TECHNICAL WORKING PAPER SERIES

SIMULATED MOMENTS ESTIMATION OF MARKOV MODELS OF ASSET PRICES

Darrell Duffie

Kenneth J. Singleton

Technical Working Paper No. 87

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1990

We are grateful for several useful conversations with Lars Hansen and Whitney Newey, and for the comments of Peter Bossaerts, Andrew Lo, Neil Pearson, and Bruce Lehmann. Singleton acknowledges funding from the National Science Foundation. Duffie acknowledges the support of a Batterymarch Fellowship. This paper is part of NBER's research programs in Economic Fluctuations and Financial Markets and Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #87
March 1990

SIMULATED MOMENTS ESTIMATION OF MARKOV MODELS OF ASSET PRICES

ABSTRACT

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.

Darrell Duffie
Graduate School of Business
Stanford University
Stanford, CA 94305

Kenneth J. Singleton
Graduate School of Business
Stanford University
Stanford, CA 94305

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. The 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. Applications include the estimation of models based on the moments of decision variables when agents' decision rules cannot be expressed in closed form, or of models to which Euler equation methods [Hansen and Singleton (1982)] are not applicable. These extensions are feasible in part because of the recent development and refinement of new algorithms for solving discrete-time general equilibrium models [Taylor and Uhlig (1989) and the references therein]. Given choices of a solution algorithm and a parameterization of a Markov state vector, the SME proposed in this paper provides a computationally tractable means of simultaneously solving and estimating an equilibrium asset pricing model.

The basic construction of the SME 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. 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 simulation, nor those imposed by McFadden (1987) and Pakes and Pollard (1987) 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 typically required. In most circumstances, however, the stationary distribution of the simulated process, as a function of the parameter choice, is unknown. Hence, the simulated process is generally non-stationary. This issue was recognized independently, although no resolution of it was proposed, in the work of Lee and Ingram (1989). 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 non-stationarity 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. The exact nature of our conditions will gradually become apparent as the paper unfolds.

The remainder of the paper is organized as follows. Section 2 uses a simple asset pricing setting to illustrate 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

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

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. This is equivalent to maximization of share market value [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

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

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

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

where $\{\epsilon_t\}$ 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

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

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 $\{\hat{\epsilon}_t\}$ that is identically and independently distributed of $\{\epsilon_t\}$; that is,

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

It kinda looks like matching the mean simulated moment

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

Next, for some chosen observation function f, in each period t an observation $f_t^* \equiv f(Y_t, Y_{t-1}, \ldots, Y_{t-\ell+1})$ 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 $\{f_t^{\beta}\}_{t=1}^{T}$ and the sample mean of $\{f_t^*\}_{t=1}^{T}$, where T is the actual number of periods of data.

Several considerations motivate the use of simulated moments in estimating β_0. First, dynamic asset pricing models like (2.1)-(2.5) typically do not yield analytic representations of the decision rules or laws of motion for capital and equity returns nor for the population moments of f_t^*. Consequently, outside of certain linear-quadratic examples [as, for example, in Hansen and Sargent (1980)], maximum likelihood estimation of the parameters of agents' decision rules is often not feasible. Instead, the SME of β_0 is chosen to minimize the discrepancy between certain moments of the joint distributions of the ℓ-histories of Y_t and those of the corresponding simulated values from the model, where the moments of interest are specified through the choice of f.

There are also several potential advantages of simultaneously solving and estimating dynamic models of asset prices over estimating the implied Euler equations. 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. Furthermore, estimation of asset pricing models using Euler equations [as, for example, in Hansen and Singleton (1982)] is not always feasible. For instance, assuming that the representative consumer's decisions are interior to the choice set, the first order conditions to problem (2.3)-(2.4) can be expressed as

$$p_t c_t^{-\alpha} u_t = E_t \left(\delta c_{t+1}^{-\alpha} u_{t+1} [p_{t+1} + d_{t+1}] \right), \tag{2.6}$$

$$c_t^{-\alpha} u_t = \delta E_t \left(c_{t+1}^{-\alpha} u_{t+1} [z_{t+1} k_{t+1}^{\phi-1} + \mu] \right), \tag{2.7}$$

where E_t denotes conditional expectation given the consumer's information at date t. The presence of the unobserved taste shock u_t in (2.6) and (2.7) precludes the application of instrumental variables estimators based on these conditional moment restrictions.

Temporal aggregation provides an additional motivation for using the SME. In the context of the present example, consumption is measured as the cumulative flow over a month or quarter. If the agent's decision interval is shorter than these sampling intervals, then measured consumption is not the observed counterpart of c_t in these Euler equations. Hence, GMM estimation may give inconsistent estimators of β_0 [Hall (1988), Hansen and Singleton (1989)]. In contrast, temporal aggregation is easily accommodated using the SME. After solving and simulating the equilibrium consumption sequence at intervals corresponding to the decision interval of the agents, the simulated series can be time-averaged in the manner corresponding to the construction of the data before calculating f_t^β.

The first step in implementing the SME is solving for the equilibrium of the asset pricing model. Several alternative numerical methods for solving discrete-time dynamic rational expectations models have recently been proposed in the literature. A useful summary of these methods and their properties for a stochastic growth model that is similar to (2.1)-(2.5) is provided in Taylor and Uhlig (1989) and the references cited therein. Many of the algorithms discussed involve approximations to either the distributions of the forcing variables or the model itself. For instance, the value-function-iteration approach used by Donaldson and Mehra (1984) as well as the quadrature solutions to the integral equations implied by Euler equations used by Mehra and Prescott (1983) and Tauchen (1986), among others, require discrete state-space approximations to the continuous distributions of the forcing process $\{X_t\}$. These approximations affect the large sample properties of the SME since, as sample size increases, one obtains a consistent estimator of the approximate model. For some of these solution methods, it seems apparent that the approximation error becomes negligible if the accuracy of the approximation is improved as the sample size increases. Likewise, as in Example 2 of Section 4 (to follow), one can approximate continuous-time Markov-state processes with various types of discrete-time Markov processes. In future research, we plan to explore the rate at which the approximations must be adjusted in order to assure that the asymptotic distribution of the SME is not affected by certain types of approximations. The methods described in this paper apply to the approximate model if approximations are used to solve for equilibrium asset prices.

