4

Two Difficulties in Interpreting Vector Autoregressions

by Lars Peter HANSEN and Thomas J. SARGENT

Introduction

The equilibrium of a typical dynamic rational expectations model is a covariance stationary $(n \times 1)$ vector stochastic process $z(t)$. This stochastic process determines the manner in which random shocks to the environment impinge over time on agents' decisions and ultimately upon market prices and quantities. *Surprises*, i.e., random shocks to agents' information sets, prompt revisions in their contingency plans, thereby impinging on equilibrium prices and quantities.

Every $(n \times 1)$ covariance stationary stochastic process $z(t)$ can be represented in the form of a vector autoregression (of any finite order). Consequently, it is natural to represent the equilibrium of a dynamic rational expectations model in terms of its vector autoregression. A vector autoregression induces a vector of innovations which yields characterizations of the vector stochastic process via the "innovation accounting" techniques invented by Sims (1980).

In interpreting these innovation accountings, it is useful to understand the connections between the innovations recovered by vector autoregressions, on the one hand, and the random shocks to agents' information sets, on the other hand. From the viewpoint of interpreting vector autoregressions that are estimated without imposing restrictions from formal economic theories, it would be desirable if the innovations recovered by a vector autoregression could generally be expected to equal either the random shocks to agents' information sets, or else some simply interpretable functions of these random shocks. This paper describes two important settings in which no such simple connections exist. In these settings, without explicitly imposing the restrictions implied by the economic theory, it is impossible to make correct inferences about the shocks impinging on agents' information sets. In addition to describing these situations, we briefly indicate in each case how the economic theory can be used to deduce correct inferences about the shocks impinging on agents' information sets.

Let $z(t)$ be an $(n \times 1)$ vector, covariance stationary stochastic process. Imagine that $z(t)$ is observed at discrete points in time separated by the sampling interval Δ. A *vector autoregression* is defined by the projection equation

$$(1) \qquad z(t) = \sum_{j=1}^{\infty} A_j^{\Delta} z(t - \Delta j) + a(t), \quad t = 0, \pm\Delta, \pm 2\Delta, \ldots$$

where $a(t)$ is an $(n \times 1)$ vector of population residuals from the regression with $Ea(t)a(t)^T = V$, and where the A_j^{Δ}'s are $(n \times n)$ matrices that, in general, are uniquely determined by the orthogonality conditions (or normal equations)

$$(2) \qquad E\, z(t - \Delta j) a(t)^T = 0, \qquad j \geq 1 .$$

The A_j^{Δ}'s in general are "square summable," that is, they satisfy

$$(3) \qquad \sum_{j=1}^{\infty} \text{trace } A_j^{\Delta} A_j^{\Delta T} < +\infty$$

where the superscript T denotes matrix transposition. Equations (1)–(3) imply two important properties of $a(t)$. First, (1) and (2) imply that

$$E\, a(t) a(t - \Delta j)^T = 0, \qquad j \neq 0 ,$$

so that $a(t)$ is a vector white noise. Second, (1) and (3) imply that $a(t)$ is in the closed linear space spanned by $\{z(t),\ z(t-\Delta),\ z(t-2\Delta), \ldots\}$. Further, by successively eliminating all lagged $z(t)$'s from (1), we obtain the *vector moving-average representation*

$$(4) \qquad z(t) = \sum_{j=0}^{\infty} C_j^{\Delta}\, a(t - \Delta j)$$

where the C_j^{Δ}'s are $n \times n$ matrices that satisfy

$$-\sum_{j=0}^{\infty} C_j^{\Delta} A_{s-j}^{\Delta} = \begin{cases} I & s = 0 \\ 0 & s \neq 0 \end{cases}$$

where $A_0^{\Delta} \equiv -I$. The C_j^{Δ}'s in (4) satisfy

$$(5) \qquad \sum_{j=0}^{\infty} \text{trace } C_j^{\Delta} C_j^{\Delta T} < +\infty .$$

Equations (4) and (5) imply that $z(t)$ is in the closed linear space spanned by $(a(t),\ a(t-\Delta),\ a(t-2\Delta), \ldots)$. Thus, the closed linear space spanned by $(z(t),\ z(t-\Delta), \ldots)$ equals the closed linear space spanned by $(a(t),\ a(t-\Delta), \ldots)$. In effect, $a(t)$ is a stochastic process that forms an orthogonal basis for the stochastic process $z(t)$, and which can be constructed from $z(t)$ via a *Gram-Schmidt* process. The property of the vector white noise $a(t)$ that it is contained in the linear space spanned by current and lagged $z(t)$'s is said to mean that "$a(t)$ is a *fundamental* white noise for the $z(t)$ process."

It is a moving-average representation for $z(t)$ in terms of a fundamental white noise that is automatically recovered by vector autoregression.[1] However, there are in addition a variety of *other* moving-average representations for $z(t)$ of the form

$$(6) \qquad z(t) = \sum_{j=0}^{\infty} \tilde{C}_j^{\Delta}\, \tilde{a}(t - \Delta j)$$

where $\tilde{a}(t)$ is an $(n \times 1)$ vector white noise in which the linear space spanned by $(\tilde{a}(t), \tilde{a}(t-\Delta), \ldots)$ is strictly *larger* than the linear space spanned by current and lagged $z(t)$'s. Current and lagged $z(t)$'s fail to be "fully revealing" about the $\tilde{a}(t)$'s in such representations.

Representation (4) induces the following decomposition of j-step ahead prediction errors[2]

$$(7) \qquad \begin{aligned} &E(z(t) - \hat{E}_{t-j} z(t))\, (z(t) - \hat{E}_{t-j} z(t))^T \\ &= \sum_{k=0}^{j-1} C_k^{\Delta}\, V\, C_k^{\Delta T} . \end{aligned}$$

By studying versions[3] of decomposition (7), Sims has shown how the j-step ahead prediction error variance can be decomposed into parts attributable to *innovations* in particular components of the vector $z(t)$.

Sims has described methods for estimating vector autoregressions and for obtaining alternative fundamental moving-average representations. He has also created a useful method known as "innovation accounting" that is based on decomposition (7). In the hands of Sims and other skilled analysts, these methods have been used successfully to detect interesting patterns in data, and to suggest possible interpretations of them in terms of the responses of systems of people to surprise events.

This paper focuses on the question of whether dynamic economic theories readily appear in the form of a fundamental moving-average representation (4), so that the vector white noises $a(t)$ recovered by vector autoregressions are potentially interpretable in terms of the white noise impinging on the information sets of the agents imagined to populate the economic model. This question is important because it influences the ease with which one can interpret the variance decompositions (or innovation accounts) and the responses to innovations $a(t)$ that are associated with the fundamental moving average (4).

This paper is organized as follows. Section 1 describes a class of discrete-time models whose equilibria can be represented in the form

$$z_t = \sum_{j=0}^{\infty} D_j \, \varepsilon_{t-j} \tag{8}$$

where ε_t is an $(n \times 1)$ vector white noise; D_j is an $(n \times n)$ matrix for each j; and $\sum_{j=0}^{\infty}$ trace $D_j D_j^T < \infty$. Here ε_t represents a set of shocks to agents' information sets. We study how the ε_t of representation (8) are related to the $a(t)$ of the (Wold) representation (4), and how the D_j's of (8) are related to the C_j^{Δ}'s of (4). We describe contexts in which $a(t)$ fails to match up with ε_t and C_j^{Δ} fails to match up with D_j because $z(t)$ fails to be fully revealing about ε_t. Such examples were encountered earlier by Hansen and Sargent (1980a), Futia (1981), and Townsend (1983). The discussion in Section 1 assumes that the sampling interval Δ equals the sampling interval in terms of which the economic model is correctly specified.

Section 2 describes a class of continuous time models whose equilibria are represented in the form

$$z_t = \int_0^{\infty} p(\tau) w(t-\tau) d\tau \tag{9}$$

where $w(t)$ is an m-dimensional continuous time white noise and $p(\tau)$ is an $(n \times m)$ function satisfying $\int_0^{\infty}$ trace $p(\tau)p(\tau)^T d\tau < +\infty$. In (9), $w(t)$ represents shocks to agents' information set. It is supposed that economic decisions occur in continuous time according to (9), but that the econometrician possesses data only at discrete intervals of time.

Section 2 studies the relationship between the $w(t)$ of (9) and the $a(t)$ of (4), and also the relationship between $p(\tau)$ of (9) and the C_j^{Δ} of (4). In general, these pairs of objects do not match up in ways that can be determined without the imposition of restrictions from a dynamic economic theory.[4]

1. Unrevealing Stochastic Processes

We consider a class of discrete time linear rational expectations models that can be represented as the solution of the following pair of stochastic difference equations

$$\begin{aligned} H(L)y_t &= E_t \, J(L^{-1})^{-1} p x_t \\ x_t &= K(L) \, \varepsilon_t \end{aligned} \tag{1.1}$$

where

$$\begin{aligned} H(L) &= H_0 + H_1 L + \ldots + H_{m_1} L^{m_1} \\ J(L) &= J_0 + J_1 L + \ldots + J_{m_2} L^{m_2} \\ K(L) &= \sum_{j=0}^{\infty} K_j L^j, \qquad K_0 = I \\ \varepsilon_t &= x_t - E(x_t \mid x_{t-1},\, x_{t-2},\, \ldots) \end{aligned} \tag{1.2}$$

In (1.1), y_t is an $n_1 \times 1$ vector, while x_t is an $n_2 \times 1$ vector. In (1.2), J_j and H_j are $(n_1 \times n_1)$ matrices, while K_j is an $(n_2 \times n_2)$ matrix. In (1.1), p is an $(n_1 \times n_2)$ matrix. We assume that the zeroes of $\det H(z)$ lie outside the unit circle, that those of $\det J(z)$ lie inside the unit circle, and that those of $\det K(z)$ do not lie outside the unit circle.

Many discrete time linear rational expectations models are special cases of (1.1). For example, interrelated factor demand versions of Lucas-Prescott equilibrium models are special cases with $J(L^{-1}) = H(L^{-1})^T$ and with $H(L^{-1})^T \, H(L)$ being the matrix factorization of the Euler equation that is solved by the fictitious social planner (see Hansen and Sargent (1981a) and Eichenbaum (1983) for some examples). Kydland-Prescott equilibria with feedback from market-wide variables to forcing variables that individual agents face parametrically form a class of examples with $H(L^{-1})^T \neq J(L^{-1})$ (see Sargent 1981). Other examples with $H(L^{-1})^T \neq J(L^{-1})$ arise in the context of various dominant player equilibria of linear quadratic differential games (see Hansen, Epple, and Roberds 1985). Finally, market equilibrium models of the Kennan (1988)-Sargent (1987b) variety, an example of which is studied below, solve a version of (1.1) with $H(L^{-1})^T \neq J(L)$. Models of this general class are studied by Whiteman (1983).

