Econometrica, Vol. 61, No. 4 (July, 1993), 929–952

SIMULATED MOMENTS ESTIMATION OF MARKOV MODELS OF ASSET PRICES

BY DARRELL DUFFIE AND KENNETH J. SINGLETON[1]

This paper provides a simulated moments estimator (SME) of the parameters of dynamic models in which the state vector follows a time-homogeneous Markov process. Conditions are provided for both weak and strong consistency as well as asymptotic normality. Various tradeoffs among the regularity conditions underlying the large sample properties of the SME are discussed in the context of an asset-pricing model.

KEYWORDS: Monte Carlo simulation, generalized method of moments, geometric ergodicity, uniform strong law of large numbers, model estimation.

1. INTRODUCTION

THIS PAPER PROVIDES CONDITIONS for the consistency and asymptotic normality of a simulated moments estimator (SME) of the parameters of asset-pricing models with time-homogeneous Markov representations of the stochastic forcing process. SME's for economic models have been proposed by McFadden (1989) and Pakes and Pollard (1989) for i.i.d. environments, and by Lee and Ingram (1991) for a time series environment. The SME for time series models examined in this paper is as follows. The state vector Y_t that determines asset prices is assumed to follow a time-homogeneous Markov process whose transition function depends on an unknown parameter vector β_0. Asset prices, and possibly other relevant data, are observed as $f(Y_t, \beta_0)$, for some given function f of the underlying state and parameter vector. In parallel, a simulated state process $\{Y_s^\beta\}$ is generated (analytically or numerically) from the economic model and corresponding simulated observations $f(Y_s^\beta, \beta)$ are taken, for a given parameter choice β. The parameter β is chosen so as to "match moments," that is, to minimize the distance between sample moments of the data, $f(Y_t, \beta_0)$, and those of the simulated series $f(Y_t^\beta, \beta)$, in a sense to be made precise.

The proposed SME extends the generalized method-of-moments (GMM) estimator (Hansen (1982)) to a large class of asset-pricing models for which the moment restrictions of interest do not have analytic representations in terms of observable variables and the unknown parameter vector. We provide conditions on the transition function of Y_t and the observation function f under which the SME of β_0 is consistent, and characterize the normalized asymptotic distribution of the estimator. For two reasons, neither the regularity conditions underlying Hansen's (1982) analysis of GMM estimators for time-series models without

[1] We are grateful for several useful conversations with the co-editor Lars Hansen and Whitney Newey, and for the comments of Peter Bossaerts, Andrew Lo, Neil Pearson, Bruce Lehmann, and the referees. Singleton acknowledges funding from the National Science Foundation. Duffie acknowledges the support of a Battery March Fellowship.

simulation, which were also used by Lee and Ingram (1991) for their SME estimator, nor those imposed by McFadden (1989) and Pakes and Pollard (1989) for simulated moments estimation in i.i.d. environments, are applicable to the estimation problems posed in this paper. First, in simulating time series, pre-sample values of the series are typicallyrequired. In most circumstances, however, the stationary distribution of the simulated process, as a function of the parameter choice, is unknown. Hence, the initial conditions for the time series will generally not be drawn from their stationary distribution and the simulated process will generally be nonstationary. Second, functions of the current value of the simulated state depend on the unknown parameter vector both through the structure of the model (as in any GMM problem) and indirectly through the generation of data by simulation. The feedback effect of the latter dependence on the transition law of the simulated state process implies that the first-moment-continuity condition used by Hansen (1982), or the generalizations proposed by Andrews (1987), in establishing the uniform convergence of the sample to the population criterion functions are not directly applicable to the SME. Furthermore, the nonstationarity of the simulated series must be accommodated in establishing the asymptotic normality of the SME.

We address these difficulties by assuming geometric ergodicity as a condition on the state process ensuring that the simulated processes are asymptotically stationary with an ergodic distribution that is independent of starting values, and by imposing a damping condition on the feedback effect of parameter choice on the law of motion of the state process. Under these conditions, the nonstationarities associated with simulation are shown to be inconsequential for the asymptotic distribution of the SME.

The remainder of the paper is organized as follows. Section 2 uses a simple asset-pricing setting to illustrate in more detail the econometric issues that arise with estimation by simulation. The formal structure of the estimation problem and the definition of the simulated moments estimator are laid out in Section 3. Section 4 provides conditions for consistency, both weak and strong, the key ingredient being an appropriate extension of the uniform law of large numbers. Section 5 characterizes the asymptotic distribution of the SME, while Section 6 provides several extensions of the SME.

2. AN ILLUSTRATIVE ASSET-PRICING MODEL

In this section we describe a simple dynamic asset-pricing model that illustrates many of the econometric problems that arise in the use of simulation methods in estimation. The model is an extended version of the stochastic growth model studied by Brock (1980) and Michner (1984). After briefly describing the model, the use of simulation methods is given a more extensive motivation. Several econometric issues related to estimation using simulation are then introduced in the context of this model. This section is intended as an informal backdrop to the simulated moments estimator presented in Section 3 and analyzed in Sections 4 and 5.

Suppose that production of the single consumption commodity is determined by

$$F(k_t, z_t) = z_t k_t^{\phi}, \qquad 0 < \phi < 1, \tag{2.1}$$

for some function F, where k_t is the level of the capital stock at date t and z_t is a technology shock. The firm rents capital from consumers at the rental rate r_t^k and pays out the profits to the owners of its shares in the form of dividends, d_t. In each period, the firm solves the following static optimum problem (maximization of profits)

$$d_t = \arg\max_{k_t} \{z_t k_t^{\phi} - r_t^k k_t\} \tag{2.2}$$

in order to choose the level k_t of capital to rent from the consumer. In equilibrium, this is equivalent to maximization of share market value (see, for example, Duffie (1988, Section 20)).

Given the price p_t of a share of the firm, the representative consumer faces the budget constraint

$$c_t + k_{t+1} + p_t s_{t+1} = (d_t + p_t) s_t + (r_t^k + \mu) k_t, \tag{2.3}$$

where c_t and s_t denote consumption and shares of claims to the dividend stream of the firm, respectively, and $(1-\mu)$ denotes a constant depreciation rate on the capital stock. Subject to this constraint, the representative consumer chooses consumption and share holdings so as to maximize utility for the infinite-horizon consumption process $\{c_t\}$. Allowing for an unobserved (to the econometrician) taste shock $\{u_t\}$ and adopting a typical additively-separable utility criterion, the agent's problem is then

$$\max_{\{c_t, k_t\}} E\left[\sum_{t=1}^{\infty} \delta^t \frac{(c_t - 1)^{1-\alpha}}{1-\alpha} u_t\right], \qquad \alpha < 0, \tag{2.4}$$

where α is the constant coefficient of relative risk aversion and $\delta \in (0,1)$ is a subjective discount factor.

The vector $X_t' = (z_t, u_t)$ is assumed to be a Markov process satisfying

$$X_t = h(X_{t-1}, \varepsilon_t, \rho_0), \tag{2.5}$$

where $\{\varepsilon_t\}$ is a two-dimensional i.i.d. stochastic process, h is a transition function, and ρ_0 is an unknown parameter vector. For the moment, we also assume that $\{X_t\}$ does not exhibit growth over time.

In order to estimate the unknown parameter vector $\beta_0 = (\phi, \alpha, \rho_0, \mu, \delta)'$, a point in some compact parameter set Θ, we proceed as follows. The economic system (2.1)–(2.5) is solved analytically or numerically for the equilibrium transition function H generating the augmented state process $Y_t = (X_t', k_t)'$, according to

$$Y_{t+1} = H(Y_t, \varepsilon_{t+1}, \beta_0). \tag{2.6}$$

For any admissible parameter vector $\beta \in \Theta$, we can also generate a simulated state process $\{Y_t^\beta\}$ according to the same transition function H, but using a shock sequence $\{\hat{\varepsilon}_t\}$ that is identically and independently distributed of $\{\varepsilon_t\}$; that is,

$$Y_{t+1}^\beta = H\left(Y_t^\beta, \hat{\varepsilon}_{t+1}, \beta\right).$$

From this, a history $\{Y_t^\beta\}_{t=1}^{\mathscr{T}}$ of $\mathscr{T}$ simulated equilibrium states can be generated.

Next, for some chosen observation function f, in each period t an observation $f_t^* \equiv f(Y_t, Y_{t-1}, \ldots, Y_{t-l+1})$ is made of a finite "l-history" of state information. Likewise, a corresponding observation f_t^β can be formed for each l-history of simulated states. The components of f_t^β may be known analytic functions (for example, $k_t^\beta \cdot k_{t-1}^\beta$) or determined numerically as functions of the l-history of simulated states (for example, equilibrium asset prices or consumption). Finally, the SME is a value of β chosen to minimize the distance between the sample mean of $\{f_t^\beta\}_{t=1}^{\mathscr{T}}$ and the sample mean of $\{f_t^*\}_{t=1}^{T}$, where T is the number of historical observations on f_t^*.

Several considerations motivate the simultaneous solution of the model and SME estimation of β. First, solving for the stochastic equilibrium of the model permits an assessment of the goodness-of-fit directly in terms of aspects of the joint distribution of asset returns, consumption, and capital.[2] Furthermore, estimation of asset-pricing models using Euler equations (Hansen and Singleton (1982)) is not always feasible, as in the version of this model with taste shocks. Third, temporal aggregation may lead to inconsistent GMM estimators of β_0 (Hall (1988), Hansen and Singleton (1989)), but temporal aggregation can often be accommodated using the SME.