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 (1987) and Pakes and Pollard (1987). 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 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 $\{Y_t^\beta\}$ in 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 non-stationary.

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 mis-measurement of the initial stock.[1]

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 taste and technology shocks. The equilibrium processes $\{Y_t\}$ and $\{f_t^\beta\}$, and in particular the equilibrium equity return, depend on k in addition to X. Thus, we also show that geometric ergodicity of X, together with natural conditions on the depreciation of capital, are sufficient for the large sample properties of $\mathcal{T}^{-1}\sum_{t=1}^{\mathcal{T}} f_t^\beta$ to be unaffected by the choice of initial conditions for the forcing variables or the capital stock.

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

[1] See Dunn and Singleton (1986); Eichenbaum, Hansen, and Singleton (1987); and Eichenbaum and Hansen (1988) for examples of studies of Euler equations using GMM estimators in which this type of initial condition problem arises.

structure of the model.[2] 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.

To be more concrete, consider the special case of (2.1)-(2.5) 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

$$\ell n\, z_{t+1} = \varsigma_z + \xi t + \rho \ell n\, z_t + \epsilon_{t+1}, \tag{2.8}$$

for constants ς_z and ξ. Under these simplifying assumptions, the equilibrium asset pricing function and law of motion for the captial stock implied by (2.6) and (2.7) are [Michner (1984)]:

$$p_t = \frac{\delta}{(1-\delta)}(1-\phi)z_t k_t^{\phi} \tag{2.9}$$

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

$$k_{t+1} = \delta\phi z_t k_t^{\phi}. \tag{2.10}$$

In (2.9) and (2.10), p_t and k_t exhibit growth over time induced by the growth in z_t. The stock return process $\{r_t^s = (p_t + d_t)/p_{t-1}\}$ and capital growth process $\{k_t/k_{t-1}\}$ are nevertheless stationary. Thus, in implementing an SME for this model, one could simulate trend-free $\{z_t/z_{t-1}\}$, construct the implied simulated returns and growth rates of capital, and then choose f^* and f^β to depend on ℓ-histories of simulated and measured r_t^s and k_t/k_{t-1}, respectively.

Following Eichenbaum and Hansen (1988), the implied restrictions on trends in the decision variables can be imposed in estimation by appending the moment

[2] See Eichenbaum and Hansen (1988) and Eichenbaum, Hansen and Singleton (1987) 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.

conditions associated with least squares estimation of the trend equations to the moment equations involving f^* and f^β:

$$\ell n\, k_t = \varsigma_k + \xi t + \nu_{kt}, \qquad \ell n\, p_t = \varsigma_p + \xi t + \nu_{pt}, \tag{2.11}$$

where ς_k and ς_p are constants, while ν_{kt} and ν_{pt} are trend-free, stationary components of $\ell n\, k_t$ and $\ell n\, p_t$. The subsequent discussion in this paper extends immediately to this case using the large sample theory discussed in Eichenbaum and Hansen (1988) for GMM estimators of (2.11).

Two additional comments about trends are warranted. First, if the forcing variables exhibit stochastic trends (unit roots), then our estimation strategy applies only if the entire model, including the forcing variables, can be transformed to a model expressed in terms of trend-free processes. Examples of growth economies with capital accumulation for which such tranformations are possible include those in Kydland and Prescott (1982) [see Altug (1988)] and Christiano (1987). Second, stochastic trends in decision variables can arise as the endogenous outcome of agents' decisions in the presence of trend-free forcing variables, as in Eichenbaum and Singleton (1986) or King, Plosser, Stock, and Watson (1987). If the form of the trends in the decision variables is known and induced growth takes the form of unit roots in the processes, then our proposed simulation methods will often be applicable.

Henceforth, we suppress trends and their associated parameters in order to conserve on notation, though reference will occassionally be made to modifications of our discussion that would be required in the presense of trends.

3. The Estimation Problem

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

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

(ii) a measurable *observation function* $f : \mathbb{R}^{N\ell} \times \Theta \to \mathbb{R}^M$, with $M \geq Q$.

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

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

where the parameter vector β_0 is to be estimated, and $\{\epsilon_t\}$ is an *i.i.d.* sequence of $\mathbb{R}^p$-valued random variables on a given probability space $(\Omega, \mathcal{F}, P)$. The function H may be determined implicitly by the numerical solution of a discrete-time model for equilibrium asset prices, or by a discrete-time approximation of a continuous-time model. Let $Z_t = (Y_t, Y_{t-1}, \ldots, Y_{t-\ell+1})$ for some positive integer $\ell < \infty$. 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,

$$b_T = \operatorname{argmin}_{\beta \in \Theta} \left[\frac{1}{T} \sum_{t=1}^{T} f_t^* - E[f(Z_t, \beta)] \right]' W_T \left[\frac{1}{T} \sum_{t=1}^{T} f_t^* - E[f(Z_t, \beta)] \right], \tag{3.2}$$

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 $\{\hat{\epsilon}_t\}$ of random variables that is identical in distribution to, and independent of, $\{\epsilon_t\}$. Then, for any $\mathbb{R}^N$-valued initial point $\widehat{Y}_1$ and any parameter vector $\beta \in \Theta$, the *simulated state process* $\{Y_t^\beta\}$ can be constructed inductively by letting $Y_1^\beta = \widehat{Y}_1$ and

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

Likewise, the simulated observation process $\{f_t^\beta\}$ is constructed by $f_t^\beta = f(Z_t^\beta, \beta)$, where $Z_t^\beta = (Y_t^\beta, \ldots, Y_{t-\ell+1}^\beta)$. 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 $\mathcal{T} : \mathbb{N} \to \mathbb{N}$ define the simulation sample size $\mathcal{T}(T)$ that is generated for a given sample size T of actual observations, where $\mathcal{T}(T) \to \infty$ as $T \to \infty$. For any parameter vector β, let

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

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 $(H, \epsilon, \mathcal{T}, \widehat{Y}_1, W)$ to be the sequence $\{b_T\}$ given by

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

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

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 dampening, rather than exploding, variety.