Hansen and Sargent (1981a) displayed a convenient representation of the solution of models related but not identical to (1.1). To adapt their results, first obtain the partial fractions representation of $J(z^{-1})^{-1}$. We have $J(z^{-1})^{-1} = \text{adj } J(z^{-1}) / \det J(z^{-1})$. Let

$$\det J(z^{-1}) = \lambda_0(1-\lambda_1 z^{-1}) \ \ldots \ (1-\lambda_k z^{-1})$$

where $k = m_2 \cdot n_1$ and $|\lambda_j| < 1$ for $j = 1, \ldots, k$. The λ_j's are the zeroes of $\det J(z^{-1})$ which are assumed to be distinct. Then we have

$$J(z^{-1})^{-1} = \sum_{j=1}^{k} \frac{M_j}{(1-\lambda_j z^{-1})} \tag{1.3}$$

where

$$M_j = \lim_{z\to\lambda_j} J(z^{-1})^{-1}\,(1-\lambda_j z^{-1}) \,. \tag{1.4}$$

Substitute (1.3) into (1.1) to obtain

$$H(L)\, y_t = E_t \sum_{j=1}^{k} \frac{M_j}{1-\lambda_j L^{-1}}\, p\, x_t \tag{1.5}$$

Hansen and Sargent (1980a) establish that

$$E_t \frac{M_j}{1-\lambda_j L^{-1}}\, p\, x_t = M_j\, p \left(\frac{LK(L) - \lambda_j K(\lambda_j)}{L-\lambda_j}\right) \varepsilon_t \,. \tag{1.6}$$

Define the operator M by

$$M(K(L)) = \sum_{j=1}^{k} M_j\, p \left(\frac{LK(L) - \lambda_j K(\lambda_j)}{L-\lambda_j}\right) . \tag{1.7}$$

Then, using (1.5), (1.6), and (1.7) we have the representation of the solution

$$\begin{aligned} H(L)\, y_t &= M(K(L))\, \varepsilon_t \\ x_t &= K(L)\, \varepsilon_t \end{aligned} \tag{1.8}$$

A vector stochastic process $(y_t^T,\, x_t^T)$ governed by (1.8) generally has a singular spectral density at all frequencies because $(y_t^T,\, x_t^T)$ consists of $n_1 + n_2$ variables being driven by only n_2 white noises. Such a model implies that various of the first n_1 equations of the following model, which is equivalent to (1.8),[5]

$$\begin{aligned} H(L)\, y_t &= M(K(L))\; K(L)^{-1}\, x_t \\ x_t &= K(L)\, \varepsilon_t \end{aligned}$$

will fit perfectly (i.e., possess sample $\bar{R}^2$'s of 1).

To avoid this implication of no errors in various of the equations of the model, while still retaining the model, one path that has been suggested is to assume that the econometrician seeks to estimate (1.8), but that he possesses data only on a subset of the variables in $(y_t,\, x_t)$. (See Hansen and Sargent 1980a). One common procedure, but not the only one possible, is the following one described by Hansen and Sargent (1980a). Assume that (1.8) holds, but that the econometrician only has data on a subset of observations x_{2t} of x_t. Further suppose that the second equation of (1.1) can be partitioned and restricted as

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} K_1(L) & 0 \\ 0 & K_2(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} . \tag{1.9}$$

Then (1.8) assumes a special form which can be represented as

$$\begin{pmatrix} H(L) & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} y_t \\ x_{2t} \end{pmatrix} = \begin{pmatrix} M(K_1(L)) & M(K_2(L)) \\ 0 & K_2(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} . \tag{1.10}$$

The idea is to imagine that the econometrician is short of observations on a sufficient number of series, those forming x_{1t}, to make the $(y_t^T,\, x_{2t}^T)$ process described by (1.10) have a nonsingular spectral density matrix at all frequencies. To accomplish this, it will generally be sufficient that the dimension of the vector of variables of $(y_t^T,\, x_{2t}^T)$ be less than or equal to the dimension of $(\varepsilon_{1t}^T,\, \varepsilon_{2t}^T)$. For the argument below, we will consider the case often encountered in practice in which $(y_t^T,\, x_{2t}^T)$ and $(\varepsilon_{1t}^T,\, \varepsilon_{2t}^T)$ have equal dimensions. Thus we assume that x_{1t} is an $(n_1 \times 1)$ vector, so that ε_{1t} is an $(n_1 \times 1)$ vector of white noises.[6]

Equation (1.10) implies the moving-average representation for $(y_t,\, x_{2t})$

$$\begin{pmatrix} y_t \\ x_{2t} \end{pmatrix} = \begin{pmatrix} H(L)^{-1}M(K_1(L)) & H(L)^{-1}M(K_2(L)) \\ 0 & K_2(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} . \tag{1.11}$$

Equation (1.11) is a moving average that expresses $(y_t,\, x_{2t})$ in terms of current and lagged values of the white noises $(\varepsilon_{1t},\, \varepsilon_{2t})$ that are the innovations in the information sets $(x_{1t},\ x_{2t})$ of the agents in the model. Equivalently, $(\varepsilon_{1t},\ \varepsilon_{2t})$ are *fundamental* for $x_{1t},\ x_{2t}$, the one step-ahead errors in predicting $x_{1t},\ x_{2t}$ from their own pasts being expressible as linear combinations of $\varepsilon_{1t},\ \varepsilon_{2t}$.

Granted that the linear space spanned by current and lagged $(\varepsilon_{1t},\ \varepsilon_{2t})$ equals that spanned by current and lagged values of the agents' information $(x_{1t},\ x_{2t})$, there remains the question of whether this space equals that spanned by current and lagged values of the econometrician's information $(y_t,\ x_{2t})$. As is evident from the construction of (1.11), the latter space is included in the former. The question is whether they are equal. This question is an important one from the viewpoint of interpreting vector autoregressions because a vector autoregression by construction would recover a vector moving average for $(y_t,\ x_{2t})$ that is driven by a vector white noise a_t that is fundamental for $(y_t,\ x_{2t})$, i.e., one that is in the linear space spanned by current and lagged values of $(y_t,\ x_{2t})$. If this space is smaller than the one spanned by current and lagged values of agents' information $(\varepsilon_{1t},\ \varepsilon_{2t})$, then the moving-average representation recovered by the vector autoregression will in general give a distorted impression of the response of the system to surprises from agents' viewpoint.

The vector white noise $(\varepsilon_{1t},\ \varepsilon_{2t})$ is fundamental for $(y_t,\ x_{2t})$ if and only if the zeroes of

$$\det \begin{pmatrix} H(z)^{-1}M(K_1(z)) & H(z)^{-1}M(K_2(z)) \\ 0 & K_2(z) \end{pmatrix}$$
$$= \det H(z)^{-1} \cdot \det M(K_1(z)) \cdot \det K_2(z)$$

do not lie inside the unit circle. The zeroes of $\det K_2(z)$ do not lie inside the unit circle by assumption, and $\det H(z)^{-1} = 1/\det H(z)$ is a function with all its poles outside the unit circle. Therefore, the necessary and sufficient condition that $(\varepsilon_{1t},\ \varepsilon_{2t})$ be fundamental for $(y_t,\ x_{2t})$ is that

$$\det\ M(K_1(z)) = 0 \Rightarrow |z| \geq 1\ , \tag{1.12}$$

or equivalently, using (1.7),

$$\det \sum_{j=1}^{k} M_j\, p \left(\frac{zK_1(z) - \lambda_j K_1(\lambda_j)}{z - \lambda_j}\right) = 0 \Rightarrow |z| \geq 1\ . \tag{1.12$'$}$$

In general, condition (1.12) is not satisfied. For some specifications of $K_1(L)$ and $J(L^{-1})$, which determines the $(M_j,\ \lambda_j)$ via (1.3)–(1.4), condition (1.12) is met, while for others, it is not met. Hansen and Sargent (1980a) encountered a class of examples where (1.12) is not met. Furthermore, the class of cases for which (1.12) fails to be met is not thin in any natural sense. Our conclusion is that for the class of models defined by (1.1)–(1.2), the moving-average representation (1.11) that is expressed in terms of the white noises that are fundamental for agents' information sets in general cannot be expected to be fundamental for the econometrician's data set $(y_t,\ x_{2t})$. Equivalently, current and lagged values of $(y_t,\ x_{2t})$ fail to be fully revealing of current and lagged values of $(\varepsilon_{1t},\ \varepsilon_{2t})$.

For convenience, let us rewrite (1.10) as

$$S(L)\ z_t = R(L)\ \varepsilon_t \tag{1.13}$$

where

$$S(L) = \begin{pmatrix} H(L) & 0 \\ 0 & I \end{pmatrix},\ R(L) = \begin{pmatrix} M(K_1(L)) & M(K_2(L)) \\ 0 & K_2(L) \end{pmatrix}$$
$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},\ z_t = \begin{pmatrix} y_t \\ x_{2t} \end{pmatrix}.$$

The condition that ε_t be fundamental for z_t is then expressible as the condition that the zeroes of $\det R(z)$ not lie inside the unit circle. If this condition is violated, then a Wold representation for z_t, which is what is recovered via vector autoregression, will be related to representation (1.13) as follows. It is possible to show[7] that there exists a matrix polynomial $G(L)^T = \sum_{j=0}^{\infty} G_j^T L^j$ with the following properties:

(*i*) $G(L^{-1})\ G(L)^T = I$

(*ii*) $G(L)^T$ is one-sided in nonnegative powers of L.

(*iii*) $R(z)\, G(z^{-1})$ has a power series expansion with square summable coefficients and the zeroes of $\det(R(z)\, G(z^{-1}))$ are not inside the unit circle of the complex plane.

Evidently, we can represent $R(L)\, \varepsilon_t$ as

$$R(L)\varepsilon_t = R(L)\, G(L^{-1})\, G(L)^T\ \varepsilon_t$$

or

$$R(L)\varepsilon_t = R^*(L)\varepsilon_t^*\ , \tag{1.14}$$

where

$$R^*(L) = R(L)\, G(L^{-1}) \tag{1.15}$$

and

$$\varepsilon_t^* = G(L)^T \varepsilon_t \,. \tag{1.16}$$

Then a Wold representation for z_t corresponding to (1.13) is

$$S(L) z_t = R^*(L)\, \varepsilon_t^* \,. \tag{1.17}$$

By construction, the ε_t^* process defined in (1.16) is a vector white noise that lies in the space spanned by current and past values of the process z_t, in other words ε_t^* is *fundamental* for z_t. It is property (*iii*) which assures that the spanning condition is satisfied. If, in addition, $R^*(L)$ has an inverse that is one-sided in nonnegative powers of L, we can obtain a representation for ε_t in terms of current and past z_t's:

$$\varepsilon_t^* = R^*(L)^{-1}\, S(L)\, z_t \,. \tag{1.18}$$

In general, $R^*(L)$ given by (1.15) can be very different from $R(L)$, so that the impulse response of the system to ε_t (which are the innovations to *agents'* information set) can look very different from the impulse response to ε_t^* (which are the innovations to the *econometrician's* information set). Equation (1.16) and property (*ii*) above shows that ε_t^* is a one-sided distributed lag of current and past ε_t's so that the innovation ε_t^* reflects information that is *old news* to the agents in the model at t. Only in the special case that none of the zeros of $\det R(z)$ are inside the unit circle, so that $G(L)^T$ can be taken to be the identity operator, does ε_t^* given by (1.16) only respond to contemporary news ε_t possessed by the agents. In the case that $G(L)^T \neq I$, it is generally true that[8]

$$E(R_0^* \varepsilon_t^*)\,(R_0^* \varepsilon_t^*)^T > E(R_0 \varepsilon_t)\,(R_0 \varepsilon_t)^T \,. \tag{1.19}$$

This statement that the contemporaneous covariance matrix of $R_0^*\, \varepsilon_t^*$ is larger than that of $R_0\, \varepsilon_t$ concisely summarizes how ε_t^* contains less information about the z_t process than does ε_t.

We now describe a concrete hypothetical numerical example, one in which the econometrician observes no x's, only y's, so that (1.10) takes the special form

$$H(L)\, y_t = M(C_1(L))\, \varepsilon_{1t} \,. \tag{1.20}$$

The model is one of the dynamics of price and quantity in a single market, and is related to ones studied by Sargent (1987b) and Kennan (1988). Behavior by agents on the two sides of the market, supply and demand, are each described by linear Euler equations. The Euler equation for suppliers is

$$-E_t\Big\{\big[h_s + g_s(1-\beta L^{-1})\,(1-L)\big]\, q_t\Big\} + p_t = s_t \tag{1.21}$$

where p_t is price at time t, q_t is quantity supplied at time t, s_t is a supply stock at t, β is a discount factor between zero and one, and h_s and g_s are parameters of the suppliers' cost function. Equation (1.21) is typical of the kind of Euler equation that characterizes the optimum problem of a competitive firm facing adjustment costs that are quadratic in $(1-L)\, q_t$.

