Econometrica, Vol. 50, No. 5 (September, 1982)

GENERALIZED INSTRUMENTAL VARIABLES ESTIMATION OF NONLINEAR RATIONAL EXPECTATIONS MODELS[1]

BY LARS PETER HANSEN AND KENNETH J. SINGLETON

This paper describes a method for estimating and testing nonlinear rational expectations models directly from stochastic Euler equations. The estimation procedure makes sample counterparts to the population orthogonality conditions implied by the economic model close to zero. An attractive feature of this method is that the parameters of the dynamic objective functions of economic agents can be estimated without explicitly solving for the stochastic equilibrium.

1. INTRODUCTION

THE ECONOMETRIC IMPLICATIONS of dynamic rational expectations models in which economic agents are assumed to solve quadratic optimization problems, subject to linear constraints, have been analyzed extensively in [**11**, **12**, **25**, **29**, and **30**]. Linear-quadratic models lead to restrictions on systems of constant coefficient linear difference equations, which provide complete characterizations of the equilibrium time paths of the variables being studied. Hence, the parameters of these models can be estimated using the rich body of time series econometric tools developed for the estimation of restricted vector difference equations [e.g., **11**, **20**, and **32**]. Once the linear-quadratic framework is abandoned in favor of alternative nonquadratic objective functions, dynamic rational expectations models typically do not yield representations for the variables that are as convenient from the standpoint of econometric analysis. Indeed, in many models, closed-form solutions for the equilibrium time paths of the variables of interest have been obtained only after imposing strong assumptions on the stochastic properties of the "forcing variables," the nature of preferences, or the production technology. See, for example, the models in Merton [**24**], Brock [**4**], and Cox, Ingersoll, and Ross [**5**].

The purpose of this paper is to propose and implement an econometric estimation strategy that circumvents the theoretical requirement of an explicit representation of the stochastic equilibrium, yet permits identification and estimation of parameters of economic agents' dynamic (nonquadratic) objective functions, as well as tests of the over-identifying restrictions implied by the theoretical model. The procedures we propose do not require a complete, explicit representation of the economic environment and, in particular, do not require strong *a priori* assumptions about the nature of the forcing variables. Consequently, estimation and inference can be conducted when only a subset of the economic environment is specified *a priori*. While our strategy involves specifying the objective functions of a subset of agents, it is distinct from specifying the

[1]This research was supported in part by NSF Grant SES-8007016. We wish to thank Robert Engle, Robert King, Scott Richard, and two anonymous referees for helpful comments on earlier drafts of this paper, and Ravi Jagannathan for research assistance.

decision rules (e.g., dynamic demand or supply schedules) of a subset of agents without specifying the entire economic environment.[2] The latter approach is characteristic of many applications of limited information methods to conventional simultaneous equations models. Lucas [**21**] has criticized this approach by noting that, under the assumption of rational expectations, the dynamic decision rules of economic agents depend explicitly on the stochastic specification of the forcing variables and possibly the structural specification of the entire economic environment.

The basic idea underlying our estimation strategy is as follows. The dynamic optimization problems of economic agents typically imply a set of stochastic Euler equations that must be satisfied in equilibrium. These Euler equations in turn imply a set of population orthogonality conditions that depend in a nonlinear way on variables observed by an econometrician and on unknown parameters characterizing preferences, profit functions, etc. We construct nonlinear instrumental variables estimators for these parameters in the manner suggested by Amemiya [**1, 2**], Jorgenson and Laffont [**18**], and Hansen [**10**] by making sample versions of the orthogonality conditions close to zero according to a certain metric. An important feature of these estimators is that they are consistent and have a limiting normal distribution under fairly weak assumptions about the stochastic processes generating the observable time series. Also, more orthogonality conditions are typically available for use in estimation than there are parameters to be estimated and, in this sense, the models are "overidentified." The overidentifying restrictions can be tested using a procedure, justified in Hansen [**10**], that examines how close sample versions of population orthogonality conditions are to zero.

Other authors have proposed using stochastic Euler equations to estimate parameters (but not test restrictions) in the context of models containing linear-quadratic optimization problems (e.g., Hayashi [**16**] and Kennan [**19**]). By focusing on the Euler equations in these linear models, some of the restrictions implied by the model are ignored at the gain of computational simplicity (see Hansen and Sargent [**13**]). In addition to computational simplicity, there is perhaps a more compelling reason for using instrumental variables procedures in the nonlinear environments considered here. Namely, there is the added difficulty of obtaining a complete characterization of the stochastic equilibrium under weak assumptions about the forcing variables.[3] Fair and Taylor [**7**] proposed an alternative, approximate maximum likelihood procedure that can be applied to

[2] By the economic environment we mean a specification of preferences, technology, and the stochastic process underlying the forcing variables. By a decision rule we mean a rule used by economic agents to determine the current period "decision" as a function of the current "state" of the economy.

[3] The procedures discussed in this paper are also of use in quadratic optimization environments when it is important to allow conditional variances to be dependent on variables in the information set. Allowing for these dependencies complicates decision rule derivation and does not permit use of conventional asymptotic distribution theory results for instrumental variables estimators. See Section 3 for further discussion of these issues.

an equation system that includes a set of stochastic Euler equations. By imposing certainty equivalence on the nonlinear rational expectations model, they circumvent some of the difficulties in obtaining a complete characterization of the stochastic equilibrium. The instrumental variables procedure proposed here avoids their approximations for an important class of models.

The remainder of this paper is organized as follows. In Section 2 the class of stochastic Euler equations that can be used in estimation is discussed, and then an example from the literature on multi-period asset pricing is presented. In Section 3 the generalized instrumental variables estimator is formulated, and its large sample properties are discussed. In Section 4 this estimator is compared to the maximum likelihood estimator in the context of a nonlinear model of stock market returns. This discussion contrasts the orthogonality conditions exploited by the instrumental variables estimator with those exploited by the maximum likelihood estimator when a specific distributional assumption is made. In Section 5 results from applying the generalized instrumental variables estimator to this stock return model are presented. Finally, some concluding remarks are made in Section 6.