For several reasons, this illustrative estimation problem is not a special case of either Hansen's (1982) GMM estimation problem or the simulated moments problems examined by McFadden (1989) and Pakes and Pollard (1989), or Lee and Ingram (1991). The most important difference between the estimation problem with simulated time series and the GMM estimation problem discussed by Hansen (1982) lies in the parameter dependency of the simulated time series $\{f_t^\beta\}$. In the stationary, ergodic environment studied by Hansen (1982), one observes $f(Y_t, \beta_0)$, where the data generation process $\{Y_t\}$ is fixed and β_0 is the parameter vector to be estimated. In contrast, $f_t^\beta = f(Y_t^\beta, \beta)$ depends on β not only directly, but indirectly through the dependence of the entire past history of the simulated process $\{Y_t^\beta\}$ on β. In Section 4, we present versions of uniform

[2] Several alternative numerical methods for solving discrete-time dynamic rational expectations models have recently been proposed in the literature; see Taylor and Uhlig (1990), Tauchen and Hussey (1991), and the references cited therein for useful summaries. Many of the algorithms discussed involve approximations to either the distributions of the forcing variables or the model itself. Additional approximations are involved when the underlying model is expressed in continuous time and a discrete-time approximation is being estimated. These approximations affect the large sample properties of the SME since, as sample size increases, one obtains a consistent estimator of the approximate model. At a minimum, the methods described in this paper apply to the approximate model if approximations are used to solve for equilibrium asset prices. They may apply to the original model if the approximation error can be made negligible as the sample size increases.

weak and strong laws of large numbers that accommodate this parameter dependency of the data generation process for simulated time series.

Furthermore, in contrast to the simulated moments estimators for i.i.d. environments, the simulation of time series requires initial conditions for the forcing variables Y_t. Even if the transition function of the Markov process $\{Y_t\}$ is stationary (that is, has a stationary distribution), the simulated process $\{Y_t^\beta\}$ is not generally stationary since the initial simulated state Y_1^β is typically not drawn from the ergodic distribution of the process. In this case, the simulated process $\{f_t^\beta\}$ is nonstationary.

A related initial conditions problem, common to the GMM and SM estimation of asset-pricing models, occurs with capital accumulation. Specifically, the current equilibrium capital stock can typically be expressed as a function of the previous period's stock plus investment in new capital. Measurements of investment are often more reliable than measurements of the stock of capital, which may not be based on compatible assumptions about depreciation. Accordingly, in constructing a time series on the capital stock to be used in estimation, one may wish to accommodate mismeasurement of the initial stock.[3]

In Section 4, we present a set of sufficient conditions for the Markov process $\{Y_t\}$ to be geometrically ergodic, which (among other things) implies that the large-sample properties of functions of Y_t are invariant to the choice of initial conditions used in simulating both exogenous (taste and technology shocks) and endogenous (e.g., the capital stock) state variables.

Throughout this discussion we have assumed that the Markov process described by (2.5) does not exhibit growth. In fact, there is real growth in output, and hence in certain asset prices. If the technology shock $\{z_t\}$, for instance, exhibits growth over time, then the implied trends for the components of Y_t are restricted by the structure of the model.[4] Conversely, the structure of the model restricts the class of admissible trend specifications. Furthermore, accommodating these trends typically requires that the implied form of the trends in Y_t is known, and that it is possible to build an adjustment for trends directly into the function f of the data and to simulate a trend-free version of the model.

Following Eichenbaum and Hansen (1988), the implied restrictions on deterministic trends in the decision variables can be imposed in estimation by appending the moment conditions associated with least squares estimation of the trend equations to the moment equations involving f^* and f^β. The subsequent discussion in this paper extends to this case using arguments similar to those in Eichenbaum and Hansen (1988) for GMM estimators of (2.11). If the forcing variables exhibit stochastic trends (unit roots), then our estimation

[3] See Dunn and Singleton (1986); Eichenbaum, Hansen, and Singleton (1988); and Eichenbaum and Hansen (1988) for examples of studies of Euler equations using GMM estimators in which this type of initial condition problem arises.

[4] See Eichenbaum and Hansen (1988) and Eichenbaum, Hansen and Singleton (1988) for a discussion of restrictions on trends implied by Euler equations. Singleton (1987) discusses the analogous restrictions on deterministic seasonal components of agents' decision variables.

strategy applies only if the entire model, including the forcing variables, can be transformed to a model expressed in terms of trend-free processes.

3. THE ESTIMATION PROBLEM

This section defines the simulated moments estimator. The basic primitives for the model are:

(i) a measurable *transition function* H: $\mathbb{R}^N \times \mathbb{R}^p \times \Theta \to \mathbb{R}^N$, with compact parameter set $\Theta \subset \mathbb{R}^Q$, for some positive integers N, p, and Q;

(ii) a measurable *observation function* f: $\mathbb{R}^{Nl} \times \Theta \to \mathbb{R}^M$, for positive integers l and M, with $M \geqslant Q$.

A given $\mathbb{R}^N$-valued *state process* $\{Y_t\}_{t=1}^{\infty}$ is generated by the difference equation

$$(3.1) \qquad Y_{t+1} = H(Y_t, \varepsilon_{t+1}, \beta_0),$$

where the parameter vector β_0 is to be estimated, and $\{\varepsilon_t\}$ is an i.i.d. sequence of $\mathbb{R}^p$-valued random variables on a given probability space $(\Omega, \mathscr{F}, P)$. The function H may be determined implicitly by the numerical solution of a model for equilibrium asset prices. Let $Z_t = (Y_t, Y_{t-1}, \dots, Y_{t-l+1})$ for some positive integer $l < \infty$. Estimation of β_0 is based on moments of the vector $f_t^* \equiv f(Z_t, \beta_0)$.

For certain special cases of (3.1) and f, the function mapping β to $E[f(Z_t, \beta)]$ is known and independent of t. In these cases, the GMM estimator,

$$(3.2) \qquad b_T = \underset{\beta \in \Theta}{\arg\min} \left[\frac{1}{T} \sum_{t=1}^{T} f_t^* - E[f(Z_t, \beta)] \right]' W_T \left[\frac{1}{T} \sum_{t=1}^{T} f_t^* - E[f(Z_t, \beta)] \right],$$

for given "distance matrices" $\{W_T\}$, is consistent for β_0 and asymptotically normal under regularity conditions in, for example, Hansen (1982). The requirement that $\beta \mapsto E[f(Z_t, \beta)]$ is known, however, limits significantly the applicability of the GMM estimator to asset-pricing problems.

The simulated moments estimator circumvents this limitation by making the much weaker assumption that the econometrician has access to an $\mathbb{R}^p$-valued sequence $\{\hat{\varepsilon}_t\}$ of random variables that is identical in distribution to, and independent of, $\{\varepsilon_t\}$. Then, for any $\mathbb{R}^N$-valued initial point $\hat{Y}_1$ and any parameter vector $\beta \in \Theta$, the *simulated state process* $\{Y_t^\beta\}$ can be constructed inductively by letting $Y_1^\beta = \hat{Y}_1$ and

$$(3.3) \qquad Y_{t+1}^\beta = H(Y_t^\beta, \hat{\varepsilon}_{t+1}, \beta).$$

Likewise, the simulated observation process $\{f_t^\beta\}$ is constructed by $f_t^\beta = f(Z_t^\beta, \beta)$, where $Z_t^\beta = (Y_t^\beta, \dots, Y_{t-l+1}^\beta)$. Finally, the SME of β_0 is the parameter vector b that best matches the sample moments of the actual and simulated observation processes, $\{f_t^*\}$ and $\{f_t^b\}$.

More precisely, let $\mathscr{T}$: $\mathbb{N} \to \mathbb{N}$ define the simulation sample size $\mathscr{T}(T)$ that is generated for a given sample size T of actual observations, where $\mathscr{T}(T) \to \infty$ as

$T \to \infty$. For any parameter vector β, let

$$G_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} f_t^* - \frac{1}{\mathscr{T}(T)}\sum_{s=1}^{\mathscr{T}(T)} f_s^\beta \tag{3.4}$$

denote the difference in sample moments. If $\{f_t^*\}$ and $\{f_s^\beta\}$ satisfy a law of large numbers, then $\lim_T G_T(\beta) = 0$ if $\beta = \beta_0$. With identification conditions, $\lim_T G_T(\beta) = 0$ if and only if $\beta = \beta_0$. We therefore introduce a sequence $W = \{W_T\}$ of $M \times M$ positive semi-definite matrices and define the *simulated moments estimator* for β_0 given $(H, \varepsilon, \mathscr{T}, \hat{Y}_1, W)$ to be the sequence $\{b_T\}$ given by

$$b_T = \underset{\beta \in \Theta}{\arg\min}\; G_T(\beta)' W_T G_T(\beta) \equiv \underset{\beta \in \Theta}{\arg\min}\; C_T(\beta). \tag{3.5}$$

The distance matrix W_T is chosen with rank at least Q, and may depend on the sample information $\{f_1^*, \ldots, f_T^*\} \cup \{f_1^\beta, \ldots, f_{\mathscr{T}(T)}^\beta : \beta \in \Theta\}$.

Comparing (3.2) and (3.5) shows that the SME extends the method-of-moments approach to estimation by replacing the population moment $E[f(Z_t, \beta)]$ with its sample counterpart, calculated with simulated data. The latter sample moment can be calculated for a large class of asset-pricing models. Extensions of the SME are provided in Section 6.

4. CONSISTENCY

The presence of simulation in the estimator pushes one to special lengths in justifying regularity conditions for the consistency of method-of-moments estimators that, without simulation, are often taken for granted. As illustrated in Section 2, there are two particular problems. First, since the simulated state process is usually not initialized with a draw from its ergodic distribution, one needs a condition that allows the use of an arbitrary initial state, knowing that the state process converges rapidly to its stationary distribution. Second, one needs to justify the usual starting assumption of some form of uniform continuity of the observation as a function of the parameter choice. With simulation, a perturbation of the parameter choice affects not only the current observation, but also affects transitions between past states, a dependence that compounds over time. We will present a natural (but restrictive) condition directly on the state transition function guaranteeing that this compounding effect is of a damping, rather than exploding, variety.

