j)$  can be expressed in terms of the free parameters of  $C(L)$  using formula (2.2). To emphasize this, it is worthwhile substituting (2.2) into (2.38) to obtain

$$(2.39) \quad L = -\frac{nT}{2} \log 2\pi - \frac{1}{2} \sum_{j=1}^{T-1} \log \det [C(e^{-i\omega_j})C(e^{-i\omega_j})'] \\ - \frac{1}{2} \sum_{j=1}^{T-1} \text{trace } [C(e^{-i\omega_j})C(e^{-i\omega_j})'] I(\omega_j)$$

In computing (2.39), it is useful to exploit the fact that

$$\log \det [C(e^{-i\omega})C(e^{-i\omega})'] = \log \det \{C[e^{-i(2\pi-\omega)}]C[e^{-i(2\pi-\omega)}]'\} \\ (2.40) \\ \text{trace } [C(e^{-i\omega})C(e^{-i\omega})']^{-1} I(\omega) \\ = \text{trace } \{C[e^{-i(2\pi-\omega)}]C[e^{-i(2\pi-\omega)}]'\}^{-1} I(2\pi-\omega).$$

These formulas permit (2.39) to be rewritten in terms of sums over only  $T/2$  frequencies. The free parameters of  $C(L)$  are estimated by maximizing (2.39) over the free parameters of

$C(L)$ .<sup>12</sup> This is a nonlinear maximization problem, which can be solved by any of a variety of algorithms that are described by Bard (1974). Phadke and Kedem (1978) suggest a modification of this procedure in situations in which  $y$  can be represented as a finite order vector moving average that will yield exact maximum likelihood estimates.<sup>13</sup>

### 3. Applications to the Term Structure of Interest Rates

In this section we illustrate our methods with expectations models of the term structure. The techniques illustrated have important counterparts in other potential applications of linear rational expectations models as well. First, we discuss the joint modeling of interest rates of various maturities with special focus on situations in which the sampling interval is finer than the term of the shortest term interest rate. The implications of this discussion carry over to expectations models of forward and spot prices when the length of the forward contract exceeds the sampling interval. For the purposes of this discussion, it turns out to be useful to begin with a continuous time expectations model of the term structure. Second, we discuss situations in which first differences of  $y_1$  are assumed covariance stationary rather than levels. Although we use this discussion to obtain a modified term structure model, our treatment is general enough to apply to model testing whenever  $A(L^{-1})$  and  $B(L)$  are specified a priori. Third, we specify the precise form of the term structure models which we estimated and provide some details of how the estimation was carried out. Fourth, we discuss our empirical results.

a. Implications of a Continuous Time Model of the Term Structure

Let us consider a version of the expectations model of the term structure. Let  $r_t$  denote an instantaneous interest rate and  $R_{kt}$  denote a  $k$ -period interest rate realized at time  $t$ . We assume that

$$(3.1) \quad R_{kt} = \frac{1}{k} E \left[ \int_0^k r_{t+s} ds \mid \Omega_t \right]$$

where  $\Omega_t$  is the information set of economic agents which possibly includes the past continuous record of interest rates of various maturities. Relation (3.1) is assumed to hold for any maturity of length  $k$ . Suppose the econometrician observes  $p$ -period and  $q$ -period interest rates at integer points in time where both  $p$  and  $q$  are integers. Relation (3.1) implies that

$$(3.2) \quad R_{pt} = \frac{1}{p} \hat{E} \left[ \int_0^p r_{t+s} ds \mid \Phi_t \right] \\ = \frac{1}{p} \hat{E} [R_{1t} + R_{1t+1} + \dots + R_{1t+p-1} \mid \Phi_t]$$

and

$$(3.3) \quad R_{qt} = \frac{1}{q} \hat{E} \left[ \int_0^q r_{t+s} ds \mid \Phi_t \right] \\ = \frac{1}{q} \hat{E} [R_{1t} + R_{1t+1} + \dots + R_{1t+q-1} \mid \Phi_t]$$

where  $\Phi_t$  is the econometrician's information set that includes at least current and past observations of  $R_p$  and  $R_q$  at integer points in time but not necessarily  $R_1$ . Let us define

$$(3.4) \quad \hat{R}_{1t} = \hat{E}[R_{1t} | \phi_t].$$

By definition,  $\hat{R}_{1t}$  is an element of  $\phi_t$ . An iterated projection argument can be used to establish

$$(3.5) \quad R_{qt} = \frac{1}{q} \hat{E}[A_q(L^{-1}) \hat{R}_{1t} | \phi_t]$$

$$R_{pt} = \frac{1}{p} \hat{E}[A_p(L^{-1}) \hat{R}_{1t} | \phi_t]$$

where

$$(3.6) \quad A_q(L^{-1}) = \frac{1}{q} [1 + L^{-1} + \dots + L^{-q+1}].$$

$$A_p(L^{-1}) = \frac{1}{p} [1 + L^{-1} + \dots + L^{-p+1}].$$

One possible way to test the rational expectations term structure restrictions is to assume that

$$(3.7) \quad \hat{R}_{1t} = C_0(L)u_t$$

where  $u$  is a white noise vector that is fundamental for the  $y$  process consisting of variables observed by the econometrician. Recall that  $y$  contains both  $R_q$  and  $R_p$ . As in Section 1, we assume  $y$  is covariance stationary and write

$$(3.9) \quad y_t = \begin{bmatrix} R_{qt} \\ R_{pt} \\ x_t \end{bmatrix}$$

$$C(L) = \begin{bmatrix} C_1(L) \\ C_2(L) \\ C_3(L) \end{bmatrix}$$

where  $x$  is a vector of other variables observed by the econometrician. Restrictions (3.5) and (3.6) imply that

$$(3.10) \quad [A_q(L^{-1})C_0(L)]_+ = C_1(L)$$

$$[A_p(L^{-1})C_0(L)]_+ = C_2(L).$$

The term structure restrictions can be tested by first estimating the parameters of  $C_0$ ,  $C_1$ ,  $C_2$  and  $C_3$  subject to the restrictions (3.10). The maximized value of the likelihood function with these restrictions imposed should be compared to the maximized value of the likelihood function used to estimate the free parameters of  $C_1$ ,  $C_2$  and  $C_3$  without imposing restrictions (3.10). In this formulation  $\hat{R}_{1t}$  can be thought of as a "hidden variable" that is not included in  $y_t$ . However, the parameters of  $C_0$  remain estimable.<sup>14</sup>

An equivalent testing procedure is available when  $p$  is an integer multiple of  $q$ , i.e., when  $p = mq$  where  $m$  is a positive integer. Term structure model (3.1) implies that

$$(3.11) \quad R_{pt} = \hat{E}[A(L^{-1})R_{qt} | \phi_t]$$

where

$$A(L^{-1}) = \frac{1}{m}[1 + L^{-q} + \dots + L^{-(m-1)q}]$$

This specification is a special case of the general model given in Section 1. Restrictions (3.11) imply that

$$(3.12) \quad [A(L^{-1})C_1(L)]_+ = C_2(L)$$

The equivalence of testing restrictions (3.10) and restrictions (3.12) is established in Appendix B.

In situations in which more than two maturities are used in estimation, both of the strategies described above can be generalized in the obvious ways.<sup>15</sup>

b. 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  $y_1$  is covariance stationary. We maintain the model restrictions (1.4) which for convenience are written below:

$$(3.13) \quad E[A(L^{-1})y_{1t} | \Omega_t] = B(L)y_{2t}.$$

Recall that  $A(L^{-1}) = \sum_{j=0}^{\infty} a_j L^{-j}$ . Let

$$(3.14) \quad \begin{aligned} a_j^* &= \sum_{k=j}^{\infty} a_k && \text{for } j = 0, 1, \dots, \\ y_{1t}^* &= y_{1t} - y_{1t-1}, \end{aligned}$$

and

$$A^*(L^{-1}) = \sum_{j=1}^{\infty} a_j^* L^{-j}.$$

Now  $a_j = a_{j+1}^* - a_j^*$  implying that

$$(3.15) \quad A(L^{-1})y_{1t} = A^*(L^{-1})y_{1t}^* + a_0^* y_{1t}.^{16}$$

Substituting (3.15) into (3.12), we obtain

$$(3.16) \quad E[A^*(L^{-1})y_{1t}^* | \Omega_t] = B(L)y_{2t} - a_0^* y_{1t}.$$

Suppose that  $x_t$  is vector of variables not contained in  $[y_{1t}', (B(L)y_{2t} - a_0^*y_{1t})']$ . Furthermore, suppose that  $B(L)$  is specified a priori. Let

$$(3.17) \quad y_{2t}^* = \begin{bmatrix} B(L)y_{2t} - a_0^*y_{1t} \\ x_t \end{bmatrix}$$

$$b_0^* = [I, 0].$$

We can write the first difference model as

$$(3.18) \quad E[A^*(L^{-1})y_{1t}^* | \Omega_t] = b_0^*y_{2t}^*$$

and assume that  $y^{*'} = (y_1^{*'}, y_2^{*'})$  is 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 (1979a). In particular, Sargent first differenced (3.12) to obtain

$$(3.19) \quad E[A(L^{-1})y_{1t} | \Omega_t] - E[A(L^{-1})y_{1t-1} | \Omega_{t-1}] = B(L)y_{2t} - B(L)y_{2t-1}.$$