Initially we describe the concept of geometric ergodicity, a condition ensuring that the simulated state process satisifies 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, $b_T \rightarrow \beta_0$ 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, $b_T \rightarrow \beta_0$ 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 probability measure π on the state space of the process such that, for every initial point x,

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

where $\|\cdot\|_v$ is the total variation norm.[3] 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 non-linear AR(1) models, which includes our setting. Mokkadem's regularity conditions B and C are satisfied provided $\{Y_t^\beta\}$ is irreducible. For example, it is enough that the distribution of Y_2^β given Y_1^β has $\mathbb{R}^N$ as its support. With this, Mokkadem's Theorem 3 implies geometric ergodicity of $\{Y_t^\beta\}$ provided, for some $K > 0$, some $\delta < 1$, and some $q > 0$,

(i) $\| H(y, \epsilon_1, \beta) \|_q < \delta \| y \|, \quad \| y \| > K.$

(ii) $H(\cdot, \epsilon_1, \beta) : \mathbb{R}^N \to L^q$ is well defined and continuous.

We summarize with:

---

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

LEMMA 1. *Suppose, for a given $\beta \in \Theta$, that $\{Y_t^\beta\}$ is irreducible and that (H, ϵ) satisfies conditions (i) and (ii). Then $\{Y_t^\beta\}$ is geometrically ergodic and $\| Y_t^\beta \|_q$ and $\| Y_\infty^\beta \|_q$ are uniformly bounded.*

Example 1 (The Traditional AR-1 Model). Suppose $N = p = 1$ and $Y_{t+1}^\beta = \beta Y_t^\beta + \epsilon_t$, where $\Theta = [\underline{\beta}, \overline{\beta}] \subset (-1, 1)$. If ϵ_1 is in L^q, then all conditions for Lemma 1 are satisfied and $\{Y_t^\beta\}$ is geometrically ergodic.

Example 2 (The Euler Approximation to a Diffusion). For any $\mu : \mathbb{R}^N \times \Theta \to \mathbb{R}^N$ and $\sigma : \mathbb{R}^N \times \Theta \to \mathbb{R}^{N \times p}$ that are Lipschitz in the first (y) argument, a diffusion process Y_t^β is well defined by the stochastic differential equation

$$dY_t^\beta = \mu(Y_t^\beta, \beta)\, dt + \sigma(Y_t^\beta, \beta)\, dB_t, \tag{4.2}$$

where B is a standard Brownian motion in $\mathbb{R}^p$. The *Euler approximation* to Y_t^β, for $n \geq 1$ time periods per unit of time, is the process $\{Y_t^{\beta n}\}$ defined as follows. A state process $\{\widetilde{Y}_t^{\beta n}\}$ is defined by $\widetilde{Y}_{t+1}^{\beta n} = H_n\left(\widetilde{Y}_t^{\beta n}, \epsilon_{t+1}, \beta\right)$, where $\epsilon_{t+1} = B_{t+1} - B_t$ and

$$H_n(y, \epsilon, \beta) = y + \frac{1}{n}\mu(y, \beta) + \frac{1}{\sqrt{n}}\sigma(y, \beta)\epsilon. \tag{4.3}$$

We then let $Y_t^{\beta n} = \widetilde{Y}_{nt}^{\beta n}$.

Irreducibility of $\{\widetilde{Y}_t^{\beta n}\}$ follows if $N \leq p$ and σ is everywhere of full rank. Condition (ii) for geometric ergodicity follows from the continuity of μ and σ as well as the existence of all moments for the normal distribution.

As for condition (i), we know that

$$E\left[\| H_n(y, \epsilon_1, \beta) \|^2\right] = \| y \|^2 \left[\frac{\| y + \mu(y)/n \|^2 + \| \sigma(y) \|^2 /n}{\| y \|^2}\right] \equiv \| y \|^2\, \mathcal{E}(y).$$

For geometric ergodicity condition (i), it is thus enough that some $K > 0$ can be chosen so that $\sup_{\|y\|>K} \mathcal{E}(y) < 1$, which is equivalent to

$$\sup_{\|y\|>K}\ 2y'\mu(y) + \mu(y)'\mu(y)/n + \operatorname{tr}\left[\sigma(y)\sigma(y)'\right] < 0,$$

and requires that $y'\mu(y) < 0$ for $\| y \| > K$. These conditions on $\mathcal{E}$, as well as corresponding results for higher order diffusion approximations, are examined in other work planned by the authors.

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 $\left\{\{f_t^\beta\} : \beta \in \Theta\right\}$ of processes satisfies the uniform weak law of large numbers if, for each $\delta > 0$,

$$\lim_{T\to\infty} P\left[\sup_{\beta\in\Theta}\left| E(f_\infty^\beta) - \frac{1}{T}\sum_{t=1}^{T} f_t^\beta \right| > \delta\right] = 0. \tag{4.4}$$

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 (4.4). We assume 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,*

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

where $K^T = T^{-1}\sum_{t=1}^{T} K_t$ 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 $\left\{f_t^\beta\right\}$ is Lipschitz, uniformly in probability. Then $\left\{\{f_t^\beta\} : \beta \in \Theta\right\}$ satisfies the uniform weak law of large numbers.*

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

In order to provide conditions that can be verified in applications, we replace the ergodicity assumption on $\{Y_t^\beta\}$ with Mokkadem's conditions for geometric ergodicity directly on the transition function H and disturbance ϵ_t.

