Journal of Monetary Economics 9 (1982) 263–296. North-Holland Publishing Comapny

INSTRUMENTAL VARIABLES PROCEDURES FOR ESTIMATING LINEAR RATIONAL EXPECTATIONS MODELS*

Lars Peter HANSEN

Carnegie-Mellon University, Pittsburgh, PA 15213, USA

Thomas J. SARGENT

University of Minnesota, Minneapolis, MN 55455, USA

Federal Reserve Bank of Minneapolis, MN, USA

This paper illustrates how to use instrumental variables procedures to estimate the parameters of a linear rational expectations model. These procedures are appropriate when disturbances are serially correlated and the instrumental variables are not exogenous. We compare our procedures to some alternative estimators that estimate free parameters from restrictions implied by the Euler equations. The procedures are applicable to a variety of linear rational expectations models, several examples of which we cite.

1. Introduction

In a variety of linear rational expectations models, agents' decisions are supposed to depend on geometrically declining weighted sums of expected future 'forcing variables'. These forcing variables are typically described by stochastic processes that the agents view as being beyond their control. The following are examples of such models.

(i) *Cagan's model of portfolio balance.* Letting p_t be the logarithm of the price level, y_t be the logarithm of the money supply, and a_t be a stationary disturbance to portfolio balance, Cagan's model can be represented as

$$p_t = \frac{1}{1-\alpha} E\left[\sum_{j=0}^{\infty} \left(\frac{\alpha}{\alpha-1}\right)^j (y_{t+j} - a_{t+j}) \,|\, \Omega_t\right],$$

where E is the expectations operator, Ω_t is agents' information set at

*Fumio Hayashi, Robert Hodrick, Bennett McCallum, Dan Peled, Kenneth Singleton, and Robert Townsend provided some helpful comments on an earlier draft. This research was supported in part by NSF Grant SES-8007016.

time t, and $\alpha<0$ is the slope of the portfolio balance schedule.[1]

(ii) *A dynamic model of demand for factors of production.* Let n_t be the stock of a factor of production, y_t be the real wage rate of the factor, and a_t be a random shock to technology. Then a linear-quadratic version of a costly adjustment model predicts that n_t will obey

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

where $0<\lambda<1$, $0<\beta<1$, $\delta>0$, and Ω_t is the firm's information set at time t. Here β is the firm's discount factor, and δ is a parameter measuring the costs of adjustment.[2]

(iii) *The permanent income model of consumption.* Let c_t be consumption, A_t non-human assets, y_t labor income, and a_t 'transitory consumption'. Then the permanent income model of consumption can be written

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

where ρ is the interest rate and β is the marginal propensity to consume.[3]

More examples of such models in which the geometric sums $E(\sum_{j=0}^{\infty}\lambda^j y_{t+j}\,|\,\Omega_t)$ appear can be found in Sargent (1979) and Hansen and Sargent (1981). As Hansen and Sargent (1981) show, such geometric sums are important terms in a wide class of models that come from infinite horizon, linear-quadratic stochastic optimum problems. In such models, it is common to suppose that the values of the forcing variable y_t are observable both to the econometrician and the agent, but that only the agent observes the forcing variable a_t. The 'hidden variable' a_t thus becomes one source of the error in the equations fit by the econometrician [see Hansen and Sargent (1980)]. Both the y and the a processes usually are modeled as being beyond the control of the private agent. The private agent is assumed to face these processes as a 'price taker' or 'income taker'. However, for standard simultaneous equations reasons, this assumption does *not* imply that the y process will be strictly econometrically exogenous with respect to the decision variable. Indeed, the assumption that y is uncontrolled by the agent does not even imply that y fails to be Granger caused by the private agents' decision variable. However, in most of the technical literature published to

[1] For further exposition of this example, see Sargent (1977, 1979).
[2] This and related examples are described in Sargent (1979).
[3] For a description of this example, see Hall (1978) and Sargent (1978).

date,[4] estimation of linear rational expectations models has been treated either under the assumption that y is strictly exogenous, or under the weaker assumption that y is not Granger-caused by the private agents' decision variable.[5]

The purpose of the present paper is to describe optimal estimation procedures in the case in which y is not strictly exogenous, in which the agents' decision variable in general Granger-causes the forcing variable y and in which full-blown maximum likelihood procedures are thought to be undesirable or inapplicable. For applications, this is an important extension to existing estimation procedures. Thus the theoretical presumption for each of the examples given above is probably in favor of dynamic feedback from market-wide measures of the decision variable on the left-hand side of the equation to the y process on the right-hand side.[6]

This paper proposes estimators that can be interpreted as instrumental variables estimators. The basic idea of this paper is to carry out identification and estimation of the model's free parameters from the projections of the decision variables and the forcing variables on instruments, and the projection of the instruments on their own lagged values. These projections are characterized by a set of cross-equation restrictions involving the free parameters of the model, restrictions that are often stringent enough to permit identification of the model's free parameters. It is significant that the instruments need not be strictly econometrically exogenous with respect to the left-hand side or decision variables. It is even permitted that the decision variables Granger-cause the instruments. Further, the disturbances in the equation are permitted to be serially correlated, though the procedures do not require the analyst explicitly to parameterize the stochastic process for the disturbances. Among other things, this paper helps clarify the relationship between Granger causality and the criterion for appropriateness of an instrument.

There are two principal virtues of the instrumental-variables-type estimators of the present paper *vis-a-vis* the maximum likelihood estimator proposed, e.g., by Hansen and Sargent (1980). First, fewer parameters need to be estimated simultaneously than are required with the maximum likelihood estimator. Second, precise parameterizations of the disturbances need not be specified with the present estimators, while they must be with maximum likelihood.

While the estimators described are applicable to a variety of linear rational expectations models, we have chosen to describe them by referring to our

[4]A few simple examples exist that use maximum likelihood estimators that explicitly take account of feedback from the left-hand side variable to y_t. For example, see Sargent (1977).

[5]For example, see Hansen and Sargent (1980).

[6]See Kydland and Prescott (1977) and Sargent (1980) for discussions of how linear dynamic competitive equilibria can be calculated when there is feedback from market-wide values or agents' decision variables to prices or incomes that individual agents view as uncontrollable.

third example, that of the permanent income consumption function. This example is one in which the failure of y to be exogenous is well known and widely described in econometrics textbooks. It will be evident how our methods apply to other examples, including those given above.

We go on to compare our methods to some related methods proposed by Kennan (1979) and Hayashi (1980) that directly estimate Euler equations, and thereby avoid dealing explicitly with geometrically declining sums of expected future forcing variables. While many of the comments we make about estimation carry over to these related methods, it turns out that these other methods ignore theoretical restrictions and therefore sacrifice statistical efficiency relative to the methods that we propose. It is convenient to make this latter point in the context of the second example, a dynamic model of demand for factors of production.

This paper is organized as follows. In section 2 we specify precisely a version of the permanent income model of consumption and discuss the econometric restrictions implied by the model. We characterize the model by projections in various directions that can be utilized econometrically. In section 3 we propose some instrumental variables estimators of the parameters of the model and discuss their large sample properties. We also indicate how the estimator of Hayashi and Sims (1982) compares with the optimal instrumental variables estimator. Section 4 contrasts our methods with Euler equation methods proposed by Kennan (1979) and Hayashi (1980) that do not work directly with geometrically declining sums of expected future forcing variables. Our conclusions are in section 5.

2. The statistical model

In this section we examine the restrictions which emerge from a permanent income model of consumption. We consider a linear model for consumption of the form

$$c_t = \beta y_{pt} + a_t, \tag{2.1}$$

where c_t is consumption at time t, a_t is 'transitory consumption' at t and y_{pt} is permanent income at t. The econometrician is assumed not to have observations on transitory consumption or on permanent income. We postpone spelling out what properties we assume for transitory consumption and temporarily focus on our working definition of permanent income. We assume that

$$y_{pt} = \frac{\rho}{1+\rho}\left[A_t + \sum_{j=0}^{\infty}(1+\rho)^{-j}E_t y_{t+j}\right], \tag{2.2}$$

where ρ is the real interest rate assumed constant over time, A_t is non-human assets at time t, y_t is after tax labor income at t,[7] and $E_t[\cdot]=E[\cdot|\Omega_t]$ is the mathematical expectation operator conditioned on a set of information Ω_t available to private agents at time t.[8] Substituting (2.2) into (2.1) and making explicit the information set Ω_t available to the consumer at t, gives

$$c_t=\beta(1-\delta)\left[A_t+\sum_{j=0}^{\infty}\delta^j Ey_{t+j}|\Omega_t\right]+a_t, \tag{2.3}$$

where $\delta=(1+\rho)^{-1}$. We assume that $\Omega_t\supset\Omega_{t-1}\supset\Omega_{t-2}\ldots$ and that $\{y_t, y_{t-1},\ldots\}\subset\Omega_t$.

To motivate definition (2.2) of permanent income, consider a setup in which infinite lived consumers face the sequence of budget constraints

$$A_{t+1}=(1+\rho)A_t+(1+\rho)(y_t-c_t), \qquad t=t_0,t_{0+1},\ldots,$$

$$A_{t_0} \text{ given.}^9 \tag{2.4}$$

It is assumed that y is a stochastic process which is beyond the control of the consumer. Solving the stochastic difference eq. (2.4) forward and imposing the terminal condition

$$\lim_{j\to\infty}(1+\rho)^{-(t+j)}E_tA_{t+j}=0 \tag{2.5}$$

gives the 'realizable' solution

$$\sum_{j=0}^{\infty}\left(\frac{1}{1+\rho}\right)^j E_tc_{t+j}=A_t+\sum_{j=0}^{\infty}\left(\frac{1}{1+\rho}\right)^j E_ty_{t+j}\equiv W_t, \tag{2.6}$$

where W_t is the consumer's total wealth, human and non-human. Eq. (2.6) states that the expected present discounted value of consumption equals the

[7]There are some important empirical issues outstanding in the literature on consumption that relate to the definitions of A_t and y_t. These issues include whether government bonds and social insurance obligations should be included in non-human wealth, and how future tax liabilities required to service these claims should be treated.