2. THE IMPLICATIONS OF RATIONAL EXPECTATIONS MODELS USED IN CONSTRUCTING ESTIMATORS

Discrete-time models of the optimizing behavior of economic agents often lead to first-order conditions of the form:

$$E_t h(x_{t+n}, b_0) = 0, \tag{2.1}$$

where x_{t+n} is a k dimensional vector of variables observed by agents and the econometrician as of date $t+n$, b_0 is an l dimensional parameter vector that is unknown to the econometrician, h is a function mapping $R^k \times R^l$ into R^m, and E_t is the expectations operator conditioned on agents' period t information set, I_t. Expectations are assumed to be formed rationally and, hence, E_t denotes both the mathematical conditional expectation and agents' subjective expectations as of date t. For the purposes of this paper, we shall think of equation (2.1) as emerging from the first-order conditions of a representative agent's utility maximization problem in an uncertain environment. Our procedures can also be applied with some modification to models of panel data in which (2.1) represents the first-order conditions associated with the optimum problems of heterogeneous agents, so long as the heterogeneity is indexed by individual characteristics observed by the econometrician. More generally, our approach to estimation is appropriate for any model that yields implications of the form (2.1) with x observed. This latter qualification does rule out some models in which the implied Euler equations involve unobservable forcing variables.

An example will be useful both for interpreting (2.1) and understanding the estimation procedure discussed in Section 3. Following Lucas [**22**], Brock [**4**], Breeden [**3**], and Prescott and Mehra [**26**], suppose that a representative consumer chooses stochastic consumption and investment plans so as to maximize

$$E_0\left[\sum_{t=0}^{\infty}\beta^t U(C_t)\right], \tag{2.2}$$

where C_t is consumption in time period t, $\beta \in (0,1)$ is a discount factor, and $U(\cdot)$ is a strictly concave function. Further, suppose that the consumer has the choice of investing in a collection of N assets with maturities M_j, $j = 1, \ldots, N$. Let Q_{jt} denote the quantity of asset j held at the end of date t, P_{jt} the price of asset j at date t, R_{jt} the date t payoff from holding a unit of an M_j-period asset purchased at date $t - M_j$, and W_t (real) labor income at date t.[4] All prices are denominated in terms of the consumption good. The feasible consumption and investment plans must satisfy the sequence of budget constraints

$$C_t + \sum_{j=1}^{N} P_{jt}Q_{jt} \leq \sum_{j=1}^{N} R_{jt}Q_{jt-M_j} + W_t. \tag{2.3}$$

The maximization of (2.2) subject to (2.3) gives the first-order necessary conditions (Lucas [**22**], Brock [**4**], Prescott and Mehra [**26**]):

$$P_{jt}U'(C_t) = \beta^{M_j}E_t\left[R_{jt+M_j}U'(C_{t+M_j})\right] \qquad (j = 1, \ldots, N). \tag{2.4}$$

If, for example, the jth asset is a default-free, zero coupon bond with term to maturity M_j, then R_{jt+M_j} in (2.4) equals the real par value of the bond at date $t + M_j$. Alternatively, if Q_{jt} denotes the quantity of shares of stock of a firm held at date t, D_{jt} denotes the dividend per share of stock j at date t, and $M_j = 1$, then $R_{jt+1} = (P_{jt+1} + D_{jt+1})$ and (2.4) becomes

$$P_{jt}U'(C_t) = \beta E_t\left[(P_{jt+1} + D_{jt+1})U'(C_{t+1})\right], \tag{2.5}$$

with P_{jt} interpreted as the exdividend price per share. Note that (2.5) is a generalization of the model studied by Hall [**9**] in which preferences were quadratic and real interest rates were assumed to be constant over time.

Estimation and testing using (2.4) or (2.5) requires that the function U be explicitly parameterized. For the moment, we assume only that preferences are described by a vector of parameters γ, $U(\cdot,\gamma)$, in order to emphasize the generality of our estimation strategy. At this level of generality, the representative agent assumption plays a critical role in the derivation of (2.4) and (2.5).

[4]This income term W_t could emerge under the assumption that labor is supplied inelastically. In this case W_t can be thought of as not being controllable by the representative consumer. Alternatively, we can introduce a period t labor supply variable L_t into the specification of U and let

$$U(C_t, L_t) = U_1(C_t) - U_2(L_t)$$

where L_t is a choice variable of the consumer. For this case, $W_t = L_t w_t$ where w_t is the real wage rate at period t.

Rubinstein [**28**] has shown, however, that (2.4) and (2.5) can implicitly accommodate certain types of heterogeneity when agents' preference functions are members of the HARA class. In Sections 4 and 5, we consider in detail the special case of (2.5) with a constant relative risk averse preference function. Using a theorem in Rubinstein [**28**], this version of (2.5) can be derived from a model in which agents are allowed to have different stochastic endowment streams.

Relation (2.4) can be used to construct the h function specified in (2.1). Suppose the econometrician has observations on P_j and R_j for a subset of m of the assets ($m \leq N$) with maturities $n_1, n_2, \ldots, n_m$, and on consumption C. Since C_t and P_{jt} are known to agents at time t, (2.4) implies

$$E_t\left[\beta^{n_j}\frac{U'(C_{t+n_j},\gamma)}{U'(C_t,\gamma)}x_{jt+n_j}-1\right]=0, \tag{2.6}$$

where $x_{jt+n_j}=R_{n_jt+n_j}/P_{n_jt}$, for $j=1,\ldots,m$. Let $n=n_m$ and $x'_{t+n}=(x_{1t+n_1}, \ldots, x_{mt+n}, C_t^{*\prime})$ where the n^* constituents of C_t^* are observable functions of C_t and the distinct values of C_{t+n_j}, $j=1,\ldots,n$. For example, $U'(C_{t+n_j})/U'(C_t)$ in (2.6) may involve the function C_{t+n_j}/C_t. Then the function $h(x_{t+n},b_0)$ in (2.1) is given by

$$h(x_{t+n},b_0)=\begin{bmatrix}\beta^{n_1}\dfrac{U'(C_{t+n_1},\gamma)}{U'(C_t,\gamma)}x_{1t+n_1}-1\\ \vdots \\ \beta^{n}\dfrac{U'(C_{t+n},\gamma)}{U'(C_t,\gamma)}x_{mt+n}-1\end{bmatrix}.$$

Notice that x has $m+n^*$ coordinates.