COROLLARY. *Suppose, for each $\beta \in \Theta$ and for $q = 1$, that H satisfies Mokkadem's conditions (i)–(ii), that $\{Y_t^\beta\}$ is irreducible, and that $E(|f_\infty^\beta|) < \infty$. Then, for each β, $\{Y_t^\beta\}$ is geometrically ergodic.*[4] *If, in addition, $\{f_t^\beta\}$ is Lipschitz, uniformly in probability, and $\beta \mapsto E(f_\infty^\beta)$ is continuous, then $\left\{\{f_t^\beta\} : \beta \in \Theta\right\}$ satisfies the uniform weak law of large numbers.*

This corollary follows immediately from Lemma 2 after applying Lemma 1 to see that $\{Y_t^\beta\}$ is ergodic and integrable.

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

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

The hypotheses of the corollary to Lemma 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).

[4] Note that geometric ergodicity is used in establishing that $\{f_t^\beta\}$ has a unique ergodic distribution and that the sample mean of $\{f_t^\beta\}$ converges in probability to $E(f_\infty^\beta)$. Any other regularity conditions that also implied these properties of $\{f_t^\beta\}$ could be substituted in the statement of this corollary.

Assumption 3 (Convergence of Distance Matrices). *Σ_0 is non-singular and $W_T \longrightarrow W_0 = \Sigma_0^{-1}$ almost surely, where (for any t)*

$$\Sigma_0 \equiv \sum_{j=-\infty}^{\infty} E([f_t^* - E(f_t^*)][f_{t-j}^* - E(f_{t-j}^*)]'). \tag{4.5}$$

For the second moments in this assumption to exist, and their sum to converge absolutely, the assumptions that $\{\| f_t^* \|_{2+\delta}: t = 1, 2, \dots\}$ 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 SMEs 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 sum of sample autocovariances of the data $\{f_t^*\}$, weighted as in Newey and West (1987)[5]. If, on the other hand, f_t^* depends on β, then a two-step procedure for estimating Σ_0 would be necessary. This would be the case, for example, if the moment equations associated with least squares estimation of the trend equations (2.11) were included. 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.

Under Assumptions 1–3, the asymptotic criterion function $C : \Theta \to I\!R$ defined by $C(\beta) = G_\infty(\beta)' W_0 G_\infty(\beta)$ constant for each β almost surely.

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 $\{b_T : T \geq 1\}$ given by (3.5).

[5] Several estimators of Σ_0 have been proposed in the literature. See, for example, Hansen and Singleton (1982), Eichenbaum, Hansen, and Singleton (1987), 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.

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 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 USLLN:

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

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

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \ell n \rho_\theta(\epsilon_t) = \alpha_\theta < 0 \quad a.s. \tag{4.6}$$

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

$$\| H(y,\beta,\epsilon_t) - H(x,\beta,\epsilon_t) \| \leq \rho_\theta(\epsilon_t) \| y - x \| .$$

In other words, for the AUC condition, $H(\cdot,\beta,\epsilon_t)$ must have a Lipschitz coefficient $\rho_\theta(\epsilon_t)$ with the property that $\prod_{s=0}^{t} \rho_\theta(\epsilon_s)$ 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 $\| \beta - \theta \| \leq \delta$, 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 $z \in \mathbb{R}^{N\ell}$, the function $f(z,\cdot) : \Theta \to \mathbb{R}^p$ has a Lipschitz constant $C_1(z)$, where C_1 satisfies a growth condition.[6] 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 $\| \beta - \theta \| \leq \delta$ implies that, for all $y \in \mathbb{R}^N$ and $\epsilon \in \mathbb{R}^p$,

$$\| H(y,\beta,\epsilon) - H(y,\theta,\epsilon) \| \leq C_2(y,\epsilon) \| \beta - \theta \|,$$

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.

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

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

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*

$$\sup_{\beta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} f_t^\beta - \frac{1}{T} \sum_{t=1}^{T} f_t^{\infty\beta} \right| \xrightarrow{\text{a.s.}} 0 \text{ as } T \to \infty, \tag{4.7}$$

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

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

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

denote the "modulus of continuity" of the process $\{f_t^{\infty\beta}\}$ at θ. Suppose:

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

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:

THEOREM 2 (STRONG CONSISTENCY). *Under Assumptions 2–5, 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 $E[\text{mod}_t(\delta, \theta)] < \infty$ 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 $\{\rho_\theta(\epsilon_t)\}$ satisfying $E[\rho_\theta(\epsilon_t)^2] < 1$ such that, for all x and y,*

$$\| H(y, \beta, \epsilon_t) - H(x, \beta, \epsilon_t) \| \leq \rho_\theta(\epsilon_t) \| y - x \| .$$

By Jensen's inequality, $\ell n E[\rho_\theta(\epsilon_t)] > E[\ell n \rho_\theta(\epsilon_t)]$, so that the L^2UC condition implies the AUC condition. Hence the lemmas preceeding Theorem 2 continue to hold under the L^2UC condition.

This strengthening of the unit circle condition leads to:

THEOREM 3. *Under Assumptions 2–4, the assumption that H and f are S-smooth, and the L^2UC 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 $\{\epsilon_t\}$. In this section we discuss several practical considerations that may influence which, if either, of these results assures consistency for SM estimation of dynamic asset pricing models.