Sargent projected both sides of (3.19) onto  $\Omega_t$  to obtain

$$(3.20) \quad \hat{E}[A(L^{-1})y_{1t}^* | \phi_{t-1}] = \hat{E}[B(L)(1-L)y_{2t} | \phi_{t-1}].$$

Although restrictions (3.20) can be tested using procedures discussed in this paper, some implications of (3.13) are lost by projecting onto  $\phi_{t-1}$  rather than  $\phi_t$ . On the other hand, (3.18) involves a projection onto  $\phi_t$  rather than  $\phi_{t-1}$  and imposes more restrictions than (3.20). It is therefore quite possible that the procedures proposed in this section can detect empirical contradictions of the hypothesis (3.13) that Sargent's procedure could not.

c. Implementation

In our empirical example, we used monthly observations of a three-month interest rate and a five-year interest rate for our y vector time series. Let  $p = 60$  and  $q = 3$ . Four different term structure models were estimated. The first two are versions of a model used by Sargent (1979a). Sargent tested the restrictions

$$(3.19) \quad \hat{E}\left[\frac{1}{m}(R_{q,t} + R_{q,t+q} + \dots + R_{q,t+(m-1)q} \mid \Phi_t)\right]$$

where  $\Phi_t = \{R_{qt}, R_{qt-1}, \dots, R_{pt}, R_{pt-1}, \dots\}$  and where  $m = 20$ . One possibility is to assume that deviations from a constant and linear time trend of  $(R_q, R_p)'$  are covariance stationary and that (3.19) applies to the detrended versions of  $R_q$  and  $R_p$ . We refer to this as Model I. A second alternative is assumed that the first differences in  $R_q$  are covariance stationary and that (3.19) applies to the levels. Following the strategy described in Section 3b, we know that

$$(3.20) \quad \begin{aligned} & \hat{E}\left[\left(\frac{m-1}{m}\right)(\Delta R_{q,t+1} + \Delta R_{q,t+2} + \dots + \Delta R_{q,t+q}) \right. \\ & \quad + \left(\frac{m-2}{m}\right)(\Delta R_{q,t+q+1} + \Delta R_{q,t+q+2} + \dots + \Delta R_{q,t+2q}) + \dots \\ & \quad + \left(\frac{1}{m}\right)(\Delta R_{q,t+(m-2)q+1} + \Delta R_{q,t+(m-2)q+2} + \dots \\ & \quad \left. + \Delta R_{q,t+(m-1)q}\right) \mid \Phi_t) \\ & = R_{p,t} - R_{q,t} \end{aligned}$$

where  $\Delta R_{q,t} = R_{q,t} - R_{q,t-1}$ . We can assume that  $(\Delta R_q, R_p - R_q)$  is covariance stationary and that restrictions (3.20) apply to this differenced time series. We refer to this as Model II.

Shiller (1979a) has used an expectations term structure model of the form

$$(3.21) \quad \hat{E} \left[ \frac{1-\gamma}{1-\gamma^m} (R_{q,t} + \gamma R_{q,t+q} + \dots + \gamma^{m-1} R_{q,t+(m-1)q}) \mid \Phi_t \right] \\ = R_{p,t}.$$

Following Shiller we set  $\gamma = .98$ . Again we have two strategies available. One is to assume the (3.21) applies to deviations about a linear time trend and that these deviations are jointly covariance stationary. We refer to this as Model III. The second strategy is assume that the first difference in  $R_q$  is covariance stationary. The first difference model takes the form

$$(3.22) \quad \hat{E} \left[ \left( \frac{1-\gamma}{1-\gamma^m} \right) (\gamma + \gamma^2 + \dots + \gamma^{m-1}) (\Delta R_{q,t+1} + \Delta R_{q,t+2} + \dots \right. \\ \left. + \Delta R_{q,t+q}) + \left( \frac{1-\gamma}{1-\gamma^m} \right) (\gamma^2 + \dots + \gamma^{m-1}) (\Delta R_{q,t+q+1} \right. \\ \left. + \Delta R_{q,t+q+2} + \dots + \Delta R_{q,t+2q}) + \dots + \left( \frac{1-\gamma}{1-\gamma^m} \right) \gamma^{m-1} \right. \\ \left. (\Delta R_{q,t+(m-2)q+1} + \Delta R_{q,t+(m-2)q+2} + \dots \right. \\ \left. + \Delta R_{q,t+(m-1)q}) \mid \Phi_t \right] \\ = R_{p,t} - R_{q,t}.$$

We can assume that  $(\Delta R_q, R_p - R_q)$  is covariance stationary and that restrictions (3.22) apply to this differenced time series. We refer to this as Model IV. Before we examine these results we need to consider a few more details about estimation.

All four models are special cases of the model discussed in Sections 1 and 2. To estimate and test these models, we need to impose finite parameterizations on the lag polynomial  $C(L)$  that also satisfy the restrictions. We parameterize the elements of  $C(L)$  as rational polynomials so that the  $y$  process is assumed to be a mixed autoregressive moving average model. Interest rate data from the U.S. are known to have what Granger (1966) has dubbed as "the typical spectral shape." That is, they are characterized by a covariogram which damps very slowly. The first differencing and autoregressive parameters can accommodate this feature of the time series.<sup>18</sup>

For each of the models we assume that

$$(3.23) \quad \begin{bmatrix} R_{qt} \\ R_{pt} \end{bmatrix} = \begin{bmatrix} \frac{\alpha_1(L)}{\beta(L)} & \frac{\alpha_2(L)}{\beta(L)} \\ \frac{\eta_1(L)}{\beta(L)} & \frac{\eta_2(L)}{\beta(L)} \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

where

$$\beta(L) = 1 + \beta_1 L + \dots + \beta_{n_0} L^{n_0}$$

$$(3.24) \quad \alpha_j(L) = \alpha_{j0} + \alpha_{j1} L + \dots + \alpha_{jn_1} L^{n_1}$$

$$\eta_j(L) = \eta_{j0} + \eta_{j1} L + \dots + \eta_{jn_2} L^{n_2}$$

$j = 1, 2$ . We impose the normalization that  $\alpha_{20} = 0$ .

It turns out that the term structure Models I-IV impose restrictions on the  $\alpha$ 's and  $\eta$ 's. In particular, if the  $\alpha$ 's are known, the  $\eta$ 's can be calculated from the restrictions. In our application we computed the  $\eta$ 's recursively. For all four models  $A(L^{-1})$  is a finite order polynomial in the lead operator  $L^{-1}$ . That is

$$(3.25) \quad A(L^{-1}) = A_0 + A_1 L + \dots + A_k L^{-k}.$$

Now

$$(3.26) \quad [A(L^{-1}) \frac{\alpha_j(L)}{\beta(L)}]_+ = \frac{\eta_j(L)}{\beta(L)}$$

or

$$(3.27) \quad A(L^{-1}) \frac{\alpha_j(L)}{\beta(L)} - \psi_j(L^{-1}) = \frac{\eta_j(L)}{\beta(L)}$$

where  $\psi_j(L^{-1}) = \psi_{j1} L^{-1} + \psi_{j2} L^{-2} + \dots + \psi_{jk} L^{-k}$ .