We can interpret

$$u_{t+n}=h(x_{t+n},b_0)$$

as the disturbance vector in our econometric estimation. The matrix Eu_tu_t' is assumed to have full rank. This assumption implicitly imposes some structure on the link between u and the "forcing variables" not observed by the econometrician that enter, for example, through the production technology. The autocovariance structure of u depends on the nature of the assets being studied. If the m assets are stocks and n_1 through n equal unity, for example, then $h(x_{t+1},b_0)$ is constructed from the Euler equations by setting $x_{jt+1}=(P_{jt+1}+D_{jt+1})/P_{jt}$. In this case, u is serially uncorrelated, since observations on x_{t-s}, $s\geq 0$, are contained in I_t and $E_t[h(x_{t+1},b_0)]=0$. On the other hand, if $n_j>1$ for some j, as in the model of the term structure of bond prices implied by (2.4), the condition $E_t[h(x_{t+n},b_0)]=0$ does not preclude serial correlation in u. This can be seen by noting that x_{t+n_j-1} is not necessarily included in I_t if $n_j>1$. The presence of serial correlation in u leads to a more complicated asymptotic covariance matrix for our proposed estimator, but it does not affect consistency (see Section 3).

Before describing our estimation strategy in detail, we briefly consider what is perhaps a more "natural" approach to estimation in order to help motivate our approach. A possible way of proceeding with estimation is: (i) to explicitly specify the rest of the economic environment, including the production technology and the stochastic properties of the forcing variables; (ii) to solve for an equilibrium representation for the endogenous variables in terms of past endogenous variables and current and past forcing variables; and (iii) to estimate the parameters of tastes, technology, and the stochastic process governing the forcing variables using a full information procedure such as maximum likelihood. This approach will yield an exact relationship among current and past endogenous variables and current and past forcing variables. To avoid an implication of a stochastic singularity among variables observed by the econometrician, it can be assumed that the econometrician does not have observations on some of the forcing variables. This approach is viable if both the form of (2.1) and the stochastic specification of the forcing variables are relatively simple. To allow for a general representation of the forcing variables, explicitly solve for an equilibrium representation of the observables, and proceed to estimate the parameters of tastes and technology together with the parameters of the forcing processes appears to be an overly ambitious task outside of linear environments.[5] For this reason we adopt an alternative estimation strategy that can be viewed as an extension to nonlinear environments of the procedures of McCallum [**23**] and Cumby, Huizinga, and Obstfeld [**6**] for estimating linear rational expectations models.

3. ESTIMATION

In this section we describe how to estimate the vector b_0 using a generalized instrumental variables procedure. The basic idea underlying our proposed estimation strategy is to use the theoretical economic model to generate a family of orthogonality conditions. These orthogonality conditions are then used to construct a criterion function whose minimizer is our estimate of b_0. This criterion function is constructed in a manner that guarantees that our parameter estimator is consistent, asymptotically normal, and has an asymptotic covariance matrix that can be estimated consistently. The orthogonality conditions also can be used to construct a test of the overidentifying restrictions implied by the theoretical model. We elaborate on each of the steps in the following discussion.

Let $u_{t+n} = h(x_{t+n}, b_0)$ and consider again the first-order conditions

$$E_t[u_{t+n}] = 0, \tag{3.1}$$

with the additional assumption that the m constituents of u_{t+n} have finite second

[5] Even in linear environments this is a nontrivial econometric endeavor. See Hansen and Sargent [**11**, **12**] and Sargent [**30**] for a discussion of these issues in linear environments.

moments. Also, let z_t denote a q dimensional vector of variables with finite second moments that are in agents' information set and observed by the econometrician; and define the function f by

$$f(x_{t+n}, z_t, b) = h(x_{t+n}, b) \otimes z_t, \tag{3.2}$$

where f maps $R^k \times R^q \times R^l$ into R^r, $r = m \cdot q$, and $\otimes$ is the Kronecker product. Then an implication of (3.1), (3.2), and the accompanying assumptions is that

$$E\left[f(x_{t+n}, z_t, b_0) \right] = 0, \tag{3.3}$$

where E is the unconditional expectations operator.[6] Equation (3.3) represents a set of r population orthogonality conditions from which an estimator of b_0 can be constructed, provided that r is at least as large as the number of unknown parameters, l.

We proceed by constructing an objective function that depends only on the available sample information $\{(x_{1+n}, z_1), (x_{2+n}, z_2), \ldots, (x_{T+n}, z_T)\}$ and the unknown parameters. Let $g_0(b) = E[f(x_{t+n}, z_t, b)]$, where $b \in R^l$ and it is assumed that the left-hand side does not depend on t. Note that (3.3) implies that g_0 has a zero at $b = b_0$. Thus, if the model underlying (3.1) is true, then the method of moments estimator of the function g_0,

$$g_T(b) = \frac{1}{T} \sum_{t=1}^{T} f(x_{t+n}, z_t, b), \tag{3.4}$$

evaluated at $b = b_0$, $g_T(b_0)$, should be close to zero for large values of T. Given this fact, and the assumption that f is continuous in its third argument, it is reasonable to estimate b_0 by selecting b_T from a parameter space $\Omega \subseteq R^l$ that makes g_T in (3.4) "close" to zero. In this paper, we follow Amemiya [**1**, **2**], Jorgenson and Laffont [**18**], and Hansen [**10**] and choose $b_T \in \Omega$ to minimize the function J_T given by

$$J_T(b) = g_T(b)' W_T g_T(b), \tag{3.5}$$

where W_T is an r by r symmetric, positive definite matrix that can depend on sample information. The choice of the weighting matrix W_T defines the metric used in making g_T close to zero.

Sufficient conditions for strong consistency and asymptotic normality of estimators constructed in this fashion are provided in Hansen [**10**].[7] A few key

[6] A researcher may wish to use a (possibly different) subvector of z to form the orthogonality conditions for each element of u_{t+n}. Modifications to the following analysis to account for this possibility are straightforward.