[8]It is the presence of the transitory consumption term that differentiates the consumption model here from the one considered by Hall (1978). Hall's short-cut econometric procedure relies critically on the absence of this transitory consumption term [see Flavin (1981)]. Our definition of permanent income differs from that used by Sargent (1978) because of our inclusion of non-human assets in our measure of permanent income.

[9]If we take (2.4) literally, it implies that a stochastic singularity exists in the joint $(c,y,A)'$ process. We assume that this singularity does not exist. Instead we implicitly assume that there are shocks to this budget constraint which might take the form of unobservable (to the econometrician) components of income.

present value of non-human assets A_t plus the expected present discounted value of labor income. For convenience, write (2.6) as

$$\sum_{j=0}^{\infty} \delta^j E_t c_{t+j} = W_t = A_t + \sum_{j=0}^{\infty} \delta^j E_t y_{t+j}. \tag{2.7}$$

Notice that the constant level $c_{t+j}=\bar{c}_t$ of planned consumption at t that satisfies (2.5) is found from

$$\bar{c}_t \sum_{j=0}^{\infty} \delta^j = W_t$$

or

$$\bar{c}_t = (1-\delta) W_t = (1-\delta)\left(A_t + \sum_{j=0}^{\infty} \delta^j E_t y_{t+j}\right)$$

$$= \frac{\rho}{1+\rho}\left[A_t + \sum_{j=0}^{\infty} \left(\frac{1}{1+\rho}\right)^j E_t y_{t+j}\right]. \tag{2.8}$$

According to one widely accepted definition, $\bar{c}_t$ defined by (2.8) is the level of permanent income: it is the rate at which the consumer expects to be able to consume indefinitely given his current total wealth. Specification (2.2) embodies this notion of permanent income.[10]

We proceed to specify properties that we assume about transitory consumption. The random process a is assumed to have mean zero, but it can be serially correlated. Further, it can be correlated with the y process. Thus, we do not assume that y is econometrically exogenous in (2.3). Obviously, however, to give content to (2.3) and to proceed with estimation, some orthogonality conditions must be imposed on a_t. We assume that there is a $(p \times 1)$ vector x_t that is included in Ω_t satisfying

$$E a_t x_{t-j} = 0, \tag{2.9}$$

$j \geqq 0$. It is important to note that a_t is allowed to be correlated with future x's

[10] An alternative way to derive the model given in (2.1) and (2.2) is to follow Hall (1978) and assume that a representative agent solves a time-separable, quadratic optimization problem subject to a lifetime budget constraint. The imposition of the lifetime budget constraint in effect imposes (2.5). Under the assumption that the subjective rate of time preference is less than ρ, the parameter β is greater than one. The transitory consumption terms could be interpreted as combinations of shocks to preferences and shocks to the budget constraint (2.4). With these assumptions, a model of the form given in (2.1) and (2.2) can be obtained in which all of the observable (to the econometrician) variables are measured as deviations from their unconditional means.

and therefore the x's do not have to be econometrically exogenous with respect to c.[11] We define a reduced information set $\Phi_t=\{x_t, x_{t-1},\dots\}$. The idea underlying our estimation strategy is to exploit the orthogonality conditions (2.9) and, in a sense, to employ the x process as an instrument for y.

Now rewrite (2.3) as

$$c_t=\frac{\beta\rho}{1+\rho}A_t+\frac{\beta\rho}{1+\rho}\sum_{j=0}^{\infty}(1+\rho)^{-j}Ey_{t+j}\,|\,\Phi_t+a_t+s_t, \tag{2.10}$$

where the 'error' s_t is given by[12]

$$s_t=\frac{\beta\rho}{1+\rho}\sum_{j=0}^{\infty}(1+\rho)^{-j}(Ey_{t+j}\,|\,\Omega_t-Ey_{t+j}\,|\,\Phi_t). \tag{2.11}$$

By construction the error s_t is orthogonal to Φ_t, since Φ_t is included in Ω_t. That is, by the law of iterated projections, we have[13]

$$E(Ey_{t+j}\,|\,\Omega_t-Ey_{t+j}\,|\,\Phi_t)\,|\,\Phi_t=Ey_{t+j}\,|\,\Phi_t-Ey_{t+j}\,|\,\Phi_t=0.$$

In order to calculate the projection of permanent income onto the reduced information set Φ_t, we need to make some precise assumptions about the forcing variables. We calculate this projection to obtain a set of cross-equation restrictions that can be exploited in estimation. We make the convenient assumption that $(y, A, x', a, s)'$ is a vector, covariance stationary, linearly indeterministic stochastic process.[14] In what follows we assume that conditional expectations and best linear predictors coincide. Alternatively, we could abstain from this coincidence assumption and instead assume that $E[\cdot\,|\,\Phi_t]$ denotes the linear least squares projection operator onto the information set Φ_t.

Since x is a covariance stationary, linearly indeterministic stochastic

[11]In fact, c would in general Granger-cause x even if (2.9) were extended to hold for all j, and not just $j\geqq 0$.

[12]The idea of replacing the information set Ω_t with a subset Φ_t, thereby adding an error term like s_t, was suggested by Shiller (1972) in his study of the term structure of interest rates, and was exploited in a related context by Hansen and Sargent (1980).

[13]The law of interated projections states that $E(y\,|\,x)=E(Ey\,|\,x,z)\,|\,x)$, where y, x, z are random variables and E is either the mathematical expectation or the linear least squares projection operator.

[14]See Rozanov (1967) for a definition of covariance stationary, linearly indeterministic. In formally verifying the large sample properties of the estimators we propose in the next section, Hansen (1980) strengthens this assumption to require strict stationarity. In order for these stationarity assumptions to be consistent with the budget constraint (2.4), it is necessary to restrict β to be greater than one. In the discussion which follows, all variables will be viewed as deviations from their unconditional means.

process, it has a Wold vector moving average representation

$$x_{t+1}=\alpha(L)e_{t+1}, \tag{2.12}$$

where $\alpha(L)=I+\alpha_1 L+\ldots$ and where

$$e_{t+1}=x_{t+1}-E[x_{t+1}\mid\Phi_t]. \tag{2.13}$$

We add the additional restrictions that

$$\sum_{j=0}^{\infty}[\operatorname{trace}\alpha_j'\alpha_j]^{\frac{1}{2}}<+\infty \tag{2.14}$$

and that the function F is bounded away from zero where

$$F(\omega)=\det[\alpha(e^{i\omega})\alpha(e^{i\omega})']. \tag{2.15}$$

The symbol $'$ denotes both transposition and conjugation. Among other things these assumptions are sufficient to imply that there exists a one-sided operator $\gamma=\alpha^{-1}$ or equivalently that x has an autoregressive representation

$$\gamma(L)x_{t+1}=e_{t+1}, \tag{2.16}$$

where $\gamma(L)=I-\gamma_1 L-\ldots$. The dimension of e is $(p\times 1)$ while α and γ are both $(p\times p)$.[15]

We write the projection of y_t on Φ_t as

$$Ey_t\mid\Phi_t=\theta(L)x_t, \tag{2.17}$$

where $\theta(L)=\theta_0+\theta_1 L+\ldots$. That is, we have the orthogonal decomposition

$$y_t=\theta(L)x_t+u_t, \tag{2.18}$$

where $Eu_t x_{t-j}=0$ for $j\geqq 0$. Substituting (2.12) into (2.18), we obtain

$$y_t=\theta(L)\alpha(L)e_t+u_t. \tag{2.19}$$

Now let the after tax labor income variable whose forecast appears in (2.8)

[15] We are also assuming that the covariance matrix of e has full rank.

be denoted

$$y_t^* \equiv \sum_{j=0}^{\infty} (1+\rho)^{-j} y_{t+j} = \sum_{j=0}^{\infty} \delta^j y_{t+j} = \left[\frac{1}{1-\delta L^{-1}}\right] y_t, \tag{2.20}$$

recalling that $\delta = (1+\rho)^{-1}$. Substituting (2.19) into (2.20) gives

$$y_t^* = \frac{\theta(L)\alpha(L)}{1-\delta L^{-1}} e_t + \frac{1}{1-\delta L^{-1}} u_t. \tag{2.21}$$

Note that u_{t+j} for $j>0$ is orthogonal to the information set $\Phi_t = \{x_t, x_{t-1}, \ldots\}$ by virtue of the orthogonality conditions $E u_t x_{t-j} = 0$ for $j \geqq 0$. We use the Wiener–Kolmogorov prediction formula to compute

$$E y_t^* \mid \Phi_t = \left[\frac{\theta(L)\alpha(L)}{1-\delta L^{-1}}\right]_+ e_t, \tag{2.22}$$

where $[\;\;]_+$ is the 'annihilation operator' that instructs us to ignore negative powers of L.[16] That is,

$$\left[\sum_{j=-\infty}^{+\infty} \mu_j L^j\right]_+ = \sum_{j=0}^{\infty} \mu_j L^j.$$

Using the lemma in Appendix A of Hansen and Sargent (1980) to evaluate the above term in $[\;\;]_+$, we obtain

$$E y_t^* \mid \Phi_t = \left[\frac{\theta(L)\alpha(L) - \delta L^{-1}\theta(\delta)\alpha(\delta)}{1-\delta L^{-1}}\right] e_t. \tag{2.23}$$

Since $\alpha(L) e_t = x_t$ and $\alpha^{-1} = \gamma$, eq. (2.23) can be written in the equivalent form

$$E y_t^* \mid \Phi_t = \left[\frac{\theta(L) - \delta L^{-1}\theta(\delta)\gamma(\delta)^{-1}\gamma(L)}{1-\delta L^{-1}}\right] x_t. \tag{2.24}$$

Substituting (2.20) and (2.24) into (2.10) gives the equation

$$c_t = \frac{\beta\rho}{1+\rho} A_t + \frac{\beta\rho}{1+\rho}\left[\frac{\theta(L) - \delta L^{-1}\theta(\delta)\gamma(\delta)^{-1}\gamma(L)}{1-\delta L^{-1}}\right] x_t + a_t + s_t. \tag{2.25}$$

[16]See Whittle (1963) for derivation of the formula and examples showing its usefulness.