Thus

$$(3.28) \quad A(L^{-1})\alpha_j(L) - \psi_j(L^{-1})\beta(L) = \eta_j(L).$$

For all of our applications,  $k$  was larger than  $n_0+1$  and  $n_1+1$ . Performing the multiplication and solving for the  $\psi_j$ 's recursively, we obtain

$$\psi_{jk} = A_k \alpha_{j0}$$

$$(3.29) \quad \psi_{jk-1} = A_{k-1} \alpha_{j0} + A_k \alpha_{j1} - \psi_{jk} \beta_1$$

·  
·  
·

$$\psi_{j1} = A_1 \alpha_{j0} + A_2 \alpha_{j1} + \dots + A_{n_1+1} \alpha_{jn_1}$$

$$- (\psi_{j2} \beta_1 + \psi_{j3} \beta_2 + \dots + \psi_{jn_0+1} \beta_{n_0}).$$

Using the  $\psi$ 's we calculated the  $\eta$ 's as follows

$$\eta_{j0} = A_0 \alpha_{j0} + A_1 \alpha_{j1} + \dots + A_{n_1} \alpha_{jn_1}$$

$$- (\psi_{j1} \beta_1 + \psi_{j2} \beta_2 + \dots + \psi_{jn_0} \beta_{n_0})$$

$$(3.30) \quad \eta_{j1} = A_0 \alpha_{j1} + A_1 \alpha_{j2} + \dots + A_{n_1-1} \alpha_{jn_1}$$

$$- (\psi_{j1} \beta_2 + \psi_{j2} \beta_3 + \dots + \psi_{jn_0-1} \beta_{n_0})$$

$$\eta_{jn_1} = A_0^{\alpha} j_{n_1}.$$

Formulas (3.30) assume that  $n_1 > n_0$ .

Using formulas (3.29) and (3.30) and a hypothetical choice of  $\alpha_1$ ,  $\alpha_2$  and  $\beta$  it is straightforward to evaluate the frequency domain approximation to the log likelihood function given in Section 2b. Estimates of the parameters of  $\alpha_1$ ,  $\alpha_2$  and  $\beta$  for the four models were obtained by maximizing this approximate log likelihood function. The asymptotic covariance matrix of the parameter estimator is estimated by minus the inverse of the Hessian of the log likelihood function evaluated at the estimated parameter values. Unrestricted models were estimated by freeing the parameters of  $\eta_1$  and  $\eta_2$  and maximizing the corresponding frequency domain approximation to log likelihood function. The restrictions were tested by comparing the maximum values of the restricted and its associated unrestricted log likelihood functions.

d. Empirical Results

To illustrate the methods discussed in this paper, we have estimated versions of Models I-IV using monthly U.S. data. In estimating Models I and III, we used observations from January 1959 through June 1971. In estimating Models II and IV, we used observations from March 1959 through June 1971. We estimated bivariate models using the five-year government bond rate and the three-month treasury bill rate. The data are point-in-time, first of month observations.<sup>19</sup> In estimating Models I and III, the yields were each regressed on a constant and linear trend. The residuals from these regressions were then taken as the data used to estimate the parameters.

Tables 1-3 report estimates for Models I-IV. These tables also contain estimates from unrestricted parameterizations within which the expectations term structure models are nested. The  $\eta$  parameters reported in these tables were calculated using the  $\alpha$  parameters and the  $\beta$  parameters in the restricted parameterizations but were not constrained in the the unrestricted runs. In the tables  $L_r$  denotes the maximum value of the log likelihood attained under the restrictions, while  $L_u$  denotes the maximum value attained under the unrestricted parameterization. The tables also report marginal confidence levels which are the probabilities under the null hypothesis that a chi square random variable with degrees of freedom equal to the number of restrictions would be less than or equal to the reported value of  $-2(L_r - L_u)$ .<sup>20</sup>

In Table 1 we report the results for Models I and III using detrended data. The lag polynomials  $\alpha_1(L)$  and  $\alpha_2(L)$  are both second order and  $\beta(L)$  is first order. The likelihood ratio test statistics indicate strong rejection of both models. However, the estimated autoregressive parameter  $\beta_1$  is -1 for both Models I and III and the unrestricted model. The quality of frequency domain approximation to the likelihood functions deteriorates when autoregressive or moving average roots get close to the unit circle. For this reason the test statistics and estimated parameters should be examined with some skepticism. This suggests that the first difference results may be of more interest.

Table 2 reports results for Models II and IV using first differenced data. Again the lag polynomials  $\alpha_1(L)$  and  $\alpha_2(L)$  are both second order and  $\beta(L)$  is first order. The likelihood ratio test statistics indicate strong rejection of both models but not as strong as in Table 1. Table 3 reports results for Models II and IV with richer specifications of the lag polynomials  $\alpha_1(L)$ ,  $\alpha_2(L)$  and  $\beta(L)$ . Both  $\alpha_1(L)$  and  $\alpha_2(L)$  are third order while  $\beta(L)$  is second order. Again the likelihood ratio test statistics provide strong evidence against the expectations term structure restrictions.

Overall, the results yield considerable stronger evidence against the expectations term structure hypothesis than was reported earlier by Sargent (1979a). One plausible explanation for this difference lies in the fact mentioned in a previous subsection that Sargent's procedure tests fewer restrictions

TABLE 1

Estimated Term Structure Models Using Detrended Data

	Model I	Model III	Unconstrained
$\beta$ (L)	1.0000 -1.0000L (.0004)	1.0000 -1.0000L (.0004)	1.0000 - .9999L
$\alpha_1$ (L)	.4272 - .0596L- .2219L <sup>2</sup> (.0053) (.0011) (.0029)	.4272 - .0593L- .2218L <sup>2</sup> (.0003) (.0048) (.0178)	.2070 - .0698L- .0002L <sup>2</sup>
$\alpha_2$ (L)	.1122L+ .1245L <sup>2</sup> (.0015) (.0020)	.1123L+ .1246L <sup>2</sup> (.0009) (.0010)	.2769L+ .0968L <sup>2</sup>
$\eta_1$ (L)	.1598 - .0030L- .0111L <sup>2</sup>	.1630 - .0036L- .0133L <sup>2</sup>	.2383 - .0393L- .0082L <sup>2</sup>
$\eta_2$ (L)	.2249 + .0056L+ .0062L <sup>2</sup>	.2226 + .0068L+ .0075L <sup>2</sup>	.1206 + .0160L+ .0456L <sup>2</sup>
L	68.1309	67.8941	114.5817
$-2(L_r - L_u)$	92.9015	93.3751	
Marginal Confidence Level	1.0000	1.0000	

TABLE 2

Estimated Term Structure Models Using First-Difference Data

	Model II	Model IV	Unconstrained
$\beta(L)$	1.0000 - .9301L (.0120)	1.0000 - .9306L (.0247)	1.0000 - .9188L (.0291)
$\alpha_1(L)$	.3095 - .2725L - .0282L <sup>2</sup> (.0263) (.0150) (.0029)	.3096 - .2727L - .0287L <sup>2</sup> (.0040) (.0078) (.0048)	.3025 - .2555L - .0329L <sup>2</sup> (.0056) (.0088) (.0046)
$\alpha_2(L)$	.1034L - .0849L <sup>2</sup> (.0063) (.0072)	.1035L - .0845L <sup>2</sup> (.0090) (.0392)	.0486L - .0555L <sup>2</sup> (.0069) (.0082)
$\eta_1(L)$	-.1310 + .0268L	-.1308 - .0270L	-.1372 - .0323L (.0013) (.0081)
$\eta_2(L)$	.2159 - .0806L	.2163 - .0794L	.2023 - .1185L (.0104) (.0141)
L	118.8009	118.4766	128.9158
$-2(L_r - L_u)$	20.2298	20.8784	
Marginal Confidence Level	.9995	.9997	

TABLE 3

## Estimated Term Structure Models Using First-Difference Data

	Model II	Model IV	Unconstrained
$\beta(L)$	1.0000 -1.0383L +.1052L <sup>2</sup>	1.0000 -1.0753L +.1404L <sup>2</sup>	1.0000 -1.0954L -.0373L <sup>2</sup>
$\alpha_1(L)$	.3089 -.2925L -.0089L <sup>2</sup> +.0014L <sup>3</sup>	.3082 -.3002L -.0042L <sup>2</sup> -.0044L <sup>3</sup>	.3062 -.2235 -.0604L <sup>2</sup> -.0001L <sup>3</sup>
$\alpha_2(L)$	-.0498L +.1120L <sup>2</sup> -.0467L <sup>3</sup>	-.0518L +.1124L <sup>2</sup> -.0452L <sup>3</sup>	.0436L -.0708L <sup>2</sup> +.0151L <sup>3</sup>
$\eta_1(L)$	-.1250 -.0252L +.0013L <sup>2</sup>	-.1239 -.0236L +.0041L <sup>2</sup>	-.1408 -.0468L -.0683L <sup>2</sup>
$\eta_2(L)$	.2112 -.1137L -.0443L <sup>2</sup>	.2116 -.1207L -.0425L <sup>2</sup>	1.881 -.0846L +.0289L <sup>2</sup>
L	122.0983	121.8576	137.7130
$-2(L_1 - L_2)$	31.2295	31.7109	
Marginal Confidence Level	1.0000	1.0000	

implied by models.<sup>21</sup> The results in this paper are consistent with findings of Shiller (1979a) and Singleton (1980a) who examined these models using variance bounds tests.