[7] Hansen [**10**] assumes that the parameter space is compact, that continuity conditions are satisfied for the $\partial h/\partial b$, and that the stochastic process $\{(x_t, z_t)\}_{t=1}^{\infty}$ is stationary and ergodic. Presumably these assumptions can be relaxed, perhaps along the dimensions proposed by White [**33**] for cross-sectional analysis. In relaxing these assumptions it is important to consider the impact not only on estimation, but on model specification as well. Many theoretical rational expectations models rely on some form of stationarity to derive time invariant relationships.

observations are important. First, it is not necessary to specify how $E_t[u_{t+n}u_{t+n}']$ depends on elements in the information set as of time t. That is, we can allow u_{t+n} to be conditionally heteroskedastic and can conduct statistical inference without explicitly characterizing the dependence of the conditional variances on the information set. In the context of the asset pricing models discussed in Section 2, for example, this feature of our estimation procedure allows the conditional variances of asset yields to fluctuate with movements of variables in the conditioning information set. Thus, we can accommodate the assumption in Cox, Ingersoll, and Ross [**5**] that the conditional variances of interest rates vary with the level of past interest rates.

Second, we can think of these estimators as instrumental variables estimators where z_t denotes the vector of instruments. For this reason, we refer to our estimators as generalized instrumental variables estimators. We require only that the z's be "predetermined" as of time period t; they need not be "econometrically exogenous." For example, current and lagged x's can be chosen. Furthermore, as noted in Section 2, u will generally be serially correlated when $n > 1$. Our estimators will be consistent even when the disturbances are serially correlated and the instruments are not exogenous. From the standpoint of obtaining a consistent estimator, a researcher is given considerable latitude in selecting z_t.

Third, the asymptotic covariance matrix for these generalized instrumental variables estimators depends on the choice of weighting matrix W_T. It is possible to choose W_T "optimally" in the sense of constructing an estimator with the smallest asymptotic covariance matrix among the class of estimators employing alternative choices of weighting matrices W_T. More precisely, assume that h is differentiable, and that the vector of instruments z_t is chosen such that the matrix

$$D_0 = E\left[\frac{\partial h}{\partial b}(x_{t+n}, b_0) \otimes z_t\right],$$

has full rank. Also, assume that the weighting matrix W_T converges almost surely to a limiting constant matrix W_0 of full rank. Let

$$S_0 = \sum_{j=-n+1}^{n-1} E\left[f(x_{t+n}, z_t, b_0) f(x_{t+n-j}, z_{t-j}, b_0)'\right],$$

with the number of population autocovariances, n, determined by the order of the moving average disturbance term u_t. Assuming that S_0 has full rank, Hansen's [**10**] Theorems 3.1 and 3.2 imply that the smallest asymptotic covariance matrix for an estimator b_T that minimizes (3.5) is obtained by letting W_0 be $W_0^* = S_0^{-1}$. The resulting asymptotic covariance matrix is $(D_0' W_0^* D_0)^{-1}$.

In order to implement this "optimal" procedure and to conduct asymptotically valid inference we need consistent estimators of D_0 and W_0^*. These matrices can be estimated using[8]

$$D_T = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial h}{\partial b}(x_{t+n}, b_T)\otimes z_t,$$

$$R_T(j) = \frac{1}{T}\sum_{t=1+j}^{T} f(x_{t+n}, z_t, b_T)f(x_{t+n-j}, z_{t-j}, b_T)', \tag{3.6}$$

$$W_T^* = \left\{R_T(0) + \sum_{j=1}^{n-1}\left[R_T(j) + R_T(j)'\right]\right\}^{-1}.$$

Note that to compute W_T^* in actual applications, a consistent estimator b_T of b_0 is needed. This can be obtained by initially using a suboptimal choice of W_T in minimizing (3.5) to obtain b_T. Then this b_T can be employed to calculate W_T^* using the formulas supplied above. Once W_T^* is calculated, b_T^* can be obtained by minimizing (3.5) with W_T^* substituted for W_T. Thus, the "optimal" estimation procedure requires two steps.

The optimal properties of the b_T^* estimator proposed above are, in one sense, weak. Nowhere did we offer any precise suggestions about how to choose z_t optimally from the information set as of time t. The solution to this problem is discussed in Hansen and Singleton [**15**].[9] The general solution for z_t may not be operational because it requires the researcher to make precise assumptions about the entire economic environment. Nevertheless, there are potentially important special cases in which the optimal choice of z_t can be used for estimation of models like (2.1).

Before concluding this section, we consider a straightforward way of testing the restrictions implied by the model. The estimation procedure sets the l linear combinations of the r orthogonality conditions associated with the first-order conditions to the minimization problem (3.5) equal to zero in estimating b_0:

$$\underset{l\times r}{\left[\frac{\partial g_T(b)'}{\partial b}\right.}\;\underset{r\times r}{\left.W_T\right]}\;\underset{r\times 1}{g_T(b)} = 0. \tag{3.7}$$

[8]If $n > 1$, then W_T^* in (3.6) is not positive definite by construction. Alternative spectral estimators that are positive definite by construction are available [e.g., **10** and **13**]; however, these estimators do not exploit the fact that the defining sum for S only involves a finite number of terms.

In practice, one may wish to remove the product of the sample means $g_T(b_T)g_T(b_T)'$ from the cross-products $f(x_{t+n}, z_t, b_T)f(x_{t+n-j}, z_{t-j}, b_T)'$ when computing $R_T(j)$ in (3.6). Under the null hypothesis, this adjustment has no effect on the asymptotic properties of the test statistics or parameter estimates. Under some alternative hypotheses, not all elements of $g_T(b_T)$ are expected to be near zero. In these circumstances the sample mean corrections may have important effects on the values of the test statistics and, hence, the power of the tests.

[9]For an extensive discussion of optimal instrument selection in linear environments, see Hansen and Sargent [**13**] and Hayashi and Sims [**17**].