The Euler equation for demanders is

$$-E_t\Big\{h_d + g_d[a(\beta L^{-1})\, a(L)]\Big\}\, q_t - p_t = d_t \tag{1.22}$$

where $a(L) = a_0 + a_1 L + a_2 L^2 + a_3 L^3 + a_4 L^4$. In (1.22), h_d and g_d are preference parameters, while the parameters $[a_0, a_1, a_2, a_3, a_4]$ characterize a technology by which purchases of q_t at t give rise to utility-generating services to the demander in subsequent periods. In (1.22), d_t is a stochastic process of disturbances to demand.

To complete the model, we specify the stochastic law of motion for the forcing processes, i.e., the demand and supply shocks. These shocks are assumed to satisfy

$$\begin{aligned} s_t &= B_s(L) w_{st} \\ d_t &= B_d(L) w_{dt} \end{aligned} \tag{1.23}$$

where $B_s(z)$ and $B_d(z)$ are scalar polynomials with zeros that are outside the unit circle. The w_{st} and w_{dt} processes are mutually uncorrelated white noises so that w_{st} is the innovation in the supply shock, s_t, and w_{dt} is the innovation in the demand shock, d_t. Economic agents are assumed to observe current and past values of both shocks and hence also the innovations in both shocks.

The typical supplier and demander are both assumed to view the stochastic processes p_t, d_t, s_t as beyond their control and to choose a stochastic process for q_t. At time t, both suppliers and demanders have the common information set of $\{p_r, d_r, s_r, q_{r-1}\ ;\ r \leq t\}$. Both suppliers and demanders choose contingency plans for q_t as a function of this information set. As discussed in Kennan (1988) and Sargent (1987b), an *equilibrium* is a stationary stochastic process for $\{q_t,\ p_t\}$ that solves the pair of difference equations (1.21)–(1.22).[9]

This model fits into our general set up (1.1) as follows. Let

$$y_t = \begin{pmatrix} q_t \\ p_t \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} w_{st} \\ w_{dt} \end{pmatrix} .$$

$$K_2(L) = 0, \qquad x_{2t} \equiv 0 .$$

$$K_1(L) = \begin{pmatrix} B_s(L) & 0 \\ 0 & B_d(L) \end{pmatrix} .$$

Define a matrix polynomial

$$E(L) = \begin{pmatrix} -(h_s + g_s(1-L)(1-\beta L^{-1})) & 1 \\ -(h_d + g_d a(L) a(\beta L^{-1})) & -1 \end{pmatrix} .$$

Then polynomial matrices $H(L)$ and $J(L)$ that are one-sided in non-negative powers of L can be found such that

$$J(L^{-1})\ H(L) = E(L)$$

and such that the zeroes of $\det J(z)$ lie inside the unit circle, while the zeroes of $\det H(z)$ lie outside the unit circle. (See Whiteman 1983, and Gohberg, Lancaster, and Rodman 1982 for proofs of the existence of such a matrix factorization, and for descriptions of algorithms for achieving the factorization.)[10]

In this model, it is possible for the demand and supply shocks to generate an information set that is strictly larger than that generated by current and past quantities and prices. So an econometrician using innovation accounts derived from observations on quantities and prices may not obtain innovations that are linear combinations of the contemporaneous innovations to the demand and supply shocks.

The equilibrium of the model has representation

$$S(L) \begin{pmatrix} q_t \\ p_t \end{pmatrix} = R(L) \begin{pmatrix} w_{dt} \\ w_{st} \end{pmatrix} \tag{1.24}$$

where $S(L) = H(L)$ and $R(L) = M(C_1(L))$, and where $S(L)$ is a (2×2) fourth-order matrix polynomial in L, with the zeroes of $\det S(z)$ outside the unit circle. The zeroes of $\det R(z)$ can be on either side of the unit circle in this example. Only when the zeroes of $\det R(z)$ are not inside the unit circle can the one-step ahead forecast errors from the vector autoregression of prices and quantities be expressed as linear combinations of the contemporaneous demand and supply shock innovations (w_{st}, w_{dt}).

We have computed a numerical example that illustrates the ideas discussed above.[11] We set the parameters of the model as follows:[12]

$$h_s = h_d = 1\,,\ g_s = 10,\ g_d = .1$$
$$a(L) = 1 + .8L + .6L^2 + .4L^3 + .2L^4$$
$$B_d(L) = (1 + .6L)\ (1 + .4L)\ (1 + .2L)$$
$$B_s(L) = (1 - .8L)\ (1 + .4L)\ (1 + .2L)$$
$$\beta = 1/1.05$$
$$E\,w_{st}^2 = .5\,,\ E\,w_{dt}^2 = 4\,,\ E\,w_{st}\,w_{dt} = 0$$

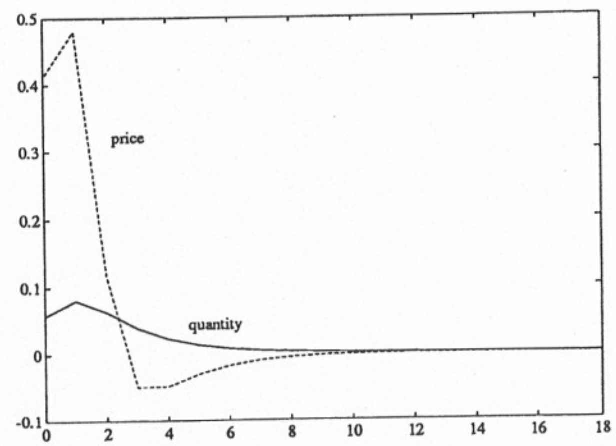

Figure 1a. Impulse response functions to first innovation in $S(L)\begin{bmatrix} q_t \\ p_t \end{bmatrix} = R(L)\varepsilon_t$.

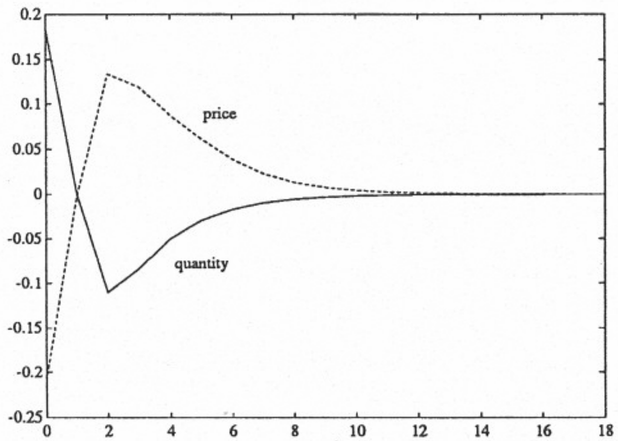

Figure 1b. Impulse response functions to second innovation in $S(L)\begin{bmatrix} q_t \\ p_t \end{bmatrix} = R(L)\varepsilon_t$.

Figure 1 displays response functions of $(q_t,\, p_t)$ to an impulse in the innovation $(w_{dt},\, w_{st})$ to agents' information set. This corresponds to representation (1.13) or (1.24). Figure 2 displays the response function of $(q_t,\, p_t)$ to a fundamental (Wold) innovation ε_t^* which corresponds to representation (1.17) or

(1.25) $$S(L)\begin{bmatrix} q_t \\ p_t \end{bmatrix} = R^*(L)\,\varepsilon_t^*\,.$$

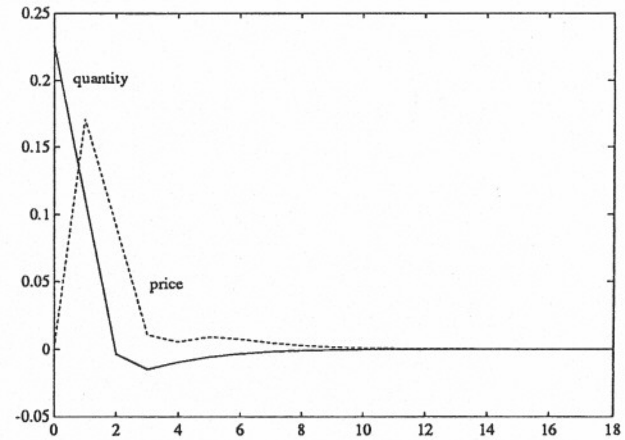

Figure 2a. Impulse response functions to first innovation in $S(L)\begin{bmatrix} q_t \\ p_t \end{bmatrix} = R^*(L)\varepsilon_t^*$. Quantity is 'first' in the orthogonalization.

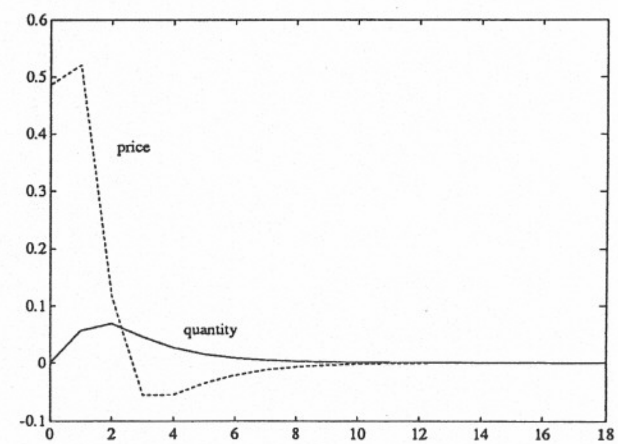

Figure 2b. Impulse response functions to second innovation in $S(L)\begin{bmatrix} q_t \\ p_t \end{bmatrix} = R^*(L)\varepsilon_t^*$. Quantity is 'first' in the orthogonalization.

In Figure 2, we have followed Sims (1980) in normalizing R_0^* and ε_t^* by selecting a version of ε_t^* which has diagonal contemporaneous covariance matrix, and for which the variance of ε_{1t}^* is maximal. This amounts to "letting q_t go first" in a Gram-Schmidt procedure that orthogonalizes the contemporaneous covariance matrix.[13] Figure 3 shows the impulse response functions in the Wold representation for which "p_t goes first" in the orthogonalization procedure.

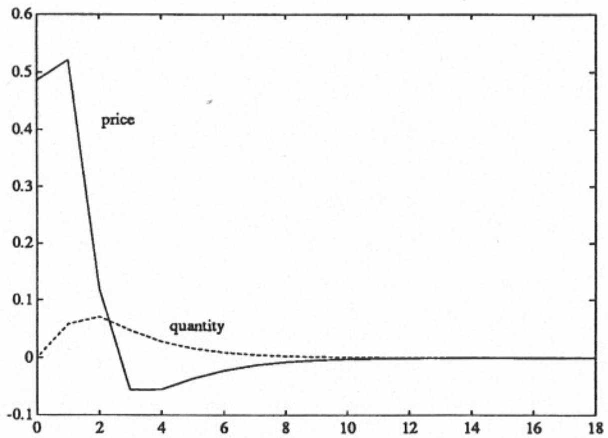

Figure 3a. Impulse response functions for

$$S(L)\begin{bmatrix} q_t \\ p_t \end{bmatrix} = R^*(L)UU'\varepsilon_t^*,$$

$UU' = I$. Price is 'first' in the orthogonalization.