As is evident from the lemmas of Section 2, we are not assured that the resulting estimated implicit moving average representations are fundamental. In checking this we verified that all of the moving representations are fundamental with the exception of the unconstrained run in Table 3. We are guaranteed, however, of an observationally equivalent fundamental representation which yields the same maximized value of the likelihood function. Therefore, the values of the likelihood ratio test statistics in Table 3 are not affected by the fact that the implied unrestricted moving average representation is not fundamental.

Beyond the conclusions from the test statistics, the main value of results such as those portrayed in Tables 1-3 is to provide statistical representations of the joint process estimated subject to the cross-equation restrictions. Such representations are valuable for at least two reasons. First, in rational expectations models that are more general than, but resemble the current models, estimating such representations is often an essential step in the process of overcoming Lucas's critique of econometric policy evaluation procedures (see Hansen and Sargent (1980)). Second, as Sims (1980) and Litterman (1979) have argued, for purposes of so-called unconditional forecasting, it may be wise to use a vector moving average constrained by even a false null hypothesis that economizes on the number of

parameters to be estimated. In this sense, the model-constrained results such as those presented in Tables 1-3 could be useful for forecasting even if one respects the evidence which our procedures have turned up against the term structure restrictions.

4. A Comparison of Variance Bounds to Tests of Constraints on the Moving Average Coefficients

An alternative procedure exists for testing the rational expectations restrictions in situations in which  $A(L^{-1})$  and  $B(L)$  are specified a priori. Leroy and Porter (1979), Shiller (1979a) and Singleton (1980a) have proposed variance bounds tests. These tests rest on the observation that the restrictions

$$E[A(L^{-1})y_{1t} | \Omega_t] = B(L)y_{2t}$$

imply that

$$\begin{aligned} (4.1) \quad & E\{[A(L^{-1})y_{1t}][A(L^{-1})y_{1t}]'\} \\ & \geq E\{[B(L)y_{2t}][B(L)y_{2t}]'\} \\ & \geq E\{\{\hat{E}[A(L^{-1})y_{1t} | \Psi_t]\}\{\hat{E}[A(L^{-1})y_{1t} | \Psi_t]\}'\} \end{aligned}$$

where  $\Psi_t \subset \Omega_t$ . Alternative choices of  $\Psi_t$  give rise to different lower bounds on the variance of  $B(L)y_{2t}$ .

The constraints which we placed on the moving average representation of  $y$  embody the variance bounds restrictions in (4.1) for a whole family of possible choices of  $\Psi_t$ . Recall that we imposed the restrictions

$$(4.2) \quad B(L)C_2(L) = [A(L^{-1})C_1(L)]_+$$

Now

$$\begin{aligned}
 (4.3) \quad & E\{[A(L^{-1})y_{1t}][A(L^{-1})y_{1t}]'\} \\
 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} A(e^{i\omega})C_1(e^{-i\omega})\{A(e^{i\omega})C_1(e^{-i\omega})\}'d\omega \\
 &\geq \frac{1}{2\pi} \int_{-\pi}^{\pi} [A(e^{i\omega})C_1(e^{-i\omega})]_+[A(e^{i\omega})C_1(e^{-i\omega})]_+'d\omega \\
 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} B(e^{i\omega})C_2(e^{-i\omega})[B(e^{i\omega})C_2(e^{-i\omega})]'d\omega \\
 &= E\{[B(L)y_{2t}][B(L)y_{2t}]'\}.
 \end{aligned}$$

Thus, the upper bound given in (4.1) is satisfied. For the lower bound we begin by choosing  $\Psi_t = \Phi_t$ . Since  $B(L)y_{2t} \in \Phi_t$ , we can claim that

$$(4.4) \quad E[A(L^{-1})y_{1t}|\Omega_t] = \hat{E}[A(L^{-1})y_{1t}|\Phi_t] = B(L)y_{2t}$$

and that

$$\begin{aligned}
 (4.5) \quad & E\{[B(L)y_{2t}][B(L)y_{2t}]'\} \\
 &= E\{\{\hat{E}[A(L^{-1})y_{1t}|\Psi_t]\}\{\hat{E}[A(L^{-1})y_{1t}|\Psi_t]\}'\}.
 \end{aligned}$$

Thus the lower bound given in (4.2) is satisfied for  $\Psi_t = \Phi_t$ . It follows immediately that the rational expectations restrictions also imply that the lower variance bound holds for any subset  $\Psi_t \subset \Phi_t$ .

This shows that in a population sense the variance bounds restrictions (4.1) are implied by the restrictions (4.2) that we have imposed on coefficients of the moving average representation. The converse, however, is not, in general, true. That is, it may be possible to satisfy the variance bounds restrictions without necessarily satisfying the restrictions on the moving average coefficients. Thus, tests of the restrictions on the moving average coefficients embody a richer set implications of rational expectations than variance bounds tests. Geweke (1979) has made essentially the same point in the context of a regression formulation of the restrictions. The testing procedures which we advocate differ from those suggested by Geweke (1979) in that we do not conduct our tests as a linear restriction on a regression equation. On the other hand, asymptotically efficient estimation of the regression parameters in Geweke's formulation can be accomplished in a manner very similar to what we have proposed in this paper.<sup>22</sup>

## 5. 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 method 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.

We carry out the discussion in the context of the following modified version of the labor demand schedule described in Section 1:

$$(5.1) \quad n_t - \lambda n_{t-1} = - \frac{\lambda}{\delta} E \left[ \sum_{j=0}^{\infty} (\lambda \gamma)^j w_{t+j} \mid \Omega_t \right] \\ + \frac{\lambda}{\delta} E \left[ \sum_{j=0}^{\infty} (\lambda \gamma)^j a_{t+j} \mid \Omega_t \right]$$

where  $0 < \lambda < 1$ ,  $0 < \gamma < 1$ ,  $\delta > 0$ ,  $n_t$  is the stock of employment,  $w_t$  the real wage, and  $a_t$  a shock to productivity.

We assume that while current and lagged  $a$ 's are in the information set of private agents at time  $t$ , they are not included in the data available to the econometrician. As in Sargent (1978b) and Hansen and Sargent (1980), this provides an interpretation of an error term in the relation to be estimated by the econometrician. Let

$$(5.2) \quad y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}$$

where

$$(5.3) \quad y_{1t} = \begin{bmatrix} a_t \\ w_t \end{bmatrix} \quad y_{2t} = \begin{bmatrix} n_t \\ x_t \end{bmatrix}$$

where  $x$  has  $r$  rows and contains the variables that help predict  $w_t$ . Further, let

$$A_1(L^{-1}) = \frac{\lambda}{\delta}(1 - \lambda\gamma L^{-1})^{-1}$$

$$(5.4) \quad A(L^{-1}) = [A_1(L^{-1}), -A_1(L^{-1})]$$

$$B(L) = [(1 - \lambda L), 0]$$

Then assumption (5.1) can be rewritten

$$(5.5) \quad E[A(L^{-1})y_{1t} | \Omega_t] = B(L)y_{2t}.$$

The methods of Sections 1 and 2 can readily be used to represent this model in moving average form and to estimate its free parameters once we make some additional assumptions about

the relationship between economic agents' information set  $\Omega_t$  and the information set  $\phi_t = \{y_t, y_{t-1}, \dots\}$ . The first set of assumptions we consider are

$$(5.6) \quad E[y_{t+1} | \Omega_t] = \hat{E}[y_{t+1} | \phi_t]$$

and that

$$(5.7) \quad \{y_{1t}, y_{1t-1}, \dots, x_t, x_{t-1}, \dots\} = \phi_t.$$

Assumption (5.6) says the expectation of  $y_{t+1}$  conditioned on  $\Omega_t$  coincides with the linear least squares projection of  $y_{t+1}$  onto the reduced information set  $\phi_t$ . Assumption (5.7) says that if we were to observe all of  $y$  including  $a$ , there would be a redundancy in the information set  $\phi_t$ . This is consistent with the notion that if an econometrician were to observe disturbance terms there would be an exact relationship among variables observed by the econometrician. In the models considered in the previous section, stochastic singularities of this nature were ruled out because of an implicit omitted information interpretation of disturbance terms. Thus in the absence of assumption (5.6), assumption (5.7) would seem to be hard to defend.

