# SIMULATION-BASED ESTIMATION OF CONTINUOUS DYNAMIC MODELS

MANUEL S. SANTOS
Arizona State University

September 17, 2003

**Abstract**

This paper presents several results on consistency properties of simulation-based estimators for a general class of continuous dynamic models. The consistency of these estimators follows from a uniform convergence property of the sample paths over the vector of parameters. This convergence property is established under certain contractivity and monotonicity conditions on the dynamics of the system.

KEYWORDS: Continuous dynamical system, Markov equilibrium, invariant probability, simulation-based estimation, consistency, random contraction, random monotone process.

## 1 INTRODUCTION

As in other applied sciences, economic theories build upon the analysis of abstract, highly-stylized models that are often simulated by numerical techniques. The estimation and testing of these models can be quite challenging because of the nonlinearities embodied in the decision rules of private and public agents in environments which may comprise time and uncertainty, and the mechanisms of allocation of commodities in these economies. Most computable dynamic models are recursive, and their analysis is usually confined to equilibrium solutions generated by a dynamical system or policy function that defines a Markov equilibrium. It becomes then of interest to characterize the invariant probabilities or steady-state solutions, which commonly determine the long-run behavior of a model. But because of lack of information about the domain and form of these invariant probabilities the model must be simulated to compute the moments and other statistics of these distributions. Therefore, the process of estimation may entail the simulation of a family of models indexed by a vector of parameters. Moreover, properties of these estimators such as consistency and asymptotic normality are going to depend on the dynamics of the system. The study of these asymptotic properties may then require methods of analysis of probability theory in its interconnection with dynamical systems.

In a remarkable paper, Dubins and Freedman (1966) established certain stability properties of invariant probabilities for some families of Markov processes. As these authors observe, a continuity condition on the Markov process defined over a compact state space guarantees the existence of an invariant probability measure. Then, within the class of continuous Markov processes Dubins and Freedman focus on two seemingly simple cases: (i) For every realization of the shock process the dynamical system is contractive, and (ii) for every realization of the shock process the dynamical system is monotone.[1] For these two separate families of models they show that under mild regularity conditions the Markov process has a unique invariant probability measure, and such probability is globally stable in that starting from any initial distribution the system will converge in a certain defined sense to the unique invariant probability.

My purpose in this essay is to present a fairly systematic study of consistency properties of some simulation-based estimators for the above two families of continuous dynamical systems singled out by Dubins and Freedman. The consistency of these estimators for contractive systems has been explored by Duffie and Singleton (1993), and for monotone systems by Santos (2003). Here, I offer several extensions and generalizations of these results. A key step in the method of proof is to establish the uniform convergence of the simulated moments to their exact values in the vector of parameters. This is a standard strategy of proof in econometrics, but such convergence property is much harder to obtain for stochastic dynamical systems, and has been largely unexplored in the context of these models. This convergence property amounts to a uniform law of large numbers over a parameterized family of stochastic processes; in contrast, the stability of an invariant probability measure refers to the convergence of a sequence of distributions generated by an individual stochastic process.

[1] For present purposes, a mapping $h : X \to X$ is contractive if $\|h(x) - h(x')\| < \|x - x'\|$, where $\|x\|$ is the max norm on $X$. A mapping $h : X \to X$ is monotone increasing if $h(x) \geq h(x')$ for $x \geq x'$ where $\geq$ is an order on $X$, and $h$ is monotone decreasing if $h(x) \geq h(x')$ for $x' \geq x$.

A broad conclusion of the present study is then that the two families of Markov processes investigated in Dubins and Freedman (1966) also have the aforementioned property of uniform convergence of the sample moments. Therefore, these models can generate consistent simulation-based estimators. Of course, several important classes of dynamic models are left out of this study. First, some recent contributions [e.g., Bhattacharya and Lee (1988), Hopenhayn and Prescott (1992), Stenflo (2001) and Bhattacharya and Majumdar (2003)] have emphasized that for a random contraction or a random monotone process there could be a unique, globally stable invariant probability measure even in the absence of the continuity assumption. Hence, an open issue is whether some non-continuous families of models may also generate consistent estimators. Second, continuous Markov models with a unique invariant probability measure have the property that such distribution is globally stable in a mean sense [e.g., see Futia (1982, p. 383)] and in some cases the convergence is geometric. Hence, within the class of continuous stochastic processes it should be of interest to characterize some other families of models that generate consistent simulated estimators. It seems plausible that the consistency of these estimators may be validated under differentiability conditions. Fundamental developments in this area [cf. Arnold (1998)] have extended some classical results in the theory of dynamical systems to stochastic dynamics. For present purposes it would be useful to have in hand an analogous version of the infinite-dimensional implicit function theorem that is now available for deterministic systems [see Araujo and Scheinkman (1977), Santos and Bona (1989) and Burke (1990)]. This implicit function theorem has become a powerful tool in the comparative study of dynamic solutions.