Repeating (2.18) and 2.16), we also have the projection equations

$$y_t = \theta(L)x_t + u_t, \tag{2.26}$$

$$x_{t+1} = \gamma^1(L)x_t + e_{t+1}, \tag{2.27}$$

where $\gamma^1(L) = \gamma_1 + \gamma_2 L + \dots$.

As is the hallmark of dynamic rational expectations models, eqs. (2.25), (2.26), and (2.27) possess a set of cross-equation restrictions, indicated by the presence of the parameters of the lag operators α and θ in (2.25). The presence of these parameters reflects that consumers are making use of the properties of the y process in forming estimates of their permanent income. The existence of these cross-equation restrictions can be used to identify and estimate the parameters of the operators γ and θ and the parameters β and ρ. Key elements in the model are the following three sets of orthogonality conditions:

$$Ex_{t-j}(a_t + s_t) = 0, \tag{2.28}$$

$$Ex_{t-j}u_t = 0, \tag{2.29}$$

$$Ex_{t-j}e'_{t+1} = 0 \tag{2.30}$$

for $j \geqq 0$. Recall that x_t is a $(p \times 1)$ random vector, so that $Ex_{t-j}(a_t + s_t)$ and $Ex_{t-j}u_t$ are $(p \times 1)$ vectors, while $Ex_{t-j}e'_{t+1}$ is a $(p \times p)$ matrix.

The orthogonality conditions (2.29) and (2.30) stem directly from the construction of (2.26) and (2.27) as projection equations. In other words, (2.26) was constructed by projecting y_t onto Φ_t and (2.27) by projecting x_{t+1} onto Φ_t. These orthogonality conditions emerge from the 'orthogonality principle' which states that forecast errors associated with best linear predictors must be orthogonal to all random variables in the information set used in constructing the forecast. The orthogonality condition (2.28) stems jointly from the definition (2.11) of s_t together with the assumption that $Ex_{t-j}a_t = 0$ for $j \geqq 0$. Orthogonality condition (2.28) states that the projection of $c_t - \beta\rho A_t/(1+\rho)$ onto Φ_t is

$$\frac{\beta\rho}{1+\rho}\left[\frac{\theta(L) - \delta L^{-1}\theta(\delta)\gamma(\delta)^{-1}\gamma(L)}{1-\delta L^{-1}}\right]x_t = \pi(L)x_t \tag{2.31}$$

and thus we can view (2.28) as a projection equation also.

The exonometric model to be estimated consists of the three sets of eqs. (2.25), (2.26) and (2.27). One may be tempted to think of these as reduced form equations since they define the projection of $c_t-\beta\rho A_t/(1+\rho), y_t$ and x_{t+1} onto Φ_t. However, this interpretation is not quite correct. The random variable $c_t-\beta\rho A_t/(1+\rho)$ is not observable to the econometrician since it involves the unknown parameters β and ρ. In the absence of knowledge of β, ρ and the cross-equation restrictions, eq. (2.25) is not well defined as a reduced form equation. More plausible candidates for the reduced form equations for our model, are the projections of c_t, A_t, y_t and x_{t+1} onto Φ_t. Since all of these variables are assumed to be observable, the coefficients of their projections onto the observable information set can be automatically identified. In fact eqs. (2.26) and (2.27) are such projections. Therefore, the parameters $\gamma_1, \gamma_2, \ldots, \theta_0, \theta_1, \ldots,$ are identified.[17] Identification of the remaining structural parameters, β and ρ, could be cast in the conventional terms of whether they can be inferred from the reduced form coefficients of the projections of c_t, A_t, y_t and x_{t+1} onto Φ_t. It turns out that we do not have to estimate all of the restricted reduced form projection equations simultaneously to achieve identification of the structural parameters. For this reason we will address the issue of identification of β and ρ in terms of the orthogonality condition (2.28) using the fact that the parameters of γ and θ are identified.

Suppose that β^* and ρ^* allow orthogonality condition (2.28) and the cross-equation restrictions to be satisfied. Let

$$\delta^*=\frac{1}{1+\rho^*}, \tag{2.32}$$

$$\eta^*=-\frac{\beta\rho}{1+\rho}+\frac{\beta^*\rho^*}{1+\rho^*}, \tag{2.33}$$

$$\pi^*(L)=\frac{\beta^*\rho^*}{1+\rho^*}\left[\frac{\theta(L)-\delta^*L^{-1}\theta(\delta^*)\gamma(\delta^*)^{-1}\gamma(L)}{1-\delta^*L^{-1}}\right]. \tag{2.34}$$

[17]It is a well-known result from linear prediction theory that the orthogonality conditions (2.29) and (2.30) uniquely define elements $v_t \in \Phi_t$ and $w_t \in \Phi_t$ such that

$$y_t-v_t=u_t, \qquad x_{t+1}-w_t=e_{t+1}.$$

The parameters of γ^1 and θ given in (2.26) and (2.27) are identified as long as the lag operators are not over parameterized in the sense that if γ^{1*} and θ^* correspond to a vector in the admissible parameter space other than the true parameter vector, then

$$\theta(L)^*x_t \neq \theta(L)x_t=v_t, \qquad \gamma^1(L)^*x_t \neq \gamma^1(L)x_t=w_t.$$

We can write

$$c_t-\frac{\beta^*\rho^*}{1+\rho^*}A_t=c_t-\frac{\beta\rho}{1+\rho}A_t-\eta^*A_t$$

$$=c_t-\left(\frac{\beta\rho}{1+\rho}+\eta^*\right)A_t. \tag{2.35}$$

Suppose the projection of A_t onto Φ_t is given by

$$A_t=\xi(L)x_t+v_t, \tag{2.36}$$

where

$$Ex_{t-j}v_t=0 \tag{2.37}$$

for $j\geqq 0$ and $\xi(L)=\xi_0+\xi_1L+\dots$, where ξ_j is $(1\times p)$ for $j\geqq 0$.[18]

Substituting (2.25), (2.31) and (2.36) into (2.35), we obtain

$$c_t-\frac{\beta^*\rho^*}{1+\rho^*}A_t=[\pi(L)-\eta^*\xi(L)]x_t-\eta^*v_t+s_t+a_t$$

$$=\pi^*(L)x_t+s_t^*+a_t^*, \tag{2.38}$$

[18]The identification and estimation strategy being discussed is a limited information one since it does not take into account the restrictions implied by the budget constraint (2.4). If we introduce a shock to this constraint that is orthogonal to Φ_t and project both sides of (2.4) onto Φ_t we obtain

$$E[A_{t+1}\,|\,\Phi_t]=(1+\rho)E[A_t\,|\,\Phi_t]+(1+\rho)E[y_t\,|\,\Phi_t]-(1+\rho)E[c_t\,|\,\Phi_t].$$

This equation imposes constraints on $\xi(L)$. Using the procedures described in section 4, it can be verified that

$$\xi(L)=\frac{\xi_0\gamma(L)}{(1-\lambda L)}+\frac{L(1+\rho)[\theta(L)-\pi(L)]}{(1-\lambda L)},$$

where $\lambda=1+(1-\beta)\rho$ and ξ_0 is a $(1\times p)$ vector of parameters. Imposing these restrictions implies the projection equation:

$$E[c_t\,|\,\Phi_t]=\frac{\beta\rho}{1+\rho}\frac{[\xi_0+\theta(\delta)\gamma(\delta)^{-1}]}{1-\lambda L}\gamma(L)x_t.$$

Estimating the projections of c_t, A_t, y_t, and x_{t+1} onto Φ_t subject to the restrictions given above imposes more information than the strategy illustrated in the paper. The estimators which we propose in section 3 can be modified in a straighforward way to accommodate these extra restrictions.

where $s_t^* + a_t^*$ satisfies orthogonality condition (2.28). Projecting (2.38) onto Φ_t, we obtain

$$\pi^*(L)x_t = [\pi(L) - \eta^* \xi(L)]x_t \tag{2.39}$$

or

$$\eta^* \xi(L)x_t = [\pi(L) - \pi^*(L)]x_t \tag{2.40}$$

for η^* and π^* given by (2.33) and (2.34), respectively. In order for our model to be identified, we assume that ξ is not of the form (2.40) for any admissible choice of β^* and ρ^*. This assumption seems innocuous since it is ruling out only singular or very special structures in ξ and π. We conclude that all of the parameters of our model are identified via the cross-equation restrictions and the orthogonality conditions (2.28), (2.29) and (2.30).