Initially we describe the concept of geometric ergodicity, a condition ensuring that the simulated state process satisfies a law of large numbers with an asymptotic distribution that is invariant to the choice of initial conditions. Then ergodicity of the simulated series is used to prove a uniform weak law of large numbers for $G_T(\beta)$ and weak consistency of the SME (that is, $b_T \to \beta_0$ in probability). Weak consistency is proved under a global modulus-of-continuity condition rather than the more usual local condition underlying proofs of strong consistency. Subsequently, we present Lipschitz and modulus of continuity conditions on the primitives (H, ε, f) that are sufficient for strong consistency (that is, $b_T \to \beta_0$ almost surely). Though weaker than the damping conditions typically used to verify near-epoch dependence (Gallant and White (1988)), these conditions nevertheless exclude an important class of geometrically ergodic processes. This fact is the primary reason for our initial focus on weak consistency. Finally, various tradeoffs in choosing among the regularity conditions leading to weak and strong consistency are discussed in the context of the illustrative model presented in Section 2.

4.1. *Geometric Ergodicity*

In order to define geometric ergodicity, let P_x^t denote the t-step transition probability for a time-homogeneous Markov process $\{X_t\}$; that is, P_x^t is the distribution of X_t given the initial point $X_0 = x$. The process $\{X_t\}$ is *ρ-ergodic*, for some $\rho \in (0,1]$, if there is a probability measure π on the state space of the process such that, for every initial point x,

$$(4.1) \qquad \rho^{-t}\|P_x^t - \pi\|_v \to 0 \quad \text{as} \quad t \to \infty,$$

where $\|\cdot\|_v$ is the total variation norm.[5] The measure π is the ergodic distribution. If $\{X_t\}$ is ρ-ergodic for $\rho < 1$, then $\{X_t\}$ is *geometrically ergodic*. In calculating asymptotic distributions, geometric ergodicity can substitute for stationarity since it means that the process converges geometrically to its stationary distribution. Moreover, geometric ergodicity implies strong (α) mixing in which the mixing coefficient $\alpha(m)$ converges geometrically with m to zero (Rosenblatt (1971), Mokkadem (1985)).

In what follows, for any ergodic process $\{X_t\}$, it is convenient for us to write "X_∞" for any random variable with the corresponding ergodic distribution. We adopt the notation $\|X\|_q = [E(\|X\|^q)]^{1/q}$ for the L^q norm of any $\mathbb{R}^N$-valued random variable X, for any $q \in (0, \infty)$. We let L^q denote the space of such X with $\|X\|_q < \infty$, and let $\|x\|$ denote the usual Euclidean norm of a vector x.

General criteria for the geometric ergodicity of a Markov chain have been obtained by Nummelin and Tuominen (1982) and by Tweedie (1982). We will review simple sufficient conditions established by Mokkadem (1985) for the special case of nonlinear AR(1) models, which includes our setting.

A key ingredient for ergodicity is positive recurrence,[6] for which a key condition is irreducibility. For a finite Markov chain, irreducibility means essentially that each state is accessible from each state, obviously a sufficient condition in this case for both recurrence and geometric ergodicity. Mokkadem (1985) uses the following convenient sufficient condition for irreducibility of a time-homogeneous Markov chain $\{X_t\}$ valued in $\mathbb{R}^N$ with t-step transition probability P_x^t.

[5] The total variation of a signed measure μ is $\|\mu\|_v = \sup_{h:\,|h(y)| \leqslant 1} \int h(y)\, d\mu(y)$.

[6] For a finite-state Markov chain, recurrence means essentially that each state occurs infinitely often from any given state. See, for example, Doob (1953) for some general definitions.

CONDITION B: *For any measurable $A \subset \mathbb{R}^N$ of nonzero Lebesgue measure and any compact $K \subset \mathbb{R}^N$, there exists some integer $t > 0$ such that*

$$\inf_{x \in K} P_x^t(A) > 0. \tag{4.2}$$

It is obviously enough that $P_x^1(A)$ is continuous in x and supports all of $\mathbb{R}^N$ for each x, but this single-period "full support" condition is too strong an assumption in a setting with endogenous state variables. For example, the process for Y_t given by (2.6) fails this single-period full-support condition because the distribution of the capital stock k_{t+1} given X_t is degenerate, but often passes the weaker Condition B. To be more concrete, consider the special case of (2.1)–(2.6) with $u_t = 1$ for all t, $\mu = 0$ (100% depreciation), and $\alpha = 1$ (logarithmic utility). Also, suppose that the law of motion for the technology shock is given by

$$\ln z_{t+1} = \zeta_z + \rho \ln z_t + \varepsilon_{t+1}, \tag{4.3}$$

for constants ζ_z and ρ. Under these simplifying assumptions, the implied equilibrium asset-pricing function and law of motion for the capital stock are (Michner (1984)):

$$p_t = \frac{\delta}{(1-\delta)}(1-\phi) z_t k_t^{\phi}, \tag{4.4}$$

$$d_t = (1-\phi) z_t k_t^{\phi},$$

$$k_{t+1} = \delta \phi z_t k_t^{\phi}. \tag{4.5}$$

If $\{\varepsilon_t\}$ is say i.i.d. normal, then $\{Y_t\}$ for this illustrative economy satisfies Condition B. More generally, Condition B is not a strong condition on models with endogenous state variables provided the endogenous state variables do not move in such a way that some states are inaccessible from others.

If the state process $\{X_t\}$ is valued in a proper subset S of $\mathbb{R}^N$, Condition B obviously does not apply, but analogous results hold if Condition B applies when substituting S everywhere for $\mathbb{R}^N$ (and relatively open sets for sets of nonzero Lebesgue measure).

A second key ingredient for ergodicity is aperiodicity. For example, the Markov chain that alternates deterministically from "heads" to "tails" to "heads" to "tails," and so on, is not geometrically ergodic, despite its recurrence.

With these definitions in hand, we can review Mokkadem's sufficient conditions for geometric ergodicity of what he calls "nonlinear AR(1) models," which includes our setting.

LEMMA 1 (Mokkadem): *Suppose $\{Y_t\}$, as defined by* (3.1), *is aperiodic and satisfies Condition B. Fix β and suppose there are constants $K > 0$, $\delta \in (0,1)$, and*

$q > 0$ *such that* $H(\cdot, \varepsilon_1, \beta)$: $\mathbb{R}^N \to L^q$ *is well defined and continuous with*

$$\|H(y, \varepsilon_1, \beta)\|_q < \delta\|y\|, \qquad \|y\| > K. \tag{4.6}$$

Then $\{Y_t\}$ *is geometrically ergodic. Moreover,* $\{\|Y_t^\beta\|_q\}$ *and* $\|Y_\infty^\beta\|_q$ *are uniformly bounded over* t.

Condition (4.6), inspired by Tweedie (1982), means roughly that $\{Y_t\}$, once outside a sufficiently large ball, heads back into the ball at a uniform rate.

4.2. *A Uniform Weak Law of Large Numbers*

Since geometric ergodicity of $\{Y_t^\beta\}$ implies α-mixing, it also implies that $\{Y_t^\beta\}$ satisfies a strong (and hence weak) law of large numbers. For consistency of the SME estimator, however, standard sufficient conditions require that a strong or weak law holds in a uniform sense over the parameter space Θ. For example, the family $\{\{f_t^\beta\}: \beta \in \Theta\}$ of processes satisfies the uniform weak law of large numbers if, for each $\delta > 0$,

$$\lim_{T \to \infty} P\left[\sup_{\beta \in \Theta}\left|E(f_\infty^\beta) - \frac{1}{T}\sum_{t=1}^{T} f_t^\beta\right| > \delta\right] = 0. \tag{4.7}$$

In our setting of simulated moments, $\{Z_t^\beta\}$ is simulated based on various choices of β, so continuity of $f(Z_t^\beta, \beta)$ in β (via both arguments) is useful in proving (4.7). We will use the following global modulus of continuity condition on $\{f_t^\beta\}$.

DEFINITION: The family $\{f_t^\beta\}$ is *Lipschitz, uniformly in probability*, if there is a sequence $\{K_t\}$ such that, for all t and all β and θ in Θ,

$$\|f_t^\beta - f_t^\theta\| \leqslant K_t\|\beta - \theta\|,$$

where $K^T = T^{-1}\Sigma_{t=1}^T K_t$ is bounded (with T) in probability.

LEMMA 2 (Uniform Weak Law of Large Numbers): *Suppose, for each* $\beta \in \Theta$, *that* $\{Y_t^\beta\}$ *is ergodic and that* $E(|f_\infty^\beta|) < \infty$. *Suppose, in addition, that the map* $\beta \mapsto E(f_\infty^\beta)$ *is continuous and the family* $\{f_t^\beta\}$ *is Lipschitz, uniformly in probability. Then* $\{\{f_t^\beta\}: \beta \in \Theta\}$ *satisfies the uniform weak law of large numbers.*

The proofs of this and all subsequent propositions in Section 4 are provided in the Appendix.

The ergodicity assumption on $\{Y_t^\beta\}$ in Lemma 2 can be replaced with Mokkadem's conditions for geometric ergodicity on the transition function H and disturbance ε_t, summarized in Lemma 1.

4.3. *Weak Consistency*

Next, we summarize several important assumptions that are used in our proofs of both consistency and asymptotic normality of the SME.

ASSUMPTION 1 (Technical Conditions): *For each $\beta \in \Theta$, $\{\|f_t^\beta\|_{2+\delta}: t = 1, 2, \ldots\}$ is bounded for some $\delta > 0$. The family $\{f_t^\beta\}$ is Lipschitz, uniformly in probability, and $\beta \mapsto E(f_\infty^\beta)$ is continuous.*

ASSUMPTION 2 (Ergodicity): *For all $\beta \in \Theta$, the process $\{Y_t^\beta\}$ is geometrically ergodic.*

The hypotheses of Lemmas 1 and 2 are sufficient for Assumptions 1 and 2 provided Mokkadem's conditions apply for some $q > 2$.

We impose the following condition on the distance matrices $\{W_T\}$ in (3.5).

ASSUMPTION 3 (Convergence of Distance Matrices): *Σ_0 is nonsingular and $W_T \to W_0 = \Sigma_0^{-1}$ almost surely, where (for any t)*

$$\Sigma_0 \equiv \sum_{j=-\infty}^{\infty} E\left(\left[f_t^* - E(f_t^*)\right]\left[f_{t-j}^* - E(f_{t-j}^*)\right]'\right). \tag{4.8}$$

For the second moments in this assumption to exist, and their sum to converge absolutely, the assumptions that $\{\|f_t^*\|_{2+\delta}: t = 1, 2, \ldots\}$ is bounded for some $\delta > 0$ and geometric ergodocity of $\{Y_t\}$ together suffice, as shown by Doob (1953, pp. 222–224). Also, as with Hansen's (1982) GMM estimator, the choice of W_0 in Assumption 3 leads to the most efficient SME within the class of SME's with positive definite distance matrices.