Thus, when $r > l$, there are $r - l$ remaining linearly independent orthogonality conditions that are not set to zero in estimation, but should be close to zero if the model restrictions are true. To test these overidentifying restrictions, we invoke a theorem in Hansen [**10**] that implies that T times the minimized value of (3.5) for an optimal choice of W_T is asymptotically distributed as a chi-square with $r - l$ degrees of freedom. It follows that the minimized value of the second step objective function can be used to test the nonlinear rational expectations model (3.1). The particular $r - l$ orthogonality conditions underlying this test are determined by the choice of the matrix premultiplying $g_T(b)$ in (3.7), which in turn was chosen to make b_T "optimal." If one has *a priori* reasons for believing that another set of r linear combinations of $g_T(b)$ should hold, then the overidentifying restrictions associated with these conditions can also be tested using the results from Hansen's [**10**] Lemma 4.1.

4. COMPARISON TO MAXIMUM LIKELIHOOD

Amemiya [**2**] has noted in a somewhat different context that maximum likelihood estimators will, in general, be asymptotically more efficient than nonlinear instrumental variables procedures if the distributional assumptions are specified correctly. On the other hand, maximum likelihood estimates may fail to be consistent if the distribution of the observable variables is misspecified. In this section we illustrate the second point in the context of the stock return model in Section 2.

For the purposes of this illustration and the empirical example in Section 5, we assume that preferences are of the constant relative risk averse type,

$$U(C_t) = \frac{(C_t)^\gamma}{\gamma}, \qquad \gamma < 1.$$

In this case, the marginal utility is given by

$$U'(C_t) = (C_t)^\alpha, \qquad \alpha \equiv \gamma - 1.$$

If the m assets considered in Section 2 are stocks, then (2.6) simplifies to

$$E_t[\beta(x_{kt+1})^\alpha x_{jt+1}] = 1 \qquad (j = 1, 2, \dots, m), \tag{4.1}$$

where x_{kt+1} is the ratio of consumption in time period $t + 1$ to consumption in time period t, and the one-period real return x_{jt+1} is given by $(P_{jt+1} + D_{jt+1})/P_{jt}$. Grossman and Shiller [**8**] have discussed some empirical implications of this specification of the stock price model.

Maximum likelihood estimation requires that additional distributional assumptions be imposed on the model. Suppose that the stochastic process $x' = (x_1, x_2, \dots, x_m, x_k)$ is lognormally distributed, and that $X = \log x$ is stationary, indeterministic, and of full rank. Further, suppose for simplicity that the econometrician chooses to estimate α and β in (4.1) using observations only on x or, equivalently, X. Then asymptotically efficient parameter estimates can be obtained from a restricted version of the Wold moving average representation

$$X_{t+1} = \Psi(L)V_{t+1} + \mu$$

where $\Psi(L)$ is a $k \times k$ infinite order matrix polynomial in the lag operator L, with $\Psi(0) = I$,

$$E[V_{t+1} \mid X_t, X_{t-1}, \ldots] = 0, \qquad E[V_{t+1}V'_{t+1} \mid X_t, X_{t-1}, \ldots] = \Sigma,$$

Σ has full rank, and μ is a k dimensional vector of constants.[10] To derive the restrictions on $\Psi(L)$ and μ implied by (4.1), we first note that

$$\log[\beta(x_{kt+1})^{\alpha}x_{jt+1}] = \log\beta + \alpha X_{kt+1} + X_{jt+1}.$$

Thus, from the Wold moving average representation, it follows that, for $j = 1, 2, \ldots, m$,

$$E[\log[\beta(x_{kt+1})^{\alpha}x_{jt+1}] \mid X_t, X_{t-1}, \ldots] = \log\beta + \alpha\left[\frac{\Psi_k(L)}{L}\right]_+ V_t + \left[\frac{\Psi_j(L)}{L}\right]_+ V_t + \alpha\mu_k + \mu_j \tag{4.2}$$

where $\Psi_j(L)$ and $\Psi_k(L)$ are the jth and last rows of $\Psi(L)$, respectively, $\mu' = [\mu_1, \ldots, \mu_k]$, and $[\;\;]_+$ denotes the annihilation operator that instructs us to ignore negative powers of L. Also, the conditional variances satisfy the relationship

$$\begin{aligned} \operatorname{var}\{\log[\beta(x_{kt+1})^{\alpha}x_{jt+1}] \mid X_t, X_{t-1}, \ldots\} &= \operatorname{var}[\alpha V_{kt+1} + V_{jt+1}] \\ &= \alpha^2\sigma_{kk} + \sigma_{jj} + 2\alpha\sigma_{kj}, \end{aligned} \tag{4.3}$$

$j = 1, \ldots, m$, where $\Sigma = [\sigma_{ij}]$ and var denotes the variance operator. Finally, using an iterated expectations argument, it can be shown that (4.1) implies

$$E[\beta(x_{kt+1})^{\alpha}x_{jt+1} \mid X_t, X_{t-1}, \ldots] = 1 \tag{4.4}$$

for $j = 1, 2, \ldots, m$, as long as $X_t \in I_t$. Combining equations (4.2), (4.3), and

[10] A good reference on covariance stationary processes and Wold's Decomposition Theorem is Rozanov [**27**].

(4.4), gives the restrictions

$$\text{(4.5)}\qquad \begin{aligned} &\log \beta + \frac{\alpha^2\sigma_{kk} + \sigma_{jj} + 2\alpha\sigma_{kj}}{2} + \alpha\mu_k + \mu_j = 0, \\ &\left[\frac{\Psi_j(L)}{L}\right]_+ = -\alpha\left[\frac{\Psi_k(L)}{L}\right]_+ , \end{aligned}$$

for $j = 1, 2, \ldots, m$. We can estimate the parameters of Σ, Ψ, μ, α, and β by imposing the restrictions (4.5), using maximum likelihood with a Gaussian density function, and employing observations on $X_1, X_2, \ldots, X_{T+1}$ (see, e.g., Hansen and Singleton [**14**]).