First of all, the assumptions of geometric ergodicity and the asymptotic unit circle condition for $\{Y_t^\beta\}$ are not equivalent. 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

$$z_t = \xi + \rho z_{t-1} + \sigma \nu_{t-1}^\gamma \epsilon_t, \qquad \gamma < 1, \sigma > 0, |\rho| < 1, \tag{4.8}$$

where $\nu_t = z_t$ if $z_t \geq \eta > 0$ and $\nu_t = \eta$ otherwise, and suppose that $E(\epsilon_t) = 0$ for all t. (Equation (4.8) can also be interpreted as the discrete-time Euler approximation to a diffusion model, as in Example 2 of Section 4.1.) This representation of a shock process, which is similar to several widely studied representations of conditionally heteroskedastic processes, generally does not satisfy an L^2 unit circle condition. To see this, let $h(z, \epsilon, \beta)$ denote the right hand side of (4.8). Then

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

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

The process (4.8) 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, $\{Y_t^\beta\}$ is an irreducible process and our weak uniform continuity condition is satisfied, then weak consistency of the SME is implied by the UWLLN (Lemma 2). The fact that there is an important class of geometrically ergodic forcing processes that do not satisfy the unit circle conditions is a primary motivating reason for our analysis of weak consistency.

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. Though weaker in spirit than an L^2 unit circle condition, the uniform continuity condition implicity requires some damping of the effects of past shocks on current values of Y^β. We have not shown that processes of the form (4.8), 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.

Finally, verification of the geometric ergodicity assumption for $\{Y_t^\beta\}$ is not always straightforward. In particular, the sufficient conditions for geometric ergodicity of $\{Y_t^\beta\}$ given by Lemma 1 include the assumption that $\{Y_t^\beta\}$ is irreducible. This is a mild restriction on the process governing the exogenous forcing variables $\{X_t\}$. However, for models with endogenous state variables, verification of the irreducibility assumption may not be immediate. Fortunately, under plausible assumptions about the evolution of, say, the capital stock, a weak law of large numbers can be established without examining the irreducibility of the augmented state vector $\{Y_t\}$. We sketch a WLLN for this case in the context of the model in Section 2, since the proof highlights the robustness of our results to the choice of arbitrary initial conditions for the capital stock and forcing process.

Suppose that $\{X_t\}$ is geometrically ergodic and that the law of motion for the capital stock is given by

$$k_t = K(k_{t-1}, X_{t-1}, \beta), \tag{4.9}$$

where $\| k_t \|_2 < \infty$ for all t and $\{k_t\}$ is initialized at an arbitrary value of the capital stock. Assume that the function K satisfies an AUC condition; that is, for each $\theta \in \Theta$, there is some $\delta > 0$ and some sequence $\{\rho_\theta(X_t)\}$ of random variables with $\lim \frac{1}{T} \sum_{t=1}^T \ell n \rho_\theta(X_t) = \alpha_\theta < 0$ such that, for any k and k',

$$\| K(k, X_t, \beta) - K(k', X_t, \beta) \| \leq \rho_\theta(X_t) \| k - k' \|, \qquad \| \beta - \theta \| \leq \delta. \tag{4.10}$$

Inequality (4.10), analogous to our AUC condition on the transition function H, is satisfied by a large class of standard depreciation schemes for capital. For instance, the equilibrium law of motion for $\{\ell n\, k_t\}$ implied by (2.10) clearly satisfies (4.10).

Now, the geometric ergodicity of $\{X_t\}$ implies the existence of a stationary and ergodic process $\{X_t^\infty\}$ that satisfies (2.5). Furthermore, by an argument similar to that of Lemma 3, (4.10) implies the existence of a stationary and ergodic process $\{k_t^\infty\}$ that satisfies (4.9) with $\{X_t^\infty\}$ as the forcing process. By the ergodic theorem, $\{Y_t^{\infty\beta} = [X_t^{\infty\prime}, k_t^{\infty\beta}]'\}$ satisfies of WLLN. Thus, if $\left(\frac{1}{T}\sum_{t=1}^T Y_t^\beta - \frac{1}{T}\sum_{t=1}^T Y_t^{\infty\beta}\right)$

converges in probability to zero, then $\{Y_t^\beta\}$ satisfies a WLLN. The convergence of $\left(\frac{1}{T}\sum_{t=1}^T X_t - \frac{1}{T}\sum_{t=1}^T X_t^\infty\right)$ to zero in probability follows immediately from the geometric ergodicity of $\{X_t\}$. Also, using (4.10) and an argument similar to that in the proof of Lemma 4, it follows that

$$\begin{aligned}\left\| \frac{1}{T}\sum_{t=1}^T (k_t - k_t^\infty) \right\| &\le \frac{1}{T}\sum_{t=1}^T \| \; k_t - k_t^\infty \; \| \\ &\le \frac{1}{T}\sum_{t=1}^T \left[\prod_{j=1}^t \rho_\theta(X_{t-j}) \; \| \; k_0 - k_0^\infty \; \| \right] \longrightarrow 0 \text{ a.s. },\end{aligned}$$

as $T \to \infty$. Thus, the average difference between the two capital stock series converges in probability to zero.

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:

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.

(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 i-th row is the i-th 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,

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

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 $D_\beta f_t^\beta = \frac{d}{d\beta} f(Z_t^\beta, \beta)$ (the total derivative). Under the additional assumptions that the family $\left\{D_\beta f_t^\beta : \beta \in \Theta, t = 1, 2, \ldots\right\}$ is Lipschitz, uniformly in probability, and $E\left(D_\beta f_\infty^\beta\right)$ is a continuous function of β, Lemma 2 and Theorem 4.1.5 of Amemiya (1985) imply that $\text{plim}_T \partial G_T(b_T)/\partial\beta = D_0$. We therefore adopt:

ASSUMPTION 7. *The family* $\left\{D_\beta f_t^\beta : \beta \in \Theta, t = 1, 2, \ldots\right\}$ *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 $T/\mathcal{T}(T) \to \tau$ as $T \to \infty$. Under Assumptions 1-4, and 6-7,*

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

PROOF: From the definition of G_T,

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

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 $\sqrt{T}/\sqrt{\mathcal{T}(T)}$ to $\sqrt{\tau}$. ∎

An immediate implication of Theorem 3 is

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 (1987), Pakes and Pollard (1987), and Lee and Ingram (1989). As τ gets small, the asymptotic covariance matrix of $\{b_T\}$ approaches $\left[D_0' \Sigma_0^{-1} D_0\right]^{-1}$, 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 $\mathcal{T}(T)$ 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.