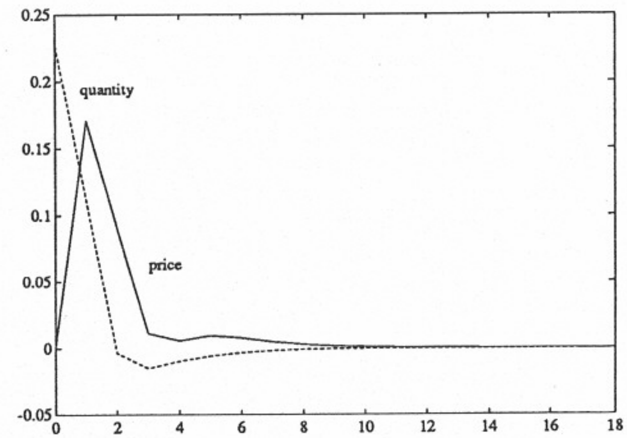

Figure 3b. Impulse response functions for

$$S(L)\begin{bmatrix} q_t \\ p_t \end{bmatrix} = R^*(L)UU'\varepsilon_t^*,$$

$UU' = I$. Price is 'first' in the orthogonalization.

The representation in Figure 3 is related to that in Figure 2 via the equation

$$S(L)\begin{bmatrix} q_t \\ p_t \end{bmatrix} = (R^*(L)\ U)\ (U^T\varepsilon_t^*)\ ,$$

where U is an orthogonal matrix that implements the Gram-Schmidt procedure that puts p_t first in the orthogonalization process.

That Figures 2 and 3 are very similar is a reflection of the fact that the contemporaneous covariance matrix of $R_0^*\varepsilon_t^*$ is nearly diagonal:

$$E\,(R_0^*\,\varepsilon_t^*)\ (R_0^*\,\varepsilon_t^*)^T = \begin{bmatrix} .0517 & -.0332 \\ -.0337 & .2354 \end{bmatrix}$$

From representation (1.24) corresponding to Figure 1, we computed

$$E\,(R_0\,\varepsilon_t)\ (R_0\,\varepsilon_t)^T = \begin{bmatrix} .0374 & -.0149 \\ -.0149 & .2121 \end{bmatrix}$$

Thus, inequality (1.19) holds strictly for our example.

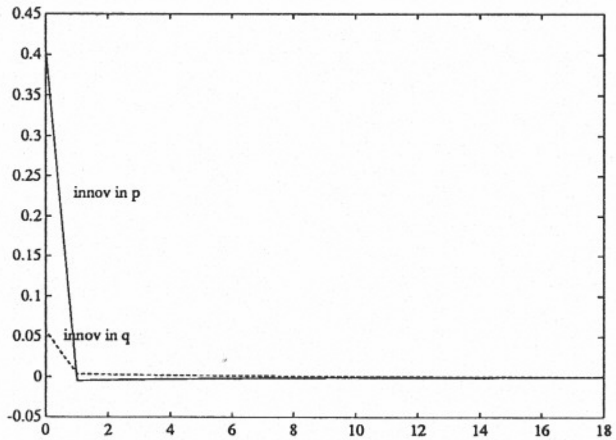

Figure 4a. Impulse response functions for $\varepsilon_t^* = G(L)'\varepsilon_t$. Response of ε_t^* to a demand innovation.

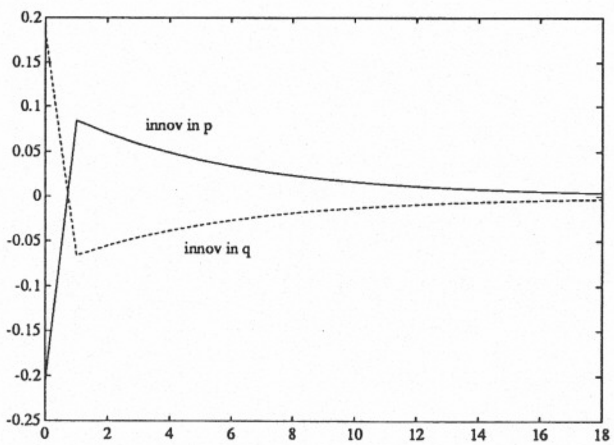

Figure 4b. Impulse response functions for $\varepsilon_t^* = G(L)'\varepsilon_t$. Response of ε_t^* to a supply innovation.

Figure 4 depicts the response of ε_t^* to the innovations $(w_{dt},\ w_{st})$ in agents' information sets, which corresponds to representation (1.16), namely $\varepsilon_t^* = G(L)^T\,\varepsilon_t$. Notice that the demand innovation w_{dt} gives rise almost entirely to a contemporary response in both the innovation in p and the innovation in q, with the numerically much larger response being in the innovation to p. However, the supply innovation w_{st} gives rise to distributed responses in the innovations both to p and to q, with numerically the larger one being in the q innovation. Figure 4 is consistent with the approximation that the innovation in p from representation (1.24) reflects the innovation to demand in a timely manner, but that the innovation to q is a distributed lag mainly of the innovation in supply.

These interpretations bear up when we compare Figures 1 and 2. Notice how much Figure 1a resembles Figure 2b, which is consistent with the interpretation of the innovation in p as an innovation in the demand shock. However, Figure 1b does not very much resemble Figure 2a. Indeed, Figure 1b more closely resembles Figure 4b, which depicts the response of ε_t^* to the supply innovation. This is understandable in view of the nearly entirely contemporaneous response of q to its own innovation depicted in Figure 2a.

1a. Remedies in Discrete Time

The preceding difficulty can be circumvented if a sufficiently restrictive dynamic economic theory is imposed during estimation. Hansen and Sargent (1980a) describe methods for estimating $S(L)$ and $R(L)$ subject to extensive cross-equation restrictions of the rational expectations variety. The approach is to use the method of maximum likelihood to estimate free parameters of preferences and constraint sets, of which the parameters of $S(L)$ and $R(L)$ are in turn functions. These methods do not require that the zeroes of $\det R(z)$ be restricted, and in particular are capable of recovering good estimates of $R(z)$ even when some of the zeroes of $\det R(z)$ are inside the unit circle.

Simply take representation (1.13) and operate on both sides by the two-sided inverse $R(L)^{-1}$ to obtain

$$\varepsilon_t = R(L)^{-1}\,S(L)\,z_t\,. \tag{1.26}$$

When some of the zeros of $\det R(z)$ are inside the unit circle, $R(L)^{-1}$ is two-sided, so that (1.26) expresses ε_t as a linear function of past, present, and future z_t's. By using this equation together with estimates of the model's parameters, it would be possible to construct estimates of ε_t as functions of an observed record on $\{z_t\}$. Equation (1.26) once again illustrates how ε_t fails to lie in the linear space spanned by current and lagged z's.

2. Time Aggregation

Consider a linear economic model that is formulated in continuous time, and which can be represented as

$$z(t) = \int_0^{\infty} p(\tau)\, w(t-\tau) d\tau \tag{2.1}$$

where $z(t)$ is an $(n \times 1)$ vector stochastic process, $w(t)$ is an $(m \times 1)$ vector white noise with $Ew(t)w(t-s)^T = \delta(t-s)I$, δ is the Dirac delta generalized function, and $p(\tau)$ is an $(n \times m)$ matrix function that satisfies $\int_0^\infty$ trace $p(\tau)p(\tau)^T\, d\tau < +\infty$. We let $P(s) = \int_0^\infty e^{-s\tau} p(\tau) d\tau$, i.e., $P(s)$ is the Laplace transform of $p(\tau)$. Sometimes we shall find it convenient to write (2.1) in operator notation

$$z(t) = P(D)w(t) \tag{2.2}$$

where D is the derivative operator. We shall assume that det $P(s)$ has no zeroes in the right half of the complex plane. This guarantees that square integrable functionals of $(z(t-s),\ s \geq 0)$ and of $(w(t-s),\ s \geq 0)$ span the same linear space, and is equivalent to specifying that (2.1) is a Wold representation for $z(t)$.

A variety of continuous time stochastic linear rational expectations models have equilibria that assume the form of the representations (2.1) or (2.2). Hansen and Sargent (1981d) provide some examples. In these examples, the continuous time white noises $w(t)$ often have interpretations as innovations in the uncontrollable processes that agents care about forecasting, and which stochastically drive the model. These include processes that are imagined to be observable to both the econometrician and the private agent (e.g., various relative prices and quantities) and also those which are observable to private agents but are hidden from the econometrician (e.g., random disturbances to technologies, preferences, and maybe even particular factors of production such as *effort* or capital of specific kinds). The $w(t)$ process is economically interpretable as the continuous time innovation to private agents, because a forecast error of the variables in the model over any horizon $t+\tau$ which the private agents are assumed to make at t can be expressed as a weighted sum of $w(s)$, $t < s \leq t+\tau$. Thus, to private agents the $w(t)$ process represents *news* or *surprises*.

In rational expectations models, typically there are extensive restrictions across the rows of $P(D)$. In general these restrictions leave open the possibility that the current and lagged values of the $w(t)$ process span a larger linear space than do current and lagged values of the $z(t)$ process. This outcome can possibly occur even if the dimension m of the $w(t)$ process is less than or equal to the dimension n of the $z(t)$ process. This is the continuous time version of the phenomenon that we treated for discrete time in the previous section. In the present section, we ignore this phenomenon, by assuming that det $P(s)$ has no zeroes in the right half of the complex plane.

For this continuous time specification, there exists a discrete time moving-average representation

$$z_t = C(L)a_t \tag{2.3}$$

where $C(L)$ is an infinite order, $(n \times n)$ polynomial in the lag operator L, where a_t is a vector white noise with $Ea_t a_t^T = W$, and where $a_t = z_t - \hat{E}\left[z_t | , z_{t-1}, \ldots\right]$. The operator $C(L)$ and the positive semi-definite matrix W solve the following equation, subject to the side condition that the zeroes of det $C(z)$ do not lie inside the unit circle:[14]

$$C(e^{-i\omega})W\, C(e^{i\omega})^T = \sum_{j=-\infty}^{+\infty} P(i\omega + 2\pi i j)\, P(-i\omega - 2\pi i j)^T\,. \tag{2.4}$$

When z_t has a discrete time autoregressive representation, the discrete time innovations a_t are related to the $w(t)$ process by the formula

$$a_t = C(L)^{-1}\, P(D)\, w(t)$$

or

$$a_t = V(L)\ P(D)\ w(t) = V(L) \int_0^\infty p(\tau)\ w(t-\tau)d\tau \tag{2.5}$$

where we have defined $V(L) = C(L)^{-1} = \sum_{j=0}^\infty V_j L^j$, $V_0 = I$. Here $-V_j$ is the $n \times n$ matrix of coefficients on the j^{th} lag in the vector autoregression for z. It follows directly upon writing out (2.5) that

$$a_t = \int_0^\infty f(\tau) w(t-\tau) d\tau \tag{2.6}$$

where[15]

$$f(\tau) = \sum_{j=0}^{\infty} V_j\, p(\tau - j) \tag{2.7}$$

It also follows from (2.6) and the identity for integer t, $C(L)\, a_t = P(D)\, w_t$, that

$$p(\tau) = \sum_{j=0}^{\infty} C_j\, f(\tau - j)\ . \tag{2.8}$$

Equations (2.6) and (2.7) show how the discrete time innovation a_t in general reflects all past values of the continuous time innovation $w(t)$.