We now proceed to discuss issues that emerge in estimating the model parameters. We begin by indicating that estimators such as generalized least squares and versions of maximum likelihood are not easily applicable to estimating (2.25), (2.26) and (2.27). Maximum likelihood estimation requires that more auxiliary assumptions be made about the temporal covariances of $(s+a), u, A, y$. and x than have been made above. Notice that the preceding construction in general produces $(s+a)$ and u processes that are serially correlated. Furthermore, the nature of the s process in general depends on the time series properties of elements in the broad information set Ω_t which private agents are permitted to see but which the econometrician has not necessarily been assumed to see. In addition, in neither (2.25) nor (2.26) are x's strictly econometrically exogenous. That is, in general the disturbances are correlated with future values of the x's. This implies that attempts to 'correct' for serial correlation in the disturbances via the implicit use of filters as occurs in time series versions of generalized least squares will result in estimators that are statistically inconsistent. The reason for this is that simply filtering (2.25) and (2.26) will distort the orthogonality conditions required for consistency, since the x's are not strictly exogenous.

It is of some interest to note that not only do the x's fail to be strictly econometrically exogenous in (2.25), but also in general c will Granger-cause x. That is, given lagged x's, lagged c's will help to predict x_t. This is so in spite of the orthogonality conditions (2.28), (2.29) and (2.30). In general, s_t and a_t are both correlated with future x's so that c_t contains information that marginally helps to predict future x's. The upshot of these remarks is that in the present context, failing a Granger–Sims test for the null hypothesis that c fails to Granger cause x does not necessarily signal model misspecification; in particular, it has no bearing on whether or not the orthogonality condition (2.28) is appropriate.

It is also worth mentioning that from the point of view of extracting good estimates of the 'structural' parameters β and ρ, it is not appropriate to search for a specification of x_t (or Φ_t) that predicts y_t^* as well as possible. Given two specifications for $\Phi_t=\{x_t, x_{t-1},\ldots\}$, the one that minimizes $E(y_t^* - Ey_t^* \mid \Phi_t)^2=\sigma_{y*}^2$ is not necessarily to be preferred. The reason is that for extracting consistent estimates of the structural parameters, the orthogonality conditions $Ex_{t-j}(a_t+s_t)=0$, $j\geqq 0$, are relied upon. The value of the prediction error variance σ_{y*}^2 has no bearing on which of the two competing specifications for Φ_t more nearly satisfies the orthogonality conditions (2.28). Indeed, the motivation of the procedures in this paper is the presumption that current and lagged values of y itself perhaps should be excluded from Φ_t because such a specification would violate (2.28). This is true in spite of the presumption that including lagged y's in Φ_t would usually decrease the prediction error variance.

In conclusion, in this section we have derived a statistical model of the consumption function and have shown how orthogonality conditions (2.28), (2.29) and (2.30) can be used to identify the free parameters of (2.25), (2.26) and (2.27). We have yet to suggest estimation procedures other than to indicate that asymptotically efficient estimation requires something different from the serial correlation corrections implicit in time series versions of generalized least squares. In the next section we discuss some correct procedures for estimating the parameters of the model. The first procedure we discuss involves estimating the parameters by using method of moments estimators and choosing admissible parameter values that minimize a weighted average of a specified number of the sample counterparts of the population orthogonality conditions (2.28), (2.29) and (2.30). The weighting scheme is chosen with a view to achieving the minimum asymptotic covariance matrix for estimators that exploit the same fixed set orthogonality conditions. We describe the details involved in executing this estimation strategy. In turns out that these 'generalized method of moments' (GMM) estimators are not asymptotically as efficient as maximum likelihood, although they are computationally more convenient and, in a sense, more robust. The GMM estimators described above use only a fixed finite number of orthogonality conditions independent of sample size. For this reason, we also investigate the question of how to use all of the available orthogonality conditions, which are infinite in number.

3. Construction of consistent GMM estimators

3.1. Estimators using a fixed number of orthogonality conditions

This section indicates in some detail how to construct consistent GMM estimators described by Hansen (1982) and Hansen and Sargent (1980) to estimate the parameters of the model specified in (2.25)–(2.30). We begin by imposing finite parameterizations for the operators γ and θ. In particular, we assume that $\gamma(L)=I-\gamma_1 L-\ldots-\gamma_q L^q$ and $\theta(L)=\theta_0+\theta_1 L+\ldots+\theta_r L^r$. By virtue of the covariance stationarity assumption, the zeroes of det $\gamma(z)$ lie outside the circle. For the purpose of this exposition, we shall set $r=q-1$, although it will be evident how the estimation procedure is to be modified if $q-1\neq r$. Performing a series of calculations similar to those applied by Hansen and Sargent (1980) in a different context, it is possible to derive explicitly the coefficients of the polynomial in L that appears in (2.24). For convenience, let us define that polynomial as $\psi(L)=\sum_{j=0}^{r}\psi_j L^j$, so that

$$\psi(L)=\left[\frac{\theta(L)-\delta L^{-1}\theta(\delta)\gamma(\delta)^{-1}\gamma(L)}{1-\delta L^{-1}}\right]. \tag{3.1}$$

It can be shown that

$$\begin{aligned}
&\psi_0=\theta(\delta)\gamma(\delta)^{-1},\\
&\psi_j=\delta\theta(\delta)\gamma(\delta)^{-1}(\gamma_{j+1}+\delta\gamma_{j+2}+\ldots+\delta^{r-j}\gamma_{r+1})\\
&\qquad+(\theta_j+\delta\theta_{j+1}+\ldots+\delta^{r-j}\theta_r),\qquad j=1,\ldots,r.
\end{aligned} \tag{3.2}$$

The expressions in (3.2) provide us with a convenient explicit representation for the restrictions across the parameters of (2.25), (2.26), and (2.27).

Solving (2.25), (2.26), and (2.27) for (a_t+s_t), u_t, and e_{t+1} and substituting into the population orthogonality conditions (2.28), (2.29), and (2.30), respectively, gives

$$Ex_{t-\tau}\left(c_t-\frac{\beta\rho}{1+\rho}A_t-\frac{\beta\rho}{1+\rho}\sum_{j=0}^{r}\psi_j x_{t-j}\right)=0, \tag{3.3}$$

$$Ex_{t-\tau}\left(y_t-\sum_{j=0}^{r}\theta_j x_{t-j}\right)=0, \tag{3.4}$$

$$Ex_{t-\tau}\left(x'_{t+1}-\sum_{j=0}^{r}x'_{t-j}\gamma'_{j+1}\right)=0 \tag{3.5}$$

for $\tau=0,1,\ldots,P$, where P is the number of lagged x's used in the orthogonality conditions. Denote by z_t the vector of observables $(c_t, A_t, y_t, x'_{t+1})'$. Let the free parameters of the model β, ρ, $\gamma_1,\ldots,\gamma_{r+1}$,

$\theta_0,\ldots,\theta_r$ be denoted by the Q dimensional vector ζ_0. Let

$$d_t=\begin{bmatrix} s_t+a_t \\ u_t \\ e_{t+1} \end{bmatrix}=\lambda(L;\zeta_0)z_t, \tag{3.6}$$

where $\lambda(L;\zeta_0)=\lambda_0(\zeta_0)+\lambda_1(\zeta_0)L+\ldots+\lambda_{r+1}(\zeta_0)L^{r+1}$,

$$\lambda_0(\zeta_0)=\begin{bmatrix} 1 & \dfrac{-\beta\rho}{1+\rho} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\lambda_j(\zeta_0)=\begin{bmatrix} 0 & 0 & 0 & \dfrac{-\beta\rho}{1+\rho}\psi_{j-1}(\zeta_0) \\ 0 & 0 & 0 & -\theta_{j-1} \\ 0 & 0 & 0 & -\gamma_j \end{bmatrix}, \qquad j=1,\ldots,r+1 \tag{3.7}$$

and the ψ's are defined in (3.2). The parameters of $\lambda(L;.)$ depend on the parameters ζ_0 via the cross-equation restrictions exhibited in (2.25), (2.26), (2.27), 3.1) and (3.2). It is convenient to have notation for the functions of the data and the parameters whose mathematical expectations are zero according to the orthogonality conditions (2.28), (2.29), and (2.30). So we define the vector function

$$f_t(\zeta)=\lambda(L;\zeta)z_t\otimes\begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-P} \end{bmatrix}, \tag{3.8}$$

where '$\otimes$' denotes the Kronecker product and where ζ is an element in the admissible parameter space containing the true parameter vector ζ_0. Expression (3.8) defines $f_t(\zeta_0)$ as a $p(P+1)(p+2)=R$ vector of random variables whose expected values are restricted to be zero by (3.3), (3.4) and (3.5). The content of the theory in (2.28), (2.29) and (2.30) can now be succinctly stated as $E[f_t(\zeta_0)]=0$.

Suppose the investigator has a sample of observations on z_t for $t=-P+1,\ldots,T$. Then for each parameter ζ in the admissible parameter space, one can view

$$\frac{1}{T}\sum_{t=1}^{T} f_t(\zeta)=g_T(\zeta) \tag{3.9}$$

as an estimator $Ef_t(\zeta)$. Since $Ef_t(\zeta_0)=0$, we can think of estimating ζ_0 by finding the element ζ_T in the parameter space that makes $g_T(\zeta)$ small in some sense. To be more precise, we choose a 'distance' or weighting matrix S that is R by R and positive definite, and let ζ_T be a minimizer of

$$J(S,\zeta)=g_T(\zeta)'S^{-1}g_T(\zeta). \tag{3.10}$$

We will describe appropriate procedures for selecting the distance matrix shortly.