Theorem 4 presumes that $\{Y_t^\beta\}$ is a geometrically ergodic process. As noted in Section 4, this may not be easily verified in the presence of endogenous state variables. As with consistency, however, for many economic models our results still obtain in the presence of endogenous state variables as long as a depreciation condition like (4.10) holds. Pursuing this example of depreciable capital, an argument analogous to the discussion following (4.10) implies that $T^{-1/2}\sum_t(k_t - k_t^\infty)$ converges in probability to zero. Furthermore, geometric ergodicity of $\{X_t\}$ implies that $T^{-1/2}\sum_t(X_t - X_t^\infty)$ converges in distribution to zero. Hence, asymptotic normality of the SME follows from convergence in distribution of $T^{-1/2}\sum_t[Y_t^{\infty\beta_0} - E(Y_\infty^{\beta_0})]$ which, in turn, follows from the central limit theorem for stationary and ergodic processes.

All of 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 identify 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

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

can be calculated numerically. For large values of the simulation size T, $D(\beta)$ is approximately equal to $\frac{\partial E(f_t^\beta)}{\partial\beta}$. 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. Following are three obvious extensions:

(1) incorporating measurement errors on the observation vector f_t^*,

(2) letting f_t^* be a function of β, and

(3) avoiding the explicit calculation of security prices, or other observations defined by conditional moments.

In order to accommodate these extensions, we need one additional primitive, a measurable observation function $g : I\!R^{N\bar{\ell}} \times \Theta \to I\!R^M$, where $\bar{\ell}$ is the number of periods of states entering into the observation $g[(Y_t, \ldots, Y_{t-\bar{\ell}+1}), \beta]$ at time t. We can always assume without loss of generality that $\bar{\ell} = \ell$. We replace the observation f_t^* on the actual state process used in the SME with the observation $g_t^{\beta_0} \equiv g(Z_t, \beta_0)$. This leads us to consider the difference in sample moments:

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

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

For example, suppose that, instead of observing f_t^* itself, the contaminated series $g_t^{\beta_0} = f(Z_t, \beta_0) + u_t$, where $\{u_t\}$ is an ergodic, mean-zero $\mathbb{R}^M$-valued measurement error, is observed. 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 $\{g_t^\beta\}$. In this case, we replace Σ_0 defined by (4.5) with the weighted covariance matrix, for some positive scalar weight τ,

$$\Sigma_{f,g,\tau} = \tau\Sigma_0 + \Sigma_1, \tag{6.2}$$

where

$$\Sigma_1 = \sum_{j=-\infty}^{\infty} E\left([g_t^{\beta_0} - E(g_t^{\beta_0})][g_t^{\beta_0} - E(g_t^{\beta_0})]'\right). \tag{6.3}$$

Assuming that the families $\{f_t^\beta\}$ and $\{g_t^\beta\}$ satisfy the technical conditions of Assumption 1,[7] 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, under the same assumptions as in Theorem 4, $\sqrt{T}(b_T - \beta_0)$ converges in distribution to a normal random vector with mean zero and covariance matrix

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

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.

As a second extension, suppose one is interested in moments of functions of the data that depend on the unknown parameter vector. For instance, in an asset pricing

[7] 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).

setting, 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. In this case, both f_t^β and g_t^β will depend on β. Furthermore,

$$D_0 = \left[E\left(\frac{\partial g_t^{\beta_0}}{\partial \beta}\right) - E\left(\frac{\partial f_t^{\beta_0}}{\partial \beta}\right)\right] = E\left[-\left(\frac{\partial f_t^{\beta_0}}{\partial z_t^\beta}\right)\left(\frac{\partial z_t^\beta}{\partial \beta}\right)\right],$$

because (under the null hypothesis) the expected partial derivatives of g_t^β and f_t^β with respect to their second arguments are equal.

An important special case of this extension 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.

Finally, in some cases, and we are thinking specifically of asset prices, one of the coordinate functions, say g_j, on the actual state observation may be of the form

$$g_j[(Y_t, Y_{t-1}, \ldots, Y_{t-\ell+1}), \beta] = E[h_j(Y_t, \ldots, Y_{t+\ell-1}, \beta) | Y_t, Y_{t-1}, \ldots, Y_{t-\ell+1}],$$

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, by the law of iterated expectations, the observation $f_j(Z_t^\beta, \beta) = h_j(Z_t^\beta, \beta)$ is feasible, and has the same mean as $g_j(Z_t^\beta, \beta)$.

The obvious case is that of an asset claiming at time t the future dividends $d_1(Y_{t+1}, \beta)$, $d_2(Y_{t+2}, \beta), \ldots, d_{\ell-1}(Y_{t+\ell-1}, \beta)$ for the next respective $\ell - 1$ periods. In a representative-agent style economy, as in Lucas (1978), the price of the asset at time t is of the form

$$S_t = E\left(h_j[(Y_{t+\ell-1}, \ldots, Y_t), \beta] \mid Y_t\right),$$

where

$$h_j[(Y_{t+\ell-1},\ldots,Y_t),\beta] = \frac{1}{\mu(Y_t,\beta)} \sum_{s=t+1}^{t+\ell-1} [\Pi_{k=1}^{s}\rho(Y_{t+k})]\mu(Y_{t+s},\beta)d_s(Y_{t+s}^{\beta},\beta),$$

and where μ is the representative agent's marginal indirect utility for wealth and ρ is the representative agent's subjective discount factor. The security in question might be a coupon bond maturing in period $t+\ell-1$, or perhaps a European call option on another asset whose price in period s is $\pi(Y_s,\beta)$. In the latter case, for instance, where $\bar{s}$ is the time to expiry and K is the exercise price, we have $d_s = 0$, $s \neq \bar{s}$, and $d_{\bar{s}}(y,\beta) = [\pi(y,\beta) - K]^{+}$.