Analyses of vector autoregressions often proceed by summarizing the shape of $C(L)$ in various ways, and attempting to interpret that shape. The innovation accounting methods of Sims, based on decomposition (7), are good examples of procedures that summarize the shape of $C(L)$. From the viewpoint of interpreting discrete time vector autoregressions in terms of the economic forces acting on individual agents, it would be desirable if the discrete time and continuous time moving-average representations were to match up in some simple and interpretable ways. In particular, the following two distinct but related features would be desirable. First, it would be desirable if the discrete time innovations a_t closely reflected the behavior of $w(s)$ near t. Probably the most desirable outcome would be if a_t could be expressed as

$$a_t = \int_0^1 f(\tau)\, w(t-\tau) d\tau\ , \tag{2.9}$$

so that in (2.6), $f(\tau) = 0$ for $\tau > 1$. In that case, a_t would be a weighted sum of the continuous time innovations over the unit forecast interval. It would be even more desirable if (2.9) were to hold with $f(\tau) = p(\tau)$, for then a_t would equal the one step ahead forecast error from the continuous time system. Second, assuming a smooth $p(\tau)$ function, it would be desirable if the discrete time moving-average coefficients $\{C_0, C_1, C_2, \ldots\}$ resemble a sampled version of the continuous time moving average kernel $\{p(\tau),\ \tau \geq 0\}$. This is desirable because the pattern of the C_j's would then faithfully reflect the response of the system to innovations in continuous time. We shall consider each of these *desiderata* in turn.

We first study conditions under which $f(\tau) = 0$ for $\tau > 1$. Consider

the decomposition

$$\begin{aligned} a_t &= z(t) - \hat{E}\,[z(t) \mid w(t-s),\ s \geq 1] \\ &\quad + \hat{E}\,[z(t)|w(t-s),\ s \geq 1] - \hat{E}\,[z_t|z_{t-1}, \ldots] \\ &= \int_0^1 p(\tau) w(t-\tau) d\tau + \int_1^\infty p(\tau) w(t-\tau) d\tau \\ &\quad - \hat{E}\,[\int_1^\infty p(\tau) w(t-\tau) d\tau | z_{t-1}, \ldots]\ . \end{aligned}$$

This last equality implies that if (2.9) is to hold it must be the case that

$$\hat{E}[z(t)|w(t-s),\ s \geq 1] = \hat{E}[z_t|z_{t-1}, \ldots]\ , \tag{2.10}$$

which in turn implies that $p(\tau) = f(\tau)$ for $0 \leq \tau \leq 1$. The interpretation of requirement (2.10) is that the discrete time and continuous time forecasts of $z(t)$ over a unit time interval coincide.

When condition (2.9) is met, the link between $P(D)$ and $C(L)$ is particularly simple. Using $f(\tau) = 0$ for $\tau > 1$, equation (2.8) becomes

$$p(\tau) = C_j f(\tau - j) \quad \text{for} \quad j \leq \tau < j + 1\ . \tag{2.11}$$

Equation (2.11) implies that for the particular class of continuous time processes for which $f(\tau) = 0$ for $\tau > 1$, the continuous time moving-average coefficients are completely determined by the discrete time moving-average coefficients and the function $f(\tau)$ defined on the unit interval. The aliasing problem is manifested in this relationship because $f(\tau)$ cannot be inferred from discrete time data. In the absence of additional restrictions, all functions $f(\tau)$ that satisfy

$$\int_0^1 f(\tau) f(\tau)^T\, d\tau = W$$

are observationally equivalent. Relation (2.11) also implies that in general, without some more restrictions on $p(\tau)$, condition (2.9) does not place *any* restrictions on the discrete time moving-average coefficients.

However, in many (if not most) applications, it is usual to impose the additional requirement that the continuous time moving-average coefficients be a continuous function of τ.[16] This requirement together with (2.11) then imposes a very stringent restriction on the discrete time moving-average representation. In particular, (2.11) then implies that

$$C_j\, f(0) = C_{j-1}\, f(1) \tag{2.12}$$

where $f(\tau)$ is now a continuous function on the unit interval. When $w(t)$ and $z(t)$ have the same dimension ($m = n$) and $f(0)$ is nonsingular, relation (2.12) implies that

$$C_j = [f(1)f(0)^{-1}]^j$$

and

$$C(L) = [I - f(1)f(0)^{-1}L]^{-1} .$$

This implies that if (2.9) is to hold, the discrete time process must have a first order autoregressive representation. We have therefore established that condition (2.9) and the continuity requirement on $p(\tau)$ substantially restrict not only the admissible continuous time moving-average coefficients but the admissible discrete time moving-average coefficients as well.

Thus, with a continuous $p(\tau)$ function, in general, relation (2.9) does not hold. Instead, a_t given by (2.6) is a function of all current and past $w(t)$'s, a function whose nature can pose problems in several interrelated ways for interpreting a_t in terms of the continuous time noises $w(t)$ that are imagined to impinge on agents in the model. First, as in the discrete time case, the process $w(t)$ need not be fundamental for $z(t)$ in continuous time. Second, the matrix function $f(\tau)$ in (2.6) is not usually diagonal, so that each component of a_t in general is a function of all of the components of $w(t)$. This is a version of what Geweke (1978) has characterized as "contamination," which occurs in the context of the aggregation over time of several interrelated distributed lags. It is also related to the well-known phenomenon that aggregation over time generally leads to Granger-causality of discrete sampled y to x even when y fails to Granger-cause x in continuous time. Third, the matrix function $f(\tau)$ in (2.5) in general is nonzero for all values of $\tau > 0$, so that a_t in general depends on values of $w(t-\tau)$ in the remote past.

We now turn to our second *desideratum*, namely that the sequence $\{C_j\}_{j=0}^{\infty}$ resemble a sampled version of the function $p(\tau)$. For studying this matter, we set $m = n$, because we are interested in studying circumstances under which $\{C_j\}$ fails to reflect $p(\tau)$ even when the number of white noises n in a_t equals the number m in $w(t)$. We can represent most of the issues here with a univariate example, and so set $m = n = 1$ in most of our discussion. It is also convenient to study the case in which z_t has a rational spectral density in continuous time. Thus we assume that

$$\theta(D)z_t = \psi(D)\, w(t) \tag{2.13}$$

where z_t is a scalar stochastic process, and $\theta(s) = (s-\lambda_1)\,(s-\lambda_2)\ldots(s-\lambda_r)$, $\psi(s) = \psi_0 + \psi_1\, s + \ldots + \psi_{r-1}\, s^{r-1}$. We assume that the real parts of $\lambda_1, \ldots, \lambda_r$, which are the zeroes of $\theta(s)$, are less than zero, but that the real parts of the zeroes of $\psi(s)$ are unrestricted. Only if the real parts of the zeroes of $\psi(s)$ are less than zero do current and past values of $z(t)$ and $w(t)$ span the same linear space. If any zeroes of $\psi(s)$ have real parts that exceed zero, then current and lagged $w(t)$ span a larger space than do current and lagged $z(t)$. The above equation can be expressed as

$$z_t = P(D)\, w(t) \tag{2.14}$$

where $P(D) = \psi(D)/\theta(D)$. A partial fraction representation of $P(D)$ is

$$P(D) = \sum_{j=1}^{r} \frac{\delta_j}{D-\lambda_j} \tag{2.15}$$

where

$$\delta_j = \lim_{s\to\lambda_j} P(s)\,(s-\lambda_j)\, . \tag{2.16}$$

We therefore have

$$p(\tau) = \sum_{j=1}^{r} \delta_j\, e^{\lambda_j \tau}\, . \tag{2.17}$$

Thus, the weighting function $p(\tau)$ in the continuous time moving-average representation is a sum of r exponentially decaying functions. Our object will now be to get an analogous expression to (2.17) for the discrete time coefficients B_k.

It is known that the discrete time process z_t implied by (2.13) is an r^{th} order autoregressive, $(r-1)$ order moving average process. Let this be $z_t = \frac{c(L)}{d(L)}\, a_t$ where $c(L) = \sum_{j=0}^{r-1} c_j\, L^j$, $d(L) = \sum_{j=0}^{r} d_j\, L^j$. To find this representation, we must use (2.4). A.W. Phillips (1959) and Hansen and Sargent (1983b) show that for the process (2.13), the term on the right side of (2.4) can be represented

$$\sum_{j=-\infty}^{\infty} P(i\omega + 2\pi i j)\, P(-i\omega - 2\pi i j) = \sum_{j=1}^{r} \left[\frac{w_j}{(1-e^{\lambda_j}\, e^{-i\omega})} + \frac{w_j\, e^{\lambda_j}\, e^{+i\omega}}{(1-e^{\lambda_j}\, e^{+i\omega})}\right]$$

where

$$w_j = \lim_{s \to \lambda_j} P(s)\, P(-s)\, (s - \lambda_j)\ .$$

Letting $z = e^{-i\omega}$, to find the required mixed moving-average autoregressive representation we must solve

$$\frac{c(z)c(z^{-1})}{d(z)d(z^{-1})} = \sum_{j=1}^{r} \left[\frac{w_j}{1 - e^{\lambda_j} z} + \frac{w_j e^{\lambda_j} z^{-1}}{1 - e^{\lambda_j} z^{-1}} \right] \tag{2.18}$$

subject to the condition that the zeroes of $c(z)$ and $d(z)$ all lie outside the unit circle. The term on the right side of (2.18) can be expressed as

$$\begin{aligned} &\frac{\sum_{j=1}^{r} w_j \,\Pi_{k \neq j}^{r} (1 - \alpha_k z)\, \Pi_{k=1}^{r} (1 - \alpha_k z^{-1})}{\Pi_{j=1}^{r} (1 - \alpha_j z)\, \Pi_{k=1}^{r} (1 - \alpha_k z^{-1})} + \\ &\frac{\sum_{j=1}^{r} w_j \,\alpha_j \,\Pi_{k=1}^{r} (1 - \alpha_k z)\, \Pi_{k \neq j}^{r} (1 - \alpha_k z^{-1})\, z^{-1}}{\Pi_{j=1}^{r} (1 - \alpha_j z)\, \Pi_{k=1}^{r} (1 - \alpha_k z^{-1})} \end{aligned} \tag{2.19}$$

where $\alpha_j \equiv e^{\lambda_j}$. Note that $|\alpha_j| < 1$ by virtue of the assumption that $\mathrm{real}\,(\lambda_j) < 0$. Thus, the denominator is already factored as required, so that

$$d(z) = \Pi_{j=1}^{r} (1 - \alpha_j z)\ . \tag{2.20}$$

The numerator must be factored to find $c(z)$. Standard procedures to find the zeroes of scalar polynomials can be used to achieve this factorization, as described by Hansen and Sargent (1981a).

Thus we have that

$$z_t = \frac{c(L)}{d(L)}\, a_t \equiv C(L)\, a_t\ . \tag{2.21}$$

Proceeding in a similar fashion as we did for the continuous time moving-average representation, we can find a partial fraction representation for $C(L)$, namely

$$C(L) = \sum_{j=1}^{r} \frac{\gamma_j}{1 - \alpha_j L} \tag{2.22}$$

where

$$\gamma_j = \lim_{z \to \alpha_j^{-1}} C(z)\, (1 - \alpha_j z)\ . \tag{2.23}$$

Recalling that $\alpha_j = e^{\lambda_j}$, equation (2.22) implies that

$$C_k = \sum_{j=1}^{r} \gamma_j\, e^{\lambda_j k}. \tag{2.24}$$

Collecting and comparing the key results, we have that

$$p(\tau) = \sum_{j=1}^{r} \delta_j\, e^{\lambda_j \tau}\ , \qquad \tau \in [0, \infty)\ . \tag{2.17}$$

$$C_k = \sum_{j=1}^{r} \gamma_j\, e^{\lambda_j k}\ , \qquad k = 0, 1, 2, \ldots \tag{2.24}$$

Equations (2.17) and (2.24) imply that C_k will be (proportional to) a sampled version of $p(\tau)$ if and only if $\gamma_j/\delta_j = \gamma_1/\delta_1$ for all $j = 2, \ldots, r$. It can be shown directly by using (2.17) and (2.24) in (2.7) and (2.8) that this condition will not be met for any $r \geq 2$. Thus, only if $z(t)$ is a first-order autoregressive process does C_k turn out to be a sampled version of $p(\tau)$.