The restrictions given in (4.5) together with relation (4.2) imply a logarithmic form of relation (4.4), namely,

$$\text{(4.6)}\qquad \begin{aligned} &E\left[\log \beta + \alpha \log x_{kt+1} + \log x_{jt+1} \mid X_t, X_{t-1}, \ldots\right] \\ &\quad = \frac{-\alpha^2\sigma_{kk} - \sigma_{jj} - 2\alpha\sigma_{kj}}{2} . \end{aligned}$$

By the law of iterated expectations, it follows that the random variable

$$U_{jt+1} = \log \beta + \alpha X_{kt+1} + X_{jt+1} + \frac{\left(\alpha^2\sigma_{kk} + \sigma_{jj} + 2\alpha\sigma_{kj}\right)}{2}$$

satisfies the orthogonality conditions

$$\text{(4.7)}\qquad E\left[U_{jt+1}\right] = 0; \qquad E\left[U_{jt+1}X_{t-s}\right] = 0;$$

for $j = 1, 2, \ldots, m$ and $s \geq 0$.

From the first-order conditions of the likelihood function, it can be seen that the method of maximum likelihood implicitly uses the logarithmic orthogonality conditions (4.7). The validity of these orthogonality conditions is crucially dependent on x being lognormally distributed. For other distributional assumptions, the logarithmic form of the orthogonality conditions (4.7) will generally not hold. Consequently, the maximum likelihood estimators of α and β obtained from (4.7) under the assumption of lognormality will generally not be consistent if this distributional assumption is incorrect. In contrast, the procedures which we proposed in Section 3 do not require that the distribution of x be specified *a priori* and, in particular, they do not require that the logarithmic form of the orthogonality conditions hold. Instead, we work directly with the orthogonality conditions implied by (4.1).

5. EMPIRICAL RESULTS

To illustrate the use of the generalized instrumental variables estimator, we estimated the parameters of preferences, α and β, for the model of stock prices discussed in Section 4. Two different measures of consumption were considered:

nondurables plus services (NDS) and nondurables (ND).[11] The monthly, seasonally adjusted observations on aggregate real consumption of nondurables and services were obtained from the Federal Reserve Board. Real per capita consumption series were constructed by dividing each observation of these series by the corresponding observation on population, published by the Bureau of Census. Each measure of consumption was paired with three sets of stock returns: the equally-weighted average return on all stocks listed on the New York Stock Exchange (EWR), the value-weighted average of returns on the New York Stock Exchange (VWR), and equally-weighted average returns on the stocks of three two-digit SEC industries. The industries chosen were chemicals (SEC code 28), transportation and equipment (SEC code 37), and other retail trade (SEC codes 50–52 and 54–59). The aggregate return data were obtained from the CRSP tapes and the industry return data were obtained from Stambaugh [**31**]. Nominal returns were converted to real returns, which appear in (4.1), by dividing by the implicit deflator associated with the measure of consumption.

Following the notation adopted in Section 4, we let

$$x'_{t+1} = \left[\frac{P_{1t+1} + D_{1t+1}}{P_{1t}} \quad \frac{C_{t+1}}{C_t} \right]$$

and

$$h(x_{t+1}, b_0) = \beta (x_{2t+1})^{\alpha} x_{1t+1} - 1,$$

where $b'_0 = (\alpha, \beta)$. The vector of instruments z_t was formed using lagged values of x_{t+1}. For this specification of the process $\{(x_{t+1}, z_t);\ t = 1, 2, \dots\}$, the stationarity assumption accommodates certain types of real growth in consumption. The number of lagged values of x_{t+1} included in z_t, NLAG, was chosen to be 1, 2, 4, or 6. As NLAG is increased, more orthogonality conditions are employed in the estimation. Furthermore, the asymptotic covariance matrix becomes smaller, and the number of overidentifying restrictions being tested increases.[12]

Table I displays the parameter estimates obtained using the aggregate return series for the period February, 1959 through December, 1978.

[11] Using a separation argument like that for labor supply discussed in note 4, it is possible to argue that the restrictions hold for a measure of a subset of aggregate consumption. For instance, suppose that C_{1t} and C_{2t} are two different components of consumption at time period t and that the function U is given by

$$U(C_{1t}, C_{2t}) = \frac{(C_{1t})^{\gamma}}{\gamma} + U_2(C_{2t}).$$

In this case, the restrictions we test are appropriate when C_1 is used as the measure of consumption. The two choices for C_1 considered correspond to two potentially different assumptions about the separability of U.

[12] To be more precise, the asymptotic covariance matrix will not increase as more orthogonality conditions are used. Using more orthogonality conditions may, at some point, lead to estimators with less desirable small sample properties.

TABLE I

INSTRUMENTAL VARIABLE ESTIMATES FOR THE PERIOD 1959:2–1978:12

Cons	Return	NLAG	$\hat{\alpha}$	$\widehat{SE}(\hat{\alpha})$	$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$	χ^2	DF	Prob
NDS	EWR	1	−.9457	.3355	.9931	.0031	4.9994	1	.9746
NDS	EWR	2	−.9281	.2729	.9929	.0031	7.5530	3	.9438
NDS	EWR	4	−.7895	.2527	.9925	.0031	9.1429	7	.7574
NDS	EWR	6	−.8927	.2138	.9934	.0030	15.726	11	.8484
NDS	VWR	1	−.9001	.3130	.9979	.0025	1.1547	1	.7174
NDS	VWR	2	−.8133	.2298	.9981	.0025	3.2654	3	.6475
NDS	VWR	4	−.6795	.1855	.9973	.0024	6.3527	7	.5008
NDS	VWR	6	−.7958	.1763	.9980	.0023	14.179	11	.7767
ND	EWR	1	−.9737	.1245	.9922	.0031	5.9697	1	.9854
ND	EWR	2	−.9664	.1074	.9919	.0031	8.9016	3	.9694
ND	EWR	4	−.9046	.0926	.9918	.0031	11.084	7	.8650
ND	EWR	6	−.9466	.0793	.9422	.0030	15.663	11	.8459
ND	VWR	1	−.8985	.1057	.9971	.0025	1.5415	1	.8756
ND	VWR	2	−.8757	.0856	.9974	.0025	3.2654	3	.6475
ND	VWR	4	−.8174	.0742	.9967	.0024	7.8776	7	.5008
ND	VWR	6	−.8514	.0629	.9973	.0024	14.938	11	.8147

The estimates of α range from $-.95$ to $-.68$ when ND is used as the measure of consumption, and from $-.97$ to $-.82$ when ND is used as the measure of consumption. The estimated standard errors for α, $\widehat{SE}(\hat{\alpha})$, are smaller when consumption is measured as ND than when consumption is measured as NDS. As expected, all of the estimates of β exceed .99 but are less than unity. The chi-square tests are also displayed in Table I, where the number of overidentifying restrictions is indicated by DF and Prob is the probability that a $\chi^2(\mathrm{DF})$ random variate is less than the computed value of the test statistic under the hypothesis that the restrictions (3.1) are satisfied. These tests provide greater evidence against the model when EWR is included as the return, and when the instrument vector is formed from a small number of lagged values of x.