Notice that Σ_0 in Assumption 3 is a function of the moments of $\{f_t^*\}$ alone; in particular, Σ_0 depends neither on β nor on the moments of the simulated process $\{f_t^\beta\}$. Thus, Σ_0 can be estimated using, for instance, the approaches discussed by Andrews (1991).[7] Given the definition of Σ_0 and the fact that geometric ergodicity implies α-mixing, it follows that the Newey-West estimator is consistent for Σ_0 in our environment.

Alternatively, Σ_0 could be estimated using simulated data $\{f_t^\beta\}$. Since the rate of convergence of spectral estimators is slow and one has control over the size $\mathcal{T}(T)$ of the simulated sample, this alternative may be relatively advantageous. A two-step procedure for estimating Σ_0 is required, however, so in establishing consistency of a simulated estimator of Σ_0 one would need to account both for dependence of $\{f_t^\beta\}$ on an estimated value of β and the parameter dependence of simulated series. One approach to establishing consistency would be to

[7] Several estimators of Σ_0 have been proposed in the literature. See, for example, Hansen and Singleton (1982), Eichenbaum, Hansen, and Singleton (1988), and Newey and West (1987). In general, $E[f_t^* - Ef_t^*)(f_{t-j}^* - Ef_{t-j}^*)']$ is nonzero for all j in (4.8) and the Newey-West estimator is appropriate.

extend the discussion of consistent estimation of spectral density functions using estimated residuals without simulation, found in Newey and West (1987) and Andrews (1991), to the case of simulated residuals.

Under Assumptions 1–3, the criterion function $C_T(\beta)$ converges almost surely to the asymptotic criterion function C: $\Theta \to \mathbb{R}$ defined by $C(\beta) = G_\infty(\beta)' W_0 G_\infty(\beta)$.

ASSUMPTION 4 (Uniqueness of Minimizer): $C(\beta_0) < C(\beta)$, $\beta \in \Theta$, $\beta \neq \beta_0$.

Our first theorem establishes the consistency of the SME $\{b_T: T \geqslant 1\}$ given by (3.5).

THEOREM 1 (Consistency of SME): *Under Assumptions* 1–4, *the SME* $\{b_T\}$ *converges to* β_0 *in probability as* $T \to \infty$.

4.4. *Strong Consistency*

The Uniform Weak Law of Large Numbers (UWLLN) underlying the discussion in Sections 4.2 and 4.3 maintained the uniform continuity condition in Assumption 1. In this subsection we provide primitive conditions on H, ε, and f for a local modulus of continuity condition with simulation, and thereby explore in more depth the nature of the requirements in simulation environments for $\{f_t^\beta\}$ to satisfy the Uniform Strong Law of Large Numbers (USLLN):

$$\sup_{\beta \in \theta} \left| \frac{1}{T} \sum_{t=1}^{T} f_t^\beta - E\left(f_\infty^\beta\right) \right| \xrightarrow{\text{a.s.}} 0 \quad \text{as} \quad T \to \infty.$$

The basic nature of the conditions are of three forms: continuity conditions, growth conditions, and a contraction (or "damping") condition on the transition function H that we call an "asymptotic unit-circle (AUC) condition."

Our proof of strong consistency of the SME proceeds in three steps.[8] First, we introduce the AUC condition, which assures that current shocks have a damping effect on future simulated observations. Under the AUC condition, it is shown that, for each β, there exists a stationary and ergodic process $\{Y_t^{\infty\beta}\}$ that satisfies (3.1) and can be substituted for $\{Y_t^\beta\}$ in proving consistency (and asymptotic normality) of the SME. Second, we show that the AUC condition and certain continuity and growth conditions imply a version of Hansen's (1982) modulus of continuity condition for simulation environments. Strong consistency of the SME then follows from results in Hansen (1982).

DEFINITION (The Asymptotic Unit-Circle Condition): The transition function H and shock process ε satisfy the *Asymptotic Unit-Circle Condition* if, for each

[8] The strategy of using a unit-circle condition with a Lipschitz coefficient that changes geometrically toward zero in proving strong consistency of the SME was suggested to us by Lars Hansen in his discussion of an earlier version of this paper.

$\theta \in \Theta$, there is some $\delta > 0$ and a sequence of positive random variables $\{\rho_\theta(\varepsilon_t)\}$ satisfying

$$(4.9) \qquad \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \ln \rho_\theta(\varepsilon_t) = \alpha_\theta < 0 \quad \text{a.s.}$$

such that, whenever $\|\beta - \theta\| \leqslant \delta$, for any x and y,

$$\|H(y, \beta, \varepsilon_t) - H(x, \beta, \varepsilon_t)\| \leqslant \rho_\theta(\varepsilon_t)\|y - x\|.$$

In other words, for the AUC condition, $H(\cdot, \beta, \varepsilon_t)$ must have a Lipschitz coefficient $\rho_\theta(\varepsilon_t)$ with the property that $\Pi_{s=0}^{t} \rho_\theta(\varepsilon_s)$ declines geometrically toward zero as $t \to \infty$. This is a weaker requirement than the unit-circle condition used by Gallant and White (1988) to verify near-epoch dependence of a process.

We say that f is Θ-locally Lipschitz if, for each $\theta \in \Theta$, there is a δ and a constant k such that, whenever $\|\beta - \theta\| \leqslant \delta$, the function $f(\cdot, \beta)$ has the Lipschitz constant k. Next, we define f to be S-smooth (sufficiently smooth) if f is Θ-locally Lipschitz and, for each $z \in \mathbb{R}^{Nl}$, the function $f(z, \cdot)$: $\Theta \to \mathbb{R}^p$ has a Lipschitz constant $C_1(z)$, where C_1 satisfies a growth condition.[9] Obviously, if f is Lipschitz, then f is S-smooth, but a Lipschitz condition is unnecessarily strong and is not satisfied in many applications. (Take, for example, $f(z, \beta) = \beta z$.) We say that H is S-smooth if, for each $\theta \in \Theta$, there is a δ small enough that $\|\beta - \theta\| \leqslant \delta$ implies that, for all $y \in \mathbb{R}^N$ and $\varepsilon \in \mathbb{R}^p$,

$$\|H(y, \beta, \varepsilon) - H(y, \theta, \varepsilon)\| \leqslant C_2(y, \varepsilon)\|\beta - \theta\|,$$

where C_2 satisfies a growth condition.

The smoothness assumption on f and the AUC condition imply that the nonstationarity induced by the initial conditions problem can be ignored when studying the large sample properties of the SME. We establish this result in the following two lemmas.

LEMMA 3: *If (H, ε) satisfies the AUC condition, then for each β in Θ there exists a stationary and ergodic process $\{Y_t^{\infty\beta}\colon -\infty < t < \infty\}$ such that, for all t, $Y_t^{\infty\beta}$ is measurable with respect to $\{\hat{\varepsilon}_{t-s}\colon s \geqslant 0\}$ and $Y_{t+1}^{\infty\beta} = H(Y_t^{\infty\beta}, \hat{\varepsilon}_{t+1}, \beta)$.*

Next we argue that $\{Y_t^\beta\}$, simulated with an arbitrary initial condition, can be replaced by $\{Y_t^{\infty\beta}\}$ for the purpose of proving a USLLN.

LEMMA 4: *If f is S-smooth and (H, ε) satisfies the AUC condition, then*

$$(4.10) \qquad \sup_{\beta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} f_t^\beta - \frac{1}{T} \sum_{t=1}^{T} f_t^{\infty\beta} \right| \xrightarrow{\text{a.s.}} 0 \quad \text{as} \quad T \to \infty,$$

where $f_t^{\infty\beta} = f[(Y_t^{\infty\beta}, Y_{t-1}^{\infty\beta}, \ldots, Y_{t-l+1}^{\infty\beta}), \beta]$.

[9] A real-valued function F on a Euclidean space satisfies a growth condition if there exist constants k and K such that for x, $|F(x)| \leqslant k + K\|x\|$.

The final step in proving strong consistency of the SME is showing that $\{f_t^{\infty\beta}\}$ satisfies a USLLN. Toward this end, for each θ in Θ and $\delta > 0$, let

$$\operatorname{mod}_t(\delta, \theta) \equiv \sup\left\{\|f_t^{\infty\beta} - f_t^{\infty\theta}\| : \|\beta - \theta\| < \delta, \beta \in \Theta\right\}$$

denote the "modulus of continuity" of the process $\{f_t^{\infty\beta}\}$ at θ, defined ω by ω. Consider the following:

ASSUMPTION 5: *For each $\theta \in \Theta$, there is a $\delta > 0$ such that $E[\operatorname{mod}_t(\delta, \theta)] < \infty$.*

With this, combined with our earlier assumptions, Hansen's (1982) Theorem 2.1 implies that $\{f_t^{\infty\beta}\}$ satisfies a USLLN and that $\{b_T\}$ is a strongly consistent estimator of β_0. We summarize with the following theorem.

THEOREM 2 (Strong Consistency): *Under Assumptions* 3–5, *the AUC condition, and the assumption that f is S-smooth, the SME $\{b_T\}$ converges to β_0 almost surely as $T \to \infty$.*

The assumption in Theorem 2 that $E[\operatorname{mod}_t(\delta, \theta)] < \infty$ is not known to be implied by the AUC condition. However, by strengthening the statement of the AUC condition, Assumption 5 becomes redundant. Specifically, we introduce the following strong AUC condition:

DEFINITION (L^2 Unit-Circle Condition): The transition function H and the shock process ε satisfy the L^2 *Unit-Circle condition* if, for each $\theta \in \Theta$, there is some $\delta > 0$ and a sequence of positive random variables $\{\rho_\theta(\varepsilon_t)\}$ satisfying $E[\rho_\theta(\varepsilon_t)^2] < 1$ such that, whenever $\|\beta - \theta\| \leqslant \delta$, for all x and y,

$$\|H(y, \beta, \varepsilon_t) - H(x, \beta, \varepsilon_t)\| \leqslant \rho_\theta(\varepsilon_t)\|y - x\|.$$

By Jensen's inequality, $\ln E[\rho_\theta(\varepsilon_t)] > E[\ln \rho_\theta(\varepsilon_t)]$, so that the L^2 Unit-Circle Condition (L^2 UC condition) implies the AUC condition. Hence the lemmas preceeding Theorem 2 continue to hold under the L^2 UC condition.