With these assumptions we know that the fundamental representation for  $y$  can be written

$$(5.8) \quad \begin{bmatrix} a_t \\ w_t \\ y_{2t} \end{bmatrix} = \begin{bmatrix} C_{11}(L) & C_{12}(L) \\ C_{21}(L) & C_{22}(L) \\ C_{31}(L) & C_{32}(L) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

where  $C_{j1}$  has one column and  $C_{j2}$  has  $r$  columns for  $j = 1, 2, 3$  and  $u_1$  is a one-dimensional subvector and  $u_2$  is an  $r$  dimensional subvector of the vector white noise  $u$ . Consistent with assumption (5.7) the dimension of  $u$  is one less than the dimension of  $y$ . The assumption that  $u$  is fundamental for  $y$  carries with it the implication that at least one of the  $r+1$  dimensional minors of

$$(5.9) \quad C(z) = \begin{bmatrix} C_{11}(z) & C_{12}(z) \\ C_{21}(z) & C_{22}(z) \\ C_{31}(z) & C_{32}(z) \end{bmatrix}$$

do not have zeroes inside the unit circle. To have hope of identifying all of the parameters in  $C(L)$ ,  $\lambda$ ,  $\delta$ , and  $\gamma$ , we have to make an additional assumption about how  $a$  interacts with the remaining  $y$  vector. We assume that

$$(5.10) \quad \hat{E}[a_{t+1} | a_t, a_{t-1}, \dots] = \hat{E}[a_{t+1} | \phi_t]$$

that is, we assume that no other variables in  $\phi_t$  Granger-cause  $a$ . Assumption (5.10) allows us to set

$$(5.11) \quad C_{12}(L) = 0$$

and claim that the one-step-ahead forecast error in forecasting  $a_{t+1}$  is a scalar multiple of  $u_{1t+1}$ . Restrictions (5.11) and (5.6) imply that

$$(5.12) \quad [A_1(L^{-1})C_{11}(L)]_+ + [-A_1(L^{-1})C_{21}(L)]_+ = B(L)C_{31}(L)$$

$$[-A_1(L^{-1})C_{22}(L)]_+ = B(L)C_{22}(L).$$

The identification questions addressed by Lemma 1 and Lemma 2 become more complicated in this environment because of the assumption that  $C_{12}(L) = 0$  and that  $u$  is  $r+1$  dimensional while  $y$  is  $r+2$  dimensional. From a practical standpoint both of these complications result in making the possibility of multiple peaks in the likelihood function much less likely.

Since  $a$  is not observable, the moving average representation for the process observable to the econometrician is obtained by removing the first row from (5.8) and writing

$$(5.13) \quad \begin{bmatrix} w_t \\ y_{2t} \end{bmatrix} = \begin{bmatrix} C_{21}(L) & C_{22}(L) \\ C_{31}(L) & C_{32}(L) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \\ = C^*(L)u_t.$$

Although  $u$  is assumed to be fundamental for  $y$ ,  $u$  is not in general fundamental for  $(w, y_2)'$ . That is the restrictions hold for a moving average representation that is not necessarily

fundamental for the process observed by the econometrician. There is no guarantee that these restrictions will also hold across a moving average representation that is fundamental for  $(w, y_2')$ '.<sup>23</sup> In this sense, Lemma 2 does not carry over to these inexact models. A related point is that restriction (5.12) involves  $C_{11}$ . This indicates that the serial correlation parameters of  $a$  can be identified and estimated without observations on  $a$ .

We now consider relaxing assumptions (5.6) and (5.7). As noted above, in the absence of assumption (5.6), assumption (5.7) seems hard to defend. That is, if the relevant information set of economic agents is larger than  $\phi_t$  then it seems hard to defend the idea that there is an exact relationship among the variable in  $y$ . Indeed it seems likely that the omitted information variables would remove the stochastic singularity. Although, an iterated projection argument implies that

$$(5.14) \quad \hat{E}[A(L^{-1})y_{1t} | \phi_t] = E[A(L^{-1})y_{1t} | \Omega_t] = B(L)y_{2t}$$

the fact that the fundamental noise vector  $u$  is allowed to be  $r+2$  dimensional substantially hinders prospects for identification. This is true even if assumption (5.11) holds. To hope to identify and estimate the parameter of  $C$  and  $\lambda$ ,  $\gamma$  and  $\delta$ , it is necessary to make some additional assumptions once (5.7) and (5.8) are relaxed. These questions are investigated in Hansen and Sargent (1980, 1981b).

One should not be too discouraged by the comments made above. Anytime unobserved components are introduced into time series econometric models, some assumptions have to be made about the dynamic interaction of these unobserved components with variables observed by the econometrician in order to achieve identification. The discussion above is meant to help characterize the type of assumptions which will yield identification of structural parameters and permit reasonable interpretations of disturbance terms.

## 6. Conclusions

As is indicated by the range of examples given in Section 1, the procedures described in this paper are applicable to a variety of linear rational expectations models. Therefore, the characterization of identification in exact rational expectations models, the frequency domain estimation methods, and the relationship depicted between the variance bounds and restricted moving average representations are useful tools for guiding a variety of empirical applications. While this paper has focused on exact linear rational expectations models, the procedures described here can be employed with minor modifications in studying some exact nonlinear rational expectations models when the underlying observed variables are assumed to be lognormally distributed.<sup>24</sup> This latter point is a topic of future research.

Appendix A

Proof of Lemma 1:

The proof of this lemma is an immediate consequence of the fact that

$$(A1) \quad [A(L^{-1})C_1(L)]_+ D' = [A(L^{-1})C_1(L)D']_+$$

for any  $(n \times n)$  matrix  $D$ .

Proof of Lemma 2:

Assuming that  $C(L)$  satisfies (2.7), we know that

$$(A2) \quad E[A(L^{-1})y_{1t} | \Sigma_t] = B(L)y_{2t},$$

where  $\Sigma_t$  is the linear space generated by current and past  $u$ 's.

Using the fact that  $\Sigma_t^* \subset \Sigma_t$ , it follows that

$$\begin{aligned} (A3) \quad \hat{E}[A(L^{-1})y_{1t} | \Sigma_t^*] &= \hat{E}\{\hat{E}[A(L^{-1})y_{1t} | \Sigma_t] | \Sigma_t^*\} \\ &= \hat{E}[B(L)y_{2t} | \Sigma_t^*] \\ &= \hat{B}(L)y_{2t}, \end{aligned}$$

since  $B(L)y_{2t} \in \Sigma_t^*$ . Thus,  $C^*(L)$  satisfies (2.7).

Proof of Lemma 3:

Let  $C^*(L)$  be the operator associated with some other not necessarily fundamental representation of  $y$ . Furthermore, suppose that a common denominator polynomial for the elements of  $C^*(z)$  is  $\mu(z)$  and that

$$(A4) \quad \det C^*(z) = \frac{\rho_1(z)\rho_2^*(z)}{\gamma(z)}$$

where  $\rho_2(z)$  is a finite order polynomial. By assumption  $y_t = C(L)u_t = C^*(L)u_t^*$  where  $u^*$  is the white noise vector associated with  $C^*(L)$ . This implies that

$$(A5) \quad C(z)C(z^{-1})' = C^*(z)C^*(z^{-1})'$$

except possibly at some isolated singularities where the notation "' " denotes transpose. It follows that

$$(A6) \quad \rho_2(z)\rho_2(z^{-1}) = \rho_2^*(z)\rho_2^*(z^{-1})$$

with the exception of  $z = 0$ . If  $\theta$  is zero of  $\rho_2^*(z)$ , by assumption  $\theta \neq 0$  and it must be the case that either  $\theta$  or  $1/\theta$  is a zero of  $\rho_2(z)$ . Since

$$(A7) \quad \det C(z) = \frac{\rho_1(z)\rho_2(z)}{\gamma(z)}$$

and  $C(L)$  is an operator associated with a fundamental representation of  $y$ , we claim that the zeroes of  $\rho_2(z)$  are not

inside the unit circle. By assumption the zeroes of  $\rho_2(z)$  are not zeroes of  $b_2(z)$  and  $\gamma(z)$ . Also the zeroes of  $b_2(z)$  and  $\gamma(z)$  are assumed to be outside the unit circle. Together all of these facts imply that if  $\theta$  is a zero of  $\rho_2^*(z)$ , then neither  $\theta$  nor  $1/\theta$  are zeroes of  $b_2(z)$  or  $\gamma(z)$ . Let  $\rho_2^*(z)$  be given by

$$(A8) \quad \rho_2^*(z) = \theta_0(z - \theta_1) \dots (z - \theta_p).$$

In general, we are not able to impose any restrictions on the magnitude of  $\theta_1, \dots, \theta_p$ . We consider some special cases.

Case (i): Suppose  $|\theta_j| \geq 1$  for  $j = 1, \dots, p$ . This implies that  $u^*$  is fundamental for  $y$ . Using Lemma 1 we are guaranteed that the restrictions are satisfied for  $C^*(L)$ .