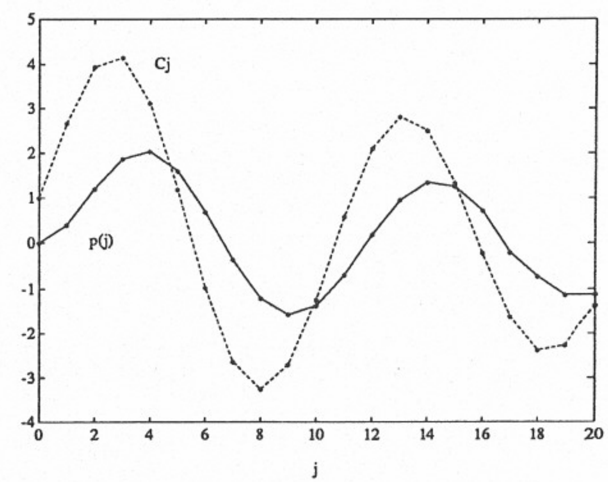

Figure 5. Continuous time ($p(j)$)and discrete time (C_j) moving average kernels.

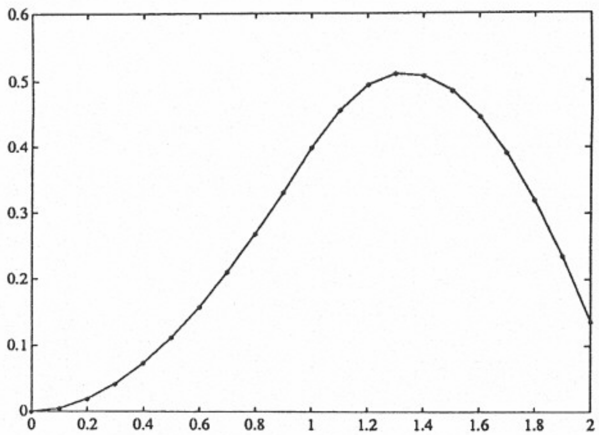

Figure 6. The function $f(s)$.

Table 1 and Figure 5 present a numerical example that illustrates the preceeding ideas. For the univariate process $(D^3 + .6D^2 + .4D + .2)\ z(t) = w(t)$, we have calculated $p(\tau)$, $f(\tau)$, $c(L)$, $d(L)$, $B(L) = c(L)/d(L)$, δ_j, γ_j for $j = 1, 2, 3$. In this example, we have that $\gamma_j/\gamma_1 \neq \delta_j/\delta_1$ for $j \geq 2$, so that the shapes of the moving averages in continuous and discrete time, $p(\tau)$ and C_k, respectively, are different. We plot C_k and $p(\tau)$ for integer values of τ in Figure 5. We also plot $f(\tau)$ in Figure 6. Notice that $f(\tau) \neq 0$ for some τ's greater than 1. In particular, notice that $f(\tau)$ is larger in absolute value over most of the interval [1, 2] than it is over the interval [0, 1]. The failure of $f(\tau)$ to be concentrated on [0, 1] and the failure of B_k to resemble a sampled version of $p(\tau)$ are both consequences of the fact that this is a third order autoregressive system in continuous time, rather than a first order one.

The preceding results and the example generalize readily to the case of a vector stochastic process z_t. Matrix versions of (2.17) and (2.24) hold, where the λ_j's are the zeroes of $\det \theta(s)$ and the δ_j's and the γ_j's are $(n \times n)$ matrices given by (2.16) and (2.23).

Table 1
An Example of Aggregation Over Time

$\psi(D) = 1$

$\theta(D) = .2 + .4D + .6D^2 + D^3$

λ_j (zeroes of $\theta(s)$): $-.5424$, $-.0288 \pm .6066$

δ_j in Partial Fraction Representation of $\psi(D)/\theta(D)$:

			δ_j/δ_1	
j	Real (δ_j)	Imaginary (δ_j)	Real	Imaginary
1	1.5831	0	1.00	0
2	−.7915	.6701	−.50	.423
3	−.7915	.6701	−.50	−.423

Zeroes of Spectral Factorization of Numerator Polynomial ($c(L)$):

	Real Part of Zero	Imaginary Part of Zero	Modulus
1	−.0441	0	.044
2	−.4359	0	.436

γ_j in Partial Fraction Representation of $C(L)$:

			γ_j/γ_1	
	Real (γ_j)	Imaginary (γ_j)	Real	Imaginary
1	1.7984	0	1.000	0
2	−.3992	2.0310	−.222	1.129
3	−.3992	−2.0310	−.222	1.129

Discrete Time Mixed Moving-Average, Autoregressive Representation:

$d(L) = 1 - 2.1779L + 1.8722L^2 - .5485L^3$

$c(L) = 1 + .4800L + .0192L^2$

Table 1 (continued)

τ	$f(\tau)$	$p(\tau)$	C_k
0	0	0	1.000000
.100	.004900	.004900	
.200	.019198	.010108	
.300	.042288	.042288	
.400	.073563	.073563	
.500	.112414	.112414	
.600	.158231	.153231	
.700	.210404	.210404	
.800	.268324	.268324	
.900	.331386	.331386	
1.000	.398987	.398987	2.657971
1.100	.457506	.470529	
1.200	.494395	.545421	
1.300	.510679	.623079	
1.400	.507397	.702926	
1.500	.485602	.784396	
1.600	.446360	.866935	
1.700	.390751	.949999	
1.800	.319860	1.033059	
1.900	.234786	1.115601	
2.000	.136629	1.197125	3.935901
3.000	−.073263	1.860267	4.144677
4.000	.032542	2.029242	3.116763
5.000	−.014212	1.593759	1.188521
6.000	.006197	.692895	−.972014
7.000	−.002701	−.361072	−2.631591
8.000	.001178	−1.208944	−3.259333
9.000	−.000513	−1.576723	−2.705194
10.000	.000224	−1.368770	−1.233866
11.000	−.000098	−.692635	.588609
12.000	.000043	.188765	2.107332
13.000	−.000019	.956663	2.810459
14.000	.000008	1.350008	2.498675
15.000	−.000004	1.252755	1.336741
16.000	.000002	.725963	−.224260
17.000	−.000001	−.020340	−1.619749
18.000	.000000	−.722582	−2.374213
19.000	−.000000	−1.131496	−2.261452
20.000	.000000	−1.124345	−1.369232

2a. Locally Unpredictable Processes and Linear Quadratic Models

The stochastic process $z(t)$ in Table 1 is mean square differentiable,[17] as evidenced by the fact that $p(0) = 0$. A stochastic process of the form (2.1) can be shown to be j times mean square differentiable if $p(0) = p'(0) = p''(0) = \ldots = p^{(j-1)}(0) = 0$ (see Sargent (1983) for a proof). Consequently, the process $(D^3 + .6D^2 + .4D + .2)z(t) = w(t)$ can be verified to be twice (but not three times) mean square differentiable. It is the smoothness and proximity to zero near $\tau = 0$ of $p(\tau)$ that makes it difficult for C_j to resemble a sampled version of $p(\tau)$, and that makes $a(t)$ a poor estimator of $\int_0^1 p(\tau)w(t-\tau)d\tau$.

Sims (1984) argued that there is a class of economic variables that are best modeled as failing to be mean square differentiable. For these processes, $p(0) \neq 0$. Processes of the form (2.1) in which $p(0) \neq 0$ are said to be *locally unpredictable* because if $p(0) \neq 0$, then

$$\lim_{\delta\to 0} \frac{E(x(t+\delta) - \hat{E}_t x(t+\delta))^2}{E(x(t+\delta) - x(t))^2} = 1 \,. \tag{2.25}$$

Here $\hat{E}_t$ is the linear least squares projection operator, conditioned on $\{x(t-s), s \geq 0\}$. Now condition (2.25) can readily be shown to imply that

$$\lim_{\delta\to 0} \frac{E(x(t+\delta) - \hat{E}_t x(t+\delta))^2}{E(x(t+\delta) - \hat{E}(x(t+\delta)|x(t),\, x(t-\delta),\, x(t-2\delta),\, \ldots))^2} = 1 \tag{2.26}$$

In (2.26), $\hat{E}_t\, x(t+\delta)$ is the linear least squares projection of $x(t+\delta)$ conditioned on $(x(t-s), s \geq 0)$, while $\hat{E}(x(t+\delta)|x(t),\, x(t-\delta), \ldots)$ is the projection of $x(t+\delta)$ on the discrete time sample $x(t)$, $x(t-\delta), \ldots$. Condition (2.26) holds for any locally unpredictable process, and states that for small enough sampling interval δ, the δ-ahead projection error from the continuous time process is close in the mean square error sense to the δ-ahead projection error from the δ-discrete time data. Thus, when $p(0) \neq 0$, for small enough δ, the innovation a_t in the δ-counterpart to (2.21) is arbitrarily close to $\int_0^\delta p(s)w(t-s)ds$ in the mean square sense.

Now suppose that $z(t)$ is given by (2.1), with $p(0) = 0$, so that $z(t)$ is mean square differentiable. Following Sims (1984), suppose that the economist is interested in studying the expectational variable $x^*(t)$ given by

$$x^*(t) = \hat{E}\Big[\int_0^\infty e^{\rho s} z(t+s)ds|(z(t-\tau),\ t \geq 0)\Big] \tag{2.27}$$

where $\rho < 0$. Hansen and Sargent (1981d) showed that

$$\begin{aligned} x^*(t) &= \Big[\frac{-P(D)+P(-\rho)}{D+\rho}\Big]\, w(t) \equiv G(D)w(t) \\ &= \int_0^\infty g(s)w(t-s)ds\ , \end{aligned} \tag{2.28}$$

where $P(s) = \int_0^\infty e^{-\tau s} p(\tau)d\tau$ is the Laplace transform of $p(\tau)$. Now if $G(s)$ is the Laplace transform of $g(\tau)$, with support $[0, \infty)$, the initial value theorem for Laplace transforms states that

$$g(0) = \lim_{s\to\infty} s\, G(s)\ .$$

Using the initial value theorem together with (2.28), we find that

$$g(0) = \lim_{s\to\infty} s\Big[\frac{-P(s)+P(-\rho)}{s+\rho}\Big] = P(-\rho) \neq 0\ .$$

(We know that $P(-\rho) \neq 0$ because $P(s)$ is assumed to have no zeroes in the right half of the complex plane by the assumption that $p(\tau)$ is the kernel associated with a Wold representation for $z(t)$.) Therefore, even if $p(0) = 0$, $g(0) \neq 0$, so that the geometric expectational variable $x^*(t)$ fails to be mean square differentiable and therefore is locally unpredictable. For such expectational variables, (2.26) holds. Therefore, for such variables, for small enough sampling interval δ, the discrete time innovation $a(t)$ corresponding to (2.21) is close to $\int_0^\delta p(s)w(t-s)ds$ in the mean squared sense.

These results imply that for a variable $x^*(t)$ and sufficiently small sampling interval δ, the situation is not as bad as is depicted by the example in Table 1. As Sims has pointed out, there are theories of consumption and asset pricing which imply that consumption or asset prices behave like $x^*(t)$ and are governed by a version of (2.27). For example, with $x^*(t)$ being consumption and $z(t)$ income, (2.27) is a version of the permanent income theory. Alternatively, with $x^*(t)$ being a stock price and $z(t)$ being the dividend process, (2.27) is a simple version of an asset-pricing formula.