For the option example, the SME does not map directly into the setting of this paper, since $[\pi(y,\beta) - K]^{+}$ is not everywhere differentiable. However, Marcet and Singleton (1989) have extended the results in this paper to a class of nondifferentiable payoff functions which includes this function. Also, this application is limited by the fact that we cannot directly treat higher moment information on asset prices. For example, one cannot merely replace h with h^2 as the observation function corresponding to simulated second moments of asset prices, for the obvious reason that, typically, $S^2_{t-\ell+1} \neq E[h(Z_t,\beta)^2|Y_{t-\ell+1}]$. To some extent, observations on options substitute for second moment information on the underlying asset price given the non-linearity of the option payoff.

As the option pricing example illustrates, there are interesting economic models that do satisfy the regularity conditions imposed in this paper. Marcet and Singleton (1989) extend the results in this paper to certain models with continuous, but not everywhere differentiable GMM criterion functions. An interesting topic for future research is the investigation of models with discontinuous and nondifferentiable criterion functions for time series models. These are the time series counterparts of the models studied by McFadden (1989) and Pakes and Pollard (1989).

Appendix

PROOF OF LEMMA 2:[A1] Since Θ is compact it can be partitioned, for any n, into n non-overlapping neighborhood $\Theta_1^n, \Theta_2^n, \ldots, \Theta_n^n$ in such a way that the distance between any two points in each Θ_i^n goes to zero as $n \to \infty$. Let $\beta_1, \beta_2, \ldots, \beta_n$ be an arbitrary sequence of vectors such that $\beta_i \in \Theta_i^n$, $i = 1, \ldots, n$. Then, for any $\epsilon > 0$,

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

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.

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

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

[A1] 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 (1989) presents a more extensive discussion of sufficient conditions for uniform convergence in probability.

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

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

By continuity of $\beta \mapsto E(f_\infty^\beta)$, we can choose n once and for all so that $|\beta-\beta_i|B + |E(f_\infty^\beta) - E(f_\infty^{\beta_i})| < \frac{\epsilon}{4}$ 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. ▮

PROOF OF THEOREM 1: By the triangle inequality,

$$\begin{aligned}
&\left|\left(\frac{1}{T}\sum_{t=1}^{T} f_t^* - \frac{1}{T}\sum_{s=1}^{T} f_s^\beta\right) - \left[E(f_\infty^*) - E(f_\infty^\beta)\right]\right| \\
&\leq \left|E(f_\infty^*) - \frac{1}{T}\sum_{t=1}^{T} f_t^*\right| + \left|E(f_\infty^\beta) - \frac{1}{T}\sum_{s=1}^{T} f_s^\beta\right|.
\end{aligned} \tag{A.4}$$

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

$$\begin{aligned}
\delta_T(\beta) &= \left| G_T(\beta)'W_T G_T(\beta) - \left[E(f_\infty^*) - E\left(f_\infty^\beta\right)\right]' W_0 \left[E(f_\infty^*) - E\left(f_\infty^\beta\right)\right]\right| \\
&\leq \left| G_T(\beta) - \left[E(f_\infty^*) - E\left(f_\infty^\beta\right)\right]\right|' \, |W_T| \; |G_T(\beta)| \\
&\quad + |E(f_\infty^*) - E(f_\infty^\beta)|' \, |W_T - W_0| \, |G_T(\beta)| \\
&\quad + |E(f_\infty^*) - E(f_\infty^\beta)|' \, |W_0| \left| G_T(\beta) - \left[E(f_\infty^*) - E(f_\infty^\beta)\right]\right|.
\end{aligned} \tag{A.5}$$

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

$$\sup_{\beta\in\Theta} \delta_T(\beta) \leq \ell_T \, |W_T| \, [\phi_0 + \ell_T] + \phi_0 \, |W_T - W_0| \, [\phi_0 + \ell_T] + \phi_0 \, |W_0| \, \ell_T\,, \tag{A.6}$$

where $\phi_0 \equiv \max\left\{\left| E(f_\infty^*) - E(f_\infty^\beta)\right| : \beta \in \Theta\right\}$ 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, $\text{plim}_T \left[\sup_{\beta\in\Theta} \delta_T(\beta)\right] = 0$. 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. ∎

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

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

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

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

where

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

Hence,

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

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

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

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

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

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 $\{\epsilon_{t-s} : s \geq 0\}$. Since $\{\epsilon_t\}$ is an *i.i.d.* sequence, the stationarity and ergodicity of $\{Y_t^{\infty\beta}\}$ follows immediately. ∎

PROOF OF LEMMA 4: Fix $\theta \in \Theta$ and without loss of generality set $\ell = 1$. For any $\beta \in \Theta$ such that $\| \beta - \theta \| < \delta_\theta$,

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

where $k(\theta)$ is given by the S-smoothness assumption. The AUC condition implies that $\frac{1}{T}\sum_{t=1}^{T}\left[\prod_{j=0}^{t}\rho_\theta(\epsilon_j)\right]$ 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

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

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

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 $\Delta_t^{\beta} \leq \eta, T \geq T^*$, for all β in Θ, which leads to (4.7). ∎

PROOF OF THEOREM 4: As noted above, the L^2UC 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 $\theta \in \Theta, E[\mathrm{mod}_t(\delta,\theta)] < \infty$ for some $\delta > 0$. As before, we write "ϵ_t" for "$\hat{\epsilon}_t$."