This strengthening of the unit-circle condition leads to the following theorem.

THEOREM 3: *Under Assumptions* 3–4, *the assumption that H and f are S-smooth, and the L^2 UC condition, the SME is a strongly consistent estimator of β_0.*

4.5. *Regularity Conditions and Dynamic Asset-Pricing Models*

Weak consistency was established by assuming that the simulated processes are geometrically ergodic and that $\{f_t^\beta\}$ satisfies a uniform Lipschitz condition in β. In contrast, strong consistency was established assuming a unit-circle condition on the transition function H and an i.i.d. shock process $\{\varepsilon_t\}$. Thus, the AUC condition substitutes in part for the Lipschitz condition in Assumption 1 and in part for geometric ergodicity in Assumption 2. Indeed, the L^2 UC condition implies geometric ergodicity. On the other hand, there is an important class of geometrically ergodic processes that do not satisfy the L^2 UC condition, and this is a primary motivating reason for our analysis of weak consistency.

In order to see this, consider again the example in Section 2 and suppose that the law of motion of the technology shock is given by

$$z_t = \xi + \rho z_{t-1} + \sigma \nu_{t-1}^{\gamma} \varepsilon_t, \qquad \gamma < 1, \quad \sigma > 0, \quad |\rho| < 1, \tag{4.11}$$

where $\nu_t = z_t$ if $z_t \geqslant \eta > 0$ and $\nu_t = \eta$ otherwise, and suppose that $E(\varepsilon_t) = 0$ for all t. This representation of a shock process is similar to several widely studied representations of conditionally heteroskedastic processes. Let $h(z, \varepsilon, \beta)$ denote the right hand side of (4.11). Then

$$\| h(z, \varepsilon, \beta) - h(z', \varepsilon, \beta) \|_2 = \left\| \rho + \sigma \varepsilon \frac{(\nu^{\gamma} - \nu'^{\gamma})}{(z - z')} \right\|_2 \| z - z' \|.$$

The ratio $(\nu^{\gamma} - \nu'^{\gamma})/(z - z')$ can be made arbitrarily large, as $\nu_t \to \eta$ for small η, in which case the factor of proportionality for $\|z - z'\|$ exceeds unity. Similarly, if ρ, σ, and the variance of ε are sufficiently large, then the unit-circle condition may be violated. This is the case, for example, if $\gamma = 1$ and $\|\rho + \sigma\varepsilon\|_2 > 1$. Furthermore, from the proofs of Lemmas 3 and 4, it is apparent that this process will not in general satisfy the AUC condition used to prove Theorem 2.

The process (4.11) is nevertheless geometrically ergodic. This can be verified easily by noting that $|\rho| < 1$ and $\|z^{\gamma}\|/\|z\|$ can be made arbitrarily small for large enough z when $\gamma < 1$. Thus, the process $\{z_t\}$ satisfies strong and weak laws of large numbers. If, in addition, $\{Y_t^{\beta}\}$ satisfies Condition B and our weak uniform continuity condition is satisfied, then weak consistency of the SME is implied by the UWLLN (Lemma 2).

Though the geometric ergodicity assumption accommodates more general processes than the AUC condition, our consistency proof based on the former requires the imposition of a uniform Lipschitz condition. This uniform continuity condition implicitly requires some damping of the effects of past shocks on current values of Y^{β}. We have not shown that processes of the form (4.11), for example, satisfy our uniform Lipschitz condition. Verifying this condition may well narrow the gap between the classes of models encompassed by the sets of regularity conditions used to prove weak and strong consistency of the SME.

5. ASYMPTOTIC NORMALITY

Under the unit-circle conditions introduced in Section 4.4, the stationary and ergodic process $\{Y_t^{\infty\beta}\}$ can be substituted for $\{Y_t^{\beta}\}$ in deducing the asymptotic distribution of the SME. Thus, the asymptotic normality of $\{b_T\}$ follows immediately under suitably modified versions of the regularity conditions imposed by Hansen (1982). If, instead, the regularity conditions used to prove weak consistency in Section 4.3 are adopted, then Hansen's (1982) conditions are no longer directly applicable because of the nonstationarity of $\{Y_t^\beta\}$. Therefore, our discussion of asymptotic normality focuses on the case of geometrically ergodic forcing processes that may not satisfy an AUC condition. The final characterization of the limiting distribution of the SME is, of course, the same for either set of regularity conditions.

In deriving the asymptotic distribution of $\{\sqrt{T}(b_T-\beta_0)\}$, we use an intermediate-value expansion of $G_T(\beta)$ about the point β_0. Accordingly, we will adopt the following assumption.

ASSUMPTION 6:
(i) β_0 *and the estimators* $\{b_T\}$ *are interior to* Θ.
(ii) f_t^β *is continuously differentiable with respect to* β *for all* t, ω *by* ω.
(iii) $D_0 \equiv E[\partial f_\infty^{\beta_0}/\partial\beta]$ *exists, is finite, and has full rank.*

Expanding $G_T(b_T)$ about β_0 gives

$$G_T(b_T)=G_T(\beta_0)+\partial G^*(T)(b_T-\beta_0), \tag{5.1}$$

where (using the intermediate value theorem) $\partial G^*(T)$ is the $M\times Q$ matrix whose ith row is the ith row of $\partial G_T(b_T^i)/\partial\beta$, with b_T^i equal to some convex combination of β_0 and b_T. Premultiplying (5.1) by $[\partial G_T(b_T)/\partial\beta]'W_T$, and applying the first order conditions for the optimization problem defining b_T,

$$\left[\frac{\partial G_T(b_T)}{\partial\beta}\right]'W_TG_T(b_T)=0=\left[\frac{\partial G_T(b_T)}{\partial\beta}\right]'W_TG_T(\beta_0)+J_T(b_T-\beta_0), \tag{5.2}$$

where

$$J_T=\left[\frac{\partial G_T(b_T)}{\partial\beta}\right]'W_T\,\partial G^*(T).$$

Equation (5.2) can be solved for $b_T-\beta_0$ if J_T is invertible for sufficiently large T. This invertibility is given by Assumption 5 (iii) provided $\partial G_T(b_T)/\partial\beta$ converges in probability to D_0. For notational ease, let $D_\beta f_t^\beta=(d/d\beta)f(Z_t^\beta,\beta)$ (the total derivative). Under the following additional assumptions, Lemma 2 and Theorem 4.1.5 of Amemiya (1985) imply that $\operatorname{plim}_T \partial G_T(b_T)/\partial\beta=D_0$.

ASSUMPTION 7: *The family* $\{D_\beta f_t^\beta\colon \beta\in\Theta,\ t=1,2,\dots\}$ *is Lipschitz, uniformly in probability. For all* $\beta\in\Theta$, $E(|D_\beta f_\infty^\beta|)<\infty$, *and* $\beta\mapsto E(D_\beta f_\infty^\beta)$ *is continuous.*

Under these assumptions, the asymptotic distribution of $\sqrt{T}(b_T-\beta_0)$ is equivalent to the asymptotic distribution of $(D_0'\Sigma_0^{-1}D_0)^{-1}\sqrt{T}\,G_T(\beta_0)$. The following theorem provides the limiting distribution of $\sqrt{T}\,G_T(\beta_0)$.

THEOREM 4: *Suppose $T/\mathscr{T}(T) \to \tau$ as $T \to \infty$. Under Assumptions* 1–4, *and* 6–7,

$$\sqrt{T}\, G_T(\beta_0) \Rightarrow N[0, \Sigma_0(1+\tau)]. \tag{5.3}$$

PROOF: From the definition of G_T,

$$\sqrt{T}\, G_T(\beta_0) = \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} [f_t^* - E(f_\infty^*)] \right) - \frac{\sqrt{T}}{\sqrt{\mathscr{T}(T)}} \left(\frac{1}{\sqrt{\mathscr{T}(T)}} \sum_{s=1}^{\mathscr{T}(T)} [f_s^{\beta_0} - E(f_\infty^{\beta_0})] \right). \tag{5.4}$$

We do not have stationarity, but the proof of asymptotic normality of each term on the right-hand side of (5.4) follows Doob's (1953) proof of a central limit theorem (Theorem 7.5), which uses instead the stronger geometric ergodicity condition. In particular, we are using the assumed bounds on $\|f_t^{\beta}\|_{2+\delta}$ to conclude that asymptotic normality of f_t^* and $f_t^{\beta_0}$ (suitably normalized) follows from the geometric ergodicity of $\{Y_t\}$ and $\{Y_t^{\beta_0}\}$. (Note that, although Doob's Theorem 7.5 includes his condition D_0 as a hypothesis, the geometric ergodicity property is actually sufficient for its proof.) Our result then follows from the independence of the two terms in (5.4) and the convergence of $\sqrt{T}/\sqrt{\mathscr{T}(T)}$ to $\sqrt{\tau}$. *Q.E.D.*

An immediate implication of Theorem 4 is the following corollary.

