Journal of Economic Dynamics and Control 2 (1980) 7–46.

FORMULATING AND ESTIMATING DYNAMIC LINEAR RATIONAL EXPECTATIONS MODELS*

Lars Peter HANSEN
Carnegie-Mellon University, Pittsburgh, PA 15213, USA

Thomas J. SARGENT
University of Minnesota, and Federal Reserve Bank, Minneapolis, MN 55455, USA

This paper describes methods for conveniently formulating and estimating dynamic linear econometric models under the hypothesis of rational expectations. An econometrically convenient formula for the cross-equation rational expectations restrictions is derived. Models of error terms and the role of the concept of Granger causality in formulating rational expectations models are both discussed. Tests of the hypothesis of strict econometric exogeneity along the lines of Sims's are compared with a test that is related to Wu's.

1. Introduction

This paper describes research which aims to provide tractable procedures for combining econometric methods with dynamic economic theory for the purpose of modeling and interpreting economic time series. That we are short of such methods was a message of Lucas' (1976) criticism of procedures for econometric policy evaluation. Lucas pointed out that agents' decision rules, e.g., dynamic demand and supply schedules, are predicted by economic theory to vary systematically with changes in the stochastic processes facing agents. This is true according to virtually any dynamic theory that attributes some degree of rationality to economic agents, e.g., various versions of 'rational expectations' and 'Bayesian learning' hypotheses. The implication of Lucas' observation is that instead of estimating the parameters of decision rules, what should be estimated are the parameters of agents' objective functions and of the random processes that they faced historically. Disentangling the parameters governing the stochastic processes that agents face from the parameters of their objective functions would enable the econometrician to predict how agents' decision rules would change across

*This research was supported by the Federal Reserve Bank of Minneapolis. 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. Helpful comments on an earlier draft were made by John Taylor.

alterations in their stochastic environment. Accomplishing this task is an absolute prerequisite of reliable econometric policy evaluation. The execution of this strategy involves estimating agents' decision rules jointly with models for the stochastic processes they face, subject to the cross-equation restrictions implied by the hypothesis of rational expectations.[1] However, even for very simple models, these cross-equation restrictions are of a complicated form, so that in applications substantial technical problems exist even about the best way of expressing these restrictions mathematically. This paper aims to extend what is known about conveniently characterizing these restrictions and estimating models subject to them.

Our work here involves a setup in which the environment and agents' decision rules can be modeled as time invariant linear stochastic difference equations. Such setups are attractive because they are ones for which the dynamic stochastic optimization theory is tractable analytically, and because it is convenient for econometric reasons to remain within the well developed domain of time invariant linear stochastic difference equations. In this paper, we adopt the device of carrying out the entire discussion in terms of a simple example, that of a firm devising a contingency plan for the employment of a single factor of production subject to quadratic costs of adjustment and uncertain technology and factor rental processes. This has the advantage of keeping the discussion simple and concrete, while setting aside several technical complications that arise in more general settings, e.g., models with multiple factors. Virtually every issue we deal with here appears in the more complicated setups. Included among the topics treated in this paper are the following:

(1) *Derivation of a convenient expression for the decision rule.* Having tractable expressions for the restrictions across the parameters of stochastic processes that agents face and their decision rules is necessary in order to make rational expectations modeling applicable to problems of even moderate dimension. Success in this part of our work will in effect extend the size of rational expectations systems that are manageable.

(2) *Delineation of the natural role played by 'Granger causality' in these models.* Formulating and estimating models of this type naturally requires use of the concept of 'Granger causality' (1969). In dynamic problems, agents' decision rules typically involve predictions of future values of the stochastic processes, say w_t, that they care about but can't control, e.g., in competitive models output prices and/or input prices. Theory asserts that current and past values of any stochastic processes that help predict w_t belong in the

[1]Examples of such cross-equation restrictions in simple setups are in Lucas (1972), Sargent (1978a, b), and Taylor (1978, 1979).

decision rules. This is equivalent with saying that all processes agents see and that 'Granger cause' w_t belong in agents' decision rules. Further, given the appropriate conditioning set or universe with respect to which Granger causality is defined, it is usually correct to assume that the decision variable of a competitive firm fails to Granger cause w_t. It is for these reasons that analysis of such models naturally leads to heavy utilization of the concept of Granger causality. However, it should be recognized that in some settings w_t fails to be Granger caused by the firm's decision variable only when the firm's information set used to forecast w_t includes market-wide totals of the firm's decision variable. This occurs, for example, when market-wide employment contributes to the determination of the factor wage w_t.

(3) *Delineation of the relationship between Granger causality and econometric exogeneity.* Sime (1972) has shown that if x_t is to be strictly exogenous in a behavioral relationship expressing n_t as a one sided distributed lag of x_t, then n_t must fail to Granger cause x_t. So failure of n_t to Granger cause x_t is a necessary condition for x_t to be strictly exogenous. It is not a sufficient condition, however, which will be evident in the context of this paper. Below we develop a statistical test of a stronger sufficient condition which is applicable to situations in which the economic behavioral relationship in question is a decision rule expressing n_t as a one-sided distributed lag of x_t. The restrictions that the assumption of rational expectations imposes across the decision rule and the stochastic process for x_t are essential in making the test feasible. This test is related to Wu's (1973) test for exogeneity so that a useful by-product of this paper is to clarify the relationship between Wu's test for exogeneity and Sims' test.

(4) *Development of models of the error terms in estimated decision rules.* This paper develops two different models of the error terms in behavioral equations. Both models use versions of the assumption that private agents observe and respond to more data than the econometrician possesses. Each model imposes substantial structure on the error term and limits the freedom of the econometrician in certain respects to be described. Together with variants of 'errors in variables' models, these models are about the only plausible models of the error processes that we can imagine. The rational expectations or equilibrium modelling strategy virtually forces the econometrician to interpret the error terms in behavioral equations in one of these ways. The reason is that the dynamic economic theory implies that agents' decision rules are *exact* (non-stochastic) functions of the information they possess about the relevant state variables governing the dynamic process they wish to control. The econometrician must resort to *some* device

to convert the exact equations delivered by economic theory into inexact (stochastic) equations susceptible to econometric analysis.

(5) *Development of estimation strategies for rational expectations models.* The discussion of this topic will draw heavily on each of the preceding four topics as we discuss methods for the tractable, consistent, and asymptotically efficient estimation of rational expectations models.

It should be emphasized that the techniques we describe are applicable to an entire class of problems of which our factor demand example is only one member. Other setups that involve identical conceptual and estimation problems include linear–quadratic versions of Lucas and Prescott's (1971) model of investment under uncertainty [e.g. Blanco (1978) or Sargent (1979)], versions of Cagan's model of portfolio balance during hyperinflations [e.g. Salemi and Sargent (1980) and Sargent (1977)], and rational expectations versions of Friedman's permanent income theory of consumption [e.g. Hall (1978) or Sargent (1978a)]. The essential characteristic of these examples is that each can be reduced to a problem in which an agent is choosing a linear dynamic contingency plan for a single variable. Extensions to multivariable dynamic choice problems are deferred to a sequel to this paper.[2]

2. Formulas for decision rules

A firm employing a single factor of production chooses a contingency plan for that factor n_t to maximize its expected present value

$$\lim_{N\to\infty} E_t \sum_{j=0}^{N} \beta^j[(\gamma_0 + a_{t+j} - w_{t+j})n_{t+j} - (\gamma_1/2)n_{t+j}^2 - (\delta/2)(n_{t+j} - n_{t+j-1})^2], \tag{1}$$

subject to n_{t-1} given, where n_t is employment of the factor at time t, w_t is the real factor rental, and a_t is a random shock to technology which is seen by the firm but unobserved by the econometrician.[3] We shall think of n_t as

[2]Aspects of estimating models under rational expectations are discussed by McCallum (1976), Shiller (1972), Wallis (1980), Taylor (1978, 1979), and Revankar (1980). While our estimation problems share many common features with the ones treated in these papers, our setup tends to impose more structure on the estimation problem because the process that we estimate is in effect a 'closed-loop system' resulting from the solution of an optimum problem by a private agent or by a fictitious planner. The paper by Kennan (1979) estimates parameters of a model similar to ours by estimating the stochastic Euler equation.

[3]The price of the output good has not been included in our formulation. One can view this as a world in which there is only one consumption good. Alternatively, we could formally introduce the output price into the analysis. However γ_1 would have to be set to zero so that no third-order terms enter into the objective function.

being employment of the single factor labor and w_t as the real wage, but it would be equally appropriate to regard n_t as the stock of capital and w_t as the real rental on capital.[4] In (1) γ_0, γ_1 and δ are each positive constants, while the constant discount factor β satisfies $0<\beta<1$. The notation $E_t(y)$ denotes the mathematical expectation of the random variable y, conditioned on information available at time t, an information set to be specified shortly. The firm faces a stochastic process for a_t of the form

$$a_t=\alpha_1 a_{t-1}+\ldots+\alpha_q a_{t-q}+v_t^a,$$

or

$$\alpha(L)a_t=v_t^a,$$

where L is the lag operator and where $\alpha(L)=1-\alpha_1 L-\ldots-\alpha_q L^q$, α_j being a scalar for all j. We assume that w_t is the first element of a vector autoregressive process x_t that satisfies

$$\zeta(L)x_t=v_t^x,$$

where x_t and v_t^x are each $(p\times 1)$, and where

$$\zeta(L)=I-\zeta_1 L-\ldots-\zeta_r L^r.$$

The matrix ζ_j is $p\times p$ and so is conformable in dimension with the vector x_t for $j=1,\ldots,r$. Here (v_t^a, v_t^x) are the innovations for the joint (a_t, x_t) process. More particularly, we assume

$$v_t^a=a_t-E[a_t|a_{t-1}, a_{t-2},\ldots, x_{t-1}, x_{t-2},\ldots],$$
$$v_t^x=x_t-E[x_t|x_{t-1}, x_{t-2},\ldots, a_{t-1}, a_{t-2},\ldots].$$

[4]In particular, we could assume that the firm chooses capital k_t to maximize

$$V_t=E_t\sum_{j=0}^{\infty}\beta^j[(\gamma_0+a_{t+j})k_{t+j}-(\gamma_1/2)k_{t+j}^2 - J_{t+j}(k_{t+j}-k_{t+j-1})-(\delta/2)(k_{t+j}-k_{t+j-1})^2],$$

subject to k_{t-1} given. Here J_t is the relative price of capital at time t, relative to the firm's output. Using 'summation by parts' it is easy to verify that the above expression is equivalent with

$$V_t=E_t J_t k_{t-1}+E_t\sum_{j=0}^{\infty}\beta^j[(\gamma_0+a_{t+j})k_{t+j}-(\gamma_1/2)k_{t+j}^2 - (J_{t+j}-\beta E_{t+j}J_{t+j+1})k_{t+j}-(\delta/2)(k_{t+j}-k_{t+j-1})^2].$$

Since k_{t-1} is given, the same decision rule for k_t will be found by maximizing (1) in the text with $k_t=n_t$ and $J_t-\beta E_t J_{t+1}=w_t$. Here $J_t-\beta E_t J_{t+1}$ can be interpreted as the rental rate on capital.