However, there is a wide class of generalized adjustment cost models discussed by Hansen and Sargent (1981a, 1981d) in which observable variables are such smoothed versions of $x^*(t)$ that they *are* mean square continuous. In adjustment cost models, decisions are driven by convolutions of $x^*(t)$, not by $x^*(t)$ alone. For example, the stochastic Euler equation for a typical quadratic adjustment cost problem is

$$(D-\rho)\, k(t) = E_t\, \Big(\frac{1}{D+\rho}\Big)\, z(t)$$

where $\rho > 0$, or

$$(D-\rho)\, k(t) = x^*(t).$$

Here $k(t)$ is *capital*. The solution for capital is then

$$k(t) = \frac{1}{D-\rho}\, x^*(t)$$

or

$$k(t) = \Big(\frac{1}{D-\rho}\Big)\, \Big[\frac{-P(D)+P(-\rho)}{D+\rho}\Big]\, w(t)$$

where $z(t) = \int_0^\infty \rho(s)\, w(t-s)$. Let

$$k(t) = \int_0^\infty h(\tau)w(t-\tau)d\tau$$

and

$$H(s) = \int_0^\infty e^{-\tau s} h(\tau)d\tau\ .$$

Then

$$H(s) = \Big(\frac{1}{s-\rho}\Big)\, \Big(\frac{-P(s)+P(-\rho)}{s+\rho}\Big)\ .$$

Using the initial value theorem to calculate $h(0)$, we have

$$h(0) = \lim_{s\to\infty} sH(s) = 0\ .$$

Thus, $k(t)$ is mean square differentiable and so is locally predictable. (The convolution integration required to transform $x^*(t)$ to $k(t)$ *smooths* $k(t)$ relative to $x^*(t)$.)

More generally, the endogenous dynamics of adjustment cost models typically lead to mean square differentiable endogenous variables, provided that the agent is posited to be facing mean square differentiable forcing processes $(z(t))$. This means that for such models, the difficulties of interpretation that are illustrated in Table 1 cannot be eluded by appealing to an approximation based on the limit (2.26).

2b. Remedies in Continuous Time Analyses

The preceding problems of interpretation are results of estimating vector autoregressions while foregoing the imposition of any explicit economic theory in estimation. These problems can be completely overcome if a sufficiently restrictive and reliable dynamic model economy is available to be imposed during estimation. For example, Hansen and Sargent (1980b, 1981d) have described how the function $p(\tau)$ can be identified and estimated from observations on discrete time data in the context of a wide class of linear rational expectations models. The basic idea is that the rich body of cross-equation restrictions that characterize dynamic linear rational expectations models can be used to identify a unique continuous time model from discrete time data.

If an estimate of $p(\tau)$ is available, then by using only discrete time data on $\{z_t\}$, it is even possible to recover an estimate of the one-step prediction error that agents are making in continuous time. This is accomplished by treating the continuous time forecast error as a hidden variable whose covariances with the discrete time process $\{z_t\}$ are known. Thus, given estimates of $p(\tau)$, let us define the one-step ahead prediction error from continuous time data as $e_t^* = \int_0^1 p(\tau)\, w(t-\tau)\, d\tau$. Then it is straightforward to calculate the following second moments:

$$E(z_t z_{t-j}^T) = \int_0^\infty p(\tau+j)p(\tau)^T d\tau = \sum_{k=0}^{\infty} C_{k+j} W C_k^T,\ j \geq 0$$

$$E(e_t^*\, z_{t+j}^T) = \begin{cases} \int_0^1 p(\tau)p(\tau+j)^T d\tau & j \geq 0 \\ 0 & j < 0\,. \end{cases}$$

We can estimate the projection $\sum_{j=-m_1}^{m_2} D_j\ z_{t-j}$ in the projection equation

$$e_t^* = \sum_{j=-m_1}^{m_2} D_j\ z_{t-j} + u_t$$

where u_t is orthogonal to z_{t-j} for all $j = -m_1\ \ldots,\ m_2$. The D_j's can be computed from the normal equations

$$E\ (e_t^*\ z_{t+k}^T) = \sum_{j=m_1}^{m_2} D_j\ E\ z_{t-j}\ z_{t+k}^T,\ k = -m_2,\ \ldots,\ m_1\,.$$

These calculations could be of use if one's aim were truly to extract and to interpret estimates of the forecast errors made by agents. In continuous time versions of various models, such as those of Lucas (1973) or Barro (1977), agents' forecasting errors are an important source of impulses, so that it is of interest to have this method for characterizing their stochastic properties and estimating them.

3. Concluding Remarks

Subsequent chapters will treat aspects of the issues that we have studied in this chapter. The next chapter by Hansen, Roberds, and Sargent describes some discrete time tax and consumption models in which the history of innovations in a vector autoregression fails to equal the history of information possessed by agents. This poses problems in testing a key feature of the models, namely, a form of present value budget balance. For these particular models, the paper describes and implements a testing strategy that is an alternative to the *remedy in discrete time* described above.

Analysis of the issues raised in Section 2 on continuous time model is taken up and extended in Chapter 10 by Marcet, who relaxes our assumption that the continuous time spectral density is rational. This lets him study continuous time processes that have Wold representations with discontinuous moving average kernels. The remaining Chapters 7, 8, and 9 contain a variety of technical results that would be useful in implementing the *remedy in continuous time* described above.

Appendix

This appendix describes the recursive methods by which Figures 1–4 were computed.

We computed the objects described in Section 1 by mapping the model into the class of economies described by Hansen and Sargent (1990), and by using the computer programs that they describe. Hansen and Sargent (1990) describe a class of economies whose equilibrium allocations solve the social planning problem: choose stochastic processes $\{c_t,\ s_t,\ \ell_t,\ g_t,\ k_t,\ h_t\}_{t=0}^\infty$ to maximize

$$(A1) \qquad -\left(\frac{1}{2}\right) E_0 \sum_{t=0}^{\infty} \beta^t\ [(s_t - b_t)\ \cdot\ (s_t - b_t) + \ell_t^2]$$

subject to

$$
(A2)\qquad
\begin{aligned}
\Phi_c\, c_t + \Phi_i\, i_t + \Phi_g\, g_t &= \Gamma k_{t-1} + d_t \\
g_t \cdot g_t &= \ell_t^2 \\
k_t &= \Delta_k\, k_{t-1} + \Theta_k\, i_t \\
h_t &= \Delta_h\, h_{t-1} + \Theta_h\, c_t \\
s_t &= \Lambda\, s_{t-1} + \Pi\, c_t \\
z_{t+1} &= A_{22}\, z_t + C_2\, w_{t+1} \\
b_t &= U_b\, z_t\ ,\ d_t = U_d\, z_t\ .
\end{aligned}
$$

In (A1), s_t is a vector of consumption service flows, b_t is a vector of preference shocks, ℓ_t is labor supply, c_t is a vector of consumption rates, i_t is a vector of investment rates, g_t is a vector of "intermediate goods," k_{t-1} is a vector of capital stocks, d_t is a vector of "endowment shocks," h_t is a vector of consumer durables, $\{w_{t+1}\}$ is a vector white noise with $E\, w_{t+1}\, w_{t+1}^T = I$, and z_t is an exogenous state vector of information. The planner maximizes A1 subject to A2 by choosing contingency plans for $\{c_t,\ i_t,\ \ell_t,\ g_t,\ k_t,\ h_t\}$ as functions of information known at t, namely $x_t^T \equiv [h_{t-1}^T,\ k_{t-1}^T,\ z_t^T]$.

There is a potential source of notational confusion because several sets of notations in $(A1)$–$(A2)$ were used to denote different objects in the model of Section 1. To avoid confusion, in this appendix we simply place a $(\bar{\ })$ above any variable in the model of Section 1 that might be confused with a similarly named variable in $(A1)$–$(A2)$.

We begin by eliminating p_t from equations (1.21) and (1.22) to obtain a single Euler equation in q_t:

$$
(A3)\qquad
\begin{aligned}
E_t\Big\{[h_d + g_d\, a(\beta L^{-1})\, a(L) + h_s + g_s(1-\beta L^{-1})\,(1-L)]\, q_t\Big\} \\
+\, \bar{d}_t + \bar{s}_t = 0
\end{aligned}
$$

To obtain a version of the social planning problem $(A1)$–$(A2)$ whose quantities (and shadow prices) solve the model of Section 1, our strategy is to choose the objects in $(A2)$ so that the Euler equations for problem $(A1)$–$(A2)$ match up with $(A3)$.

Let $f_1,\ f_2,\ f_4,\ f_5,\ f_6,\ f_7$ be undetermined scalars which we shall eventually set in order to match $(A3)$ with the Euler equation for problem $(A1)$–$(A2)$. Then we propose the following settings for the objects in $(A2)$:

$$
\begin{aligned}
s_{1t} &= f_2\, q_t \qquad\qquad (q_t = c_t) \\
s_{2t} &= f_1\, [a(L)\, q_t] \\
&= f_1\, a_0\, q_t + f_1\, [a_1\ a_2\ a_3\ a_4] \begin{bmatrix} q_{t-1} \\ q_{t-2} \\ q_{t-3} \\ q_{t-4} \end{bmatrix} \\
b_{2t} &= 0 \\
b_{1t} &= f_4\, \bar{d}_t
\end{aligned}
$$

Thus, we set

$$
\underbrace{\begin{bmatrix} s_{1t} \\ s_{2t} \end{bmatrix}}_{s_t} = \underbrace{f_1 \begin{bmatrix} 0 & 0 & 0 & 0 \\ a_1 & a_2 & a_3 & a_4 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} q_{t-1} \\ q_{t-2} \\ q_{t-3} \\ q_{t-4} \end{bmatrix}}_{h_{t-1}} + \underbrace{\begin{bmatrix} f_2 \\ f_1\, a_0 \end{bmatrix}}_{\Pi} \underbrace{q_t}_{c_t}
$$

$$
\underbrace{\begin{bmatrix} q_t \\ q_{t-1} \\ q_{t-2} \\ q_{t-3} \end{bmatrix}}_{h_t} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\Delta_h} \underbrace{\begin{bmatrix} q_{t-1} \\ q_{t-2} \\ q_{t-3} \\ q_{t-4} \end{bmatrix}}_{h_{t-1}} + \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\Theta_h} q_t
$$

As for the technology, we specify

$$
\begin{aligned}
g_{1t} - f_5\, q_t &= f_6\, \bar{s}_t \\
g_{2t} - f_7\, i_t &= 0 \\
c_t &= k_t \\
k_t &= k_{t-1} + i_t
\end{aligned}
$$

To implement these specifications, we set

$$
\underbrace{\begin{bmatrix} 1 \\ -f_5 \\ 0 \end{bmatrix}}_{\Phi_c} c_t + \underbrace{\begin{bmatrix} -1 \\ 0 \\ -f_7 \end{bmatrix}}_{\Phi_i} i_t
$$

$$
+ \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\Phi_g} \begin{bmatrix} g_{1t} \\ g_{2t} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}_{\Gamma} k_{t-1} + \begin{bmatrix} 0 \\ f_6\, \bar{s}_t \\ 0 \end{bmatrix}
$$

where $k_t = 1 \cdot k_{t-1} + 1 \cdot i_t$, so that $\Delta_k = \Theta_k = 1$. We set $d_{2t} = f_6\, \bar{s}_t$.
With these settings, we have that

$$(s_t - b_t)\cdot(s_t - b_t) = (f_2 q_t - f_4 \bar{d}_t)^2 + (f_1\, a(L)\, q_t)^2$$
$$= f_2^2 q_t^2 + f_4^2 \bar{d}_t^2 - 2 f_2 f_4 q_t \bar{d}_t + f_1^2 \left[a(L)\, q_t\right]^2 .$$