For comparison, we present some results in Table II from estimating α and β using the method of maximum likelihood under the assumption that x is lognormally distributed. They were obtained using the procedure described in Section 4 assuming that log x has a sixth-order vector autoregressive representation. The corresponding estimates of α and β from the two methods of estimation are similar. However, the estimated standard errors of α and β from the instrumental variables procedure are smaller than the corresponding standard errors from the maximum likelihood procedure.[13] Three possible explanations for this result are that the asymptotic standard errors are being estimated imprecisely, the economic model of stock returns is misspecified, or the auxiliary assumptions underlying the maximum likelihood procedure are incorrect. The maximum likelihood procedure assumes that x is lognormally distributed and that the lag length specification of the vector autoregression is correct. Since the

[13] Analytical differentiation, as opposed to numerical differentiation, was used to calculate the standard errors.

TABLE II
MAXIMUM LIKELIHOOD ESTIMATES FOR THE PERIOD 1959:2–1978:12

	Nondurables Plus Services		Nondurables	
	Equally Weighted Returns	Value Weighted Returns	Equally Weighted Returns	Value Weighted Returns
$\hat{\alpha}$	−.5194 (.4607)	−.9349 (.3341)	−.7643 (.2203)	−.9327 (.1579)
$\hat{\beta}$	.9957 (.0037)	.9995 (.0028)	.9957 (.0036)	.9988 (.0028)
χ^2	12.5854	17.8716	14.4456	18.8846
DF	11	11	11	11
Prob.	.6787	.9154	.7907	.9368

sum of lognormally distributed random variables is not lognormally distributed, it seems implausible that both (and perhaps either) the value weighted and (or) the equally weighted stock return indices are lognormally distributed. Also, the lag length specification does not emerge from any theoretical consideration. The generalized instrumental variables procedure does not require that either one of these auxiliary assumptions be satisfied.[14] On the other hand, the maximum likelihood estimates provide a more complete characterization of the stochastic process x when the lognormality and the lag length specifications are correct.

Table II also presents the likelihood ratio tests of the restrictions implied by the lognormal version of the model, where the unrestricted model is an unrestricted sixth-order vector autoregression. In contrast to the instrumental variables results, the maximum likelihood results provide more evidence against the restrictions when VWR is used than when EWR is used. For a more complete description of the maximum likelihood estimation and a more comprehensive set of results, see Hansen and Singleton [**14**].

If the Euler equation (4.1) holds for a given measure of consumption and all of the stocks listed on the NYSE, then versions of (4.1) must also hold simultaneously for the equally- and value-weighted aggregate returns. In Table III we present the results from estimating α and β using the orthogonality conditions implied by the respective versions of (4.1) for EWR and VWR, with the z vector formed using lagged values of EWR, VWR, and the consumption ratio. Note that these estimates of α are smaller than the corresponding estimates in Table I. Also, the estimated standard errors for $\hat{\alpha}$ are smaller than the corresponding

[14] Another attractive feature of the generalized instrumental variables procedure is that it requires a numerical search over a smaller parameter space than is required from the maximum likelihood procedure.

TABLE III

INSTRUMENTAL VARIABLES ESTIMATION WITH MULTIPLE RETURNS

Equally- and Value-Weighted Aggregate Returns 1959:2–1978:12								
Cons.	NLAG	$\hat{\alpha}$	$\widehat{SE}(\hat{\alpha})$	$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$	χ^2	DF	Prob.
NDS	1	−.6875	.2372	.9993	.0023	17.804	6	.9933
NDS	2	−.3624	.1728	.9995	.0022	24.230	12	.9811
NDS	4	−.3502	.1540	.9989	.0021	39.537	24	.9760
ND	1	−.7211	.0719	.9989	.0023	19.877	6	.9971
ND	2	−.5417	.1298	.9988	.0022	24.421	12	.9822
ND	4	−.5632	.1038	.9982	.0021	40.176	24	.9795
Three Industry-Average Stock Returns 1959:2–1977:12								
Cons.	NLAG	$\hat{\alpha}$	$\widehat{SE}(\hat{\alpha})$	$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$	χ^2	DF	Prob.
NDS	1	−.9993	.2632	.9941	.0028	19.591	13	.8941
NDS	4	−.4600	.1388	.9961	.0024	82.735	49	.9982
ND	1	−.9557	.0898	.9935	.0028	22.302	13	.9491
ND	4	−.8085	.0506	.9962	.0023	82.013	49	.9978

estimates in Table I when consumption is measured as NDS, and they are comparable for the models using ND.[15] Finally, there is considerably more evidence against the models when the overidentifying restrictions associated with EWR and VWR are tested simultaneously than when separate tests are conducted.

Table III also presents the results from estimating α and β using the three industry-average returns simultaneously. In these cases, the coordinates of x_{t+1} are the three real stock returns and the consumption ratio, the z_t vector was formed using lagged values of x_{t+1}, and the sample period was February, 1959 through December, 1977. The point estimates for α obtained using industry returns are smaller than the corresponding estimates in Table I. At the same time, the industry return data provides more evidence against the restrictions, since the models with NLAG $= 4$ are rejected at the one per cent significance level.

To summarize, all of the models considered yield economically plausible estimates of α and β. The test results for the single-return models are mixed, with marginal confidence levels close to unity occurring for small values of NLAG but not for large values. In contrast, the test statistics for the multiple-return models all have probability values of .95 or larger, except one. Thus, the latter results provide considerably more evidence against the stock pricing model with constant relative risk averse preferences.