In spite of all these possible extensions it should be stressed that there could be important families of models for which the aforementioned property of uniform convergence of the simulated moments in the vector of parameters may not be satisfied. The analysis centers on a system of stochastic difference equations of the following form

$$x_{t+1} = \xi(x_t, z_t, \varepsilon_{t+1}, \theta) \tag{1.1}$$

$$z_{t+1} = \psi(z_t, \varepsilon_{t+1}, \theta_2) \quad t = 0, 1, 2, \cdots .$$

These equations frequently arise in economic applications as Markovian equilibrium solutions of dynamic models. Here, $x_t$ is a vector of endogenous state variables that may represent investment decisions or the corresponding levels of the capital stocks, $z_t$ is a vector of exogenous state variables that may represent some indices of productivity, or intensity of tastes and population, and $\varepsilon_t$ is a vector of stochastic perturbations to the economy realized at the beginning of every time period $t$ and which follows an *iid* process. The vector $\theta = (\theta_1, \theta_2)$ specifies the model's parameters such as those parameters defining the utility and production functions. Observe that in this framework the vector of parameters $\theta_2$ characterizing the evolution of the exogenous state variables $z$ may influence the law of motion of the endogenous variables $x$, but this endogenous process may also be influenced by some additional parameters $\theta_1$. Functions $\xi$ and $\psi$ may represent the exact solution of a dynamic model or some numerical approximation. One should realize that the assumptions underlying these functions may be of a different economic significance, since $\xi$ governs the law of motion of the vector of endogenous variables $x$ and $\psi$ represents the evolution of the exogenous process $z$.

For a given notion of distance the estimation problem may be defined as follows: Find a parameter vector $\theta^0$ such that a selected set of the model's predictions are best matched with those of the data generating process. An estimator is a rule that yields a sequence $\{\hat{\theta}_T\}$ of candidate solutions for $\theta^0$ from finite samples of model's simulations and data. It is generally agreed that a reasonable estimator should possess the following consistency property: As sampling and simulation errors vanish the sequence of estimated values $\{\hat{\theta}_T\}$ should converge to the optimal solution $\theta^0$.

Since a change in $\theta$ may feed into the dynamics of the system in rather complex ways, traditional (data-based) estimators are of limited applicability for non-linear dynamic models. These estimators are just defined over data samples, and hence can only be applied to fully-fledged, structural dynamic models under fairly specific conditions. For instance, maximum likelihood posits a probability law for the process $(x_t, z_t)$ with explicit dependence on the parameter vector $\theta$. Likewise, standard non-linear least squares [e.g., Jennrich (1969)] and other generalized estimators [cf., Newey and McFadden (1994)] presuppose that functions $\xi$ and $\psi$ have analytical representations. Along these lines, one should consider the estimation procedures for continuous-time models of Ait-Sahalia (1996) and Hansen and Scheinkman (1995). All these methods postulate a closed-form representation for the process of state variables in the vector of parameters. This condition is particularly restrictive for the law of motion of the endogenous state variables: Only under rather especial circumstances one obtains a closed-form representation for the solution of a non-linear dynamic model.[2]

[2]The analysis of these estimators may be extended to numerical approximations in which functional evaluations can be performed by a computer program or by some other algorithmic method. But as stressed below data-based estimators can only be applied under certain functional restrictions. The analysis may break down in the presence of *latent* variables or some private information not available to the econometrician.

An alternative route to the estimation of non-linear dynamic models is via the Euler equations [e.g., see Hansen and Singleton (1982)] where the vector of parameters is determined by a set of orthogonality conditions conforming the first-order conditions or Euler equations of the optimization problem. A main advantage of this approach is that one does not need to model the shock process or to know the functional dependence of the law of motion of the state variables on the vector of parameters, since the objective is to find the best fit for the Euler equations over available data samples, within the admissible region of parameter values. The estimation of the Euler equations can then be carried out by standard non-linear least squares or by some other generalized estimator. However, model estimation via the Euler equations under traditional statistical methods is not always feasible. These methods are only valid for convex optimization problems with interior solutions in which the decision variables outnumber the parameters; moreover, the objective and feasibility constraints of the optimization problem must satisfy certain strict separability conditions along with the process of exogenous shocks. Sometimes the model may feature some latent variables or some private information which is not observed by the econometrician (e.g., shocks to preferences); lack of knowledge about these components of the model may preclude the specification of the Euler equations. An even more fundamental limitation is that the estimation is confined to orthogonality conditions generated by the Euler equations, whereas it may be of more economic relevance to estimate or test a model along some other dimensions such as those including certain moments of the invariant distributions or the process of convergence to