Case (ii): Suppose that there exists exactly one  $\theta_j$ , namely  $\theta_1$ , such that  $|\theta_1| < 1$ . Since  $\theta_1$  is a zero of  $\rho_2^*(z)$ , it follows that

$$(A9) \quad \det C^*(\theta_1) = 0.$$

Thus there exists an orthogonal matrix  $U$  such that all of the elements in the first column of  $C^*(\theta_1)U$  are zeroes. This allows us to assert that the first column of rational functions of  $C^*(z)U$  all have zeroes at  $\theta_1$ . Let us premultiply  $C^*(z)U$  by the diagonal Blaschke matrix  $G(z)$  where

$$(A10) \quad G(z) = \begin{bmatrix} \frac{1 - \theta_1 z}{z - \theta_1} & 0 \\ 0 & I \end{bmatrix}$$

to obtain  $\bar{C}(z) = C^*(z)UG(z)$ . Now

$$(A11) \quad \det \bar{C}(z) = \frac{\rho_1(z)\theta_0(1 - \theta_1 z)(\theta_2 - z) \dots (\theta_p - z)}{\gamma(z)} .$$

The rational matrix  $\bar{C}(z)$  is not singular at  $\theta_1$  but instead is singular at  $\frac{1}{\theta_1}$ . It is easily verified that

$$(A12) \quad \bar{C}(z)\bar{C}(z^{-1})' = C^*(z)C^*(z^{-1})'$$

for all  $|z| = 1$ . We conclude that  $C(L)$  is an appropriate operator for a fundamental moving average representation of  $y$ . All of the elements in the first column of  $\bar{C}(\frac{1}{\theta_1})$  are zero since  $\frac{1}{\theta_1}$  is not a zero of  $\gamma(z)$ .

Employing Lemma 1 we know that the restrictions are satisfied for  $\bar{C}(L)$ . The question is whether we can use this information to ascertain that the restrictions are satisfied for  $C^*(L)$ . It suffices for us to show that the restrictions are satisfied for  $C^*(L)U$ , and thus without loss of generality we assume that  $U = I$ . Recall that we can represent  $A(z^{-1})$  and  $B(z)$  as

$$(A13) \quad A(z^{-1}) = \frac{A_1(z)}{A_2(z)}$$

$$B(z^{-1}) = \frac{B_1(z)}{B_2(z)}$$

where  $A_1(z)$  and  $B_1(z)$  are finite order matrix polynomials,  $A_2(z)$  is a scalar polynomial with zeroes inside the unit circle and  $B_2(z)$  is scalar polynomial with zeroes outside the unit circle. It is assumed that the maximum order of polynomial elements of  $A_1(z)$  does not exceed the order of  $A_2(z)$ .

Using notation that is somewhat inconsistent with the text, let  $\bar{C}_1(z)$  denote the vector rational function obtained from the first column of  $\bar{C}(z)$ . Note that  $\bar{C}_1(\frac{1}{\theta_1}) = 0$ . We partition  $\bar{C}_1(z)$  in the same manner that  $y$  is partitioned to obtain

$$(A14) \quad \bar{C}_1(z) = \begin{bmatrix} \bar{C}_{11}(z) \\ \bar{C}_{21}(z) \end{bmatrix} .$$

Employing our knowledge that the restrictions are satisfied for  $\bar{C}(z)$ , we can write

$$(A15) \quad B(z)\bar{C}_{21}(z) \frac{(z - \theta_1)}{(1 - \theta_1 z)} = [A(z)\bar{C}_{11}(z)]_+ \frac{(z - \theta_1)}{(1 - \theta_1 z)} .$$

Also note that

$$\bar{C}_{11}(z) \frac{(z - \theta_1)}{(1 - \theta_1 z)} = C_{11}^*(z)$$

(A16)

$$\bar{C}_{12}(z) \frac{(z - \theta_1)}{(1 - \theta_1 z)} = C_{12}^*(z)$$

where

$$(A17) \quad \begin{bmatrix} C_{11}^*(z) \\ C_{12}^*(z) \end{bmatrix} = C_1^*(z)$$

and  $C_1^*(z)$  is a vector rational function formed from the first column of  $C^*(z)$ . Thus the restrictions hold for  $C^*$  if

$$(A18) \quad \left[ \frac{A_1(z)}{A_2(z)} \bar{C}_{11}(z) \right]_+ \frac{(z - \theta_1)}{(1 - \theta_1 z)} = \left[ \frac{A(z)}{A_2(z)} \bar{C}_{11}(z) \frac{(z - \theta_1)}{(1 - \theta_1 z)} \right]_+$$

i.e., if it is permissible to bring the Blaschke factor inside the annihilation operator. We proceed to verify that (A18) is satisfied. Using a lemma from Hansen and Sargent (1980), we know that

$$(A19) \quad \left[ \frac{A_1(z)}{A_2(z)} \bar{C}_{11}(z) \right]_+ = \frac{A_1(z)}{A_2(z)} \bar{C}_{11}(z) - \frac{M(z)}{A_2(z)}$$

Here  $M(z)$  is a polynomial vector of order one less than  $A_2(z)$ .

Now  $\bar{C}_1(\frac{1}{\theta_1}) = 0$ . Thus

$$\frac{A_1\left(\frac{1}{\theta_1}\right)}{A_2\left(\frac{1}{\theta_1}\right)} \bar{C}_{11}\left(\frac{1}{\theta_1}\right) = 0$$

(A20)

$$\frac{B_1\left(\frac{1}{\theta_1}\right)}{B_2\left(\frac{1}{\theta_1}\right)} \bar{C}_{21}\left(\frac{1}{\theta_1}\right) = 0.$$

These equalities make use of the fact that  $A_2\left(\frac{1}{\theta_1}\right)$  and  $B_2\left(\frac{1}{\theta_1}\right)$  are not zero. Since the restrictions are satisfied, it follows that

$$(A21) \quad M\left(\frac{1}{\theta_1}\right) = 0.$$

Thus, both  $M(z)$  and  $\bar{C}_{11}(z)$  have factors of the form  $(1 - \theta_1 z)$ . Multiplying both sides of (A2) by the Blaschke factor

$$(A22) \quad \frac{z - \theta_1}{1 - \theta_1 z}$$

we obtain

$$(A23) \quad \left[ \frac{A_1(z)}{A_2(z)} \bar{C}_{11}(z) \right]_+ \frac{(z - \theta_1)}{(1 - \theta_1 z)} = \frac{A_1(z)}{A_2(z)} \bar{C}_{11}(z) - \frac{M^*(z)}{A_2(z)}.$$

Note that  $M^*(z)$  is a polynomial vector of order less than that of  $A_2(z)$ . Thus  $\frac{M^*(z)}{A_2(z)}$  has a Laurent series expansion about zero in

a region containing the unit circle that is one-sided in strictly negative powers of  $z$ . Furthermore, we are guaranteed that

$$(A24) \quad \frac{A_1(z)}{A_2(z)} C_{11}^*(z) - \frac{M^*(z)}{A_2(z)}$$

is analytic for  $|z| < 1$ . We conclude that

$$(A25) \quad \left[ \frac{A_1(z)}{A_2(z)} C_{11}^*(z) \right]_+ = \left[ \frac{A_1(z)}{A_2(z)} \bar{C}_{11}(z) \right]_+ \frac{(z - \theta_1)}{(1 - \theta_1 z)}$$

This verifies that (A18) holds and that the restrictions are satisfied for  $C_{11}^*(L)$ .

Case (iii): Suppose that more than one of the  $\theta_j$ 's are bigger than one in modulus. For this case we can repeat the argument provided under case (ii) an appropriate number of times.