It follows that $E v_t^a|\Omega_{t-1}=0$ and $E v_t^x|\Omega_{t-1}=0$, where $\Omega_{t-1}=\{a_{t-1}, a_{t-2},\ldots, x_{t-1}, x_{t-2},\ldots\}$. At time t, the firm is assumed to know $\Omega_t \cup \{n_{t-1}, n_{t-2},\ldots\}$. We further assume that a_t and x_t are jointly covariance stationary stochastic processes. Sufficient conditions for this are that the roots of $\alpha(z)=0$ and $\det\zeta(z)=0$ lie outside the unit circle. Actually, for much of our work, the assumption of covariance stationarity can be relaxed somewhat and be replaced by the assumption that a_t and x_t are of mean exponential order less than $1/\sqrt{\beta}$.[5]

We solve the firm's problem by using the discrete time calculus of variations.[6] Differentiating the objective function (1) with respect to n_{t+j}, $j=0,\ldots,N-1$, and setting each derivative equal to zero gives the system of stochastic Euler equations

$$\beta E_{t+j} n_{t+j+1} + \phi n_{t+j} + n_{t+j-1} = (1/\delta)(w_{t+j} - a_{t+j} - \gamma_0), \quad j=0,\ldots,N-1, \tag{2}$$

where

$$\phi = -((\gamma_1/\delta)+1+\beta).$$

Differentiating with respect to the last term n_{t+N} gives the terminal or transversality condition

$$\lim_{N\to\infty} E_t \beta^N[\gamma_0 + a_{t+N} - w_{t+N} - \gamma_1 n_{t+N} - \delta(n_{t+N} - n_{t+N-1})] = 0. \tag{3}$$

To solve the Euler equations for $j=0,\ldots,\infty$, subject to the terminal condition (3) and the given initial employment n_{t-1}, first obtain the factorization of the characteristic polynomial of (2),

$$(1+(\phi/\beta)z+(1/\beta)z^2) = (1-\rho_1 z)(1-\rho_2 z).$$

Given our assumptions about the signs and magnitudes of β, γ_1, and δ, it can readily be shown that $0<\rho_1<1$ and that $\rho_2 = 1/\beta\rho_1$.[7] It follows that the unique solution of the Euler equations that satisfies the transversality condition is

$$n_t = \rho_1 n_{t-1} - (\rho_1/\delta)\sum_{j=0}^{\infty} \lambda^j E_t[w_{t+j} - a_{t+j} - \gamma_0], \tag{4}$$

[5]Sufficient conditions for this requirement of mean exponential order less than $1/\sqrt{\beta}$ are that the zeroes of $\alpha(z)$ and $\det\zeta(z)$ be greater than $\sqrt{\beta}$ in absolute value. This is the general condition on uncontrolled random processes required for the class of problems we are studying.

[6]See Sargent (1979) for an exposition of this technique and application to some simple models.

[7]See Sargent (1979).

where $\lambda \equiv \rho_2^{-1} = \beta\rho_1$. Eq. (4) is derived from the Euler eq. (2) by solving the stable root backwards, and the unstable root forwards in order to satisfy the transversality condition. See Sargent (1979) for more details. Eq. (4) exhibits the certainty equivalence or separation property. That is, the same solution for n_t would emerge if we had maximized the criterion formed by replacing (a_{t+j}, w_{t+j}) by $(E_t a_{t+j}, E_t w_{t+j})$ and dropping the operator E_t from outside the sum in the objective function (1).

Eq. (4) is not yet a decision rule, for the terms $E_t w_{t+j}$ and $E_t a_{t+j}$ must be eliminated by expressing them as functions of variables known by agents at time t. We shall use the classic Wiener–Kolmogorov formulas to derive a closed form for the decision rule.[8] To derive the decision rule we shall tentatively restrict our specification of the stochastic processes for w_t and a_t to require that the roots of $\det \zeta(z) = 0$ and of $\alpha(z) = 0$ be outside the unit circle. These conditions on roots guarantee that a_t and w_t are covariance stationary, thereby justifying the use of the Wiener–Kolmogorov prediction formulas. It will turn out, however, that while our formulas were initially derived on the assumption that these roots are outside the unit circle, the formulas remain valid under wider conditions, in fact under the conditions on the exponential orders of w_t and a_t indicated in footnote 5.

In appendix A we shall derive expressions for the terms on the right-hand side of (4) using complex analysis. That derivation seems worthwhile to us in its own right, since it illustrates a method useful in other contexts and also carries insights into the nature of the 'annihilation operator' used in solving linear least squares prediction problems. Here we shall derive the formula in a technically less demanding way using tools more familiar to economists. We desire formulas for the terms

$$U \sum_{j=0}^{\infty} \lambda^j E_t x_{t+j} = \sum_{j=0}^{\infty} \lambda^j E_t w_{t+j} \quad \text{and} \quad \sum_{j=0}^{\infty} \lambda^j E_t a_{t+j},$$

where U is the $(1 \times p)$ unit row vector with 1 in the first place and zeroes elsewhere. The moving average representations for a_t and x_t, which exist by the assumptions on the zeroes of $\alpha(z)$ and $\det \zeta(z)$, are

$$a_t = \alpha(L)^{-1} v_t^a = \chi(L) v_t^a = \left[\sum_{j=0}^{\infty} \chi_j L^j \right] v_t^a,$$

and

$$x_t = \zeta(L)^{-1} v_t^x = \xi(L) v_t^x = \left[\sum_{j=0}^{\infty} \xi_j L^j \right] v_t^x.$$

[8] These formulas are derived by Whittle (1963).

The Wiener–Kolmogorov prediction formula is

$$E_t x_{t+k} = [\xi(L)/L^k]_+ v_t^x,$$

where $[\ ]_+$ is the annihilation operator that instructs us to ignore negative powers of L. In other words,

$$E_t x_{t+k} = \left[\sum_{j=k}^{\infty} \xi_j L^{j-k}\right] v_t^x.$$

Then we have

$$\sum_{k=0}^{\infty} \lambda^k E_t x_{t+k} = \left[\sum_{k=0}^{\infty} \lambda^k \sum_{j=k}^{\infty} \xi_j L^{j-k}\right] v_t^x = \psi(L) v_t^x,$$

where

$$\psi(L) = \sum_{k=0}^{\infty} \sum_{j=k}^{\infty} \lambda^k \xi_j L^j L^{-k}.$$

Interchanging orders of summation gives

$$\begin{aligned}
\psi(L) &= \sum_{j=0}^{\infty} \sum_{k=0}^{j} \lambda^k \xi_j L^j L^{-k} \\
&= \sum_{j=0}^{\infty} \xi_j L^j \sum_{k=0}^{j} \lambda^k L^{-k} \\
&= \sum_{j=0}^{\infty} \xi_j L^j [(1 - \lambda^{j+1} L^{-j-1})/(1 - \lambda L^{-1})] \\
&= \left(\sum_{j=0}^{\infty} \xi_j L^j - \lambda L^{-1} \sum_{j=0}^{\infty} \xi_j \lambda^j\right) \Big/ (1 - \lambda L^{-1}) \\
&= (\xi(L) - \lambda L^{-1} \xi(\lambda))/(1 - \lambda L^{-1}).
\end{aligned}$$

Even though the above expression for $\psi(L)$ contains both positive and negative powers of L, by construction the polynomial in the lag operator $\psi(L)$ contains only non-negative powers of L. In summary it has been shown that

$$\sum_{k=0}^{\infty} \lambda^k E_t x_{t+k} = [(\xi(L) - L^{-1}\lambda\xi(\lambda))/(1 - \lambda L^{-1})] v_t^x. \tag{5}$$

Eq. (5) is a closed form that is especially useful for estimation in the frequency domain. The corresponding formula for a_t is

$$\sum_{k=0}^{\infty} \lambda^k E_t a_{t+k} = [(\chi(L) - L^{-1}\lambda\chi(\lambda))/(1 - \lambda L^{-1})] v_t^a. \tag{6}$$

For time domain estimation, it is desirable to replace the right-hand sides of eqs. (5) and (6) with equivalent expressions in terms of current and past values of x_t and a_t, respectively. Using

$$v_t^x = \zeta(L)x_t = \xi(L)^{-1}x_t,$$

we can substitute into eq. (5) to obtain

$$\begin{aligned}\sum_{k=0}^{\infty} \lambda^k E_t x_{t+k} &= [(\xi(L) - L^{-1}\lambda\xi(\lambda))/(1-\lambda L^{-1})]\xi^{-1}(L)x_t \\ &= [(I - L^{-1}\lambda\xi(\lambda)\xi^{-1}(L))/(1-\lambda L^{-1})]x_t \\ &= [(I - L^{-1}\lambda\zeta(\lambda)^{-1}\zeta(L))/(1-\lambda L^{-1})]x_t,\end{aligned}$$

Now calculate $\zeta(L)/(1-\lambda L^{-1})$ by using polynomial long division on $[-L^r\zeta_r - L^{r-1}\zeta_{r-1} - \ldots - L\zeta + I]/(1-\lambda L^{-1})$ to obtain

$$\begin{aligned}&\zeta(L)/(1-\lambda L^{-1}) \\ &= -\zeta_r L^r - (\zeta_{r-1} + \lambda\zeta_r)L^{r-1} - \ldots - (\zeta_1 + \lambda\zeta_2 + \ldots + \lambda^{r-1}\zeta_r)L \\ &\quad + (I - \lambda\zeta_1 - \ldots - \lambda^r\zeta_r)/(1-\lambda L^{-1}).\end{aligned}$$

It follows that

$$\begin{aligned}&L^{-1}\lambda\zeta(\lambda)^{-1}\zeta(L)/(1-\lambda L^{-1}) \\ &= L^{-1}\lambda\zeta(\lambda)^{-1}[-\zeta_r L^r - (\zeta_{r-1} + \lambda\zeta_r)L^{r-1} \\ &\quad - \ldots - (\zeta_1 + \lambda\zeta_2 + \ldots + \lambda^{r-1}\zeta_r)L] + L^{-1}\lambda I/(1-\lambda L^{-1}).\end{aligned}$$

Therefore

$$\begin{aligned}&(I - L^{-1}\lambda\zeta(\lambda)^{-1}\zeta(L))/(1-\lambda L^{-1}) \\ &= L^{-1}\lambda\zeta(\lambda)^{-1}[\zeta_r L^r + (\zeta_{r-1} + \lambda\zeta_r)L^{r-1} \\ &\quad + \ldots + (\zeta_1 + \lambda\zeta_2 + \ldots + \lambda^{r-1}\zeta_r)L] \\ &\quad + I/(1-\lambda L^{-1}) - L^{-1}\lambda I/(1-\lambda L^{-1}) \\ &= I + \zeta(\lambda)^{-1}[\lambda\zeta_r L^{r-1} + (\lambda\zeta_{r-1} + \lambda^2\zeta_r)L^{r-2} \\ &\quad + \ldots + (\lambda\zeta_1 + \lambda^2\zeta_2 + \ldots + \lambda^r\zeta_r)].\end{aligned}$$

Thus we have

$$\begin{aligned}&(I - L^{-1}\lambda\zeta(\lambda)^{-1}\zeta(L))/(1-\lambda L^{-1}) \\ &= \zeta(\lambda)^{-1}[\zeta(\lambda) + (\lambda\zeta_1 + \lambda^2\zeta_2 + \ldots + \lambda^r\zeta_r) \\ &\quad + \lambda\zeta_r L^{r-1} + (\lambda\zeta_{r-1} + \lambda^2\zeta_r)L^{r-2} \\ &\quad + \ldots + (\lambda\zeta_2 + \lambda^2\zeta_3 + \ldots + \lambda^{r-1}\zeta_r)L].\end{aligned}$$

Recalling that $\zeta(\lambda)=I-\zeta_1\lambda-\zeta_2\lambda^2-\ldots-\zeta_r\lambda^r$, we have

$$(I-L^{-1}\lambda\zeta(\lambda)^{-1}\zeta(L))/(1-\lambda L^{-1})$$
$$=\zeta(\lambda)^{-1}[I+(\lambda\zeta_2+\lambda^2\zeta_3+\ldots+\lambda^{r-1}\zeta_r)L$$
$$+\ldots+(\lambda\zeta_{r-1}+\lambda^2\zeta_r)L^{r-2}+\lambda\zeta_r L^{r-1}],$$

or

$$(I-L^{-1}\lambda\zeta(\lambda)^{-1}\zeta(L))/(1-\lambda L^{-1})$$
$$=\zeta(\lambda)^{-1}\left[I+\sum_{j=1}^{r-1}\left(\sum_{k=j+1}^{r}\lambda^{k-j}\zeta_k\right)L^j\right]. \tag{7}$$

Using an analogous argument we have

$$(1-L^{-1}\lambda\alpha(\lambda)^{-1}\alpha(L))/(1-\lambda L^{-1})$$
$$=\alpha(\lambda)^{-1}\left[1+\sum_{j=1}^{q-1}\left(\sum_{k=j+1}^{q}\lambda^{k-j}\alpha_k\right)L^j\right]. \tag{8}$$

Substituting from (5), (6), (7) and (8) into (4) gives a closed form expression for the decision rule for n_t,

$$n_t=\rho_1 n_{t-1}-\frac{\rho_1}{\delta}U\zeta(\lambda)^{-1}\left[I+\sum_{j=1}^{r-1}\left(\sum_{k=j+1}^{r}\lambda^{k-j}\zeta_k\right)L^j\right]x_t$$
$$+\frac{\rho_1}{\delta}\alpha(\lambda)^{-1}\left[1+\sum_{j=1}^{q-1}\left(\sum_{k=j+1}^{q}\lambda^{k-j}\alpha_k\right)L^j\right]a_t+\frac{\rho_1\gamma_0}{\delta}\left(\frac{1}{1-\lambda}\right), \tag{9}$$

where

$$\alpha(L)a_t=v_t^a, \qquad \zeta(L)x_t=v_t^x.$$

Eq. (9) is a convenient closed form that expresses the restrictions imposed across the decision rule and the parameters of the stochastic processes for x_t and a_t. Notice that current and $(r-1)$ lagged values of x_t are in the decision rule, while current and $(q-1)$ lagged values of a_t appear. So the numbers of lagged x's and a's in the decision rule are one less than the orders of the autoregressive processes for x_t and a_t, respectively. Further, notice that current and lagged x's appear in the decision rule because they help predict future values of the wage. Thus, any stochastic processes that both Granger cause the wage and that are included in the firm's information set appear in the decision rule for n_t. The above derivation took place under the assumptions that x_t had an rth order vector autoregressive representation and a_t a qth order univariate autoregressive representation. Analogous formulas have been obtained when x_t and a_t are permitted to be mixed

autoregressive moving average processes and when other variables observed by the firm are useful in forecasting future values of a_t [see Hansen and Sargent (1979)].