For a given value of the matrix S, minimizing (3.10) is a standard nonlinear minimization problem. In practice, an 'acceptable gradient' method could be used to minimize (3.10) with respect to ζ.[19] From the viewpoint of this minimization, formula (3.2) is a great help, since it explicitly characterizes the complicated cross-equation restrictions that are embedded in $\lambda(L;\cdot)$. This means that hill-climbing methods using analytical gradients of (3.10) are feasible. Also $g_T(\zeta)$ can be expressed in terms of

$$\frac{1}{T}\left[\lambda_0(\zeta)\sum_{t=1}^{T} z_t\otimes x_{t-j}+\ldots+\lambda_{r+1}(\zeta)\sum_{t=1}^{T} z_{t-r-1}\otimes x_{t-j}\right] \tag{3.11}$$

for $j=0,\ldots,P$. Thus, the vectors of sample moments

$$\frac{1}{T}\sum_{t=1}^{T} z_{t-k}\otimes x_{t-j} \tag{3.12}$$

for $j=0,\ldots,P$ and $k=0,\ldots,r$ need only be computed once and stored in the numerical minimization of (3.10).

The estimator described above is of the same form as the nonlinear instrumental variables estimators considered by Amemiya (1974, 1977) and Jorgenson and Laffont (1974). However, the results from those papers do not apply to the estimation environment considered in this paper because in our model disturbance terms are possibly serially correlated and instruments are not necessarily strictly exogenous. Hansen (1982) provides a treatment of the large sample properties of GMM estimators under regularity conditions that allow for serially correlated disturbances and instruments that are not strictly exogenous. He establishes the consistency and asymptotic normality of estimators in a class that includes the estimators considered in this paper. It turns out that the asymptotic covariance matrix is dependent on the choice

[19] See Bard (1974) for a description of such methods.

of distance matrix S. However, it is possible to determine an optimal choice of S that will yield an estimator with the smallest asymptotic covariance matrix among the class of estimators that use the same set of orthogonality conditions. Hansen demonstrates that the optimal choice of S is given by

$$S_f = \sum_{j=-\infty}^{+\infty} R_f(j), \tag{3.13}$$

where

$$R_f(j) = E[f_t(\zeta_0) f_{t-j}(\zeta_0)'].$$

Note that S_f is the spectral density matrix of the random vector $f(\zeta_0)$ at frequency zero. Under the more special assumptions that the z process is Gaussian or is a stationary process whose fourth-order cumulants are zero, S_f has an alternative representation

$$\begin{aligned} S_f &= \sum_{j=-\infty}^{+\infty} R_d(j) \otimes R_x^p(j) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} S_d(-\omega) \otimes S_x^p(\omega)\, \mathrm{d}\omega, \end{aligned} \tag{3.14}$$

where

$$R_x^p(j) = E \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p} \end{bmatrix} [x'_{t-j}, x'_{t-j-1}, \ldots, x'_{t-j-p}], \tag{3.15}$$

$$S_x^p(\omega) = \sum_{j=-\infty}^{+\infty} e^{i\omega j} R_x^p(j),$$

$$R_d(j) = E[d_t d'_{t-j}],$$

$$S_d(\omega) = \sum_{j=-\infty}^{+\infty} e^{i\omega j} R_d(j).$$

If we let

$$R_x(j) = E[x_t x'_{t-j}],$$

$$S_x(\omega) = \sum_{j=-\infty}^{+\infty} e^{i\omega j} R_x(j) \tag{3.16}$$

then it follows that

$$S_x^P(\omega)=\begin{bmatrix} S_x(\omega) & e^{i\omega}S_x(\omega) & \dots & e^{iP\omega}S_x(\omega) \\ e^{-i\omega}S_x(\omega) & S_x(\omega) & \dots & e^{i(P-1)\omega}S_x(\omega) \\ \vdots & \vdots & & \vdots \\ e^{-iP\omega}S_x(\omega) & e^{-(P-1)\omega}S_x(\omega) & \dots & S_x(\omega) \end{bmatrix}. \tag{3.17}$$

There are some cases in which a researcher may wish to avoid making assumptions that justify using the representation of S_f given in (3.14). For instance, consider situations in which conditional expectations and best linear predictors do not coincide. A researcher may wish to assume that the projection eqs. (2.26) and (2.27) define the linear least squares projections of y_t and x_{t+1} onto Φ_t, but avoid making the claim that these equations define the conditional expectations of y_t and x_{t+1} given current and past x's. In such situations unless fourth-order cumulants are zero, the representation of S_f given in (3.14) is inappropriate and representation (3.13) should be used.

Obviously, S_f is not a matrix that the researcher can specify correctly *a priori*. In order to obtain an estimator that is optimal in the sense described above, it is only required that S_f be estimated consistently. This can be accomplished by using an initial consistent estimator $\zeta_{T,1}$, forming the sample values $f_t(\zeta_{T,1})$ or $\lambda(L;\zeta_{T,1})z_t$, and then estimating S_f using a procedure appropriate for estimating spectral density matrices consistently employing either formula (3.13) or formula (3.14). With this estimator of S_f, which we denote S_f^T, for the distance matrix, we can obtain an optimal estimator of ζ_0 by minimizing $J(S_f^T,\zeta)$ by choice of ζ. We denote this minimizer by $\zeta_{T,2}$. A consistent estimator of the asymptotic covariance matrix for $\zeta_{T,2}$ is

$$[D_T'(S_f^T)^{-1}D_T]^{-1}, \tag{3.18}$$

where

$$D_T=\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t(\zeta_{T,2})}{\partial\zeta}. \tag{3.19}$$

The initial consistent estimator $\zeta_{T,1}$ can be obtained by minimizing $J(S,\cdot)$ using a non-optimal choice for S, e.g., the identity matrix.

This estimation procedure uses R orthogonality conditions to estimate Q parameters. For most applications R is greater than Q. Estimation of the Q parameters in essence sets Q linear combinations of the sample orthogonality conditions to zero, via the first-order conditions for (3.10). This leaves $R-Q$ independent linear combinations of the orthogonality conditions that are not set to zero in estimation but that should be 'close to zero' if the restrictions implied by the model are true. This provides us with a scheme for testing these restrictions. Hansen (1982) shows that $TJ(S_f^T, \zeta_{T,2})$ is asymptotically distributed as a chi square with $R-Q$ degrees of freedom under the null hypothesis that the restrictions are true. Since $J(S_f^T, \zeta_{T,2})$ is the minimized value of the criterion function $J(S_f^T, \cdot)$ for the second step optimal estimator, this test statistic can be computed easily.

The estimation scheme described above does exploit the serial correlation properties of the disturbances to construct an optimal estimator; however, a researcher is free to adopt a relatively general specification of the temporal covariances of these disturbances. This is an important advantage of this procedure over maximum likelihood procedures. Maximum likelihood requires a more precise specification of the temporal covariance structure of the instrumental variables and disturbances. There is an additional computational advantage in that one can estimate the parameters β, ρ, θ, and γ by numerically searching over a smaller parameter space using this instrumental variables procedure than is required by maximum likelihood procedures.

3.2. Estimators using all available orthogonality conditions

In constructing the GMM estimators described above, a fixed number R of orthogonality conditions was employed independent of sample size. On the other hand, there is an infinite number of orthogonality conditions available to use in estimation, as is indicated by (2.28), (2.29) and (2.30). While the GMM estimators described above use the R orthogonality conditions optimally, the only justification for restricting attention to these orthogonality conditions is computational simplicity. Typically by adding additional orthogonality conditions to the list used in estimation, it is possible to construct an estimator with a smaller asymptotic covariance matrix. For this reason we now discuss how to use all available orthogonality conditions 'optimally'. This discussion takes place under the special assumptions that were used to justify representation (3.14) of S_f.

Let W_j be an $(n \times p)$ matrix lag operator for $j=1,2,\dots,Q$, where $n=p+2$. We impose the restriction that W_j be one-sided for each j, that is,

$$W_j(L) = W_j^0 + W_j^1 L + \dots. \tag{3.20}$$

Furthermore, assume that the elements of $\{W_j^k\}_{k=0}^{\infty}$ are square summable for each j. We can think of estimating ζ_0 from the Q orthogonality conditions

$$Ed_t'[W_j(L)x_t] = 0 \tag{3.21}$$

for $j=1,2,\dots,Q$. These orthogonality conditions are implied by the orthogonality conditions (2.28), (2.29) and (2.30). We can write

$$d_t'[W_j(L)x_t] = d_{1t}[W_{j1}(L)x_t] + d_{2t}[W_{j2}(L)x_t] + \ldots + d_{nt}[W_{jn}(L)x_t], \tag{3.22}$$

where d_{kt} is the kth element of d_t and W_{jk} is the kth row of W_j. Let us stack the 'weighting' lag operators W_{jk} into a matrix W as follows:

$$W(L) = [W_{jk}(L)], \tag{3.23}$$

i.e., $W(L)$ is a partitioned matrix polynomial in the lag operator with $W_{jk}(L)$ in the jth row and kth column partition. Note that $W(L)$ is $(Q \times N)$ where $N = pn$.

In order to estimate ζ_0 using W, we can think of finding ζ_T in some admissible parameter space that satisfies the non-linear equations

$$\frac{1}{T}\sum_{t=1}^{T} [\lambda(L;\zeta_T)z_t]' W_j(L)x_t = 0 \tag{3.24}$$

for $j = 1, 2, \ldots, Q$. Recall that $d_t = \lambda(L;\zeta_0)z_t$ so that (3.24) is just a sample version of (3.21). Practically speaking, there are difficulties in implementing this strategy. The lag polynomial W is allowed to be an infinite order polynomial and thus (3.24) may involve observations that are not available. Criterion function (3.24) can be approximated by letting x be equal to zero for all time periods in which observations are not available. It turns out that this has a negligible impact on the asymptotic distribution of the estimator. Hansen and Singleton (1982) establish consistency and asymptotic normality of estimators of the form specified in (3.24) with relatively arbitrary choices for W. As is true for the finite orthogonality condition case, the asymptotic covariance matrix of the estimator is dependent on the choice of the weighting operator W. Our purpose here is to describe the optimal choice of W and suggest ways to construct optimal estimators in practice.