Appendix B

In this appendix we verify the equivalence of the two testing procedures when  $p = mq$  proposed in Section 3a. First of all, suppose that

$$(B1) \quad R_{qt} = (1/q)\hat{E}[\hat{R}_{1t} + \hat{R}_{1t+1} + \dots + \hat{R}_{1t+q-1} | \phi_t]$$

and

$$(B2) \quad R_{pt} = (1/p)\hat{E}[\hat{R}_{1t} + \hat{R}_{1t+1} + \dots + \hat{R}_{1t+p-1} | \phi_t].$$

Grouping terms in (B2), we have

$$(B3) \quad R_{pt} = (1/p)\hat{E}[\hat{R}_{1t} + \hat{R}_{1t+1} + \dots + \hat{R}_{1t+q-1} | \phi_t] \\ + (1/p)\hat{E}[\hat{R}_{1t+q} + \hat{R}_{1t+q+1} + \dots + \hat{R}_{1t+2q-1} | \phi_t] \\ + \dots + (1/p)\hat{E}[\hat{R}_{1,t+(m-1)q} + \hat{R}_{1,t+(m-1)q+1} \\ + \dots + \hat{R}_{1,t+mq-1} | \phi_t].$$

Substituting from (B1) and using an iterated projections argument, we obtain

$$\begin{aligned}
 (B4) \quad R_{pt} &= (1/m)R_{qt} + (1/m)\hat{E}[R_{qt+q} | \phi_t] + \dots \\
 &\quad + (1/m)\hat{E}[R_{q,t+(m-1)q} | \phi_t] \\
 &= (1/m)\hat{E}[R_{qt} + R_{qt+q} + \dots + R_{q,t+(m-1)q} | \phi_t] \\
 &= \hat{E}[A(L^{-1})R_{qt} | \phi_t].
 \end{aligned}$$

This establishes that (3.5) implies (3.11).

In establishing the converse, we impose some extra restrictions. Let

$$(B5) \quad \begin{bmatrix} R_{qt} \\ R_{pt} \\ x_t \end{bmatrix} = \begin{bmatrix} C_1(L) \\ C_2(L) \\ C_3(L) \end{bmatrix} u_t$$

be a time invariant fundamental representation. We assume that  $C_1(z)$  is analytic in  $|z| < R$  where  $R > 1$  except possibly at a finite number of points denoted  $\mu_1, \mu_2, \dots, \mu_K$ , where  $|\mu_j| = 1$ . We assume that  $\mu_k \neq \gamma_j$  where  $\gamma_j = e^{2\pi i j/q}$  for any choice of  $j = 1, 2, \dots, q-1$  and any choice of  $k = 1, 2, \dots, K$ . Furthermore, we assume that  $C_1(z)$  has finite order poles at  $\mu_1, \mu_2, \dots, \mu_K$ . We allow  $C_1(z)$  to have poles on the unit circle in order that we can analyze some borderline nonstationary stochastic processes along with covariance stationary processes. When  $C_1(z)$  has poles on the unit circle, we adopt the interpretation that the  $y$  process starts up at some time  $\tau$  and set  $u_t = 0$  for  $t < \tau$ .

We assume that the restrictions

$$(B6) \quad [A(L^{-1})C_1(L)]_+ = C_2(L)$$

hold where

$$(B7) \quad A(L^{-1}) = (1/m)[1 + L^{-q} + \dots + L^{-(m-1)q}].$$

Let

$$(B8) \quad A_q(z^{-1}) = \frac{1 - z^{-q}}{1 - z^{-1}} = \frac{z^q - 1}{z^{q-1}(z - 1)}$$

and

$$(B9) \quad A_p(z^{-1}) = \frac{1 - z^{-p}}{1 - z^{-1}} = \frac{z^p - 1}{z^{p-1}(z - 1)}.$$

The vector function

$$(B10) \quad F(z) = \frac{C_1(z)(z - 1)z^{q-1}}{z^{q-1}}$$

is analytic in region  $|z| < R$  except at  $\mu_1, \mu_2, \dots, \mu_K, \gamma_1, \gamma_2, \dots, \gamma_{q-1}$  and has a removable singularity at  $z = 1$ . Let

$$(B11) \quad C_0(z) = F(z) - \sum_{j=1}^{q-1} \frac{P_j}{z - \gamma_j}$$

where

$$(B12) \quad P_j = \lim_{z \rightarrow \gamma_j} F(z)(z - \gamma_j).$$

We wish to verify that  $C_0(L)$  satisfies

$$(B13) \quad [C_0(L)A_q(L^{-1})]_+ = C_1(L).$$

To do this, first we establish that

$$(B14) \quad \left[ \frac{A_q(L^{-1})P_j}{L - \gamma_j} \right]_+ = 0.$$

Note that

$$(B15) \quad \frac{A_q(z^{-1})P_j}{z - \gamma_j} = \frac{(z^q - 1)P_j}{z^{q-1}(z-1)(z-\gamma_j)}$$

$$= \frac{1}{z^{q-1}}(g_{j0} + g_{j1}z + \dots + g_{jq-2}z^{q-2}).$$

From here, result (B14) follows immediately. Second, we establish that

$$(B16) \quad [F(L)A_q(L^{-1})]_+ = C_1(L).$$

This follows from the definition of  $F(z)$  since

$$(B17) \quad F(z)A_q(z^{-1}) = C_1(z).$$

Taken together, (B14) and (B16) imply restrictions (B13).

We are finished establishing the converse once we show that

$$(B17) \quad [C_0(L)A_p(L^{-1})]_+ = C_2(L)$$

given that (B6) holds. Now

$$(B18) \quad F(z)A_p(z^{-1}) = C_1(z)A(z^{-1}).$$

Therefore,

$$[F(L)A_p(L^{-1})]_+ = [C_1(L)A(L^{-1})]_+ = C_2(L).$$

It remains for us to show that

$$(B19) \quad \left[ \frac{L^j}{L^{-\gamma_j}} A_p(L^{-1}) \right]_+ = 0.$$

However,

$$(B20) \quad \frac{L^j}{z^{-\gamma_j}} A_p(z^{-1}) = \frac{L^j(z^P - 1)}{z^{P-1}(z-1)(z-\gamma_j)}$$

and  $\gamma_j^P = 1$ . Therefore,

$$(B21) \quad \frac{P_{jA}(z^{-1})}{z - \gamma_j} = \frac{1}{z^{p-1}}(h_{j0} + h_{j1}z + \dots + h_{jp-2}z^{p-2})$$

from which it follows that (B19) is true. Thus with some extra assumptions (3.11) implies (3.5).

It is worthwhile to note that our choice of  $C_0$  was not the only possible one. We chose  $C_0$  so that  $C_0(z)$  does not have poles at  $\gamma_1, \gamma_2, \dots, \gamma_{q-1}$ . In cases in which it is assumed that  $C_1(z)$  does not have poles in  $|z| < R$ , this choice of  $C_0(z)$  will also not have poles in  $|z| < R$  and will be the only choice of  $C_0(z)$  that does not have poles in  $|z| < R$ .

Notes

1. For some of the discussion presented in this paper, this assumption could clearly be relaxed. We make it in order to insure the existence of various time invariant representations.
2. A useful reference for the linear least squares prediction tools we are employing is Rozanov (1967).
3. Wold's Decomposition Theorem does not guarantee the existence of a representation where  $u$  and  $y$  have the same dimension. However, we are making this extra assumption for the sake of convenience. Many of our results could be extended to cover cases in which  $u$  has smaller dimension than  $y$  but at the cost of making our discussion more tedious.
4. The Wiener-Kolmogorov prediction formulas are discussed by Whittle (1963). Sargent (1979b, p. 292-3) mentioned the idea of using these formulas to get a compact representation of the cross-equation restrictions implied by the linear rational expectations model of the term structure.
5. Whittle (1963) discusses this operator. Hansen and Sargent (1980) use the operator in contexts related to those in this paper.
6. For example, see Samuelson (1965).
7. For some additional examples, see Sargent (1979b, Chapters 12, 13, 14 and 16).
8. Blaschke factors are described by Saks and Zygmund (1971, p. 221)], and used in a somewhat related context by Hansen and Sargent (1980).

9. See Shiller (1972) and Sargent (1979b, Chapter 10) for expositions of the law of iterated projections.
10. Dividing out common factors amounts to modifying the value of a meromorphic function at removable singularities. As long as these singularities are not on the unit circle, this modification has no impact on the Laurent series expansion of the function in a region containing the unit circle.
11. These authors also propose procedures that involve concentrating the likelihood function to conserve on the number of parameters being estimated iteratively. Because of the form of the rational expectations restrictions, we do not concentrate the likelihood function. In order to justify using maximum likelihood with a Gaussian density function, we have to strengthen our assumptions about the  $y$  process. For instance, an assumption that  $y$  is a general linear process is sufficient for this justification.
12. A potential advantage of the frequency domain approximation is that it accommodates the idea of fitting time series models at subsets of frequencies that a particular economic theory is designed to explain. This point has been noted by Robinson (1977). Sims (1974), in proposing closely related procedures, suggests that seasonal frequencies should be ignored in fitting time series models. Hodrick and Prescott (1980) suggest that low frequency movements of economic time series should be ignored in business cycle modeling. A natural question that surfaces, however, is the extent to which dynamic economic theory can explain successfully

movements at some subset of frequencies without considering the "spilling over" impact onto other frequencies. This is especially a concern when the economic theory relies on assumptions that economic agents forecast optimally. Without an answer to this question, we are reluctant to recommend that researchers ignore some frequencies when using frequency domain procedures to estimate rational expectations models.