The solution method leading to (9) is usefully compared with the standard dynamic programming algorithm for computing the optimal decision rule [e.g. Bertsekas (1976) or Kwakernaak and Sivan (1972)]. It is straightforward to show that our problem is a linear optimal regulator problem with a system representation that is detectable and stabilizable.[9] The optimal value function of the problem can be determined by solving the matrix 'algebraic Riccati' equation, from which the optimum decision rule is directly calculable. The algebraic Riccati equation is solved either by iterating on the Riccati matrix difference equation until convergence is obtained, or else by Vaughan's (1970) procedure of calculating the eigenvalues and eigenvectors of the state–costate transition matrix. Such procedures do not lead to closed forms but require the use of iterative procedures either to solve for the stationary solution of a matrix difference equation or else to calculate eigenvalues of the state–costate transition matrix. Evidently, the solution method leading to (9) dominates these dynamic programming procedures both in terms of speedier computation, and in terms of delivering expressions for the decision rules which can conveniently be differentiated with respect to the free parameters. Each of these features is of substantial practical importance since the decision rule and its derivatives with respect to the free parameters $\{\gamma_0, \gamma_1, \delta, \alpha(L), \zeta(L)\}$ will have to be evaluated repeatedly in the course of non-linear maximum likelihood estimation.

It should be mentioned that we are able to obtain a closed form solution in (9) because the costs of adjustment have a simple first-order form, permitting us analytically to factor the characteristic polynomial $(1+(\phi/\beta)z+(1/\beta)z^2)$. In models with higher-order characteristic polynomials, which result either when there are higher order adjustment costs or when there are interrelated costs of factor adjustment, the characteristic polynomial cannot be factored analytically. In such models, one cannot obtain a completely closed form expression for the decision rule. Still, the method leading to eq. (9) remains useful in such models, and enables one to get 'as close as possible' to a closed form solution. The application of our method of solution to models with higher-order dynamics will be described in a sequel to this paper.

3. Restrictions on the error process

In this section we illustrate how our methods can be used to provide guidance for interpreting the disturbance or error term in a regression

[9]But not controllable. See Kwakernaak and Sivan (1972).

equation. We shall take the view that a_t is a random process that is observed by private agents but is not observed by the econometrician. This indeed is a well known way for justifying the presence of a disturbance term. Under this interpretation, eq. (9) imposes substantial structure on the error process in the equation to be fit by the econometrician. Recalling that $a_t = \alpha(L)^{-1} v_t^a$, the disturbance in eq. (9) is given by

$$e_t = \frac{\rho_1}{\delta}\alpha(\lambda)^{-1}\left[1 + \sum_{j=1}^{q-1}\left(\sum_{k=j+1}^{q}\lambda^{k-j}\alpha_k\right)L^j\right]\alpha(L)^{-1}v_t^a.$$

We can rewrite this equation as

$$\alpha(L)e_t = \pi(L)v_t^a, \tag{10}$$

where

$$\pi(L) = \frac{\rho_1}{\delta}\alpha(\lambda)^{-1}\left[1 + \sum_{j=1}^{q-1}\left(\sum_{k=j+1}^{q}\lambda^{k-j}\alpha_k\right)L^j\right].$$

Here v_t^a is the serially uncorrelated random process of innovations in a_t, i.e., v_t^a is 'fundamental' for a_t.[10]

Eq. (10) shows that the error term in the decision rule (9) is a mixed moving average, autoregressive process with autoregressive order q and moving average order $q-1$. The parameters of the autoregressive component $\alpha(L)$ are inherited from the qth order Markov specification for the technology shock a_t. The moving average part $\pi(L)$ is entirely determined by the parameters of $\alpha(L)$ and the parameters of the objective function (1). Furthermore the roots of $\pi(z)$ can in general be on either side of the unit circle.[11] This means that the moving average polynomial $\pi(L)$ may not have a stable inverse in non-negative powers of L. Consequently, even though v_t^a is fundamental for a_t, it is not necessarily fundamental for e_t. That is, v_t^a need not lie in the space spanned by square-summable linear combinations of current and lagged e's.[12]

It is of interest to contrast our specification for e_t with other time series specification strategies. Unrestricted moving average components of a mixed autoregressive moving average process have multiple representations in the

[10]Some authors impose the additional requirement that the contemporaneous covariance matrix of the serially uncorrelated process be the identity matrix as one of the conditions for being 'fundamental'. In our exposition we do not impose this additional requirement.

[11]Throughout this paper we will continually make substitutions of 'z' for 'L' and vice versa. It should be remembered that $\pi(L)$ is an operator defined on the space of stochastic processes while $\pi(z)$ is an analytic function.

[12]We have produced examples in which some of the roots of $\pi(z)$ are inside the unit circle. This happens only if $q \geqq 3$. For $q=2$, $\pi(L)$ turns out to be $(1-\hat{\alpha}_1\hat{\alpha}_2\lambda L)$, where $(1-\alpha_1 L-\alpha_2 L^2) = (1-\hat{\alpha}_1 L)(1-\hat{\alpha}_2 L)$. Since $|\hat{\alpha}_1|<1$, $|\hat{\alpha}_2|<1$, it follows that the zero of $(1-\hat{\alpha}_1\hat{\alpha}_2\lambda z)$ is outside the unit circle.

sense that different moving average specifications imply the same covariance structure of the process. A common strategy in this situation is to achieve identification by focusing on the moving average specification that is invertible, i.e., the specification for which the zeroes of the moving average polynomial do not lie inside the circle. In our setup we have restrictions across the autoregressive and moving average parameters. In particular, we are not free to assume that the zeroes of $\pi(z)$ lie outside the circle because the parameters of $\pi(L)$ are completely determined by the other parameters of our model. This has important implications which will be discussed below. Before proceeding to the discussion of these implications, we should emphasize that the restrictions which we have derived on the disturbance term depend critically on the assumption that agents use current and past observations of only the technology shock to forecast future values of the technology shock, i.e., no other processes observable to agents Granger cause a_t. As noted previously, it is possible to relax this assumption and instead operate under the notion that the firm observes a vector b_t whose first element is a_t and whose other elements are useful in forecasting future values of a_t. Unfortunately, the parameters governing the b_t process will not necessarily be identifiable. This can create problems in identifying the criterion function parameters of the firm's optimization problem except when the b_t process is orthogonal to the x_t process.

Some widely used estimation procedures for models with mixed moving average, autoregressive errors, such as those of Box and Jenkins (1970), require that the error term be written in a form for which the moving average component is invertible. If such an estimation strategy is to be used, then it is required to rewrite e_t in terms of a new process v_t^e such that

$$\alpha(L)e_t = \theta(L)v_t^e,$$

where v_t^e is fundamental for e_t. The condition that v_t^e be fundamental for e_t amounts to choosing $\theta(L)$ so that

$$\pi(z)\pi(z^{-1}) = \theta(z)\theta(z^{-1}) \quad \text{for} \quad |z| = 1,$$

where $\theta(z)$ has no zeroes inside the unit circle. To be more specific, if $z_1, \ldots, z_j$ are the zeroes of $\pi(z)$ that are inside the unit circle, then by multiplying $\pi(z)$ by Blaschke factors, we obtain[13]

$$\theta(z) = \pi(z)[(1 - z_1 z)/(z - z_1)] \ldots [(1 - z_j z)/(z - z_j)],$$

[13]The factor multiplying $\pi(z)$ is an example of the 'Blaschke factors' described, for example, by Saks and Zygmund (1971, p. 221). The Blaschke factors that we employ differ from the standard form by a constant and a conjugation. We have left out the constant because it has modulus equal to one, and we have left out the conjugation since the complex zeroes of $\pi(z)$ come in conjugate pairs.

which satisfies the above requirements. Other estimation procedures are available that do not require that the moving average polynomial for the error process e_t in (10) be invertible. Such procedures are directly applicable without need to replace $\pi(L)v_t^a$ by $\theta(L)v_t^e$. We shall describe these procedures in section 5.

An alternative to the model of the error term described above can be derived by simply positing that the error term e_t in the decision rule (9) is an mth order autoregressive process, and then to work backwards and determine what assumptions are implied about the a_t process in the objective function. In particular, suppose that

$$e_t = \varepsilon_1 e_{t-1} + \ldots + \varepsilon_m e_{t-m} + v_t^e,$$

or

$$\varepsilon(L)e_t = v_t^e, \tag{11}$$

where $\varepsilon(L) = 1 - \varepsilon_1 L - \ldots - \varepsilon_m L^m$, e_t is covariance stationary and v_t^e is the innovation in the e_t process. We then can invert (11), to get

$$e_t = \varepsilon(L)^{-1} v_t^e,$$

since $\varepsilon(z)$ has its zeroes outside the unit circle by the assumption of covariance stationarity.

Let

$$\Delta(L) = (\varepsilon(L)^{-1}(L-\lambda) + \lambda\varepsilon(0)^{-1})/L.$$

It is easily verified that $\Delta(L)$ is one-sided, i.e., that the expansion for $\Delta(z)$ about zero contains only non-negative powers of L. Define

$$a_t = \Delta(L)v_t^e,$$

and assume that v_t^e is contained in agents' information set at time t. We know that the technology shock a_t must be related to the disturbance e_t in the decision rule by

$$e_t = \frac{\rho_1}{\delta} \sum_{j=0}^{\infty} \lambda^j E_t a_{t+j} = [L\Delta(L)/(L-\lambda)]_+ v_t^e. \tag{12}$$

Using the formula (A.3) in appendix A we have that

$$\begin{aligned}[L\Delta(L)/(L-\lambda)]_+ &= (L\Delta(L) - \lambda\Delta(\lambda))/(L-\lambda) \\ &= (\varepsilon(L)^{-1}(L-\lambda) + \lambda\varepsilon(0)^{-1} - \lambda\varepsilon(0)^{-1})/(L-\lambda) \\ &= \varepsilon(L)^{-1}.\end{aligned}$$

Substituting into (12) we have that

$$e_t = \varepsilon(L)^{-1} v_t^e,$$

and therefore eq. (11) has been verified. We should note that $\Delta(z)$ is not necessarily invertible. Consequently, v_t^e might not be recoverable from current and past observations on a_t, so that for this model of the error term it must be assumed that agents observe the v_t^e process itself and not just the a_t process.

A disadvantage of this model for e_t as a purely autoregressive process is that it requires that the covariance structure of the technology shock a_t be linked in a particular way to parameters in the firm's criterion function. In our view, this is not plausible; thus we prefer the model (10) for e_t.