Before representing an optimal estimator, we provide an expression for the asymptotic covariance matrix of an estimator that uses a relatively arbitrary choice of W. Let $D[W]$ be the $(Q \times Q)$ matrix given by

$$D[W] = E\begin{bmatrix} [W_1(L)x_t]' \dfrac{\partial \lambda(L;\zeta_0)}{\partial \zeta} z_t \\ [W_2(L)x_t]' \dfrac{\partial \lambda(L;\zeta_0)}{\partial \zeta} z_t \\ \vdots \\ [W_Q(L)x_t]' \dfrac{\partial \lambda(L;\zeta_0)}{\partial \zeta} z_t \end{bmatrix}. \tag{3.25}$$

We restrict our attention to choices of W for which $D[W]$ is non-singular.[20] Assume that S_d can be represented as

$$S_d(-\omega)=\kappa(e^{i\omega})\kappa(e^{i\omega})', \tag{3.26}$$

where

$$\kappa(e^{i\omega})=\sum_{j=0}^{\infty}\kappa_j e^{i\omega j}, \tag{3.27}$$

$$\sum_{j=0}^{\infty}[\operatorname{trace}(\kappa_j\kappa_j')]^{\frac{1}{2}}<+\infty,$$

$$\det\kappa(z)\neq 0, \qquad |z|\leqq 1.$$

We let

$$S(W)=\frac{1}{2\pi}\int_{-\pi}^{\pi}W(e^{i\omega})[S_d(-\omega)\otimes S_x(\omega)]W(e^{i\omega})'\,d\omega \tag{3.28}$$

Hansen and Singleton demonstrate that the asymptotic covariance matrix of an estimator using weighting scheme W is given by

$$D(W)^{-1}S(W)D(W)^{-1\prime}. \tag{3.29}$$

In order to obtain an optimal choice of W, we can minimize (3.29) by choice of a one-sided matrix lag operator W. Hansen and Singleton solve this optimization problem and obtains an explicit characterization of the solution. The one-sided constraint on W is imposed because we do not assume that the instruments are strictly exogenous. This constraint is, in general, binding.

To obtain an explicit solution to this optimization problem, let

$$E\left[\frac{\partial\lambda(L;\zeta_0)}{\partial\zeta}z_t\,\middle|\,\Phi_t\right]=[B_1(L)x_t B_2(L)x_t\ \dots\ B_Q(L)x_t]. \tag{3.30}$$

The optimal weighting lag operator W^* is given by

$$W_j^*(L)=\kappa(L)^{-1\prime}[\kappa(L^{-1})^{-1}B_j(L)\gamma(L)^{-1}]_+\gamma(L) \tag{3.31}$$

for $j=1,\dots,Q$. The optimal weighting scheme is dependent on the serial correlation properties of the disturbance via κ, the serial correlation properties of the instruments via γ, and the projection given in (3.30) via B_j.

[20] Alternatively, we could avoid imposing this restriction, and whenever $D[W]$ is singular, interpret the asymptotic covariance matrix of the estimator as infinite.

The asymptotic covariance matrix of this optimal estimator is provided below. Let

$$H_j(L)=[\kappa(L^{-1})^{-1}B_j(L)\gamma(L^{-1})]_+=\begin{bmatrix} H_{j1}(L) \\ H_{j2}(L) \\ \vdots \\ H_{jn}(L) \end{bmatrix} \tag{3.32}$$

for $j=1,2,\ldots,Q$, where $H_{jk}(L)$ is $(1\times p)$. Let

$$H(L)=[H_{jk}(L)]. \tag{3.33}$$

The asymptotic covariance matrix of an optimal estimator is given by

$$\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} H(\mathrm{e}^{i\omega})[I\otimes V]H(\mathrm{e}^{i\omega})'\,\mathrm{d}\omega\right]^{-1}=D(W^*)^{-1}=S(W^*)^{-1}, \tag{3.34}$$

where $V=E[e_te_t']$.

In order to construct an optimal estimator, it is necessary to have consistent estimators of W^*, or equivalently consistent estimators of κ, γ and B_j for $j=1,2,\ldots,Q$.[21] To accomplish this, the fixed, finite orthogonality condition GMM procedure discussed earlier can be used to obtain a consistent estimator $\zeta_{1,T}$. Using this estimator, estimated disturbances

$$d_{t,T}=\lambda(L;\zeta_{T,1})z_t \tag{3.35}$$

can be formed for $t=1,2,\ldots,T$. These disturbances can in turn be used to estimate the parameters of κ. One practical strategy is to estimate κ by running a finite order forward vector autoregression. A key point is that these estimators are constructed so that misspecification of the serial correlation properties of d does not damage consistency but only the optimality of the resulting estimator. To achieve optimality, the choice of order of this vector autoregression should be an explicit function of sample size in cases in which d is allowed to have an infinite order vector autoregression representation. Hansen and Singleton (1982) discuss this issue.

A consistent estimator of γ is embedded in $\zeta_{1,T}$ since a subset of the parameters of ζ_0 are the parameters of γ. Estimates of B_j for $j=1,\ldots,Q$ can be obtained from $\zeta_{1,T}$ and estimates of ξ where

[21]In general, κ and B_j for $j=1,2,\ldots,Q$ are infinite order lag polynomials. The sense in which the coefficients of these lag polynomials have to be consistently estimated is discussed in Hansen and Singleton (1982).

$$E[A_t \mid \Phi_t] = \xi(L)x_t. \tag{3.36}$$

One possibility is to estimate ξ with a finite lag approximation using ordinary least squares. The lag length could be treated as fixed *a priori* or as an explicit function of sample size. A second strategy is to estimate

$$A_t = \xi(L)x_t + v_t \tag{3.37}$$

jointly with (2.25), (2.26), and (2.27).[22] Eq. (3.37) has the associated orthogonality condition

$$E[v_t x_{t-j}] = 0, \tag{3.38}$$

$j \geq 0$. The computational advantage of using the procedures we propose over maximum likelihood is that the parameters of κ are not estimated simultaneously with the rest of the parameters of the model.[23] Our procedures can avoid estimating κ altogether or in cases in which asymptotic optimality is desired, initial consistent estimators of the parameters of κ are employed.

Before concluding this section, it is useful to compare the estimators we are proposing with estimators suggested by Hayashi and Sims (1982). It turns out that this comparison will provide us with a useful interpretation of (3.31). Hayashi and Sims suggest that instrumental variables estimators be constructed by first filtering the disturbance term forward to remove serial correlation. In other words, apply $\kappa(L^{-1})^{-1}$ to d to obtain

$$w_t = \kappa(L^{-1})^{-1} d_t, \tag{3.39}$$

where w_t is a white noise and is orthogonal to all future d's. Now we can think of estimating ζ_0 using orthogonality conditions of the form

$$E[w_t \otimes x_{t-j}] = 0 \tag{3.40}$$

[22] If $\xi(L)_2 = \xi_0 + \xi_1 L + \ldots + \xi_r L^r$, if eqs. (3.38), (2.25), (2.26), and (2.27) are estimated jointly, and if the disturbances are specified as an mth order vector autoregressive process, the optimal finite orthogonality condition GMM estimator described in section 3.1 is optimal in the broader infinite orthogonality sense when P is chosen to equal $r+m$. When a version of budget constraint (2.4) is added to the equations used to derive the restrictions, it is no longer appropriate to assume that $\xi(L)$ is a finite order polynomial (see footnote 18). In such circumstances, presample x's can be set zero and estimation can proceed even though $\xi(L)$ is infinite order [see Hansen and Singleton (1982)].

[23] The fact that e is a white noise imposes restrictions on the spectral density of d. For computational purposes, it is convenient to ignore these restrictions in estimating κ.

for $j \geq 0$. Since w is obtained from d using a forward filter, orthogonality condition (3.40) is implied by orthogonality conditions (2.28), (2.29) and (2.30). Hayashi and Sims discuss estimation of models that are linear in parameters and variables using a finite number of the orthogonality conditions like those in (3.40). They compare these forward filtered estimators to ones which employ a fixed finite number of orthogonality conditions without forward filtering and illustrate some advantages of forward. filtering. They also investigate the limiting behavior of the asymptotic covariance of both estimators as the number of orthogonality conditions employed gets large.[24]

Let us now consider this form of the optimal forward filtered estimator using all of the orthogonality conditions. In particular, we consider estimators that use orthogonality conditions of the form

$$E[w_t'[C_j(L)x_t]]=0 \tag{3.41}$$

for $j=1,2,\ldots,Q$ and wish to choose $C_1, C_2,\ldots, C_Q$ optimally. Since the w's are linear combinations of the current and future d's, the optimal forward filtered estimator has asymptotic covariance matrix (3.34). In fact from (3.30), it is evident that an optimal choice of C's is

$$C_j^*(L)=[\kappa(L^{-1})^{-1}B_j(L)\gamma(L)^{-1}]_+\gamma(L). \tag{3.42}$$

Using this optimal choice of C's will result in an estimator that is asymptotically equivalent to an estimator constructed using the W_j^*'s given in (3.31). Eq. (3.42) turns out to provide the solution to an optimal prediction problem. Using the Wiener–Kolmogorov prediction formula we can verify that

$$E\left[\kappa(L^{-1})^{-1}\frac{\partial\lambda(L;\zeta_0)z_t}{\partial\zeta_j}\,\middle|\,\Phi_t\right]=C_j(L)x_t. \tag{3.43}$$

Thus, if we first filter the eqs. (2.28), (2.29) and (2.30) forward to remove serial correlation and then project the partial derivatives onto the set of instrumental variables Φ_t, we can obtain an optimal set of instruments to use in estimation. This result is consistent with more conventional instrumental

[24] Hayashi and Sims (1982) provide the interpretation given below of the optimal weighting scheme, but they do not explicitly characterize it. Rather than discussing how to construct optimal extimators, Hayashi and Sims illustrate that by driving the number of orthogonality conditions to infinity, the asymptotic covariance matrices of the estimators approach a limiting covariance matrix like (3.35).

variables estimators.[25] The application of the forward filter $\kappa(L^{-1})^{-1}$, however, requires the solution of a non-trivial prediction problem (3.43) to obtain an optimal estimator which is essentially the same computation as is used in calculating W^*.