We also have that

$$g_t \cdot g_t = (f_5 q_t + f_6 \bar{s}_t)^2 + f_7^2\, i_t^2$$
$$= f_5^2 q_t^2 + f_6^2 \bar{s}_t^2 + 2 f_5 f_6 q_t \bar{s}_t + f_7^2\, (k_t - k_{t-1})^2 .$$

Thus, the social planning criterion $(A1)$ can be expressed as

$$(A4) \qquad -.5\, E \sum_{t=0}^{\infty} \beta^t J_t$$

where

$$J_t = \Big\{ f_2^2\, q_t^2 + f_4^2\, \bar{d}_t^2 - 2 f_2 f_4 \bar{d}_t + f_1^2 \left[a(L)\, q_t\right]^2 + f_5^2\, q_t^2 + f_6^2 \bar{s}_t^2 + 2 f_5 f_6\, q_t \bar{s}_t + f_7^2\, (k_t - k_{t-1})^2 \Big\}$$

Applying techniques described in Sargent (1987b, ch. IX and XVI), the Euler equation for $(A4)$ is evidently

$$(A5) \qquad E_t \Big\{ f_2^2\, q_t - f_2 f_4 q_t \bar{d}_t + f_1^2 \left[a(\beta L^{-1})\, a(L)\right] q_t + f_5^2\, q_t + f_5 f_6 \bar{s}_t + f_7^2\, (1 - \beta L^{-1})\, (1 - L)\, q_t \Big\} = 0 .$$

To make $(A5)$ match $(A3)$, it suffices to use the following settings for the f_j's:

$$f_1 = (g_d)^{1/2} \qquad f_5 = (h_s)^{1/2}$$
$$f_2 = (h_d)^{1/2} \qquad f_6 = (h_s)^{-1/2}$$
$$f_4 = -(h_d)^{-1/2} \qquad f_7 = (g_s)^{1/2} .$$

As for the information specification, we set

$$z_t^T = [w_{st},\, w_{st-1},\, w_{st-2},\, w_{st-3},\, w_{dt},\, w_{dt-1},\, w_{dt-2},\, w_{dt-3}]$$

$$A_{22} = \begin{bmatrix} C_4 & O_4 \\ O_4 & C_4 \end{bmatrix} , \quad C_2^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

where C_4 is the (4×4) companion matrix of $[0\ 0\ 0\ 0]$, and O_4 is the (4×4) zero matrix. We set

$$U_b = f_4^{-1} \begin{bmatrix} bs & O_{1\times 4} \\ O_{1\times 8} & \end{bmatrix} , \quad U_d = f_6^{-1} \begin{bmatrix} O_{1\times 8} & \\ O_{1\times 4} & bd \\ O_{1\times 8} & \end{bmatrix}$$

where

$$bs = [B_{s0},\, B_{s1},\, B_{s2},\, B_{s3}]$$
$$bd = [B_{d1},\, B_{d2},\, B_{d3},\, B_{d4}]$$

where the B_{sj} and B_{dj} are defined in (1.23). We set U_b and U_d as above, because we want to implement $b_{1t} = f_4\, \bar{d}_t$ and $d_{2t} = f_6\, \bar{s}_t$ or equivalently, $\bar{d}_t = f_4^{-1}\, b_{1t}$ and $\bar{s}_t = f_6^{-1}\, d_{2t}$.

A solution of the social planning problem is computed by mapping A1 – A2 into a discounted dynamic programming problem. The solution is represented in the form of the stochastic difference equation

$$(A6) \qquad x_{t+1} = A^o\, x_t + C\, w_{t+1} ,$$

where

$$x_t = \begin{bmatrix} h_{t-1} \\ k_{t-1} \\ z_t \end{bmatrix} .$$

Hansen-Sargent (1990) show how to decentralize the social planning problem via a competitive equilibrium. It turns out that the spot price of consumption in the general equilibrium model of Hansen-Sargent equals p_t of the model of Section 1.[18] It also turns out that p_t is simply a linear function of the state x_t, as is q_t. The marginal utility of consumption ($q_t = c_t$) for this model can be shown to be given by

$$-E_t\, \{h_d + g_d\, a(\beta L^{-1})\, a(L)\}\, q_t - \bar{d}_t ,$$

which equals p_t by virtue of equation (1.22). Hansen and Sargent (1990) compute the marginal utility of consumption in the form $M_c\, x_t$, where M_c is a matrix comformable to x_t. Thus, we compute the price for our model simply as $p_t = M_c\, x_t$. Similarly, the level of consumption, q_t, is represented as $q_t = S_c\, x_t$, where Hansen and Sargent (1990) describe how to compute S_c.

Hansen-Sargent's computations can be used to represent the solutions of the model in the state space form

$$(A7) \qquad \begin{aligned} x_{t+1} &= A^o\, x_t + C\, w_{t+1} \\ z_t &= G\, x_t \end{aligned}$$

where $z_t = \begin{bmatrix} q_t \\ p_t \end{bmatrix}$ and where $G = \begin{bmatrix} S_c \\ M_c \end{bmatrix}$. Representation A7 is a state-space version of representation (1.13). We computed the impulse response function from (A7) to create Figure 1, which gives the impulse response of $S(L)\, z_t = R(L)\, \varepsilon_t$.

Using the Kalman filter, we obtained the following *innovations representation* associated with (A7):

$$\begin{aligned} \hat{x}_{t+1} &= A^o\, \hat{x}_t + K a_t \\ z_t &= G\, \hat{x}_t + a_t \end{aligned} \tag{A8}$$

where $a_t = z_t - E\,[z_t \mid z_{t-1}, z_{t-2}, \ldots]$, $\hat{x}_t = E[x_t | z_{t-1}, z_{t-2}, \ldots]$, and K is the Kalman gain. Representation (A8) corresponds to the Wold (fundamental) moving average representation for z_t given by representation (1.17). The innovation a_t in (A8) equals $R_0^*\, \varepsilon_t^*$ of (1.17). We created Figures 2 and 3 by computing impulse response functions from (A8).

To create Figure 4, we used (A7)–(A8) as a coupled system. First, we *turned around* A8 to achieve the *whitener*

$$\begin{aligned} \hat{x}_{t+1} &= (A^o - KG)\, \hat{x}_t + K\, z_t \\ a_t &= \; -G\, \hat{x}_t + z_t\,. \end{aligned} \tag{A9}$$

Then we created the coupled system formed by taking the *output* z_t of (A7) as the *input* of system (A9). We then used the coupled system to compute the impulse response functions of a_t (i.e., $R_t^*\, \varepsilon_t^*$) with respect to w_t (i.e., ε_t) that are reported in Figure 4.

Notes

1. This is, after all, the construction used in Wold's decomposition theorem.

2. Throughout this paper, we use $\hat{E}$ to denote the linear least squares projection operator.

3. Representations of the moving average in (1) are not in general unique once one relaxes the restriction in (1) that $A_0^\Delta = -I$, which in turn implies that $C_0^\Delta = I$. If this restriction is relaxed, then any representation generated by slipping a UU^T in between C_j^Δ and $a(t - \Delta j)$ in (4), where U is a unitary matrix ($UU^T = I$), is also a fundamental moving-average representation. That is,

$$z(t) = \sum_{j=0}^{\infty} (C_j^\Delta U)\; (U^T a(t - \Delta j))$$

is also a fundamental moving-average representation, since $U^T\; a(t)$ spans the same linear space as $a(t)$. In terms of such a representation, the decomposition of prediction error covariance becomes

$$\begin{aligned} &E(z(t) - \hat{E}_{t-j} z(t))\; (z(t) - \hat{E}_{t-j} z(t))^T \\ &\quad = \sum_{k=0}^{j-1} C_j^\Delta\; U\; V\; U^T\; C_j^{\Delta T}, \end{aligned}$$

which is altered by alternative choices of U. Sims' choice of orthogonalization order amounts to a choice of U.

4. An earlier version of this paper considered four classes of examples, the other two being nonlinearities and aggregation across agents. Due to length constraints, we decided to restrict this paper to the two classes of examples studied here.

5. Danny Quah has conveyed to us the viewpoint that implicit in the desire to match the $\{\varepsilon_t\}$ process of the economic model (1.8) with the $\{a_t\}$ process of the vector autoregression (1) must be a decision problem that concerns the data analyst. For example, on the basis of variance decompositions based on (7), the analyst might want to predict the consequences of *interventions* in the form of alterations in various diagonal elements of the innovation covariance matrix V, interpreting these alterations, e.g., as changes in the predictability by agents of various economic process, such as the money supply.

6. This assumption is made in the interest of providing the best possible chance that the processes $\{a_t\}$ and $\{\varepsilon_t\}$ match up. If ε_{1t} is a vector of dimension greater than n_1, then in general current and lagged values of $(\varepsilon_{1t}, \varepsilon_{2t})$ span a larger linear space than do current and lagged values of (y_t, x_{2t}).

7. See Rozanov (1967).

8. This inequality means that $E(R_0^* \varepsilon_t^*)(R_0^* \varepsilon_t^*)^T - E(R_0 \varepsilon_t)(R_0 \varepsilon_t)^T$ is a positive definite matrix.

9. The appendix maps the current model into a social planning problem that is a special case of the one studied by Hansen and Sargent (1990). By using Hansen and Sargent's interpretation of their setup as a general competitive equilibrium, it is possible to produce a general equilibrium interpretation of the model in the text.

10. It is interesting to note that although this system is one in which there are not strictly econometrically exogenous variables, or even any variables that are not Granger-caused by any others, its parameters are in principle identifiable. Identification is achieved through the cross-equation restrictions. Even when (w_{st}, w_{dt}) lie in the space spanned by the one-step ahead errors in predicting (q_t, p_t) from their own pasts, it is necessary to know the structural parameters of the model in order to deduce the former from the latter innovations.

11. The computational methods are described in the appendix. Briefly, we proceeded by mapping the problem into the general setup of Hansen and Sargent (1990), and using their computer programs.

12. By increasing the absolute values of the zeros of the polynomials $B_d(L)$ and $B_s(L)$ we were able to generate more "spectacular" examples of the phenomenon under discussion, in the sense that the discrepancy between the two covariance matrices in inequality (1.19) was even larger.

13. See Sims (1980) for a treatment of orthogonalization orders. Different "orthogonalization orders" in the sense of Sims amount to different triangular choices of the orthogonal matrix U that appears in footnote 3. If U^T is chosen to be upper triangular, then the first component of a_t corresponds to the first component of the new (basis) fundamental noise $U^T a_t$. On the other hand, if U^T is chosen to be lower triangular, the last component of a_t gets to go first in the Gram-Schmidt process that is used to create $U^T a_t$.

14. Practical methods for solving this equation for the case in which $P(s)$ is rational are discussed by Phillips (1959), Hansen and Sargent (1980b), and Christiano (1980).

15. An alternative derivation of (2.7) uses operational calculus. Setting $L = e^{-D}$, express (2.5) as $a_t = V(e^{-D})\ P(D)\ w(t) \equiv f(D)\ w(t)$. Here the function $f(\tau)$ is the inverse Fourier transform of $F(i\omega)$, which is defined by

$$F(i\omega) = C(e^{-i\omega})^{-1}\ P(i\omega)\ .$$

Equation (2.7) follows from the above equation by the convolution property of Fourier transforms.

16. For example, the function $p(\tau)$ will be continuous whenever $P(D)$ is rational, a common specification in applied work. The functions $p(\tau)$ and $f(\tau)$ are only defined up to an L^2 equivalence. Consequently, we can only impose continuity on one version of the continuous time moving average coefficients.

17. See Sargent (1983) for definitions of mean square continuity and mean square differentiability.

18. The spot price of consumption is p_t^t in the language of Hansen-Sargent.