4. Granger causality and econometric exogeneity

Let us write the dynamic demand schedule for labor, i.e., the firm's decision rule for n_t, as

$$n_t = \rho_1 n_{t-1} + \mu(L) x_t + \pi(L) a_t,$$

where

$$\mu(L) = -\frac{\rho_1}{\delta} U \zeta(\lambda)^{-1} \left[I + \sum_{j=1}^{r-1} \left(\sum_{k=j+1}^{r} \lambda^{k-j} \zeta_k \right) L^j \right],$$

$$\pi(L) = \frac{\rho_1}{\delta} \alpha(\lambda)^{-1} \left[1 + \sum_{j=1}^{q-1} \left(\sum_{k=j+1}^{q} \lambda^{k-j} \alpha_k \right) L^j \right].$$

For the sake of simplicity, the constant term has been omitted. Solving for n_t as a function of current and past x's and a's we have that

$$n_t = (1-\rho_1 L)^{-1} \mu(L) x_t + (1-\rho_1 L)^{-1} \pi(L) a_t. \tag{13}$$

This provides us with an expression for the firm's demand schedule for n_t as a sum of distributed lags of current and past x's and current and past a's. Recalling that $x_t = \zeta(L)^{-1} v_t^x$ and $a_t = \alpha(L)^{-1} v_t^a$, we can substitute into eq. (13) to obtain

$$n_t = (1-\rho_1 L)^{-1} \mu(L) \zeta(L)^{-1} v_t^x + (1-\rho_1 L)^{-1} \pi(L) \alpha(L)^{-1} v_t^a. \tag{14}$$

Since (v_t^a, v_t^x) are the innovations in the joint (a_t, x_t) process, it follows that v_t^a and v_t^x are serially uncorrelated and that $E v_t^a v_{t-j}^x = 0$ for $j \neq 0$. Contemporaneous correlation between v_t^a and v_t^x cannot in general be ruled out.

Let us introduce a new process c_t such that

$$c_t = v_t^a - v v_t^x,$$

where $E\, c_t v_t^x = 0$ and v is a $(1 \times p)$ row vector. This defines $v v_t^x$ as the linear least squares predictor of v_t^a given v_t^x. In the case in which v_t^a and v_t^x are uncorrelated, v is equal to zero and $c_t = v_t^a$. Substituting into eq. (14) we have

$$\begin{aligned} n_t = {} & (1-\rho_1 L)^{-1}[\mu(L)\zeta(L)^{-1} + \pi(L)\alpha(L)^{-1}v]v_t^x \\ & + (1-\rho_1 L)^{-1}\pi(L)\alpha(L)^{-1}c_t. \end{aligned} \tag{15}$$

As argued in the previous section, $\pi(z)$ may not be invertible. Thus if we define a disturbance term

$$d_t = (1-\rho_1 L)^{-1}\pi(L)\alpha(L)^{-1}c_t, \tag{16}$$

c_t may not be fundamental for d_t. Using the transformation with Blaschke factors described in section 3, there exists a $\theta(L)$ such that

$$\pi(z)\pi(z^{-1}) = \theta(z)\theta(z^{-1}) \quad \text{for} \quad |z| = 1,$$

where $\theta(z)$ does not have any zeroes inside the unit circle. This allows us to define a new serially uncorrelated process v_t^d that is fundamental for d_t such that

$$d_t = (1-\rho_1 L)^{-1}\theta(L)\alpha(L)^{-1}v_t^d. \tag{17}$$

Since $E\, c_t v_{t-j}^x = 0$ for all j it follows that $E\, d_t v_{t-j}^x = 0$ and consequently $E\, v_t^d v_{t-j}^x = 0$ for all j. We can substitute eqs. (17) and (16) into (15) and determine that

$$\begin{aligned} n_t = {} & (1-\rho_1 L)^{-1}[\mu(L)\zeta(L)^{-1} + \pi(L)\alpha(L)^{-1}v]v_t^x \\ & + (1-\rho_1 L)^{-1}\theta(L)\alpha(L)^{-1}v_t^d. \end{aligned} \tag{18}$$

Eq. (18) together with the fact that $x_t = \zeta(L)^{-1}v_t^x$ provide us with the representation of the joint (n_t, x_t) process given below

$$\begin{bmatrix} n_t \\ x_t \end{bmatrix} = \begin{bmatrix} (1-\rho_1 L)^{-1}\theta(L)\alpha(L)^{-1} & (1-\rho_1 L)^{-1}\{\mu(L)\zeta(L)^{-1} + \pi(L)\alpha(L)^{-1}v\} \\ 0 & \zeta(L)^{-1} \end{bmatrix} \times \begin{bmatrix} v_t^d \\ v_t^x \end{bmatrix}. \tag{19}$$

This system expresses (n_t, x_t) as one-sided square summable moving averages of the serially uncorrelated processes v_t^d and v_t^x which satisfy $E\, v_t^x v_{t-j}^{d\bullet} = 0$ for all j. The joint (v_t^d, v_t^x) process is fundamental for the joint (n_t, x_t) process. Thus (19) provides us with a Wold moving average representation of the joint (n_t, x_t) process. Note that we have a zero restriction in this representation in that x_t is not dependent upon v_t^d. The triangular character of this moving average representation together with Sims' theorem 1 (1972) imply that n_t fails to Granger cause x_t.[14]

Sims' theorem 2 informs us that if n_t fails to Granger cause x_t then there exists a representation of the form

$$n_t = \eta(L)x_t + u_t, \tag{20}$$

where

$$\eta(L) = \sum_{j=0}^{\infty} \eta_j L^j.$$

The coefficients of $\eta(L)$, i.e., the η_j's, are square summable matrices and u_t is a covariance stationary stochastic process obeying the orthogonality conditions

$$E\, x_t u_{t-j} = 0 \quad \text{for all} \quad j.$$

These orthogonality conditions say that x_t is strictly exogenous in (20) and that the projection of n_t onto the entire x_t process is one-sided on present and past x's. So Sims' theorem 2 informs us that if n_t fails to Granger cause x_t then there exists a regression of n_t on a one-sided distributed lag of x_t in which x_t is strictly exogenous. A candidate for the representation guaranteed by Sims' theorem is the dynamic labor demand schedule (13).[15] Our purpose here is to indicate that this schedule need not be the representation in which x_t is strictly exogenous. The upshot is that econometric exogeneity of x_t in the firm's decision rule for n_t is a stricter condition than the Granger noncausality of x_t by n_t. As we have seen, this latter condition is an implication of the assumptions we have used in our model derivation.

It is useful to substitute $v_t^x = \zeta(L)x_t$ and eq. (17) into eq. (18) yielding

$$n_t = (1-\rho_1 L)^{-1}[\mu(L) + \pi(L)\alpha(L)^{-1}\nu\zeta(L)]x_t + d_t. \tag{21}$$

[14]Formally Sims' theorems are for a bivariate process; however, they readily generalize to a partitioned vector process.

[15]We are identifying the labor demand schedule as the representation of n_t in terms of current and past x's and a's.

Since $E\, d_t v^x_{t-j}=0$ for all j, it follows that $E\, d_t x_{t-j}=0$ for all j. Letting

$$u_t=d_t,$$

and

$$\eta(L)=(1-\rho_1 L)^{-1}[\mu(L)+\pi(L)\alpha(L)^{-1}v\zeta(L)],$$

we see that eq. (21) is the representation insured by Sims' theorem. In other words, x_t is strictly exogenous in eq. (21). Comparing eqs. (13) and (21) it is immediately apparent that eq. (21) is the labor demand schedule if and only if $v=0$. The condition $v=0$ is equivalent to the requirement that $E\, v^x_t v^a_t=0$.

Summarizing our argument, Sims' theorem 1 indicates that the failure of n_t to Granger cause x_t is manifest in the existence of a triangular moving average representation for the joint (n_t, x_t) process. Equation system (19) displays such a representation for our model. The existence of this triangular moving average representation is a necessary condition for x_t to be strictly exogenous in the firm's demand schedule for labor as Sims' theorems 1 and 2 show, but it is not a sufficient condition. Sufficient conditions are both that (a) there exists a triangular moving average representation, i.e., n_t does not Granger cause x_t, and (b) the vector of regression parameters $v=0$, i.e., $E\, v^x_t v^a_t=0$. Thus the conditions under which x_t is exogenous in the labor demand schedule are more stringent than the conditions under which n_t fails to Granger cause x_t.

The hypothesis that n_t fails to Granger cause x_t is a key one in formulating the model. Further, any variable that Granger causes the real wage w_t ought to be included in the vector x_t, at least if there is a presumption that that variable was observable by the firm. Using the standard tests of Granger and Sims, these specifications can be subjected to empirical checks, before proceeding with estimation of the model. Now it happens that the parameters of the model, i.e., the parameters of the firm's objective function and the Markov processes governing x_t and a_t, are all identifiable without imposing the exogeneity assumption that $v=0$. Consistent estimators of these parameters can be constructed imposing only the Granger non-causality of x_t by n_t, but leaving v unrestricted in (21). Then, under the maintained hypotheses that n_t does not Granger cause x_t and that the other specifications of the model are correct, the null hypothesis that $v=0$ can be tested.

This latter test is similar in spirit to the test for exogeneity proposed by Wu (1973) in a classical simultaneous equations setting. The idea of Wu's test for exogeneity was to examine the covariance between the disturbance of the reduced form equation for a stochastic regressor appearing in a particular structural equation and the disturbance term in that same structural equation. In the context of our model, this is analogous to testing

whether $E v_t^x v_t^a = 0$, which is equivalent to $v = 0$. It should be pointed out, however, that the estimation environment which we are considering differs somewhat from the one which Wu considered, in that we are allowing for serial correlation in disturbance terms and that we achieve parameter identification via non-linear cross-equation restrictions implied by the hypothesis of rational expectations.

The preceding amounts to a description of the representation theory underlying our proposed test for strict exogeneity under the maintained hypothesis of the model. That is, we have shown that the hypothesis of strict econometric exogeneity in the labor demand schedule in terms of current and past n's and a's translates into the hypothesis that $v = 0$ in the population Wold representation (19). We now briefly describe two statistical procedures for testing the null hypothesis $v = 0$, each of which has a justification in terms of asymptotic distribution theory. The first procedure involves first estimating all of the free parameters of (19) by a quasi-maximum likelihood procedure,[16] including v among the free parameters. Then the model (19) is re-estimated imposing the strict exogeneity assumption $v = 0$. On the null hypothesis that $v = 0$, the likelihood ratio statistic is asymptotically distributed as chi-square with p degrees of freedom, where v is a $(p \times 1)$ vector.

A second testing procedure is closely related. It involves estimating (19) with v a free parameter by a quasi-maximum likelihood procedure, and then using the asymptotic covariance matrix of the estimated v to test the null hypothesis that $v = 0$.

The preceding representation of the hypothesis of strict exogeneity maintains the hypothesis that the specification of the model is correct. As it happens, the model often imposes overidentifying restrictions on the lag distribution of x_t on the right-hand side of (21). This means that (21) is a regression equation (i.e., $E d_t x_{t-j} = 0$ for $j \geqq 0$) only if the overidentifying restrictions are true. One can compute various specification tests for the model, which are closely related to exogeneity tests, and are based on estimates of the sample moments corresponding to $E d_t x_{t-j}$ for $j \geqq 0$. Rejection of the hypothesis that these moments are zero can be viewed as evidence against the specification of the model. Failing such a test would leave open whether the specification is faulty because of an incorrect criterion function being attributed to the maximizing agent, or because of a failure of x_t to be Granger non-caused by n_t, as the model assumes. Presumably, careful application of Granger causality tests in the 'model-free' setting of Granger and Sims could be used to help isolate the source of failure.

[16]For a precise definition of the term 'quasi-maximum likelihood procedure', see section 5.