---

Moreover, these estimators search for the best fit of the equilibrium solution, but do not target directly the moments of the model's invariant distributions or some other quantitative properties of the equilibrium dynamics.

such stationary solutions.

Therefore, traditional data-based estimators usually search for a best fit of the equilibrium law of motion –or of the corresponding Euler equations and equilibrium conditions– from data samples, and can be implemented whenever these equations are explicitly written down. These estimation methods are not intended to evaluate the moments of the model's invariant distributions or some other aspects of the dynamics. Even if the model admits a closed-form solution, the statistics of an invariant distribution may not have an analytical representation and must be computed by numerical simulation. At a more practical operational level, these estimation methods may be infeasible in cases in which the minimization of the likelihood function –or any distance function involved in the estimation– is computationally costly or cannot be achieved by standard optimization routines. This problem may occur if the optimization involves a large number of parameters, local minima, or highly pronounced non-linearities.

The aforementioned limitations of traditional, data-based estimation methods for non-linear systems along with advances in computing have fostered the more recent use of estimation and testing based upon simulations of the model. Estimation by model simulation offers more scope to evaluate the behavior of the model by computing statistics of its invariant distributions that can be compared with their data counterparts. But this greater flexibility inherent in simulation-based estimators entails a major computational cost: Extensive model's simulations may be needed to sample the entire parameter space. Relatively little is known about the family of models in which simulation-based estimators would have good asymptotic properties such as consistency and normality. These properties would seem a minimal requirement for a rigorous application of estimation methods under the rather complex and delicate techniques of numerical simulation in which approximation errors may propagate in unexpected ways.

For establishing consistency of a simulation-based estimator the following major analytical difficulty arises. Each vector of parameters is manifested in a different dynamical system and so the proof of consistency has to cope with a continuous family of invariant distributions defined over the parameter space. In contrast, in data-based estimation there is only a unique distribution generated by the data process, and such distribution is not influenced by the vector of parameters. Then, the proof of consistency for a prototypical data-based estimator builds upon a uniform convergence argument over the parameter space under a fixed empirical process. For extensive accounts of work in this area, see Pollard (1984) and van der Vaart and Wellner (2000). In Dehardt (1971) the proof of uniform convergence relies on the monotonicity of a family of functions under a fixed invariant distribution. Also, Billingsley and Topsoe (1967) prove various uniform convergence results for compact classes of functions. All these results fall short of what is generally required to substantiate consistency for simulation-based estimators.

For some recent applications of simulation-based estimation, see Feinberg and Keane (2002), Gourinchas and Parker (2002), Hall and Rust (2002), and the collection of papers in Mariano, Schuermann and Weeks (1999). The present research should also be of interest to provide theoretical foundations for some efficient methods such as indirect inference

[Gourieroux, Monfort and Renault (1993)] and score methods [Gallant and Tauchen (1996)] and for the estimation of numerical approximations under continuity properties of invariant distributions [cf., Gaspar and Judd (1997), Krusell and Smith (1998), Williams (2002) and Santos and Peralta-Alva (2003)]. At this point, it is worth pointing out another strand of the literature concerned with simulation-based estimation in microeconomic settings [e.g., McFadden (1989), Pakes and Pollard (1989) and Rust (1994)]. This latter work is not suitable for the estimation of Markov models of the form (1.1) that one usually sees in macroeconomic applications.

Section 2 presents a simulation-based estimator along with the basic underlying assumptions. This estimator was proposed by Lee and Ingram (1991), and has been further analyzed by Duffie and Singleton (1993). It should be stressed that the methods of analysis developed here are not particularly tailored to this estimator, and hence these methods are of interest for the consistency of other simulation-based estimators. Consistency is to be understood in a strong sense, since familiar versions of the ergodic theorem for stochastic processes deal with almost sure convergence.