4. A comparison to estimators constructed from Euler equations

In this section we examine three alternative instrumental variables methods for estimating the parameters of dynamic quadratic objective functions of economic agents. The first method is one proposed by Kennan (1979) and Hayashi (1980) that estimates the parameters directly from the Euler equations implied by the optimization problems of economic agents. The second method is one proposed by Hansen and Sargent (1980) that solves the Euler equations, exploits the symmetry between the feedforward and feedback portions of this solution, and imposes restrictions across the feedforward portion of the solution and the stochastic specification of the observable forcing variables. It turns out that this second method ignores some restrictions across the feedback part of the solution and the stochastic specification of the observable forcing variables. For this reason we consider a third method that imposes all of these restrictions. While the first method is computationally simple and requires that less be said about the economic environment *a priori*, it also ignores restrictions and consequently results in parameter estimators that are asymptotically less efficient than the estimators that emerge from the second and third methods.

The proposals made in section 3 about estimating the parameters of the consumption model can be modified in a straightforward way to accommodate any of the three methods. For this reason, we will not say very much about estimation here, but instead we will describe the restrictions used by each of the methods. To accomplish this, it is convenient to shift from the consumption function example used in sections one and two to the factor demand example mentioned in the introduction.

Following Hansen and Sargent (1980) we assume that a competitive firm employing a single factor of production chooses a contingency plan for the factor to maximize its expected present value

$$E\left\{\sum_{t=0}^{\infty} \beta^t[(a_t-y_t)n_t-(\varepsilon/2)n_t^2-(\delta/2)(n_t-n_{t-1})^2]\,|\,\Omega_0\right\}, \tag{4.1}$$

subject to n_{-1} given, where n_t is employment of the factor at time t, y_t is the

[25]See Amemiya (1977). This link is pointed out by Hayashi and Sims (1982) in the context of models that are linear in parameters and variables.

real factor rental at t, and a_t is the time t technology shock observed by the firm but not by the econometrician.[26] Here ε and δ are positive parameters. As in section 2, assume that there is a $(p \times 1)$ vector x_t that is included in agents' time period t information set Ω_t and that satisfies

$$E[a_t x_{t-j}] = 0, \tag{4.2}$$

$j \geqq 0$. Again, a has mean zero but can be serially correlated. The random variable a_t can be correlated with y_t and future x's and still satisfy (4.2).

The stochastic Euler equation for optimization problem (4.1) is

$$En_{t+1} \mid \Omega_t + \sigma_1 n_t + \sigma_2 n_{t-1} = \sigma_3 y_t - \sigma_3 a_t, \tag{4.3}$$

where

$$\sigma_1 = -[(\varepsilon/\delta) + 1 + \beta]/\beta,$$

$$\sigma_2 = 1/\beta,$$

$$\sigma_3 = 1/\delta\beta.^{27} \tag{4.4}$$

Following a suggestion of McCallum (1976), we can add $n_{t+1} - En_{t+1} \mid \Omega_t$ to both sides of (4.3) to obtain

$$n_{t+1} + \sigma_1 n_t + \sigma_2 n_{t-1} = \sigma_3 y_t - \sigma_3 a_t + n_{t+1} - En_{t+1} \mid \Omega_t \tag{4.5}$$

Associated with (4.5) is the orthogonality condition

$$E[(-\sigma_3 a_t + n_{t+1} - En_{t+1} \mid \Omega_t) x_{t-j}] = 0, \tag{4.6}$$

$j \geqq 0$. Orthogonality condition (4.6) is implied by condition (4.2) and the assumption that x_t is an element of Ω_t.

The Euler equation approach suggested by Kennan (1979) and Hayashi (1980) applied to this example entails constructing estimators of σ_1, σ_2 and σ_3 from the orthogonality condition (4.6). Estimators of δ, ε, and β can then be obtained from the estimators of σ_1, σ_2, and σ_3 by using the three equations in (4.4). An advantage of this procedure is that closed form expressions can be obtained for the estimators of δ, ε, and β, and that numerical search procedures are not required to calculate the parameter

[26]We suppressed the linear term in the objective function since we are assuming that the random variables all have mean zero.

[27]See Sargent (1979) and Hansen and Sargent (1980).

estimates. Although this method does not require that the projections of n_{t+1}, n_t, n_{t-1}, and y_t onto the reduced information set $\Phi_t=\{x_t, x_{t-1},\ldots\}$ be parameterized, it does implicitly assume that such projections are time invariant.[28] The alternative two methods parameterize these projections and obtain further restrictions across them. Notice that the Kennan–Hayashi estimator based on (4.6) ignores the transversality condition, which is among the first-order necessary conditions for the optimum problem. The alternative two methods incorporate the restrictions implied by the transversality condition.

We proceed to characterize these alternative methods of estimation. As in section 2, we assume that

$$y_t=\theta(L)x_t+u_t, \tag{4.7}$$

$$x_{t+1}=\gamma^1(L)x_t+e_{t+1}, \tag{4.8}$$

where

$$Eu_t x_{t-j}=0, \tag{4.9}$$

$$Ee_{t+1}x_{t-j}=0 \tag{4.10}$$

for $j\geqq 0$. Following Hansen and Sargent (1980), we solve the Euler equation, subject to the transversality condition, to obtain

$$n_t=\lambda n_{t-1}-(\lambda/\delta)\sum_{j=0}^{\infty}(\lambda\beta)^j E(y_{t+j}-a_{t+j})\,|\,\Omega_t, \tag{4.11}$$

where

$$\lambda=\frac{-\sigma_1-\sqrt{\sigma_1^2-4\sigma_2}}{2\sigma_1}. \tag{4.12}$$

Notice that λ is less than one and that decision rule (4.11) possesses a symmetry property since λ is the feedback coefficient and enters into the feedback geometric sum. Using a strategy analogous to that employed in section 2, we rewrite (4.11) as

$$n_t=\lambda n_{t-1}-(\lambda/\delta)\sum_{j=0}^{\infty}(\lambda\beta)^j Ey_{t+j}\,|\,\Phi_t+s_t+a_t^*, \tag{4.13}$$

[28] In light of White's (1982) work on instrumental variables estimators in cross-sectional analysis, this is a bit of an overstatement.

where

$$s_t=(\lambda/\delta)\sum_{j=0}^{\infty}(\lambda\beta)^j(Ey_{t+j}\,|\,\Phi_t-Ey_{t+j}\,|\,\Omega_t),$$
$$a_t^*=(\lambda/\delta)\sum_{j=0}^{\infty}(\lambda\beta)^jEa_{t+j}\,|\,\Omega_t. \tag{4.14}$$

Solving the prediction problem in (4.13) we see that

$$n_t=\lambda n_{t-1}+\pi(L)x_t+s_t+a_t^*, \tag{4.15}$$

where

$$\pi(L)=-(\lambda/\delta)\left[\frac{\theta(L)-\lambda\beta L^{-1}\theta(\lambda\beta)\gamma(\lambda\beta)^{-1}\gamma(L)}{1-\lambda\beta L^{-1}}\right],$$

$$\gamma(L)=I-L\gamma^1(L).^{29} \tag{4.16}$$

Eq. (4.16) summarizes the restrictions across the feedforward part of the decision rule and the law of motion for x. Using the definitions of s_t and a_t^* in (4.14) and an iterated projection argument we obtain the orthogonality condition

$$E[(s_t+a_t^*)x_{t-j}]=0 \tag{4.17}$$

for $j\geqq 0$. Thus $\pi(L)x_t$ is the projection of $n_t-\lambda n_{t-1}$ onto the reduced information set Φ_t. Modifying a strategy proposed in Hansen and Sargent (1980), one can construct estimators of the underlying parameters δ, ε, and β together with the parameters of θ and γ^1 from the orthogonality conditions (4.9), (4.10), and (4.11).[30] These estimators do not have closed form representations, and numerical search procedures are needed to compute them.

Since the second method is computationally more difficult than the first method, it is important to ascertain whether additional restrictions are exploited by the second method. To answer this question, observe from (4.5), (4.6), and (4.7) that the first method exploits the restrictions implied by

$$E[(L^{-1}+\sigma_1+\sigma_2L)n_t\,|\,\Phi_t]=\sigma_3\theta(L)x_t. \tag{4.18}$$

The operator $(L^{-1}+\sigma_1+\sigma_2L)$ can be factored to obtain

$$(L^{-1}+\sigma_1+\sigma_2L)=(L^{-1}-(1/\lambda\beta))(1-\lambda L), \tag{4.19}$$

[29]This can be established using the lemma in Appendix A of Hansen and Sargent (1980).