COROLLARY 3.1: *Under the assumptions of Theorem* 4, $\sqrt{T}(b_T - \beta_0)$ *converges in distribution as $T \to \infty$ to a normal random vector with mean zero and covariance matrix*

$$\Lambda = (1+\tau)\left(D_0' \Sigma_0^{-1} D_0\right)^{-1}. \tag{5.5}$$

The form of the asymptotic covariance matrix Λ is familiar from the results of McFadden (1989), Pakes and Pollard (1989), and Lee and Ingram (1991). As τ gets small, the asymptotic covariance matrix of $\{b_T\}$ approaches $[D_0'\Sigma_0^{-1}D_0]^{-1}$, the covariance matrix obtained when an analytic expression for $E(f_\infty^{\beta})$ as a function of β is known a priori. The proposed SM estimator uses a Monte Carlo generated estimate of this mean, which permits consistent estimation of β_0 for circumstances in which the functional form of $E(f_\infty^{\beta})$ is not known. In general, knowledge of $E(f_\infty^{\beta})$ increases the efficiency of the method of moments estimator of β_0. If, however, the simulated sample size $\mathscr{T}(T)$ is chosen to be large relative to the size T of the sample of observed variables $\{f_t^*\}$, then there is essentially no loss in efficiency from ignorance of this population mean. Thus, the proposed simulated moments estimator extends the class of Markov processes that can be studied using method-of-moment estimators beyond those considered previously, with potentially negligible loss of efficiency.

These results presume that the model is identified. The rank condition for the class of models considered here is Assumption 6 (iii). In many GMM problems, verifying that the choice of moment conditions identifies the unknown parameters under plausible assumptions about the correlations among the variables in the model is straightforward. However, inspection of the moment conditions used in simultaneously solving and estimating dynamic asset-pricing models may give little insight into whether Assumption 6 (iii) is satisfied. This may be especially relevant when the model is solved numerically for some of the elements of $\{Y_t^\beta\}$ as functions of the state and parameter vectors. Indeed, in this case, it may be difficult to gain much insight into which moment conditions will shed light on the values of specific parameters. We recommend that, in practice, the sensitivity of the estimates to various choices of moment conditions be examined.

Fortunately, some information about the validity of this assumption can be obtained in our environment using the simulated state $\{Y_t^\beta\}$. At a given value of β, the partial derivative matrix

$$D(\beta)=\frac{\partial\left[\dfrac{1}{\mathscr{T}}\sum_{t=1}^{\mathscr{T}} f_t^\beta\right]}{\partial\beta} \tag{5.6}$$

can be calculated numerically. For large values of the simulation size $\mathscr{T}$, $D(\beta)$ is approximately equal to $\partial E(f_t^\beta)/\partial\beta$. An orthogonalization of $D(\beta)$ can be examined at various values of β in order to gain some insight into whether the first order conditions defining the SME form a relatively ill-conditioned system of equations at certain points in the parameter space, including at the SME estimator of β_0.

6. EXTENSIONS AND CONCLUSIONS

The SME proposed in this paper can be extended along a variety of different dimensions. One obvious extension is to let f_t^* be a function of β. In order to accommodate this extension, we need one additional primitive, a measurable observation function $g\colon \mathbb{R}^{NL}\times\Theta\to\mathbb{R}^M$, where L is the number of periods of states entering into the observation $g[(Y_t,\dots,Y_{t-L+1}),\beta]$ at time t. We can always assume without loss of generality that $L=l$. We replace the observation f_t^* on the actual state process used in the SME with the observation $g_t^{\beta_0}\equiv g(Z_t,\beta_0)$, and assume that $E[g_t^{\beta_0}-f_t^{\beta_0}]=0$. This leads us to consider the difference in sample moments:

$$G_T(\beta)=\frac{1}{T}\sum_{t=1}^{T} g_t^\beta-\frac{1}{\mathscr{T}(T)}\sum_{s=1}^{\mathscr{T}(T)} f_s^\beta. \tag{6.1}$$

We once again introduce a sequence $\{W_T\}$ of positive semi-definite distance matrices, and define the criterion function $C_T(\beta)=G_T(\beta)'W_TG_T(\beta)$ as well as the extended simulated moments estimator $\{b_T\}$ of β_0, just as in (3.5).

In this case, we replace Σ_0 defined by (4.8) with the weighted covariance matrix, for some positive scalar weight τ,

$$(6.2) \qquad \Sigma_{f,g,\tau} = \tau\Sigma_0 + \Sigma_1,$$

where

$$(6.3) \qquad \Sigma_1 = \sum_{j=-\infty}^{\infty} E\left(\left[g_t^{\beta_0} - E\left(g_t^{\beta_0}\right)\right]\left[g_{t-j}^{\beta_0} - E\left(g_{t-j}^{\beta_0}\right)\right]'\right).$$

Assuming that the families $\{f_t^\beta\}$ and $\{g_t^\beta\}$ satisfy the technical conditions of Assumption 1,[10] and that $W_T \to W_0 = \Sigma_{f,g,\tau}^{-1}$ almost surely, the weak consistency of this extended SME follows from an argument almost identical to the proof of Theorem 1. Furthermore, replacing Assumption 6 (iii) by the assumption that $D_0 \equiv E[\partial g_t^{\beta_0}/\partial\beta - \partial f_\infty^{\beta_0}/\partial\beta]$ exists, is finite, and has full rank, Theorem 4 implies that $\sqrt{T}(b_T - \beta_0)$ converges in distribution to a normal random vector with mean zero and covariance matrix

$$(6.4) \qquad \Lambda_{f,g,\tau} = \left(D_0'\Sigma_{f,g,\tau}^{-1}D_0\right)^{-1}.$$

The new rank condition on D_0 is an identification condition which, among other things, rules out trivial sources of underidentification such as g_t^β and f_∞^β having the multiplicative representations $g^1(z_t, \beta^1)\psi(z_t, \beta^2)$ and $f^1(z_\infty, \beta^1)\psi(z_\infty, \beta^2)$, with β^1 and β^2 being distinct. Also, in contrast to the matrix Λ in (5.5), consistent estimation of $\Lambda_{f,g,\tau}$ must typically be accomplished in two steps, using both simulated and observed data.

Allowing the observation function g_t^β to depend on β is useful in many asset-pricing problems. For instance, one may wish to compare the sample mean of the intertemporal marginal rate of substitution of consumption in the data to the mean of the corresponding simulated series.

A second example arises when one or more of the coordinate functions defining g, say g_j, has the property that $h_j(\beta) = E[g_j(Z_\infty, \beta)]$ defines a known function h_j of β. If this calculation cannot be made for every j, one can mix the use of calculated and simulated moments by letting $f_j(z, \beta) = h_j(\beta)$ for all z, for any j for which h_j is known. This substitution of calculated moments for sample moments improves the precision of the simulated moments estimator, in that the covariance matrix $\Lambda_{f,g,\tau}$ is smaller than the covariance matrix Λ obtained when all moments are simulated. Errors in measurement of f_t^* are accommodated by letting $g_t^{\beta_0} = f(Z_t, \beta_0) + u_t$, where $\{u_t\}$ is an ergodic, mean-zero $\mathbb{R}^M$-valued measurement error. Note that the asymptotic efficiency of the SME is increased by ignoring the measurement error in simulation and comparing sample moments of the simulated $\{f(z_t^\beta, \beta)\}$ and $\{g_t^\beta\}$.

[10] Note that the uniform-in-probability Lipschitz condition for $\{g_t^\beta\}$ is qualitatively weaker than the same condition for $\{f_t^\beta\}$, since g_t^β depends only directly on β (that is, Y_t is not dependent on β).

Finally, one of the coordinate functions of the actual state observations, say g_j, may be of the form

$$g_j[(Y_t, Y_{t-1}, \dots, Y_{t-l+1}), \beta]$$
$$= E\left[h_j(Y_{t+1}, \dots, Y_{t+l+2}, \beta) | Y_t, Y_{t-1}, \dots, Y_{t-l+1}\right],$$

for some h_j. It may be infeasible to calculate the function g_j explicitly, in which case the simulated observation $g_j(Z_t^\beta, \beta)$ is not available, except perhaps by numerical approximation. On the other hand, the observation of $f_j(Z_t^\beta, \beta) = h_j(Z_t^\beta, \beta)$ is often feasible and, by the law of iterated expectations, has the same mean as $g_j(Z_t^\beta, \beta)$. An important illustration of such a function g_j arises in the option pricing literature, where the European option price $g_j(Z_t^\beta, \beta)$ is the conditional expectation of the option's payoff at maturity discounted by an appropriate factor.

Grad. School of Business, Stanford University, Stanford, CA 94305-5015, U.S.A.

Manuscript received January, 1990; final revision received December, 1992.

APPENDIX

PROOF OF LEMMA 2:[11] Since Θ is compact it can be partitioned, for any n, into n disjoint neighborhoods $\Theta_1^n, \Theta_2^n, \dots, \Theta_n^n$ in such a way that the distance between any two points in each Θ_i^n goes to zero as $n \to \infty$. Let $\beta_1, \beta_2, \dots, \beta_n$ be an arbitrary sequence of vectors such that $\beta_i \in \Theta_i^n$, $i = 1, \dots, n$. Then, for any $\varepsilon > 0$,

$$\text{(A.1)} \quad P\left[\sup_{\beta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} \left(f_t^\beta - E(f_\infty^\beta)\right) \right| > \varepsilon \right]$$
$$\leqslant P\left[\bigcup_{i=1}^{n} \left\{ \sup_{\beta \in \Theta_i^n} \left| \frac{1}{T} \sum_{t=1}^{T} \left(f_t^\beta - E(f_\infty^\beta)\right) \right| > \varepsilon \right\}\right]$$
$$\leqslant \sum_{i=1}^{n} P\left[\sup_{\beta \in \Theta_i^n} \left| \frac{1}{T} \sum_{t=1}^{T} \left(f_t^\beta - E(f_\infty^\beta)\right) \right| > \varepsilon \right]$$
$$\leqslant \sum_{i=1}^{n} P\left[\left| \frac{1}{T} \sum_{t=1}^{T} \left(f_t^{\beta_i} - E(f_\infty^{\beta_i})\right) \right| > \frac{\varepsilon}{2} \right]$$
$$+ \sum_{i=1}^{n} P\left[\frac{1}{T} \sum_{t=1}^{T} \sup_{\beta \in \Theta_i^n} \left| f_t^\beta - f_t^{\beta_i} \right| + \sup_{\beta \in \Theta_i^n} \left| E(f_\infty^\beta) - E(f_\infty^{\beta_i}) \right| > \frac{\varepsilon}{2} \right],$$

where the last inequality follows from the triangle inequality. For fixed n, since $\{Y_t^{\beta_i}\}$ is ergodic and $E(|f_t^{\beta_i}|) < \infty$, the first term on the right-hand side of (A.1) approaches zero as $T \to \infty$ by the weak law of large numbers for ergodic processes.