Section 3 derives several consistency properties of the estimator under certain contractivity conditions on the dynamics. This section extends work by Duffie and Singleton (1993) in several directions. Our assumptions are easier to check in macroeconomic applications, and the contractivity conditions are further weakened in cases in which alternative estimates of $\theta_2$ are available. Also, there is a third group of results concerned with the convergence of the estimated values from numerical approximations to the true vector of parameters, as the approximation errors of these numerical solutions converge to zero.

For a random contraction, each orbit converges exponentially to a fixed-point solution [e.g., see Schmalfuss (1996)]. Hence, one way to proceed in the proof of consistency of the estimator is to focuss on such fixed-point solution defined over the parameter space. The analytical framework is then formally equivalent to the more familiar problem of consistency of a traditional estimator for which this asymptotic property can be established by well known methods. Therefore, the consistency of a simulated estimator for a random contraction is ensured by the dampening behavior of the dynamics which leaves little scope for the propagation of small perturbations over time and guarantees the uniform convergence of the simulated moments. Contractivity conditions, however, are difficult to check for laws of motion of endogenous variables, and may appear rather restrictive for several economic applications.

Section 4 validates analogous consistency properties of the estimator under monotonicity conditions on the dynamics. These systems also preserve the uniform convergence of the simulated moments over the parameter space through an interaction of continuity and order-preserving properties, but an intuitive explanation for this result may seem now rather convoluted. The proof relies on the construction of local majorizing and minorizing mappings that bound the dynamics within small neighborhoods of parameter values. This type of local *sandwich* argument is familiar from the literature on empirical processes [e.g., Jennrich (1969) and Dehardt (1971)], and it is extended here to stochastic dynamical systems under the aforementioned continuity and order-preserving properties. Certain technical difficulties are involved in the method of proof such as the validity of a law of large numbers for the family of local majorizing and minorizing functions that may contain multiple invariant distributions.

Section 5 is devoted to a discussion of the main assumptions in the context of the one-sector neoclassical growth model, but several other economic applications are covered by the present results. Finally, let me conclude this long introduction with a word of caution about this research. Some simple dynamic economic models are hard to compute [e.g., see Ortigueira and Santos (2002)], and the application of standard numerical methods requires certain mathematical conditions. Consequently, one should expect that the assumptions under which these models may generate consistent simulation-based estimators are even more restrictive. Therefore, a primary objective of this line of research is to characterize those families of models that can be estimated under the powerful methods of numerical analysis. Simulation-based estimation offers an attractive framework to expose economic models to the data. Traditional, data-based estimation may constrain the analysis of an economic model and such estimators are not well suited to perform policy experiments.

## 2 A SIMULATION-BASED ESTIMATOR

As already pointed out, the analysis will focus on a simulation-based estimator proposed by Lee and Ingram (1991) and later analyzed by Duffie and Singleton (1993). This estimation method allows the researcher to assess the behavior of the model along various dimensions. Indeed, the conditions characterizing the estimation process may involve some moments of the model's invariant distributions or some other features of the dynamics on which the desired vector of parameters must be selected. There is, however, a major computational cost associated with this estimation exercise as extensive model's simulations may be required over representative samples of the parameter space.

### *2.1 Assumptions*

Let X denote the space of endogenous state variables $x$, and let $Z$ be the space of exogenous state variables $z$. For the sake of simplicity, both $X$ and $Z$ are compact domains that belong to some Euclidean space. The vector of shocks $\varepsilon_t$ follows an *iid* process with base space $\mathcal{E}$. The set $\Theta \equiv \Theta_1 \times \Theta_2$ denotes the region of parameter vectors $\theta = (\theta_1, \theta_2)$. The set $\Theta$ is also a compact domain.

Let $S = X \times Z$ and $\varphi = (\xi, \psi)$. Then, $s = (x, z)$ denotes an element in $S$, and $\|s\|$ is the max norm of vector $s$. Also, $\|\varphi\| = \sup_{(s,\varepsilon,\theta)\in S\times\mathcal{E}\times\Theta} \|\varphi(s, \varepsilon, \theta)\|$.

**(A.1)** Function $\varphi : S \times \mathcal{E} \times \Theta \to S$ is bounded.

**(A.2)** For every $(s, \theta)$, the mapping $\varphi(s, \cdot, \theta) : \mathcal{E} \to S$ is measurable.

**(A.3)** For every $(\varepsilon, \theta)$, the mapping $\varphi(\cdot, \varepsilon, \theta) : S \to S$ is continuous.

**(A.4)** For all $(s, \varepsilon)$, the mapping $\varphi(s, \varepsilon, \cdot) : \Theta \to \Theta$ is uniformly continuous. (That is, for every $\delta > 0$ there exists $\eta > 0$ such that for all $(s, \varepsilon)$ if $\|\theta - \theta'\| < \eta$ then $\|\varphi(s, \varepsilon, \theta) - \varphi(s, \varepsilon, \theta')\| < \delta$.)