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

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

It follows that

$$\operatorname{mod}_t(\delta, \theta) \leq \delta \sup_{\|\beta-\theta\| \leq \delta} C_1(Y_t^{\infty\beta}) + k(\theta) \sup_{\|\beta-\theta\| \leq \delta} \| Y_t^{\infty\beta} - Y_t^{\infty\theta} \| . \qquad (A.10)$$

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

$$\alpha_t \leq \rho_\theta(\epsilon_t)\alpha_{t-1} + C_2\left(Y_{t-1}^{\theta}, \epsilon_t\right)\delta. \qquad (A.11)$$

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

$$\alpha_t \leq \prod_{s=t-T}^{t} \rho_\theta(\epsilon_s)\alpha_{t-T} + \delta \sum_{s=t-T}^{t} C_2\left(Y_s^{\theta\infty}, \epsilon_s\right) \prod_{\tau=s+1}^{t} \rho_\theta(\epsilon_\tau)$$

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

$$\| X_T\alpha_{t-T} \|_1 \leq \| X_T \|_2 \| \alpha_{t-T} \|_2 \longrightarrow_T 0,$$

so, in L^1,

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

The right hand side is independent of β, and taking expectations, using the independence of $\{\epsilon_t\}$, we have

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

using independence of $\{\epsilon_t\}$ and the Cauchy-Schwarz inequality, where $\overline{\rho} = \| \rho_\theta(\epsilon_t) \|_2 < 1$ and where K is a bound on $\| C_2(Y_s^{\infty\beta}, \epsilon_s) \|_2$ implied by the growth condition on C_2 and the fact that $\| Y_s^{\infty\beta} \|_2$ and $\| \epsilon_s \|_2$ 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 $C_1(Y_t^{\infty\beta}) \leq d_1 + d_2 \| Y_t^{\infty\beta} \|$, for constants d_1, d_2. Furthermore,

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

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. ▮

References

Altug, Sumru, 1988, Gestation Lag and The Business Cycle: An Empirical Analysis, forthcoming, *International Economic Review.*

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, 1465–72.

Brock, W., 1980, Asset Prices in a Production Economy, in J. J. McCall, ed., *The Economics of Uncertainty*, Chicago, University of Chicago Press.

Christiano, L., 1988, Why Does Inventory Investment Fluctuate So Much?, *Journal of Monetary Economics* 21, pp. .

Donaldson, J. and R. Mehra, 1984, Comparative Dynamics of an Equilibrium Intertemporal Asset Pricing Model, *Review of Economic Studies* 51, 491–508.

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., Hansen L., and Singleton K., 1988, "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics* 103, 51–78.

Eichenbaum, M. and K. Singleton, 1986, Do Equilibrium Real Business Cycle Theories Explain Post-War U. S. Business Cycles?, in S. Fischer, ed., *Macroeconomics Annual 1986*, New York: National Bureau of Economic Research.

Gallant, R. and H. White, 1988, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models,* Oxford: Basil Blackwell.

Hall, R., 1988, Intertemporal Sutstitution in Consumption, *Journal of Political Economy* 96, 339–357.

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica* 50, 1029–56.

Hansen, L. and T. Sargent, 1980, Formulating and Estimating Dynamic Linear Rational Expectations Models, *Journal of Economic Dynamics and Control* 2, 7–46.

Hansen, L. and K. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50, 1269–1286.

Hansen, L. and K. Singleton, 1989, Efficient Estimation of Linear Asset Pricing Models with Moving Average Errors, manuscript, Stanford University.

Ingram, B., 1987, "Equilibrium Modeling of Asset Prices: Rationality versus Rules of Thumb," Cornell University.

King, R., C. Plosser, J. Stock, and M. Watson, 1987, Stochastic Trends and Economic Fluctuations, manuscript, University of Rochester.

Kydland, F. and E. Prescott, 1982, Time to Build and Economic Fluctuations, *Econometrica* 50, 1345–1370.

Jennrich, R., 1969, "Asymptotic Properties of Nonlinear Least Squares Estimators," *Annals of Mathematical Statistics* 40, 633-643.

Lee, B. and B. Ingram, 1988, "Estimation by Simulation," manuscript, Cornell University.

Lucas, R., 1978, "Asset Prices in an Exchange Economy," *Econometrica* 46, 1429–1445.

Marcet, A. and K. Singleton, 1989, Optimal Portfolio and Consumption Decisions in a Dynamic Model with Incomplete Markets and Restrictions on Portfolio Transactions, manuscript, Stanford University.

Marsh, T. and E. Rosenfeld, 1983, "Stochastic Processes for Interest Rates and Equilibrium Bond Prices," *Journal of Finance* 38, 635–46.

McFadden, D., 1987, "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," manuscript, MIT.

McLeish, D. 1975, "Invariance Principles for Dependent Variables," *Zeitschrift Wahrscheinlichkeitstheorie* 32, 165–178.

McLeish, D. 1975,"A Maximal Inequality and Dependent Strong Laws," *The Annals of Probability* 3, 829–839.

Mehra, R. and E. Prescott, 1985, The Equity Premium: A Puzzle, *Journal of Monetary Economics* 15, 145–162.

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 géométrique," *Comptes Rendues Academie Scientifique Paris* 301, Série I, 889-892.

Newey, W., 1989, Uniform Convergence in Probability and Stochastic Equicontinuity, manuscript, Princeton University.

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, 1987, "The Asymptotics of Simulation Estimators," manuscript, University of Wisconsin.

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

Sims, C., 1984, "Solving Nonlinear Stochastic Equilibrium Models 'Backwards', " University of Minnesota.

Singleton, K., 1987, "Asset Prices in a Time-Series Model with Disparately Inforemd, Competitive Traders," Chapter 12 in W. Barnett and K. Singleton, editors, *New Approaches to Monetary Economics*, Cambridge University Press, pp. 249-272.

Tauchen, G., 1986, Quadrature Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models, manuscript, Duke University.

Taylor, J. and H. Uhlig, 1989 "Solving Nonlinear Stochastic Growth Models: A Comparison of Alternative Solution Methods," manuscript, February, University of Minnesota and Institute for Empirical Macroeconomics.

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 University Press.