5. Estimation of the model parameters

The system that we want to estimate is

$$
\begin{aligned}
n_t &= (1-\rho_1 L)^{-1}[\mu(L)+\pi(L)\alpha(L)^{-1}\nu\zeta(L)]x_t \\
&\quad + (1-\rho_1 L)^{-1}\pi(L)\alpha(L)^{-1}c_t \\
\zeta(L)x_t &= v_t^x,
\end{aligned} \tag{22}
$$

where

$$
\pi(L)=\frac{\rho_1}{\delta}\alpha(\lambda)^{-1}\left[1+\sum_{j=1}^{q-1}\left(\sum_{k=j+1}^{q}\lambda^{k-j}\alpha_k\right)L^j\right],
$$

$$
\mu(L)=\frac{\rho_1}{\delta}\zeta(\lambda)^{-1}\left[I+\sum_{j=1}^{r-1}\left(\sum_{k=j+1}^{r}\lambda^{k-j}\zeta_k\right)L^j\right],
$$

$$
E\,c_t c_{t-j}=0, \qquad E\,v_t^x v_{t-j}^{x\prime}=0 \qquad \text{for} \quad j\neq 0,
$$

$$
E\,c_t v_{t-j}^x=0 \quad \text{for all } j.
$$

Equation system (22) displays the cross-equation restrictions on the regression equations which emerge from our model. The underlying parameters which are to be estimated are ν, ρ_1, δ, λ and the parameters of $\alpha(L)$ and $\zeta(L)$.[17] As was noted in section 4, x_t is strictly exogenous in the first equation of this system.

Equation system (22) can be estimated by using the method of quasi-maximum likelihood with a normal density function. The word 'quasi' is included because it is not necessary to assume that the stochastic processes are Gaussian in order to obtain the desired asymptotic properties of the maximum likelihood estimates, e.g., consistency and asymptotic efficiency. Suppose that we have a sample on (n_t, x_t) for $t=1,\ldots,T$. Let us stack the observations into vectors $(\bar{n}_T, \bar{x}_T)$, where

$$
\bar{n}_T=\begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_T \end{bmatrix}, \qquad \bar{x}_T=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix}.
$$

Now write the Wold moving average representation (19) in the form

$$
\begin{pmatrix} n_t \\ x_t \end{pmatrix}=\Phi(L)\begin{pmatrix} v_t^d \\ v_t^x \end{pmatrix}, \tag{23}
$$

[17] In actuality, one may be interested in knowing β and γ_1, since β, γ_1 and δ are the parameters of agents' objective functions. The parameters β and γ_1 can easily be recovered from the other parameters being estimated.

where $\Phi(L)$ is the matrix of polynomials in L on the right-hand side of (19), and where

$$E\begin{pmatrix} v_t^d \\ v_t^x \end{pmatrix}(v_t^d v_t^{x\prime}) = V = \begin{pmatrix} V_{11} & 0 \\ 0 & V_{22} \end{pmatrix}.$$

(Recall that by construction $E\, v_t^d v_t^x = 0$.) Then the covariance generating function for the joint $(n_t,\ x_t)$ process is $\Phi(z)V\Phi(z^{-1})'$. We can use the covariance generating function for the (n_t, x_t) process to generate the population elements of the covariance matrix of the stacked random vector $(\bar{n}_T', \bar{x}_T')$ in terms of the underlying parameters of the model. Let the covariance matrix of $(\bar{n}_T', \bar{x}_T')$ be

$$\Gamma_T = E\begin{pmatrix} \bar{n}_T \\ \bar{x}_T \end{pmatrix}\begin{pmatrix} \bar{n}_T \\ \bar{x}_T \end{pmatrix}'.$$

(Recall that the mean of $(\bar{n}_T', \bar{x}_T')$ is zero by virtue of the means having been subtracted off.)

The normal log likelihood function for $(\bar{n}_T, \bar{x}_T)$ is given by

$$\mathcal{L}_T = -\tfrac{1}{2}(T + Tp)\log 2\pi - \tfrac{1}{2}\log\det \Gamma_T - \tfrac{1}{2}\,[\bar{n}_T'\bar{x}_T']\Gamma_T^{-1}\begin{bmatrix} \bar{n}_T \\ \bar{x}_T \end{bmatrix}. \qquad (24)$$

Directly maximizing the log likelihood function is difficult computationally, since Γ_T is a complicated function of the underlying parameters of the model, and since inversion of the large $(T + Tp) \times (T + Tp)$ matrix Γ_T is required for each evaluation of the likelihood function.[18] An alternative strategy is to express the normal likelihood function as a product of conditional likelihood functions and to employ recursive prediction algorithms to evaluate the conditional means and variances [see Rissanen and Caines (1979)]. This relinquishes the burden of inverting the Γ_T matrix but requires the use of recursive formulas in order to evaluate the likelihood function. By virtue of the highly non-linear nature of the cross-equation rational expectations restrictions, the likelihood function will have to be maximized via some numerical method which requires repeated evaluation of the likelihood. For this reason, we mention a pair of strategies for simplifying the calculations by approximating the likelihood function.

[18]This is a well known problem in models with moving average errors. See Anderson (1975, ch. 5) or Hannan (1970, ch. VI).

5.1. Likelihood conditional on some initial observations

It is convenient to rewrite system (19) or (22) in the regression equation form

$$\alpha(L)(1-\rho_1 L)n_t = [\alpha(L)\mu(L)+\pi(L)v\zeta(L)]x_t + \theta(L)v_t^d,$$
$$\zeta(L)x_t = v_t^x. \tag{25}$$

The equations express n_t as a mixed moving average autoregressive process with an exogenous driving process x_t. In the first equation of (25), there are $(q+1)$ lagged n's, $(q+r-1)$ lagged x's, and $(q-1)$ lagged v_t^d's.

Now represent the joint density f_J of $(\bar{n}_T', \bar{x}_T')$ by

$$f_J(\bar{n}_T, \bar{x}_T) = f_c(\bar{n}_T|\bar{x}_T)\cdot f_m(\bar{x}_T), \tag{26}$$

where f_c is the density function for $\bar{n}_T$ conditioned on $\bar{x}_T$ and f_m is the marginal density of $\bar{x}_T$. It is convenient to approximate each density on the right-hand side of (26) with a density conditioned on some initial observations. First, approximate $f_m(\bar{x}_T)$ by

$$\begin{aligned} &f_m(x_T, x_{T-1}, \ldots, x_1) \\ &\approx g_m(x_T, x_{T-1}, \ldots, x_{r+1}|x_r, x_{r-1}, \ldots, x_1) \\ &= g_c(x_T|x_{T-1}, \ldots, x_{T-r})g_c(x_{T-1}|x_{T-2}, \ldots, x_{T-r-1}) \\ &\quad \ldots g_c(x_{r+1}|x_r, x_{r-1}, \ldots, x_1), \end{aligned} \tag{27}$$

where g_m and g_c are the appropriate marginal and conditional density functions. Taking logarithms on both sides of (27) we obtain

$$\begin{aligned} &\log g_m(x_T, x_{T-1}, \ldots, x_{r+1}|x_r, x_{r-1}, \ldots, x_1) \\ &= \sum_{t=r+1}^{T} \log g_c(x_t|x_{t-1}, \ldots, x_{t-r}). \end{aligned} \tag{28}$$

Using the normal density and eq. (25) we have

$$\begin{aligned} &\log g_c(x_t|x_{t-1}, \ldots, x_{t-r}) \\ &= -\frac{p}{2}\log 2\pi - \frac{1}{2}\log\det V_{22} \\ &\quad -\frac{1}{2}\left(x_t - \sum_{j=1}^{r}\zeta_j x_{t-j}\right)' V_{22}^{-1}\left(x_t - \sum_{j=1}^{r}\zeta_j x_{t-j}\right). \end{aligned}$$

Using this expression for $\log g_c$ in (27) gives

$$\log g_m(x_T, x_{T-1}, \ldots, x_{r+1} | x_r, x_{r-1}, \ldots, x_1)$$

$$= -\frac{p(T-r)}{2}\log 2\pi - \left(\frac{T-r}{2}\right)\log\det V_{22}$$

$$-\tfrac{1}{2}\sum_{t=r+1}^{T}\left(x_t - \sum_{j=1}^{r}\zeta_j x_{t-j}\right)' V_{22}^{-1}\left(x_t - \sum_{j=1}^{r}\zeta_j x_{t-j}\right). \tag{29}$$

Next approximate the term $f_c(\bar{n}_T|\bar{x}_T)$ in (26) with

$$\begin{aligned}
&f_c(n_T, n_{T-1}, \ldots, n_1 | x_T, x_{T-1}, \ldots, x_1)\\
&\approx h_m(n_T, n_{T-1}, \ldots, n_{q+r} | n_{q+r-1}, \ldots, n_1, x_T, x_{T-1}, \ldots, x_1, v^d_{q+r-1}, \ldots, v^d_1)\\
&= h_c(n_T | n_{T-1}, \ldots, n_{T-(q+1)}, x_T, \ldots, x_{T-(q+r-1)}, v^d_{T-1}, \ldots, v^d_{T-(q-1)})\\
&\quad \cdot h_c(n_{T-1} | n_{T-2}, \ldots, n_{T-(q+2)}, x_{T-1}, \ldots, x_{T-(q+r-2)}, v^d_{T-2}, \ldots, v^d_{T-q})\\
&\quad \cdot \ldots \cdot h_c(n_{q+r} | n_{q+r-1}, \ldots, n_{r-1}, x_{q+r}, \ldots, x_1, v^d_{q+r-1}, \ldots, v_{r+1}),
\end{aligned} \tag{30}$$

where h_m and h_c are the appropriate marginal and conditional densities.[19] The normal density with (25) leads to

$$\begin{aligned}
&h_c(n_t | n_{t-1}, \ldots, n_{t-(q+1)}, x_t, \ldots, x_{t-(q+r-1)}, v^d_{t-1}, \ldots, v^d_{t-(q-1)})\\
&\sim \mathcal{N}\Big\{[(-\alpha(L)(1-\rho_1 L))/L]_+ n_{t-1} + [\alpha(L)\mu(L) + \pi(L)\nu\zeta(L)]x_t\\
&\quad + [\theta(L)/L]_+ v^d_{t-1},\ \theta_0^2 V_{11}\Big\}.
\end{aligned}$$

Substituting the normal density into (30) and taking logs leads to

$$\begin{aligned}
&\log h_m(n_T, n_{T-1}, \ldots, n_{q+r} | n_{q+r-1}, \ldots, n_1, x_T, \ldots, x_1, v^d_{q+r-1}, \ldots, v^d_1)\\
&= -((T+1-(q+r))/2)\log 2\pi - ((T+1-(q+r))/2)\log\theta_0^2 V_{11}\\
&\quad -\tfrac{1}{2}(1/\theta_0^2 V_{11})\sum_{t=q+r}^{T}\{\alpha(L)(1-\rho_1 L)n_t - [\alpha(L)\mu(L)\\
&\quad + \pi(L)\nu\zeta(L)]x_t - [\theta(L)/L]_+ v^d_{t-1}\}^2.
\end{aligned} \tag{31}$$

[19]Recall that by virtue of the invertibility built into representation (19) or (23), v^d_t is in the space spanned by current and lagged n's and x's. In effect, lagged v^d's are conditioned on in (30) as proxies for the more extensive set of lagged n's and x's whose information they fully summarize.

Adding (29) and (31) and viewing the result as an approximation to the log of (26) leads to the approximate log likelihood function

$$\begin{aligned}\mathscr{L}_T^* = \{ & -(p(T-r)/2)\log 2\pi - ((T-r)/2)\log\det V_{22} \\ & -\tfrac{1}{2}\sum_{t=r+1}^{T}(\zeta(L)x_t)'V_{22}^{-1}(\zeta(L)x_t)\} \\ & +\{-((T+1-(q+r))/2)\log 2\pi \\ & -((T+1-(q+r))/2)\log\theta_0^2 V_{11} \\ & -\tfrac{1}{2}(1/\theta_0^2 V_{11})\sum_{t=q+r}^{T}(\alpha(L)(1-\rho_1 L)n_t \\ & -[\alpha(L)\mu(L)+\pi(L)\nu\zeta(L)]x_t \\ & -[\theta(L)/L]_+ v_{t-1}^d)^2\}. \end{aligned} \tag{32}$$

This approximation to the log likelihood is to be maximized over the free parameters ν, ρ_1, δ, λ, $\alpha(L)$, $\zeta(L)$ and V subject to the cross-equation restrictions exhibited in (22). The expression (32) for the approximate log likelihood indicates that it is desirable to estimate all of the equations of (22) jointly, even though x_t is strictly exogenous in the first equation of (22). The reason for joint estimation is that the parameters of $\zeta(L)$ appear in the first equation of (22) via the assumption of rational expectations. Evidently, estimating $\zeta(L)$ by maximizing the first term in braces in (32) and then maximizing the second term in braces,[20] taking $\zeta(L)$ as given, leads to a lower value of the approximate likelihood.