Observe that $(A.1)-(A.4)$ will all be satisfied if $\varphi$ is a continuous function over a compact domain $S \times \mathcal{E} \times \Theta$. Under $(A.1)-(A.3)$ and the compactness of $S$ it follows that for each given value $\theta$ there exists an invariant distribution $\mu_\theta$ on $S$ for the mapping $\varphi(\cdot,\cdot,\theta)$. For a random contraction this invariant distribution $\mu_\theta$ is unique [e.g., see Stenflo (2001)]. Also, some simple conditions guarantee the existence of a unique invariant distribution $\mu_\theta$ for a random monotone system [e.g., Bhattacharya and Lee (1988), Dubins and Freedman (1966), and Hopenhayn and Prescott (1992)]. In what follows, it is assumed that there exists a unique invariant distribution $\mu_\theta$ corresponding to each parameter $\theta$. The uniqueness of the invariant distribution will simplify the analysis considerably, and it is necessary to obtain the global convergence results presented below.

### *2.2 The Simulated Moments Estimator (SME)*

Several elements conform the SME. First, one specifies a target function which typically would characterize a selected set of moments of the invariant distribution of the model and those of the data generating process. Second, a notion of distance is defined between the selected statistics of the model and its data counterparts. The minimum distance between these statistics is attained at some vector of parameters $\theta^0 = (\theta_1^0, \theta_2^0)$. Then, the estimation process yields a sequence of candidate solutions $\{\hat{\theta}_T\}$ over increasing finite samples of model's simulations and data so as to approximate the vector $\theta^0$.

**(A)** The *target function* $f : S \to R^p$ is assumed to be continuous. This function may represent $p$ moments of an invariant distribution $\mu_\theta$ defined as $E_\theta(f) = \int f(s)\mu_\theta(ds)$.

The expected value of $f$ over the invariant distribution of the data generating process will be denoted by $\bar{f}$.

**(B)** The *distance function* $G : R^p \times R^p \to R$ is assumed to be continuous. The minimum distance is attained at a vector of parameter values

$$\theta^0 = \arg\min_{\theta \in \Theta} \quad G(E_\theta(f), \bar{f}). \tag{2.1}$$

A typical specification of the distance function $G(E_\theta(f), \bar{f})$ is the following quadratic form

$$G(E_\theta(f), \bar{f}) = (E_\theta(f) - \bar{f}) \cdot W \cdot (E_\theta(f) - \bar{f})$$

where $W$ is a positive definite $p \times p$ matrix. Under the foregoing assumptions, one can show [cf., Santos and Peralta-Alva (2003, Th. 3.2)] that for (2.1) there exists an optimal solution $\theta^0$. Moreover, for the analysis below there is no restriction of generality to consider that $\theta^0$ is unique.

**(C)** The *estimation process* yields a sequence of estimated values $\{\hat{\theta}_T\}$ so as to approximate the solution $\theta^0$. These estimated values are obtained from associated optimization problems with finite samples of model's simulations and data.

Let $\tilde{\mathbf{s}} = \{\tilde{s}_t\}$ be a sample path of observations of the data generating process. Let $\omega = \{\varepsilon_t\}$ be a corresponding sequence of realizations of the shock process. Then, for each parameter value $\theta$ and initial condition $s_0 = (x_0, z_0)$ let $\{s_t(s_0, \omega, \theta)\}$ be the sequence generated by the dynamical system (1.1); that is, $s_{t+1}(s_0, \omega, \theta) = \varphi(s_t(s_0, \omega, \theta), \varepsilon_{t+1}, \theta)$ for all $t \geq 0$ and $\varphi \equiv (\xi, \psi)$. For a given distance function $G_T$ and a simulation rule

$\tau(T)$, an estimate $\hat{\theta}_T(s_0, \omega, \tilde{\mathbf{s}})$ is obtained as a solution to the following minimization problem

$$\hat{\theta}_T(s_0, \omega, \tilde{\mathbf{s}}) = \arg\min_{\theta \in \Theta} \quad G_T(\frac{1}{\tau(T)} \Sigma_{t=1}^{\tau(T)} f(s_t(s_0, \omega, \theta)), \frac{1}{T} \Sigma_{t=1}^{T} f(\tilde{s}_t)). \tag{2.2}$$

The rule $\tau(T)$ reflects that model's simulations may be of a different length than data samples, but it is required that $\tau(T) \to \infty$ as $T \to \infty$. The sequence of functions $\{G_T\}_{T \geq 1}$ is assumed to converge uniformly to function $G$ as $T \to \infty$.