[30]Hansen and Sargent (1980) assume that y_t is in Φ_t.

where $\lambda < 1$. Thus, relation (4.18) can be written

$$E[(1-\lambda L)n_{t+1} \mid \Phi_t] - (1/\lambda\beta)E[(1-\lambda L)n_t \mid \Phi_t] = \sigma_3\theta(L)x_t. \tag{4.20}$$

In other words, one can interpret (4.18) as a set of restrictions across the projections of $(1-\lambda L)n_{t+1}$ and $(1-\lambda L)n_t$ onto Φ_t. That is, if we let

$$E[(1-\lambda L)n_t \mid \Phi_t] = \pi(L)x_t, \tag{4.21}$$

$$E[(1-\lambda L)n_{t+1} \mid \Phi_t] = \bar{\pi}(L)x_t, \tag{4.22}$$

then (4.18) implies that

$$\bar{\pi}(L) - (1/\lambda\beta)\pi(L) = \sigma_3\theta(L). \tag{4.23}$$

Now it turns out that given the projections of $(1-\lambda L)n_t$ and x_{t+1} onto Φ_t it is possible to compute the projection of $(1-\lambda L)n_{t+1}$ onto Φ_t. To see this note that

$$\begin{aligned} E[(1-\lambda L)n_{t+1} \mid \Phi_{t+1}] &= \pi(L)x_{t+1} \\ &= \pi_0 x_{t+1} + \pi^1(L)x_t. \end{aligned} \tag{4.24}$$

Projecting onto Φ_t we see that

$$E[(1-\lambda L)n_{t+1} \mid \Phi_t] = [\pi_0\gamma^1(L) + \pi^1(L)]x_t, \tag{4.25}$$

and hence

$$\bar{\pi}(L) = \pi_0\gamma^1(L) + \pi^1(L). \tag{4.26}$$

While the first method exploits only restrictions (4.18), it can be verified that the specification of $\pi(L)$ in (4.16) satisfies both (4.18) and (4.26). Thus the second method does indeed impose more restrictions than the first method.

This raises the question of whether there are any additional restrictions that can be exploited in estimation. It turns out that there are. To see this notice that the second method works with the projection of the quasi-differenced form $(1-\lambda L)n_t$ onto Φ_t but does not exploit the link between the projections of n_t and n_{t-1} onto Φ_t. In particular, let

$$En_{t-1} \mid \Phi_t = \eta(L)x_t, \tag{4.27}$$

$$En_t \mid \Phi_t = \bar{\eta}(L)x_t.$$

The second strategy uses the fact that

$$\tilde{\eta}(L)-\lambda\eta(L)=\pi(L), \tag{4.28}$$

where $\pi(L)$ satisfies (4.16). Following the same logic as above, there exist additional restrictions that link $\tilde{\eta}(L)$ to $\eta(L)$ and $\gamma^1(L)$. More precisely, note that

$$\begin{aligned}\tilde{\eta}(L)x_t &= E[En_t \mid \Phi_{t+1} \mid \Phi_t] \\ &= E[\eta(L)x_{t+1} \mid \phi_t] \\ &= E[\eta_0 x_{t+1} + \eta^1(L)x_t \mid \Phi_t] \\ &= [\eta_0\gamma^1(L)+\eta^1(L)]x_t.\end{aligned} \tag{4.29}$$

Thus,

$$\tilde{\eta}(L)=\eta_0\gamma^1(L)+\eta^1(L). \tag{4.30}$$

Combining (4.28) and (4.30), we see that

$$\eta_0\gamma^1(L)+\eta^1(L)-\lambda\eta_0-\lambda L\eta^1(L)=\pi(L), \tag{4.31}$$

where $\pi(L)$ is given in (4.16). Solving for the operator $\eta^1(L)$, it follows that

$$\eta^1(L)=\frac{\pi(L)}{(1-\lambda L)}+\frac{\lambda\eta_0}{1-\lambda L}-\frac{\eta_0\gamma^1(L)}{1-\lambda L}. \tag{4.32}$$

Therefore

$$\begin{aligned}\eta(L) &= \eta_0 + L\eta^1(L) \\ &= \frac{L\pi(L)}{(1-\lambda L)}+\frac{\eta_0\gamma(L)}{(1-\lambda L)}.\end{aligned} \tag{4.33}$$

An estimation method that imposes more restrictions than either of the two procedures mentioned previously is to estimate the projection equations

$$n_{t-1}=\eta(L)x_t+v_t, \tag{4.34}$$

where

$$Ev_tx_{t-j}=0$$

for $j \geq 0$, (4.7), and (4.8) jointly subject to restrictions (4.16) and (4.33). The parameters to be estimated under this strategy are β, δ, ε, η_0, and the parameters of $\theta(L)$ and $\gamma^1(L)$.[31] Projection (4.34) accommodates the possibility that the projection of n_t onto current, past and future x's is two-sided. If η_0 is not zero, it follows from a theorem in Sims (1972) that the observable forcing variables x_t are not strictly exogenous in a regression of n_t onto current and past x's. Consistent with our previous proposals, this procedure uses the x's as instruments but does not assume that the x's are exogenous.

In comparing the three methods, we conclude that the Euler equation approach to estimating dynamic linear rational expectations models is computationally simpler and requires that less be specified *a priori*. On the other hand, it ignores restrictions and yields estimators that are asymptotically less efficient than estimators that exploit restrictions across the decision rule parameters and the parameters of the stochastic process assumed to generate the observable forcing variables. It is important to realize that even though the Euler equation approach does not require an explicit stochastic specification of the observable forcing variables, this does not mean the resulting instrumental variables estimators will be more robust against alterations in policy regimes that occur during the sample period. As noted previously, the Euler equation approach implicitly assumes that the projections of the variables onto the instruments have time invariant representations.[32]

5. Conclusion

In building rational expectations econometric models, a researcher is often confronted with an estimation environment in which disturbance terms are serially correlated and instruments are not strictly exogenous. This paper proposes a class of estimation procedures that are appropriate in this environment. In this paper we have shown how to construct estimators from an underlying set of orthogonality conditions implied by the econometric model. A whole class of consistent and asymptotically normal estimators has been described. A researcher can take into account the tradeoff between computational simplicity and the size of the asymptotic covariance matrix of the resulting estimators in deciding which of these procedures to employ. We

[31] The projection of n_{t-1} onto Φ_t is constrained to have a denominator term $(1-\lambda L)$. Thus hypothetical disturbances depend on infinitely many past x's. As noted in footnote 22, an argument in Hansen and Singleton (1982) shows that pre-sample period x's can be set to zero.

[32] As pointed out in footnote 28, this is a bit of an overstatement. However, it is not clear that what people have in mind when considering alterations in policy regimes is accommodated by the types of deviations from stationarity that White allows in his framework.

have also shown how to construct tests of the restrictions implied by the econometric model using these instrumental variables procedures. Although our econometric discussion took place mainly in the context of a rational expectations, permanent income consumption function model, the estimators we propose are applicable to many other examples of linear rational expectations models.

References

Amemiya, T., 1974, The nonlinear two-stage least-squares estimator, Journal of Econometrics 2, 105–110.

Amemiya, T., 1977, The maximum likelihood and nonlinear three-stage least squares estimator in the general nonlinear simultaneous equations model, Econometrica 45, 955–968.

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

Flavin, M., 1981, Adjustment of consumption to changing expectations about future income, Journal of Political Economy 89, 974–1009.

Hall, R.E., 1978, Stochastic implications of the life cycle–permanent income hypothesis: Theory and evidence, Journal of Political Economy 86, 971–987.

Hansen, L.P., 1982, Large sample properties of generalized methods of moments estimators, Econometrica, forthcoming.

Hansen, L.P. and T.J. Sargent, 1980, Formulating and estimating dynamic linear rational expectations models, Journal of Economic Dynamics and Control 2, 7–46.

Hansen, L.P. and T.J. Sargent, 1981, Linear rational expectations models for dynamically interrelated variables, in: R.E. Lucas, Jr. and T.J. Sargent, eds., Rational expectations and econometric practice (University of Minnesota Press, Minneapolis, MN).

Hansen, L.P. and K.J. Singleton, 1982, Optimal instrumental variables estimators for nonlinear time series models, Manuscript.

Hayashi, F., 1980. Estimation of macroeconomic models under rational expectations: A survey, Manuscript.

Hayashi, F. and C.A. Sims, 1982, Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments, Econometrica, forthcoming.

Huntzinger, R.L., 1979, Market analysis with rational expectations, Journal of Political Economy 10, 127–145.

Jorgensen, D.W. and J. Laffont, 1974, Efficient estimation of nonlinear simultaneous equations with additive disturbances, Annals of Economic and Social Measurement 3, 615–640.

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

Kydland, F.E. and E.C. Prescott, 1977, Rules rather than discretion: The inconsistency of optimal plans, Journal of Political Economy 88, 473–491.

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 Conference on Public Policy 1 (North-Holland, Amsterdam).

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

Rozanov, R.A., 1967, Stationary random processes (Holden-Day, San Francisco, CA).

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

Sargent, T.J., 1978, Rational expectations, econometric exogeneity and consumption, Journal of Political Economy 86, 67–700.

Sargent, T.J., 1979, Macroeconomic theory (Academic Press, New York).

Sargent, T.J., 1980, Lecture notes on linear filtering and control, Manuscript.

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.
White, H., 1982, Instrumental variables regression on cross-section data, Econometrica, forthcoming.
Whittle, P., 1963, Regulation and prediction by linear least-square methods (D. Van Nostrand, Princeton, NJ).