13. Phadke and Kedem (1978) also provide some examples that indicate that there may be substantial gains in terms of small sample proportions to using an exact likelihood approach when there are roots close to the unit circle.
14. From the standpoint of the discussion in this subsection, the covariance stationarity assumption can be relaxed. Instead we could allow elements of  $C(z)$  poles on the unit circle and assume  $u_t = 0$  for  $t \leq \tau$  for some start-up time  $\tau$ . If a differenced version of  $y$  is assumed to be covariance stationary, at least asymptotically, elements of  $C(z)$  may indeed have poles at  $z = 1$ . Identification of the parameters of  $C_0$  is a little more delicate if elements of  $C(z)$  have poles on the unit circle and if  $p = mq$  for some positive integer  $m$ . See Appendix B.
15. Shiller (1980) and Sargent (1976, fn. 13, p. 220) have noted that if data on interest rates at a sufficient number of maturities are available, then the linear rational expectations model of the term structure implies a set of linear restrictions on the vector autoregression of a vector

of rates,  $(R_1, R_2, \dots, R_p)$ . Shiller (1980) has correctly attributed the more complicated, highly nonlinear nature of the restrictions on the vector autoregressions fit in Sargent (1979a) and in the present paper as reflecting "omitted variables" from the list of maturities included in the vector autoregression. However, the iterated projections argument of Shiller (1972) that was cited by Sargent (1979a) implies that omitting those variables leads to no model misspecification, nor does it affect the validity of the econometric tests employed. Furthermore, in many contexts it will simply be impossible to possess observations of sufficiently many maturities to permit the cross-equation restrictions to assume a linear form.

16. The rearranging of terms in the infinite sum in order to obtain (3.15) is justified given our assumptions about  $A(z^{-1})$  and  $y$ .
17. Singleton (1980b) has derived a special case of these restrictions to use for variance bounds tests.
18. Strictly speaking, the first differencing is used to remove a simple type of nonstationarity. The covariogram of the original process may not be well defined.
19. The data were obtained from the Saloman Brothers publication, An Analytical Record of Yields and Yield Spreads. These data are first of month observations.
20. The log likelihood functions were maximized using the POWELL and GRADX subroutines of the Goldfeld-Quandt nonlinear estimation package. In cases in which convergence was

achieved using GRADX, numerical second derivatives were used to obtain estimates of the asymptotic standard errors. These standard error estimates are reported in Tables 1 and 2 in parentheses below the coefficient estimates.

21. Also, Sargent (1979a) used quarterly rather than monthly data. On the other hand, Sargent's sample started at an earlier date than ours.
22. In situations in which  $A(L^{-1})$  contains lead terms with powers greater than one, the regression equations which Geweke (1979) discusses have disturbances that are serially correlated and regressors that are not exogenous. While ordinary least squares remains consistent and is computationally convenient, it is not an asymptotically efficient way to estimate the regression parameters. These points are discussed in more detail in Hansen and Hodrick (1980).
23. This is an extension of a discussion in Hansen and Sargent (1980). In that paper it is assumed that  $a$  does not Granger-cause  $x$ .
24. See Hansen and Singleton (1981) for a preliminary discussion of this.

References

- Bard, Y., 1974, Nonlinear Parameter Estimation (Academic Press, New York).
- Geweke, J., 1979, "A Note on the Testable Implications of Expectations Models," University of Wisconsin, Manuscript.
- Granger, C. W. J., 1966, "The Typical Spectral Shape of an Economic Variable," Econometrica 34, 150-161.
- Granger, C. W. J., 1969, "Investigating Casual Relations by Econometric Models and Cross-Spectral Methods," Econometrica 37, 424-438.
- Hakkio, C. S., 1980, "The Term Structure of the Forward Premium," National Bureau of Economic Research, Manuscript.
- Hall, R. E., 1978, "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," Journal of Political Economy 86, 971-987.
- Hannan, E. J., 1970, Multiple Time Series (Wiley, New York).
- Hansen, L. P., 1980, "Large Sample Properties of Generalized Method of Moments Estimators," Carnegie-Mellon University, Manuscript.
- Hansen, L. P. and R. J. Hodrick, 1980, "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," Journal of Political Economy, 88, 829-853.
- Hansen, L. P. and T. J. Sargent, 1980, "Formulating and Estimating Dynamic Linear Rational Expectations Models," Journal of Economic Dynamics and Control, 2, 7-46.
- Hansen, L. P. and T. J. Sargent, 1981a, "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, MN).
- Hansen, L. P. and T. J. Sargent, 1981b, "Instrumental Variables Procedures for Estimating Linear Rational Expectations Models," Carnegie-Mellon University, Manuscript.
- Hansen, L. P. and K. J. Singleton, 1981, "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Stock Market Returns," Carnegie-Mellon University, Manuscript.
- Hodrick, R. J. and E. C. Prescott, 1980, "Post-War U.S. Business Cycles: An Empirical Investigation," Carnegie-Mellon University, Manuscript.

- Kennan, J., 1979, "The Estimation of Partial Adjustment Models with Rational Expectations," Econometrica 47, 1441-1456.
- Kohn, R., 1979, "Asymptotic Estimation and Hypothesis Testing Results for Vector Linear Time Series Models," Econometrica 47, 1441-1456.
- LeRoy, S. and R. Porter, 1979, "The Present Value Relation: Tests of Implied Variance Bounds," Board of Governors of the Federal Reserve System, Manuscript.
- Litterman, R. B., 1979, "Techniques of Forecasting Using Vector Autogressions," Ph.D. Dissertation (University of Minnesota, Minneapolis).
- Lucas, R. E., Jr., "Econometric Policy Evaluation: A Critique," in: K. Brunner and A. H. Meltzer, eds., The Phillips Curve and Labor Markets, Carnegie-Rochester Conference on Public Policy 1 (North-Holland, Amsterdam).
- Phadke, M. S. and G. Kedem, 1978, "Computation of the Exact Likelihood Function of Multivariate Moving Average Models," Biometrika 65, 511-519.
- Robinson, P. M., 1977, "The Construction and Estimation of Continuous Time Models and Discrete Approximations in Econometrics," Journal of Econometrics 6, 173-197.
- Rozanov, Y. A., 1967, Stationary Random Processes (Holden-Day, San Francisco).
- Saks, S. and A. Zygmund, 1971, Analytic Functions (Elsevier, New York).
- Saloman Brothers, 1977, An Analytical Record of Yields and Yield Spreads (Saloman, New York).
- Samuelson, P. A., 1965, "Proof That Properly Anticipated Prices Fluctuate Randomly," Industrial Management Review 6, 41-50.
- Sargent, T. J., 1976, "A Classical Macroeconometric Model for the United States," Journal Political Economy 84, 207-237.
- Sargent, T. J., 1978a, "Rational Expectations, Econometric Exogeneity, and Consumption," Journal of Political Economy 86, 673-700.
- Sargent, T. J., 1978b, "Estimation of Dynamic Labor Demand Schedules Under Rational Expectations," Journal of Political Economy 86, 1009-1044.
- Sargent, T. J., 1979a, "A Note on Maximum Likelihood Estimation of the Rational Expectations of the Term Structure," Journal of Monetary Economics 5, 133-143.

- Sargent, T. J., 1979b, Macroeconomic Theory (Academic Press, New York).
- Shiller, R., 1972, "Rational Expectations and the Structure of Interest Rates," Ph.D. Dissertation (Massachusetts Institute of Technology, Cambridge, MA).
- Shiller, R., 1979a, "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," Journal of Political Economy 87, 1190-1219.
- Shiller, R., 1979b, "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," University of Pennsylvania, Manuscript.
- Shiller, R., 1980, "Alternative Tests of Rational Expectations Models: The Case of the Term Structure," University of Pennsylvania, Manuscript.
- Singleton, K. J., 1980a, "Expectations Models of the Term Structure and Implied Variance Bounds," Journal of Political Economy 88, 1159-1176.
- Singleton, K. J., 1980b, "Testing the Present-Value Relation When the Difference of the Forecasted Variable is Stationary," Carnegie-Mellon University, Manuscript.
- Sims, C. A., 1972, "Money, Income and Causality," American Economic Review 62, 540-552.
- Sims, C. A., 1974, "Seasonality in Regression," Journal of the American Statistical Association 69, 618-626.
- Sims, C. A., 1980, "Macroeconomics and Reality," Econometrica 62, 540-552.
- Whittle, P., 1963, Regulation and Prediction by Linear Least-Square Methods (D. Van Nostrand, Princeton, NJ).