In this framework, the presumption is that the researcher has access to a random realization $\tilde{\mathbf{s}} = \{\tilde{s}_t\}$ and can perform evaluations of function $\varphi$ at any given point $(s, \varepsilon, \theta)$; later, the analysis will consider the more typical situation in which the researcher can only obtain values for a numerical approximation $\varphi^n$. Also, as is typical in numerical simulation the postulated distribution of $\{\varepsilon_t\}$ is known, but no knowledge of the actual realization of shocks $\{\varepsilon_t\}$ is required. This latter assumption is too strong but it is sometimes needed for the implementation of some data-based estimators. For the SME this assumption can be supplanted by the weaker condition that the researcher can draw sequences from a generating process $\{\hat{\varepsilon}_n\}$ that can mimic the distribution of $\{\varepsilon_t\}$. A measure $\tilde{\gamma}$ is defined over the space of sequences $\tilde{\mathbf{s}} = \{\tilde{s}_t\}$ and a measure $\gamma$ is defined over the space of random shocks $\omega = \{\varepsilon\}$. (The construction of measure $\gamma$ follows from standard arguments [e.g., see Stokey, Lucas and Prescott (1989, Ch. 8)].) Let $\lambda$ represent the product measure $\lambda = \gamma \times \tilde{\gamma}$.

Definition: *The SME is a sequence of measurable functions* $\{\hat{\theta}_T(s_0, \omega, \tilde{\mathbf{s}})\}_{T \geq 1}$ *such that each function* $\hat{\theta}_T$ *satisfies (2.2) for all* $s_0$ *and* $\lambda$*-almost all* $(\omega, \tilde{\mathbf{s}})$.

Remark: Because of the recursive structure embedded in the parameter space $\Theta$, sometimes the value $\theta_2^0$ may be known or may be estimated independently by a more efficient method. In those situations, for a fixed $\theta_2^0$ one may consider a constrained version of optimization problem (2.2) over $\Theta_1$, and define the constrained SME as $\{\hat{\theta}_{1T}(s_0, \omega, \tilde{\mathbf{s}}, \theta_2^0)\}_{T\geq 1}$.

## 3 RANDOM CONTRACTIONS

This section analyzes various consistency properties of the SME under certain contractivity conditions on the dynamics of system (1.1). For stochastic systems several contractivity properties can be found in the literature that depend on the domain of definition and on the metric or distance function. Our analysis will focus on two main contractivity conditions. The consistency of the SME is established for the whole vector of parameters $\theta$, and for the first component $\theta_1$ when the true value $\theta_2^0$ is known. Further convergence results are derived for estimates obtained from numerical approximations of function $\varphi$.

### *3.1 Consistency of the SME*

As already pointed out, the analysis rests on two alternative contractivity conditions on the dynamics of system (1.1). The first condition draws on some methods developed by Kifer (1986) who proposed a notion of characteristic exponent in metric spaces. This notion seems appropriate for non-smooth functions. Let

$$A_\delta(s, \varepsilon, \theta) = \sup_{s' \in B_\delta(s), s' \neq s} \frac{\|\varphi(s', \varepsilon, \theta) - \varphi(s, \varepsilon, \theta)\|}{\|s' - s\|} \tag{3.1}$$

where $B_\delta(s) = \{s' : \|s' - s\| < \delta\}$. Hence, $A_\delta(s, \varepsilon, \theta)$ provides an upper bound for the slope of function $\varphi$ at point $(s, \varepsilon, \theta)$ over all $s'$ in $B_\delta(s)$. If $\varphi$ is a Lipschitz function, then $A_\delta(s, \varepsilon, \theta)$ is a finite number.

**(C.1)** For every $\theta$ there is a neighborhood $V(\theta)$ such that for some $\delta > 0$ and all $\hat{\theta}$ in $V(\theta)$ there exists a measurable function $c(\varepsilon)$ with the following properties

**(i)** $\log A_\delta(\varepsilon, \hat{\theta}) < c(\varepsilon)$, where $A_\delta(\varepsilon, \hat{\theta}) = \sup_{s \in S} A_\delta(s, \varepsilon, \hat{\theta})$.