Since the initial values $v_{q+r-1}^d, \ldots, v_{r+1}^d$ are unobservable, implementing the approximation (32) requires one additional approximation be used to generate estimates of these initial values of v^d. Box and Jenkins (1970) describe several ways to select the initial v_t^d's. One permissible procedure would be to set the initial values of the v_t^d's at their unconditional means of zero. The invertibility built into the Wold representation (19) guarantees that the impact of these initial v's becomes negligible as $T \to \infty$.

5.2. *Spectral approximations*

Hannan (1970) has suggested an alternative approximation to the log likelihood function $\mathscr{L}_T$ in (24).[21] Using the compact representation (23), we

[20] This amounts to estimating the second equation of (22) for the exogenous process x_t first, and then estimating the first equation of (22) taking the estimate of $\zeta(L)$ as given.

[21] See Hannan (1970, ch. VI, secs. 4 and 5).

know that the theoretical spectral density matrix of the (n_t, x_t) process is given by

$$S(\omega)=\Phi(e^{-i\omega})V\Phi(e^{i\omega})'.$$

Let $I(\omega_j)$ be the periodogram for the (n_t, x_t) process at frequency $\omega_j=2\pi j/T$. Now make the following approximations:

$$[\bar{n}_T' \bar{x}_T']\Gamma_T^{-1}\begin{bmatrix}\bar{n}_T\\ \bar{x}_T\end{bmatrix}\approx \sum_{j=1}^{T} \operatorname{tr}[S(\omega_j)^{-1}I(\omega_j)],$$

and (33)

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

Substituting into (24), the corresponding approximate log likelihood function is

$$\mathcal{L}_T^{**} = -\tfrac{1}{2}(T+Tp)\log 2\pi - \tfrac{1}{2}\sum_{j=1}^{T}\log\{\det[S(\omega_j)]\}$$
$$-\tfrac{1}{2}\sum_{j=1}^{T}\operatorname{tr}[S(\omega_j)^{-1}I(\omega_j)]. \tag{34}$$

The computational gain in employing $\mathcal{L}_T^{**}$ instead of $\mathcal{L}_T$ is evident by the fact that evaluation of $\mathcal{L}_T$ requires inversion of the $(T+Tp)\times(T+Tp)$ matrix Γ_T whereas evaluation of $\mathcal{L}_T^{**}$ requires the inversion of the much smaller $(p+1)\times(p+1)$ matrices $S(\omega_j)$ for $j=1,2,\ldots,T$. The justification for approximating $\mathcal{L}_T$ by $\mathcal{L}_T^{**}$ relies on the sample size T being large. In contrast to $\mathcal{L}_T^{*}$ defined in expression (32), computation of $\mathcal{L}_T^{**}$ does not compel one to shift to the invertible representation for (n_t, x_t).

A note of caution about these approximate likelihood functions is pertinent. In a somewhat different context, Phadke and Kedem (1978) show that when a zero of the moving average polynomial is close to unity in modulus, maximization of approximate likelihood functions analogous to $\mathcal{L}_T^{*}$ and $\mathcal{L}_T^{**}$ can give rise to parameter estimates that are substantially inferior to the ones obtained by maximizing the actual likelihood function. They found this especially to be true for the frequency domain approximization.[22]

[22]Phadke and Kedem (1978) ran Monte Carlo simulations on estimators obtained from maximizing the approximate likelihood functions and the exact likelihood function for finite order vector moving average processes.

In situations in which the parameter vector v is specified to be zero, an alternative estimation procedure is available. The linear least squares projection of n_t on current and past x's is not dependent on the parameters of the generating equation of the a_t process. This can be witnessed by examining eq. (21). The serial correlation properties of the disturbance term of this projection can be consistently estimated by employing the residuals obtained from least squares estimation of the projection. Using a variant of generalized least squares, the projection and the autoregression for x_t can then be estimated jointly subject to the cross-equation restrictions implied by the model, thus delivering estimates of the criterion function parameters of the firm's dynamic optimization problem.

In practice, the selected approximation to the likelihood function would be maximized by using an 'acceptable gradient method'. Let Ψ_∞ be the free parameters of the model and let Ψ_0 be an initial estimate. (Where the model is overidentified, i.e., if p and r are large enough, there is a variety of ways to get an initial consistent estimate.) Then the approximate likelihood function is maximizing by iterating on

$$\Psi_{i+1} = \Psi_i - \Lambda_i Q_i G_i, \tag{35}$$

where Λ_i is a scalar step size, Q_i a positive definite matrix, and G_i the gradient of the approximate log likelihood function. The reader is referred to Bard (1974) for a detailed description of alternative procedures for choosing Λ_i and Q_i and for calculating the gradient G_i. For present purposes it suffices to point out that the explicit closed form solution (22) exhibiting the cross-equation restrictions makes it possible to calculate the gradient G_i analytically fairly directly. Similarly, for methods in which Q_i is set equal to H_i^{-1} where H_i is the second derivative matrix of the approximate likelihood, the formulas in (22) make analytical calculation of the Hessian feasible. The explicit formulas (22) are thus of potential advantage both in facilitating rapid and accurate computation of estimates, and in facilitating computation of the asymptotic covariance matrix of the estimates.

Next, we note that where $Q_i = H_i^{-1}$, asymptotically efficient estimates can be obtained by taking one-step with (35) starting from an initial consistent estimate Ψ_0. A variety of such two step estimators that exploit the locally quadratic character of the normal log likelihood function have been proposed in contexts somewhat similar to the present one.

Finally, hypotheses can be tested either using the estimated asymptotic covariance matrix of the coefficients or likelihood ratio tests. For example, where the model is overidentified, the model can be tested by nesting it within a more loosely specified model, say one that does not impose the cross-equation rational expectations restrictions, and then computing a

likelihood ratio statistic. Examples of this specification test strategy are given by Sargent (1977, 1978b).

6. An omitted information variables model for the error term

In this section, we describe an alternative model of the error term which is related to a model of the error proposed by Shiller (1972) in another context. This model of the disturbance term permits estimation of the parameters of the firm's objective function under conditions which are more stringent in some respects but less stringent in other respects than are required for the model of the error term used in the preceding parts of this paper.

Let us write the demand schedule for employment as

$$n_t = \rho_1 n_{t-1} - (\rho_1/\delta) \sum_{j=0}^{\infty} \lambda^j E\, w_{t+j} | X_t + \pi(L) a_t, \tag{36}$$

where

$$X_t = \{x_t, x_{t-1}, \ldots\}.$$

Let us partition x_t as $x_t' = (x_{1t}', x_{2t}')$. Hence x_t is a $(p \times 1)$ vector and x_{1t} a $(p_1 \times 1)$ vector. We assume that x_{1t} includes at least w_t. Let $X_{1t} = \{x_{1t}, x_{1t-1}, \ldots\}$. We make the following assumptions:

(i) The firm uses the entire information set X_t to form its expectations of future w's, so that (13) or (36) is the appropriate demand schedule for employment.
(ii) The econometrician has access only to a subset of the information $X_{1t} \subseteq X_t$ which private agents use.
(iii) The random shock a_t obeys the extensive orthogonality conditions $E\, a_t x_{1t-j} = 0$ for all j.

Notice that the model (36) can be rewritten as

$$n_t = \rho_1 n_{t-1} - (\rho_1/\delta) \sum_{j=0}^{\infty} \lambda^j E\, w_{t+j} | X_{1t}$$

$$+ \pi(L) a_t + (\rho_1/\delta) \sum_{j=0}^{\infty} \lambda^j [E\, w_{t+j} | X_{1t} - E\, w_{t+j} | X_t],$$

or

$$n_t = \rho_1 n_{t-1} - (\rho_1/\delta) \sum_{j=0}^{\infty} \lambda^j E\, w_{t+j} | X_{1t} + \pi(L) a_t + s_t,$$

where

$$s_t = (\rho_1/\delta) \sum_{j=0}^{\infty} \lambda^j [E\, w_{t+j} | X_{1t} - E\, w_{t+j} | X_t].$$

Let x_{1t} have the vector autoregressive representation

$$\zeta^1(L)x_{1t}=v_t^1,$$

where

$$v_t^1=x_{1t}-E[x_{1t}|x_{1t-1},x_{1t-2},\ldots],$$

and

$$\zeta^1(L)=I-\zeta_1^1L-\ldots\zeta_{r_1}^1L^{r_1},$$

and assume that the roots of $\det \zeta^1(z)=0$ are outside the unit circle. Then eq. (22) can be written as

$$n_t=\rho_1 n_{t-1}+\mu^1(L)x_{1t}+\pi(L)a_t+s_t, \tag{37}$$

where

$$\mu^1(L)=(-\rho_1/\delta)U_1\zeta^1(\lambda)^{-1}\left[I+\sum_{j=1}^{r_1-1}\left(\sum_{k=j+1}^{r_1}\lambda^{k-j}\zeta_k^1\right)L^j\right], \tag{38}$$

and U_1 is the unit vector conformable to x_{1t} with 1 in the (first) place corresponding to w_t and zeroes elsewhere.

Now the random term a_t has been assumed orthogonal to x_{1t-j} for all integer j. Therefore $\pi(L)a_t$ is orthogonal to x_{1t-j} for all integer j. Further, by the law of iterated projections, we have for all $j\geqq 0$,

$$E[\{Ew_{t+j}|X_{1t}-Ew_{t+j}|X_t\}|X_{1t}]=Ew_{t+j}|X_{1t}-Ew_{t+j}|X_{1t}=0.$$

It follows that $Es_t|X_{1t}=0$. Therefore, we have that $Es_tx_{1t-j}=0$ for $j\geqq 0$ so that the random variable s_t is orthogonal to x_{1t-j} for $j\geqq 0$. We have therefore established that the composite error term $\pi(L)a_t+s_t$ is orthogonal to x_{1t-j} for $j\geqq 0$, i.e.,

$$E\{[\pi(L)a_t+s_t]x_{1t-j}'\}=0 \quad \text{for} \quad j\geqq 0. \tag{39}$$

However, $\pi(L)a_t+s_t$ is not in general orthogonal to n_{t-1}, since lagged values of both $\pi(L)a_t$ and s_t influence n_{t-1}, and since both $\pi(L)a_t$ and s_t may be serially correlated.

Under our current model assumptions we cannot rule out the possibility that for the joint (n_t,x_{1t}) process, n_t Granger causes x_{1t}. Even though n_t fails to Granger cause the complete block x_t, relative to the information set $\{n_t,x_{1t}\}$, n_t may Granger cause x_{1t}, for 'omitted variable' reasons [see Granger (1969)]. Thus no claim can be made that x_{1t} is exogenous in eq. (27). As noted above, the composite error term in eq. (37) may be serially

correlated. Without specifically modeling the joint covariance properties of the variables unobservable to the econometrician, fully efficient parameter estimation procedures such as quasi-maximum likelihood are not feasible. Indeed, a more richly specified model is needed in order even to write down the joint likelihood function for (n_t, x_{1t}). One possible strategy is to trace through the restrictions that our model places on the moving average representation for the joint (n_t, x_{1t}) process. Although conceptually this is a feasible approach, the restrictions are cumbersome and estimation of the moving average representation subject to these restrictions appears to be computationally impractical. Rather than pursue this line further, we consider an alternative, computationally simpler strategy. This procedure can be viewed as a generalization of the method of moments, and it exploits the orthogonality conditions implied by our model. We now examine these orthogonality conditions in more detail.