[15] One qualification to the four lag results in Table III should be mentioned. When NLAG $= 4$, twenty-six and fifty-one orthogonality conditions are used to estimate the parameters α and β for the aggregate and industry models, respectively. The quality of the consistent estimator of the weighting matrix may deteriorate when the number of orthogonality conditions being used is large. In such cases, one may wish to take the large sample approximations less seriously in evaluating the parameter estimates and the restrictions implied by the model.

6. CONCLUSIONS

In this paper we have discussed a procedure for estimating the parameters of nonlinear rational expectations models when only a subset of the economic environment is explicitly specified *a priori*. We also described how to test the over-identifying restrictions implied by the particular economic model being estimated. The advantages of these procedures are that they circumvent the need for explicitly deriving decision rules, and they do not require the specification of the joint distribution function of the observable variables. The techniques are appropriate for any dynamic model whose econometric implications can be cast in terms of a set of orthogonality conditions. As an application of these procedures, we estimated the parameters characterizing preferences in a model relating the stochastic properties of aggregate consumption and stock market returns.

Carnegie-Mellon University

Manuscript received May, 1981; revision received February, 1982.

REFERENCES

[1] Amemiya, T.: "The Nonlinear Two-Stage Least Squares Estimator," *Journal of Econometrics*, 2(1974), 105–110.

[2] ———: "The Maximum Likelihood and Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equations Model," *Econometrica*, 45(1977), 955–968.

[3] Breeden, D. T.: "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7(1979), 265–296.

[4] Brock, W. A.: "An Integration of Stochastic Growth Theory and the Theory of Finance," Manuscript, University of Chicago, 1978.

[5] Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross: "A Theory of the Term Structure of Interest Rates," Manuscript, 1978.

[6] Cumby, R., J. Huizinga, and M. Obstfeld: "Two-Step, Two-Stage Least Squares Estimation in Models with Rational Expectations," National Bureau of Economic Research, Technical Paper 11, 1981.

[7] Fair, R. C., and J. B. Taylor: "Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models," Cowles Foundation Discussion Paper No. 564, 1980.

[8] Grossman, S. J., and R. J. Shiller: "The Determinants of the Variability of Stock Market Prices," *American Economic Review*, 71(1981), 222–227.

[9] Hall, R. E.: "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86(1978), 971–987.

[10] Hansen, L. P.: "Large Sample Properties of Generalized Method of Movements Estimators," *Econometrica*, 50(1982), 1029–1054.

[11] Hansen, L. P., and T. J. Sargent: "Formulating and Estimating Dynamic Linear Rational Expectations," *Journal of Economic Dynamics and Control*, 2(1980), 7–46.

[12] ———: "Linear Rational Expectations Models for Dynamically Interrelated Variables," in *Rational Expectation and Economic Practice*, ed. by R. E. Lucas, Jr. and T. J. Sargent. Minneapolis: University of Minnesota Press, 1981.

[13] ———: "Instrumental Variables Procedures for Estimating Linear Rational Expectations Models," forthcoming *Journal of Monetary Economics*, 1981.

[14] Hansen, L. P., and K. J. Singleton: "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," Manuscript, Carnegie-Mellon University, 1982.

[15] ———: "Optimal Instrumental Variables Estimators for Nonlinear Time Series Models," Manuscript, Carnegie-Mellon University, 1982.

[16] HAYASHI, F.: "Estimation Macroeconomic Models Under Rational Expectations: A Survey," Manuscript, Northwestern University, 1980.

[17] HAYASHI, F., AND C. A. SIMS: "Nearly Efficient Estimation of Time Series Models with Predetermined, But Not Exogenous, Instruments," forthcoming in *Econometrica*.

[18] JORGENSEN, D. W., AND J. LAFFONT: "Efficient Estimation of Nonlinear Simultaneous Equations with Additive Disturbances," *Annals of Economic and Social Measurement*, 3(1974), 615–640.

[19] KENNAN, J.: "The Estimation of Partial Adjusted Models with Rational Expectations," *Econometrica*, 47(1979), 1441–1456.

[20] KOHN, R.: "Asymptotic Estimation and Hypothesis Testing Results for Vector Linear Time Series Models," *Econometrica*, 47(1979), 1441–1456.

[21] LUCAS, R. E., JR.: "Econometric Policy Evaluation: A Critique," in *The Phillips Curve and Labor Markets, Carnegie-Rochester Conference on Public Policy 1*, ed. by K. Brunner and A. H. Meltzer. Amsterdam: North Holland, 1976.

[22] ———: "Asset Prices in an Exchange Economy," *Econometrica*, 46(1978), 1429–1446.

[23] MCCALLUM, B. T.: "Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates," *Econometrica*, 44(1976), 43–52.

[24] MERTON, R. C.: "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8(1980), 323–361.

[25] NERLOVE, M., D. M. GRETHER, AND J. L. CARVALHO: *Analysis of Economic Time Series*. New York: Academic Press, 1979.

[26] PRESCOTT, E. C., AND R. MEHRA: "Recursive Competitive Equilibrium: The Case of Homogeneous Households," *Econometrica*, 48(1980), 1365–1380.

[27] ROZANOV, Y. A.: *Stationary Random Processes*. San Francisco: Holden-Day, 1967.

[28] RUBINSTEIN, M.: "An Aggregation Theorem for Securities Markets," *Journal of Financial Economics*, 1(1974), 225–244.

[29] SARGENT, T. J.: *Macroeconomic Theory*. New York: Academic Press, 1979.

[30] ———: "Interpreting Economic Time Series," *Journal of Political Economy*, 89(1981), 213–248.

[31] STAMBAUGH, R.: "Measuring the Market Portfolio and Testing the Capital Asset Pricing Model," Manuscript, University of Chicago, 1978.

[32] WALLIS, K.: "Econometric Implications of the Rational Expectations Hypothesis," *Econometrica*, 48(1980), 49–74.

[33] WHITE, H.: "Instrumental Variables Regression with Independent Observations," *Econometrica*, 50(1982), 483–499.