**(ii)** $Ec(\varepsilon) < 0$.

REMARK: Roughly speaking, $(C.1)$ asserts that over a small neighborhood $V(\theta)$ the maximum log value of the slope of function $\varphi$ with respect to $s$ is on average a negative number. For fixed $\theta$, a similar condition is stated in Kifer (1986, p. 23) and a differentiable version of this condition can be found in Schmalfuss (1996). Condition $(C.1)$ is closely related to the *Asymptotic Unit−Circle* condition of Duffie and Singleton (1993), and it is stated here in a more compact form following the work of Kifer (1986) and Schmalfuss (1996). An alternative contractivity condition that it is often easier to check in macroeconomic applications is the following:

**(C.2)** For almost all $\omega$, for every vector $\theta$ and initial condition $s_0$ we have

**(i)** There are constants $N(s_0, \omega, \theta) > 0$ and $0 < \alpha(s_0, \omega, \theta) < 1$ and a ball $B_{\delta(s_0,\omega,\theta)}(s_0) = \{s : \|s - s_0\| < \delta(s_0, \omega, \theta)\}$ such that

$$\|s_t(s, \omega, \theta) - s_t(s_0, \omega, \theta)\| \leq N(s_0, \omega, \theta)\alpha^t(s_0, \omega, \theta) \|s - s_0\| \tag{3.2}$$

for all $s$ in $B_{\delta(s_0,\omega,\theta)}(s_0)$ and all $t \geq 1$.

**(ii)** If $s_1 = \varphi(s_0, \omega, \theta)$ and $\omega^{-1} = \{\varepsilon_t\}_{t \geq 2}$, then $N(s_0, \omega, \theta) \geq N(s_1, \omega^{-1}, \theta)$, $\alpha(s_0, \omega, \theta) \geq \alpha(s_1, \omega^{-1}, \theta)$ and $\delta(s_0, \omega^{-1}, \theta) \leq \delta(s_1, \omega^{-1}, \theta)$.

In (3.2) the expression $\alpha^t(s_0, \omega, \theta)$ means constant $\alpha(s_0, \omega, \theta)$ to the power $t$, and $s_{t+1}(s, \omega, \theta)$ is defined recursively as $s_{t+1}(s, \omega, \theta) = \varphi(s_t(s, \omega, \theta), \varepsilon_{t+1}, \theta)$ for all $t \geq 1$. Hence, the first part of ($C$.2) imposes a local contractivity condition on the dynamics since

$\|s_t(s, \omega, \theta) - s_t(s_0, \omega, \theta)\| \leq N\alpha^t \|s - s_0\|$ for some constants $N > 0$ and $0 < \alpha < 1$. Then, the second part requires these local bounds to be uniform along the orbit. For models with a globally attractive invariant distribution, Condition ($C$.2) may be relevant for points $s_0$ outside the ergodic set in which constant $N$ may become arbitrarily large. For these models, ($C$.1) is very restrictive since this latter condition imposes bounds that apply for all $s_0$ in $S$. Of course, if we neglect transitional dynamic behavior then ($C$.2) is usually more stringent.

THEOREM 3.1: *Let (A.1)-(A.4) be satisfied. Then under either* ($C$.1) *or* ($C$.2)*, for all* $s_0$ *and* $\lambda$*-almost all* $(\omega, \tilde{\mathbf{s}})$ *the SME* $\{\hat{\theta}_T(s_0, \omega, \tilde{\mathbf{s}})\}_{T \geq 1}$ *converges to* $\theta^0$.

Theorem 3.1 is proved in the appendix. Two separate proofs are given corresponding to Conditions ($C$.1) and ($C$.2). The proof under ($C$.2) is relatively simple, and builds on a familiar stability argument for local contractions. The proof under ($C$.1) is more involved, and proceeds along the lines of Duffie and Singleton (1993). Under the simple Assumptions (A.1)-(A.4) one major objective in this section is to dispense with some rather technical conditions invoked by these authors. The method of proof is based on some auxiliary results of independent interest. These results will be discussed presently.

The first lemma requires an innocuous extension of the space of shocks [cf., Krengel (1985, Ch. 2)] in which $t$ ranges from $-\infty$ to $\infty$. Hence, for this result every sequence of shocks $\tilde{\omega}$ is of the form $\tilde{\omega} = (\cdots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \cdots)$.