Solving eq. (37) for $\pi(L)a_t + s_t$ and writing out the orthogonality conditions (39) for $j = 0, 1, \ldots, r_1 - 1$ gives

$$E\left[n_t - \rho_1 n_{t-1} - \sum_{k=0}^{r_1-1} \mu_k^1 x_{1t-k}\right] x'_{1t-j} = 0,$$

or

$$E\, n_t x'_{1t-j} - \rho_1 E\, n_{t-1} x'_{1t-j} - \sum_{k=0}^{r_1-1} \mu_k^1 E\, x_{1t-k} x'_{1t-j} = 0,$$

or

$$C_{nx}(j) - \rho_1 C_{nx}(j-1) - \sum_{k=0}^{r_1-1} \mu_k^1 C_{xx}(j-k) = 0, \qquad j = 0, \ldots, r_1 - 1, \tag{40}$$

where[23]

$$C_{nx}(j) = E\, n_t x'_{1t-j} \quad \text{and} \quad C_{xx}(j) = E\, x_{1t} x'_{1t-j}.$$

Each μ_k^1 is a $(1 \times p_1)$ row vector while each x_{1t-j} is a $(p_1 \times 1)$ column vector. So (39) is a system of $p_1 \cdot r_1$ equations in the $(p_1 \cdot r_1 + 1)$ parameters $(\mu_0^1, \mu_1^1, \ldots, \mu_{r_1-1}^1, \rho_1)$. Therefore, the normal equations (39) by themselves are incapable of identifying this list of parameters. But recall that not all of the parameters in this list are free. For the model imposes the extensive set of

[23]Additional orthogonality conditions can be obtained by allowing j to exceed $r_1 - 1$. In particular, it may be desirable to exploit as many orthogonality conditions as there are sample moments that can be computed. The determination of the appropriate number of these orthogonality conditions is dependent on the serial correlation properties of the composite residual. We intentionally choose not formally to specify these residual properties and instead to focus on a fixed number of orthogonality conditions, i.e., the number of orthogonality conditions we use is not dependent on sample size.

restrictions across the μ^1's and the parameters of $\zeta^1(L)$ which are summarized in formula (38). The parameters of $\zeta^1(L)$ are identified from the vector autoregression $\zeta^1(L)x_{1t}=v_t^1$ and its normal equations

$$C_{xx}(j)-\sum_{k=1}^{r_1}\zeta_k^1 C_{xx}(j-k)=0. \tag{41}$$

Given that $\zeta^1(L)$ is sufficiently rich (i.e., p_1 and r_1 are large enough – and for the current problem each can be quite small, namely one) the free parameters of $\mu^1(L)$ in (37) and ρ_1 are identified or overidentified by the population orthogonality conditions (40) and (41). It is clearly the presence of the cross-equation restrictions summarized in (38) that allows the orthogonality conditions (40) and (41) to identify the free structural parameters despite the fact that we are one orthogonality condition short in (40). Thus, (36) fails to be a regression equation, yet consistent estimation of the free parameters is still possible because of the presence of the cross-equation restrictions delivered by rational expectations.

A generalized method of moments estimator of the model parameters can be obtained by replacing the population moments in the orthogonality conditions (40) and (41) with the corresponding sample moments. When the model parameters are overidentified there are more orthogonality conditions than there are parameters to be estimated. Consequently, it is not in general possible to select parameter estimates that allow all of the sample orthogonality conditions to equal zero. This necessitates an alternative strategy for obtaining parameter estimates that allows linear combinations of the sample orthogonality conditions to be as close to zero as possible. The number of linear combinations is dictated by the number of underlying parameters to be estimated. Discussions of consistency for generalized method of moments estimators of this form when the underlying stochastic processes are stationary and ergodic are provided in Hansen (1979).[24]

Given that there is latitude as to the choice of which linear combinations of the sample orthogonality conditions to employ, a question remains as to what is the 'best' choice. In this context we mean best in sense of delivering the smallest asymptotic covariance matrix among the class of estimators under consideration. In other words, all other estimators in this class have an asymptotic covariance matrix that exceeds the 'best' estimator by a positive definite matrix. It turns out that the 'best' choice is dependent upon the serial correlation properties of the composite residual term. Consistent

[24]Sims has illustrated that a wide class of econometric estimators is encompassed in the generalized method of moments framework. He has demonstrated this in his graduate econometrics course at the University of Minnesota.

estimators of the residual covariances can be obtained from some first step consistent parameter estimates. In appendix B we show how to determine the optimal or best weighting scheme for the sample orthogonality conditions as well as how to compute the asymptotic covariance matrix for the parameter estimates. These results are developed in much more detail in Hansen (1979).

7. Conclusions

A hallmark of rational expectations models is that they typically impose restrictions across parameters in agents' decision rules and the parameters of equations that describe the uncontrollable random processes that agents face passively. These cross-equation restrictions are an important source of identification in rational expectations models, a source that helps to fill the vacuum created by the fact that in these models there are often too few exclusion restrictions of the classic Cowles commission variety to achieve identification.[25] The cross-equation restrictions play a critical role in the statistical models and tests proposed in this paper. For example, it is the presence of overidentifying cross-equation restrictions on the labor demand schedule that makes it feasible to test both the necessary and sufficient conditions that assure that x_t is strictly exogenous in the labor demand schedule. Again, the feasibility of consistent estimation with the 'omitted information variable' model of the error term rests on the presence of cross-equation restrictions which compensate for what would be a short-fall of orthogonality conditions in their absence. Again, the presence of restrictions across the parameters of the processes generating the exogenous and endogenous variables is the reason that joint estimation of the parameters of the exogenous and endogenous processes is required for statistical efficiency, in contrast to the more familiar case in which the parameters of the strictly exogenous processes can be efficiently estimated by themselves.

The methods that we have described in this paper are fairly directly applicable to a host of problems that can be reduced to that of an agent choosing a linear dynamic contingency plan for a single variable.[26] It is desirable and non-trivial to extend our methods to the case where the agent is choosing linear contingency plans for a vector of interrelated variables. An interrelated dynamic factor demand model would be a good example. This generalization will be described in a sequel to this paper [see Hansen and Sargent (1980)].

[25]The implications of rational expectations models for achieving identification with exclusion restrictions are discussed by Sims (1980), Lucas and Sargent (1978), and Sargent and Sims (1977).

[26]A linear version of the Lucas–Prescott (1971) model of investment under uncertainty fits in since their fictitious 'planner' faces a problem of this form.

Appendix A

This appendix contains a derivation of formulas (5) and (6) of the text obtained by using some analytic function theory. We use results from elementary complex analysis that can be found in many good books on the subject, e.g., Churchill (1960) and Saks and Zygmund (1971). The technique delivers a fast way of evaluating the annihilation operator $[\psi(z)]_+$ by employing a partial-fractions-like decomposition of $\psi(z)$. This method turns out to be useful in solving a variety of classical signal extraction problems in addition to our present application.

Let us begin with a two-sided lag operator $\psi(L)$ where

$$\psi(L)=\sum_{j=-\infty}^{+\infty}\psi_j L^j \quad \text{and} \quad \sum_{j=-\infty}^{+\infty}\psi_j^2<+\infty.$$

The 'z transform' of this operator is given by

$$\psi(z)=\sum_{j=-\infty}^{+\infty}\psi_j z^j=\psi^+(z)+\psi^-(z),$$

where

$$\psi^+(z)=\sum_{j=0}^{+\infty}\psi_j z^j \quad \text{and} \quad \psi^-(z)=\sum_{j=1}^{+\infty}\psi_{-j}z^{-j}.$$

$\psi^+(z)$ defines an analytic function for $|z|<1$ and $\psi^-(z)$ defines an analytic function for $|z|>1$. Furthermore

$$\lim_{z\to\infty}\psi^-(z)=0.$$

Using a result from Zygmund (1959), it follows that

$$\lim_{R\uparrow 1}\psi^+(Re^{i\omega})=\psi^+(e^{i\omega}), \qquad \lim_{R\downarrow 1}\psi^-(Re^{i\omega})=\psi^-(e^{i\omega})$$

exist for almost all $\omega\in[0,2\pi]$.[27] Thus $\psi(z)$ is at least well defined almost everywhere for $|z|=1$, and in particular

$$h(\omega)=\psi(e^{i\omega})=\psi^+(e^{i\omega})+\psi^-(e^{i\omega})$$

is the transfer function for the linear filter $\psi(L)$. The annihilation operator $[\ ]_+$ applied to z transforms is defined by

$$[\psi(z)]_+=\psi^+(z).$$

[27] See Zygmund (1959, p. 276).

In other words the annihilation operator instructs us to ignore negative powers of z. We now restrict ourselves to cases in which $\psi^-(z)$ defines an analytic function for $|z|>R$ for some $R<1$. By this we simply mean that the power series

$$\sum_{j=1}^{\infty} \psi_{-j} z^{-j}$$

is convergent for $|z|>R$. Under this additional assumption, $\psi(z)$ defines an analytic function in the annular region $R<|z|<1$. This prepares us for consideration of the following lemma:

Lemma. *Suppose $A(z)$ is a regular function in $|z|<1$ such that*

(i) $A(z)=\psi(z)$ *for* $R<|z|<1$,

(ii) $A(z)$ *has at most a finite number of singularities* $z_1, z_2, \ldots, z_k$ *in* $|z|<1$, *with* $P_1(z), P_2(z), \ldots, P_k(z)$ *denoting the corresponding principal parts of the Laurent series expansion of* $A(z)$ *at these points.*[28]

Then

$$[\psi(z)]_+ = A(z) - \sum_{j=1}^{k} P_j(z).$$

Proof. Let $B(z) \equiv A(z) - \sum_{j=1}^{k} P_j(z)$. A standard result from analytic function theory assures us that $B(z)$ is analytic in $|z|<1$.[29] Since (i) is true, $R>\max\{|z_1|, |z_2|, \ldots, |z_k|\}$. $P_j(z)$ is analytic for $|z| \neq z_j$ and

$$\lim_{z \to \infty} P_j(z) = 0.$$

Hence $D(z) \equiv \sum_{j=1}^{k} P_j(z)$ is analytic for $|z|>R$ and

$$\lim_{z \to \infty} D(z) = 0.$$

[28]Let the Laurent series expansion of $A(z)$ about z_j be given by

$$A(z) = \sum_{m=-\infty}^{\infty} A_m (z - z_j)^m.$$

The principal part is given by $\sum_{m=-\infty}^{-1} A_m (z-z_j)^m$. The function $A(z)$ is said to be regular in the region $|z|<1$ if it is analytic in that region except at isolated singularities.

[29]See Saks and Zygmund (1971, p. 146).

Using these results we have that

$$B(z)=\sum_{j=0}^{\infty} B_j z^j \quad \text{for} \quad |z|<1,$$

and

$$D(z)=\sum_{j=1}^{\infty} D_j z^{-j} \quad \text{for} \quad |z|>R.$$

Since $A(z)=B(z)+D(z)$, it follows that

$$\sum_{j=0}^{\infty} B_j z^j+\sum_{j=1}^{\infty} D_j z^{-j}$$

is the Laurent series expansion for $A(z)=\psi(z)$ in the region $R<|z|<1$. Since this expansion is unique,

$$\psi^+(z)=\sum_{j=0}^{\infty} B_j z^j=B(z).$$

This lemma provides a simple and computationally convenient formula for computing $[\psi(z)]_+$. In order to see the value of this computational technique, let us reconsider the prediction problem examined in the text. We are interested in determining the linear least squares predictor of

$$y_t=\sum_{k=0}^{\infty} \lambda^k x_{t+k}, \tag{A.1}$$

given information available up until time t. We assume that the only information useful for predicting y_t is current and past observations of x_t. We also assume that x_t is a linearly indeterministic, covariance stationary process. Wold's theorem provides us with a representation of x_t given by

$$x_t=\sum_{j=0}^{\infty} \xi_j v_{t-j}^x, \tag{A.2}$$

where v_t^x is the linear least squares one-step ahead prediction error and

$$E\, v_t^x v_{t-j}^{x\prime}=0 \quad \text{for} \quad j\neq 0.$$

Current and past v^x's encompass the same information set as current and past x's. Wold's theorem also assures us that the elements of $\{\xi_j\}$ are square summable. The z transform of $\sum_{k=0}^{+\infty} \lambda^k L^{-k}$ is

$$\sum_{k=0}^{+\infty} \lambda^k z^{-k}=1/(1-\lambda z^{-1})=z/(z-\lambda) \quad \text{for} \quad |z|>\lambda.$$

The z transform corresponding to the Wold representation is

$$\xi(z)=\sum_{j=0}^{+\infty}\xi_j z^j \quad \text{for} \quad |z|<1.$$

Combining (A.1) and (A.2) we can represent

$$y_t=\sum_{j=-\infty}^{+\infty}\psi_j v_{t-j}^x,$$

where

$$\psi(z)=\sum_{j=-\infty}^{+\infty}\psi_j z^j=z\xi(z)/(z-\lambda) \quad \text{for} \quad \lambda<|z|<1.$$

The linear least squares predictor of y_t is

$$E_t y_t=[\psi(L)]_+ v_t^x.$$

Thus we need to compute $[\psi(z)]_+$.