[11] The strategy for proving this lemma, which was suggested to us by Whitney Newey, follows the proof strategies used by Jennrich (1969) and Amemiya (1985) to prove similar lemmas. A subsequent paper by Newey (1991) presents a more extensive discussion of sufficient conditions for uniform convergence in probability.

As for the second right-hand-side term in (A.1), the Lipschitz assumption on $\{f_t^\beta\}$ implies that there exist K_t such that

$$\text{(A.2)} \qquad \sum_{i=1}^{n} P\left[\frac{1}{T}\sum_{t=1}^{T}\sup_{\beta\in\Theta_i^n}\left|f_t^\beta - f_t^{\beta_i}\right| + \sup_{\beta\in\Theta_i^n}\left|E(f_\infty^\beta) - E(f_\infty^{\beta_i})\right| > \frac{\varepsilon}{2}\right]$$

$$\leqslant \sum_{i=1}^{n} P\left[\sup_{\beta\in\Theta_i^n}|\beta-\beta_i|\frac{1}{T}\sum_{t=1}^{T}K_t + \sup_{\beta\in\Theta_i^n}\left|E(f_\infty^\beta) - E(f_\infty^{\beta_i})\right| > \frac{\varepsilon}{2}\right].$$

The assumption that $K^T = T^{-1}\Sigma_{t=1}^T K_t$ is bounded in probability implies that there is a nonstochastic bounded sequence $\{A_T\}$ such that $\operatorname{plim}(K^T - A_T) = 0$. Thus, for T larger than some T^* and some bound B, the right-hand side of (A.2) is less than or equal to

$$\text{(A.3)} \qquad \sum_{i=1}^{n} P\left[\sup_{\beta\in\Theta_i^n}|\beta-\beta_i|\,|K^T - A_T| + \sup_{\beta\in\Theta_i^n}|\beta-\beta_i|B + \sup_{\beta\in\Theta_i^n}\left|E(f_\infty^\beta) - E(f_\infty^{\beta_i})\right| > \frac{\varepsilon}{2}\right].$$

By continuity of $\beta \mapsto E(f_\infty^\beta)$, we can choose n once and for all so that $|\beta-\beta_i|B + |E(f_\infty^\beta) - E(f_\infty^{\beta_i})| < (\varepsilon/4)$ for all β in Θ_i^n and all i. Thus, the limit of (A.3) as $T\to\infty$ is zero, and the result follows. *Q.E.D.*

PROOF OF THEOREM 1: By the triangle inequality,

$$\text{(A.4)} \qquad \left|\left(\frac{1}{T}\sum_{t=1}^{T} f_t^* - \frac{1}{\mathscr{T}}\sum_{s=1}^{\mathscr{T}} f_s^\beta\right) - \left[E(f_\infty^*) - E(f_\infty^\beta)\right]\right|$$

$$\leqslant \left|E(f_\infty^*) - \frac{1}{T}\sum_{t=1}^{T} f_t^*\right| + \left|E(f_\infty^\beta) - \frac{1}{\mathscr{T}}\sum_{s=1}^{\mathscr{T}} f_s^\beta\right|.$$

Assumption 2 implies that the first term on the right-hand side of (A.4) converges to zero in probability. By Lemma 2, the second term on the right-hand side of (A.4) converges in probability to zero uniformly in β. Now $\delta_T(\beta) \equiv |C_T(\beta) - C(\beta)|$ satisfies

$$\text{(A.5)} \qquad \delta_T(\beta) = \left|G_T(\beta)'W_T G_T(\beta) - \left[E(f_\infty^*) - E(f_\infty^\beta)\right]'W_0\left[E(f_\infty^*) - E(f_\infty^\beta)\right]\right|$$

$$\leqslant \left|G_T(\beta) - \left[E(f_\infty^*) - E(f_\infty^\beta)\right]\right|'\,|W_T|\,|G_T(\beta)|$$

$$+ \left|E(f_\infty^*) - E(f_\infty^\beta)\right|'\,|W_T - W_0|\,|G_T(\beta)|$$

$$+ \left|E(f_\infty^*) - E(f_\infty^\beta)\right|'\,|W_0|\,\left|G_T(\beta) - \left[E(f_\infty^*) - E(f_\infty^\beta)\right]\right|.$$

Therefore, letting $l_T = \sup_{\beta\in\Theta}|G_T(\beta) - [E(f_\infty^*) - E(f_\infty^\beta)]|$,

$$\text{(A.6)} \qquad \sup_{\beta\in\Theta}\delta_T(\beta) \leqslant l_T|W_T|[\phi_0 + l_T] + \phi_0|W_T - W_0|[\phi_0 + l_T] + \phi_0|W_0|l_T,$$

where $\phi_0 \equiv \max\{|E(f_\infty^*) - E(f_\infty^\beta)| : \beta\in\Theta\}$ exists by the continuity condition in Assumption 1. Since each of the terms on the right-hand side of (A.6) converges in probability to zero, $\operatorname{plim}_T[\sup_{\beta\in\Theta}\delta_T(\beta)] = 0$. This implies the convergence of $\{b_T\}$ to β_0 in probability as $T\to\infty$, as indicated, for example, in Amemiya (1985, page 107). *Q.E.D.*

PROOF OF LEMMA 3: We fix β and t. For simplicity, we write "ε_t" for $\hat{\varepsilon}_t$. For each positive integer m, we define $\{Y_s^{m\beta} : t-m \leqslant s \leqslant t\}$ by the recursion $Y_{t-m}^{m\beta} = 0$ and

$$Y_{t-m+k+1}^{m\beta} = H\left(Y_{t-m-k}^{m\beta}, \beta, \varepsilon_{t-m+k+1}\right).$$

By construction, $Y_t^{m\beta}$ is measurable with respect to $\{\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-m+1}\}$. The AUC condition implies that

$$\|Y_t^{m\beta} - Y_t^{m+1,\beta}\| \leqslant \prod_{j=0}^{m} \rho_\beta(\varepsilon_{t-j}) \| H(0, \varepsilon_{t-m+1}, \beta) \|, \tag{A.7}$$

where

$$\frac{1}{m}\sum_{j=0}^{m} \ln \rho_\beta(\varepsilon_{t-j}) + \frac{1}{m}\ln\left(\max\left[1, \|H(0, \varepsilon_{t-m+1}, \beta)\|\right]\right) \xrightarrow{\text{a.s.}} \alpha_\beta < 0.$$

Hence,

$$\left[\prod_{j=0}^{m} \rho_\beta(\varepsilon_{t-j})\right]^{1/m} \|H(0, \varepsilon_{t-m+1}, \beta)\|^{1/m} \xrightarrow{\text{a.s.}} e^{\alpha_\beta} < 1. \tag{A.8}$$

This, in turn, implies that, given $\delta \in (e^{\alpha_\beta}, 1)$, there is some event Λ with $P(\Lambda) = 1$ and, for each $\omega \in \Lambda$, some integer $N(\omega, \delta)$ such that

$$\left[\prod_{j=0}^{m} \rho_\beta(\varepsilon_{t-j}(\omega))\right] \|H(0, \varepsilon_{t-m+1}(\omega), \beta)\| < \delta^m, \qquad m \geqslant N(\omega, \delta).$$

Next, at arbitrary $\omega \in \Lambda$ and $m > n \geqslant N(\omega, \delta)$,

$$\begin{aligned}
\|Y_t^{m\beta} - Y_t^{n\beta}\| &\leqslant \|Y_t^{m\beta} - Y_t^{m-1,\beta}\| + \|Y_t^{m-1,\beta} - Y_t^{m-2,\beta}\| + \cdots + \|Y_t^{n+1,\beta} - Y_t^{n\beta}\| \\
&\leqslant \prod_{j=0}^{m-1} \rho_\beta(\varepsilon_{t-j}) \|H(0, \varepsilon_{t-m}, \beta)\| + \cdots + \prod_{j=0}^{n} \rho_\beta(\varepsilon_{t-j}) \|H(0, \varepsilon_{t-n+1}, \beta)\| \\
&\leqslant \delta^{m-1} + \delta^{m-2} + \cdots + \delta^n = \frac{\delta^{n-1}(1 - \delta^{m-n+1})}{1-\delta} \leqslant \frac{\delta^{n-1}}{1-\delta}.
\end{aligned}$$

It follows that, at each $\omega \in \Lambda$, $\{Y_t^{m\beta}(\omega)\}$ is a Cauchy sequence in m. We conclude that $\lim_{m\to\infty} Y_t^{m\beta} = Y_t^{\infty\beta}$ exists almost surely. The limit process $\{Y_t^{\infty\beta}: -\infty < t < \infty\}$, constructed for each t in this manner, satisfies the difference equation (3.1) by construction and $Y_t^{\infty\beta}$ is clearly measurable with respect to $\{\varepsilon_{t-s}: s \geqslant 0\}$. Since $\{\varepsilon_t\}$ is an i.i.d. sequence, the stationarity and ergodicity of $\{Y_t^{\infty\beta}\}$ follows immediately. *Q.E.D.*

Proof of Lemma 4: Fix $\theta \in \Theta$ and without loss of generality set $l = 1$. For any $\beta \in \Theta$ such that $\|\beta - \theta\| < \delta_\theta$,

$$\begin{aligned}
\left|\frac{1}{T}\sum_{t=1}^{T} f_t^\beta - \frac{1}{T}\sum_{t=1}^{T} f_t^{\infty\beta}\right| &\leqslant k(\theta)\frac{1}{T}\sum_{t=1}^{T} \|Y_t^\beta - Y_t^{\infty\beta}\| \\
&\leqslant k(\theta)\frac{1}{T}\sum_{t=1}^{T}\left[\prod_{j=0}^{t} \rho_\theta(\varepsilon_j)\right] \|Y_0^\beta - Y_0^{\infty\beta}\|,
\end{aligned}$$

where $k(\theta)$ is given by the S-smoothness assumption. The AUC condition implies that $(1/T)\Sigma_{t=1}^T[\Pi_{j=0}^t \rho_\theta(\varepsilon_j)]$ converges almost surely to zero. Thus, given $\eta > 0$, there is an event Λ_θ with $P(\Lambda_\theta) = 1$ such that, for each ω in Λ_θ, there is some $T_\theta(\omega, \eta)$ with

$$\Delta_T^\beta \equiv \left|\frac{1}{T}\sum_{t=1}^{T} f_t^\beta - \frac{1}{T}\sum_{t=1}^{T} f_t^{\infty\beta}\right| \leqslant \eta, \qquad T \geqslant T_\theta(\omega, \eta), \tag{A.9}$$

provided $\|\beta - \theta\| \leqslant \delta_\theta$.

Since Θ is compact, it has a finite subset Θ^* defining a finite subcover of "δ_θ neighborhoods," $\theta \in \Theta^*$. Letting $\Lambda^* = \bigcap_{\theta \in \Theta^*} \Lambda_\theta$ and $T^*(\omega, \eta) = \max_{\theta \in \Theta^*} T_\theta(\omega, \eta)$, it follows that $\Delta_T^\beta \leqslant \eta$, $T \geqslant T^*$, for all β in Θ, which leads to (4.7). *Q.E.D.*

PROOF OF THEOREM 4: As noted above, the L^2 UC condition implies the AUC condition, so the conclusions of Lemmas 3 and 4 continue to hold. Thus, the consistency of $\{b_T\}$ for β_0 will be established by showing that, for each $\theta \in \Theta$, $E[\text{mod}_t(\delta, \theta)] < \infty$ for some $\delta > 0$. As before, we write "ε_t" for "$\hat{\varepsilon}_t$."

Fix $\theta \in \Theta$. For purposes of the proof, we can assume without loss of generality that $l = 1$. Since f is S-smooth, there is a $\delta > 0$ such that, for $\|\beta - \theta\| \leqslant \delta$ and for each t,

$$\begin{aligned}\left\| f(Y_t^{\infty\beta}, \beta) - f(Y_t^{\infty\theta}, \theta) \right\| &= \left\| f(Y_t^{\infty\beta}, \beta) - f(Y_t^{\infty\beta}, \theta) + f(Y_t^{\infty\beta}, \theta) - f(Y_t^{\infty\theta}, \theta) \right\| \\ &\leqslant C_1(Y_t^\beta) \left\| \beta - \theta \right\| + k(\theta) \|Y_t^{\infty\beta} - Y_t^{\infty\theta}\|.\end{aligned}$$

It follows that

$$\text{(A.10)} \qquad \underset{t}{\text{mod}}(\delta, \theta) \leqslant \delta \sup_{\|\beta - \theta\| \leqslant \delta} C_1(Y_t^{\infty\beta}) + k(\theta) \sup_{\|\beta - \theta\| \leqslant \delta} \|Y_t^{\infty\beta} - Y_t^{\infty\theta}\|.$$

Letting $\alpha_t = \|Y_t^{\infty\beta} - Y_t^{\infty\theta}\|$, the L^2 UC condition and S-smoothness of H imply that

$$\text{(A.11)} \qquad \alpha_t \leqslant \rho_\theta(\varepsilon_t)\alpha_{t-1} + C_2(Y_{t-1}^\theta, \varepsilon_t)\delta.$$

By recursively substituting α_{t-k}, using (A.11), we have for any T

$$\alpha_t \leqslant \prod_{s=t-T}^{t} \rho_\theta(\varepsilon_s)\alpha_{t-T} + \delta \sum_{s=t-T}^{t} C_2(Y_s^{\theta\infty}, \varepsilon_s) \prod_{\tau=s+1}^{t} \rho_\theta(\varepsilon_\tau).$$

Now, $X_T \equiv \prod_{s=t-T}^{t} \rho_\theta(\varepsilon_s)$ converges to zero in L^2 since $E[\rho_\theta(\varepsilon_t)^2] < 1$ and $\{\varepsilon_t\}$ is i.i.d. Since $\|\alpha_{t-T}\|_2 \leqslant \|Y_{t-T}^{\beta\infty}\|_2$ is bounded, the Cauchy-Schwarz inequality implies that

$$\|X_T\alpha_{t-T}\|_1 \leqslant \|X_T\|_2 \|\alpha_{t-T}\|_2 \to_T 0,$$

so, in L^1,

$$\alpha_t \leq \delta \lim_{T \to \infty} \sum_{s=t-T}^{t} C_2(Y_s^{\infty\theta}, \varepsilon_s) \prod_{\tau=s+1}^{t} \rho_\theta(\varepsilon_t).$$

The right-hand side is independent of β, and taking expectations, using the independence of $\{\varepsilon_t\}$ and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}E\left[\sup_{\|\beta - \theta\| \leqslant \delta} \|Y_t^{\infty\beta} - Y_t^{\infty\theta}\|\right] &\leqslant \delta E\left[\sum_{s=-\infty}^{t} C_2(Y_s^{\infty\theta}, \varepsilon_s) \prod_{\tau=s+1}^{t} \rho_\theta(\varepsilon_s)\right] \\ &\leqslant \frac{\delta K}{1 - \bar{\rho}},\end{aligned}$$

where $\bar{\rho} = \|\rho_\theta(\varepsilon_t)\|_2 < 1$ and where K is a bound on $\|C_2(Y_s^{\infty\beta}, \varepsilon_s)\|_2$ implied by the growth condition on C_2 and the fact that $\|Y_s^{\infty\beta}\|_2$ and $\|\varepsilon_s\|_2$ are bounded.

The last term in (A.10) therefore has a finite mean. To establish that the first term on the right-hand side of (A.10) has a finite mean, first note that $C_1(Y_t^{\infty\beta}) \leqslant d_1 + d_2\|Y_t^{\infty\beta}\|$, for constants d_1, d_2. Furthermore,

$$\text{(A.12)} \qquad \sup_{\|\beta - \theta\| \leqslant \delta} \|Y_t^{\infty\beta}\| \leqslant \|Y_t^{\infty\theta}\| + \sup_{\|\beta - \theta\| \leqslant \delta} \|Y_t^{\infty\beta} - Y_t^{\infty\theta}\|,$$

and both terms on the right-hand side of (A.12) have finite means.

Combining these results with Hansen's (1982) Theorem 2.1 gives the desired result. *Q.E.D.*

REFERENCES

AMEMIYA, T. (1985): *Advanced Econometrics*. Cambridge, Massachusetts: Harvard University Press.

ANDREWS, D. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1472.

——— (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

BROCK, W. (1980): "Asset Prices in a Production Economy," in *The Economics of Uncertainty*, ed. by J. J. McCall. Chicago: University of Chicago Press.

DOOB, J. (1953): *Stochastic Processes*. New York: John Wiley and Sons.

DUFFIE, D. (1988): *Security Markets: Stochastic Models*. Boston: Academic Press.

DUNN, K., AND K. SINGLETON (1986): "Modeling the Term Structure of Interest Rates under Non-separable Utility and Durability of Goods," *Journal of Financial Economics*, 17, 27–55.

EICHENBAUM, M., AND L. HANSEN (1988): "Estimating Models with Intertemporal Substitution Using Aggregate Time Series Data," manuscript, University of Chicago.

EICHENBAUM, M., L. HANSEN, AND K. SINGLETON (1988): "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics*, 103, 51–78.

GALLANT, R., AND H. WHITE (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil Blackwell.

HALL, R. (1988): "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96, 339–357.

HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica*, 50, 1029–1056.

HANSEN, L., AND K. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

——— (1989): "Efficient Estimation of Linear Asset Pricing Models with Moving Average Errors," manuscript, Stanford University.

JENNRICH, R. (1969): "Asymptotic Properties of Non-Linear Least Squares Estimators," *Annals of Mathematical Statistics*, 40, 633–643.

LEE, B., AND B. INGRAM (1991): "Simulation Estimation of Time Series Models," *Journal of Econometrics*, 47, 197–205.

MCFADDEN, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995–1026.

MICHNER, R. (1984): "Permanent Income in General Equilibrium," *Journal of Monetary Economics*, 14, 297–305.

MOKKADEM, A. (1985): "Le Modèle Non Linéaire AR(1) Général. Ergodicité et Ergodicité Géometrique," *Comptes Rendues Academie Scientifique Paris*, 301, Série I, 889–892.

NEWEY, W. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161–1167.

NEWEY, W., AND K. WEST (1987): "A Simple, Positive Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

NUMMELIN, E., AND P. TUOMINEN (1982): "Geometric Ergodicity of Harris Recurrent Markov Chains with Applications to Renewal Theory," *Stochastic Processes and Their Applications*, 12, 187–202.

PAKES, A., AND D. POLLARD (1989): "The Asymptotics of Simulation Estimators," *Econometrica*, 57, 1027–1058.

ROSENBLATT, M. (1971): *Markov Processes. Structure and Asymptotic Behavior*. New York: Springer-Verlag.

SINGLETON, K. (1987): "Asset Prices in a Time-Series Model with Disparately Informed, Competitive Traders," Chapter 12 in *New Approaches to Monetary Economics*, ed. by W. Barnett and K. Singleton. Cambridge: Cambridge University Press, pp. 249–272.

TAUCHEN, G., AND R. HUSSEY (1991): "Quadrature Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," *Econometrica*, 59, 371–396.

TAYLOR, J., AND H. UHLIG (1990): "Solving Nonlinear Stochastic Growth Models: A Comparison of Alternative Solution Methods," *Journal of Business and Economic Statistics*, 8, 1–18.

TWEEDIE, R. (1982): "Criteria for Rates of Convergence of Markov Chains, with Application to Queuing and Storage Theory," in *Probability, Statistics, and Analysis*, edited by J. F. C. Kingman and G. E. H. Reuter. Cambridge: Cambridge University Press.