Now $A(z)\equiv z\xi(z)/(z-\lambda)$ is regular in $|z|<1$ and has a simple pole at λ $A(z)=\psi(z)$ for $\lambda<|z|<1$. In the Laurent series expansion of $A(z)$ around λ, the residue of $A(z)$ is given by

$$\lim_{z\to\lambda}(z-\lambda)A(z)=\lambda\xi(\lambda).$$

This informs us that the principal part of the expansion of $A(z)$ at λ is

$$\lambda\xi(\lambda)/(z-\lambda).$$

Using the lemma we have that

$$\begin{aligned}[\psi(z)]_+ &= z\xi(z)/(z-\lambda)-\lambda\xi(\lambda)/(z-\lambda)\\ &= (\xi(z)-\lambda z^{-1}\xi(\lambda))/(1-\lambda z^{-1}).\end{aligned}$$

This agrees with eq. (5) in the text.

In the text it is claimed that the covariance stationarity assumption for a_t and x_t can be relaxed. We shall examine this claim for x_t and assert that the logic for a_t is completely analogous. We make an rth order vector Markov specification

$$\zeta(L)x_t=v_t^x \quad \text{for} \quad t\geq 0,$$

where $\zeta(L)=I-\zeta_1 L-\ldots-\zeta_r L^r$, where $x_{-1},\ldots,x_{-r}$ are given initial values, and where v_t^x is a contemporaneously uncorrelated vector white noise. We assume that the roots of $\det\zeta(z)=0$ lie outside the region $|z|<\sqrt{\beta}$.

Define

$$x_t = 0 \qquad \text{for} \quad t < -r,$$

and let

$$v_t^x = \zeta(L)x_t \qquad \text{for} \quad t < 0.$$

Using the assumption on the roots of $\det \zeta(z)$, we know that $\zeta(z)$ has an inverse with elements analytic for $|z| < \sqrt{\beta}$. Taking the Taylor series expansion about zero we have

$$\zeta(z)^{-1} = \xi(z) = \sum_{j=0}^{\infty} \xi_j z^j.$$

Since $v_t^x = 0$ for $t < -r$, $\xi(L)v_t^x$ is a well defined stochastic process and

$$x_t = \xi(L)v_t^x. \tag{A.3}$$

Even though the coefficients ξ_j may not be bounded, eq. (A.3) provides a valid representation of the x_t process to which the Wiener–Kolmogorov prediction formulas can be applied.

As was argued in the text $\lambda < \sqrt{\beta}$. Therefore

$$\psi(z) = z\xi(z)/(z-\lambda) \quad \text{is analytic for} \quad \lambda < |z| < \sqrt{\beta}.$$

Its Laurent series expansion, given by

$$\psi(z) = \sum_{j=-\infty}^{+\infty} \psi_j z^j,$$

provides a representation for y_t of the form

$$y_t = \sum_{j=-\infty}^{+\infty} \psi_j v_{t-j}^x.$$

Using the Wiener–Kolmogorov solution to the prediction problem we know that

$$E[y_t | x_t, x_{t-1}, \ldots, x_{-r}] = [\psi(L)]_+ v_{t-j}^x.$$

If we modify the domain specified in the lemma to be $|z| < \sqrt{\beta}$, we conclude that

$$[\psi(z)]_+ = (z\xi(z) - \lambda\xi(\lambda))/(z-\lambda).$$

This extends our prediction formula to non-stationary finite order Markov processes.

The lemma turns out to be handy in solving classical signal extraction problems of the kind discussed by Whittle (1963, p. 66). For example let a signal y_t be governed by an rth order Markov process

$$(1-\rho_1 L)(1-\rho_2 L)\ldots(1-\rho_r L)y_t = e_t, \qquad |\rho_j| < 1,$$

while $x_t = y_t + u_t$.

Here $E u_t e_{t-j} = 0 = E u_t = E e_t$ for all t and j, where u_t and e_t are serially uncorrelated white noises. The problem is to form the linear least squares projection $[y_t | x_t, x_{t-1}, \ldots]$. This can be solved using formula (3) of Whittle (1963, p. 66). The lemma is useful in solving this problem since it makes application of the annihilation operator $[\ \ ]_+$ fairly routine.

Appendix B

In this appendix we discuss the asymptotic distribution of the family of generalized method of moments estimators proposed in section 6. We abstain from a detailed presentation and refer the interested reader to Hansen (1979). Both Hansen's discussion and the present discussion exploit a framework developed by Sims in lectures to his graduate econometrics class at the University of Minnesota.

In this appendix we adopt different notation than is used in the text. We do this in order to facilitate the presentation. We let x_t denote the underlying observable vector process including both endogenous and exogenous variables and all relevant lags of these variables. We assume that this process is stationary and ergodic. β_∞ denotes the vector of parameters to be estimated and is of dimension k. The sample orthogonality conditions discussed in section 6 are compactly written

$$O_T(\beta) = (1/T) \sum_{t=1}^{T} f(x_t, \beta).$$

The dimensionality of O_T is assumed to be $p \geqq k$. Our model assumptions imply that

$$E f(x_t, \beta_\infty) = 0 \quad \text{for all } t,$$

and therefore

$$E O_T(\beta_\infty) = 0.$$

Let us introduce a matrix A that is $q \times p$, where $k \leqq q \leqq p$ and define β_T to be that value of β such that $|A O_T(\beta_T)|$ is as small as possible.

We assume that β_T converges in probability to β_∞. Let

$$B = E[\partial f(x_t, \beta_\infty)/\partial \beta],$$

and let

$$R_j = E[f(x_t, \beta_\infty) f(x_{t-j}, \beta_\infty)'].$$

The infinite sum

$$S = \sum_{j=-\infty}^{+\infty} R_j$$

is assumed to converge absolutely. Employing some mild side conditions Hansen obtains the result that $\sqrt{T}(\beta_T - \beta_\infty)$ converges in distribution to a normally distributed random variable with zero mean and covariance matrix

$$(B'A'AB)^{-1} B'A'ASA'AB(B'A'AB)^{-1}.$$

In order for this covariance matrix to be as small as possible, it is desirable to select A so that

$$A'A = S^{-1},$$

and the resulting covariance matrix is

$$(B'S^{-1}B)^{-1}.$$

The 'best' choice of A is dependent on S, which is not known *a priori* and must therefore be estimated. Consistent estimators of S can be obtained from a first step estimator of β_∞ employing a not necessarily 'optimal' choice of A. This allows us to obtain a sequence

$$A_T' A_T = S_T^{-1},$$

that converges in probability to S^{-1}. If we substitute A_T for A in our definition of β_T we still obtain the same asymptotic distribution for $\sqrt{T}(\beta_T - \beta_\infty)$. This provides the 'optimal' estimator alluded to in section 6.

References

Anderson, T.W., 1975, The statistical analysis of time series (Wiley, New York).

Bard, Yonathan, 1974, Nonlinear parameter estimation (Academic Press, New York).

Barro, Robert J., 1977, Unanticipated money growth and unemployment in the United States, American Economic Review 67, 101–115.

Barro, Robert J., 1978, Unanticipated money, output and the price level in the United States, Journal of Political Economy 86, 549–580.

Bertsekas, Dimitri P., 1976, Dynamic programming and stochastic control (Academic Press, New York).

Blanco, Herminio, 1978, Investment under uncertainty: An empirical analysis, Ph.D. dissertation (University of Chicago, Chicago, IL).

Box, C.E.P. and G.M. Jenkins, 1970, Time series analysis, forecasting and control (Holden-Day, San Francisco, CA).

Churchill, R.V., 1960, Complex variables and applications (McGraw-Hill, New York).

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

Graves, R. and L. Telser, 1971, Functional analysis in mathematical economics (University of Chicago Press, Chicago, IL).

Hall, Robert 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., 1979, Large sample properties of generalized method of moments estimators, Manuscript.

Hansen, L.P. and T.J. Sargent, 1979, A note on Wiener–Kolmogorov prediction formulas for rational expectations models, Manuscript.

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

Holt, C., F. Modigliani, J.F. Muth and H.A. Simon, 1960, Planning production, inventories and work force (Prentice-Hall, Englewood Cliffs, NJ).

Kennan, John, 1979, The estimation of partial adjustment models with rational expectations, Econometrica 47, 1441–1456.

Kushner, Harold, 1971, Introduction to stochastic control (Holt, Rinehart, and Winston, New York).

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

Lucas, R.E., Jr., 1972, Econometric testing of the natural rate hypothesis, in: Otto Eckstein, ed., The econometrics of price determination conference (Board of Governors of the Federal Reserve System, Washington, DC) 50–59.

Lucas, R.E., Jr., 1976, Econometric policy evaluation: A critique, in: K. Brunner and A.H. Meltzer, eds., The Phillips curve and labor markets, Carnegie-Rochester conferences on public policy 1 (North-Holland, Amsterdam).

Lucas, R.E., Jr. and E. Prescott, 1971, Investment under uncertainty, Econometrica 39, 659–681.

Lucas, R.E., Jr. and T.J. Sargent, 1978, After Keynesian macroeconomics, in: After the Phillips curve: Persistence of high inflation and high unemployment (Federal Reserve Bank of Boston, Boston, MA).

McCallum, B.T., 1976, Rational expectations and the natural rate hypothesis: some consistent estimates, Econometrica 44, 43–52.

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

Revankar, N.S., 1978, Asymptotic relative efficiency analysis of certain tests of independence in structural systems, International Economic Review 19, 165–180.

Revankar, N.S., 1980, Testing of the rational expectations hypothesis, Econometrica, forthcoming.

Rissanen, J. and P.E. Caines, 1979, The strong consistency of maximum likelihood estimators for ARMA processes, Annals of Statistics 7, 297–315.

Saks, S. and A. Zygmund, 1971, Analytic functions (Elsevier, New York).

Salemi, M. and T.J. Sargent, 1980, The demand for money during hyperinflations under rational expectations: II, International Economic Review, forthcoming.

Sargent, T.J., 1977, The demand for money during hyperinflations under rational expectations: I, International Economic Review 18, 59–82.

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., 1979, Macroeconomic theory (Academic Press, New York).
Sargent, T.J. and C.A. Sims, 1977, Business cycle modeling without pretending to have too much a priori economic theory, in: C.A. Sims, ed., New methods in business cycle research (Federal Reserve Bank of Minneapolis, MN).
Shiller, R., 1972, Rational expectations and the structure of interest rates, Ph.D. dissertation (Massachusetts Institute of Technology, Cambridge, MA).
Sims, C.A., 1972, Money, income and causality, American Economic Review 62, 540–552.
Sims, C.A., 1980, Macroeconomics and reality, Econometrica, forthcoming.
Taylor, John B., 1978, Output and price stability: An international comparison, Manuscript.
Taylor, John B., 1979, Estimation and control of a macroeconomic model with rational expectations, Econometrica 47, 1267–1286.
Tunnicliffe Wilson, G., 1973, The estimation of parameters in multivariate time series models, Journal of the Royal Statistical Societies B 35, 76–85.
Vaughan, D.R., 1970, A nonrecursive algebraic solution for the discrete Riccati equation, IEEE Transactions on Automatic Control 15, 597–599.
Wallis, Kenneth F., 1980, Econometric implications of the rational expectations hypothesis, Econometrica, forthcoming.
Whittle, P., 1963, Regulation and prediction by linear least-square methods (D. Van Nostrand, Princeton, NJ).
Wu, De-Min, 1973, Alternative tests of independence between stochastic regressors and disturbances, Econometrica 41, 733–750.
Zygmund, A., 1959, Trigonometric series, Vol. I (University Press, Cambridge).