Lemma 3.2: *Under (A.1)-(A.4) and* (**C**.1), *for almost all $\tilde{\omega}$ there exists a unique fixed-point solution $\{s_t^*(\tilde{\omega}, \theta)\}$ for $-\infty < t < \infty$ such that for each $\theta$,*

$$s_{t+1}^*(\tilde{\omega}, \theta) = \varphi(s_t^*(\tilde{\omega}, \theta), \varepsilon_{t+1}, \theta) \text{ for all } t. \tag{3.3}$$

*For each $\theta$ the mapping $s_t^*(\cdot, \theta)$ is measurable. Moreover, for every initial condition $s_0$ all sample paths $s_t(s_0, \tilde{\omega}, \theta)$ converges uniformly to $s_{t+1}^*(\tilde{\omega}, \theta)$ in $\theta$ as $t$ goes to $\infty$.*

This lemma is a extension of earlier results by Kifer (1986, Ch. 1) and Schmalfuss (1986) to the parameterized family of stochastic processes in (1.1). Then for the purposes of the proof of Theorem 3.1 it suffices to analyze the convergence properties of the sequences $\frac{1}{\tau(T)} \sum_{t=1}^{\tau(T)} f(s_t^*(\tilde{\omega}, \theta))$. Hence, standard proofs of consistency for data-based estimation [e.g., Jennrich (1969)] can be applied to the present context provided that $s_t^*(\tilde{\omega}, \theta)$ is a continuous function of $\theta$. This latter result is established in the following lemma.

Lemma 3.3: *Under the conditions of Lemma 3.2, for each $t$ the mapping $s_t^*(\tilde{\omega}, \cdot)$ is continuous on $\Theta$ for almost all $\tilde{\omega}$.*

### *3.2 Constrained Estimation*

In some applications it may be possible to get independent estimates of the true value $\theta_2^0$ by more practical estimation methods. In those situations simulation-based estimation can be restricted to the first component vector $\theta_1$. Consequently, the above contractivity conditions can be relaxed, since it is only necessary to secure the almost sure convergence of the sequence of estimates $\{\hat{\theta}_{1T}\}$ to the true value $\{\theta_1^0\}$. These contractivity conditions will now be required to hold for the law of motion of the vector of endogenous variables $x$. Let

$$H_\delta(x, z, \varepsilon, \theta_1, \theta_2^0) = \sup_{x' \in B_\delta(x), x' \neq x} \frac{\|\zeta(x', z, \varepsilon, \theta_1, \theta_2^0) - \zeta(x, z, \varepsilon, \theta_1, \theta_2^0)\|}{\|x' - x\|} \tag{3.4}$$

where $B_\delta(x) = \{x' : \|x - x'\| < \delta\}$.

**(C.1′)** For every $\theta_1$ there is a neighborhood $V(\theta_1)$ such that for some $\delta > 0$ and all $\hat{\theta}_1$ in $V(\theta_1)$ there exists a measurable function $c(z, \varepsilon)$ with the following properties

**(i)** $\log H_\delta(z, \varepsilon, \hat{\theta}_1, \theta_2^0) < c(z, \varepsilon)$, where $H_\delta(z, \varepsilon, \hat{\theta}_1, \theta_2^0) = sup_{x \in X} H_\delta(x, z, \varepsilon, \hat{\theta}_1, \theta_2^0)$.

**(ii)** $E(c(z, \varepsilon)) < 0$

In this condition the expectation $E(c(z, \varepsilon))$ is taken with respect to the invariant distribution of vector $(z, \varepsilon)$. Since $\varepsilon$ is an *iid* process, then this invariant distribution is a product measure conformed by the invariant distributions of variables $z$ and $\varepsilon$. Note that the invariant distribution of $z$ is determined by $\theta_2^0$. Also, regarding Condition $(C.2)$ the following weakened version applies to the dynamics of the vector of endogenous state variables, $x$.

**(C.2′)** For almost all $\omega$, for every vector $\theta$ and initial condition $s_0 = (x_0, z_0)$ we have

**(i)** There are constants $N(s_0, \omega, \theta) > 0$ and $0 < \alpha(s_0, \omega, \theta) < 1$ and a ball $B_{\delta(s_0,\omega,\theta)}(x_0) = \{x : \|x_0 - x\| < \delta(s_0, \omega, \theta)\}$ such that

$$\|x_t(s, \omega, \theta) - x_t(s_0, \omega, \theta)\| \leq N(s_0, \omega, \theta)\alpha^t(s_0, \omega, \theta) \|s - s_0\| \tag{3.5}$$

for all $x$ in $B_{\delta(s_0,\omega,\theta)}(x_0)$ and all $t \geq 1$.

**(ii)** If $s_1 = \varphi(s_0, \omega, \theta)$ and $\omega^{-1} = \{\varepsilon_t\}_{t \geq 2}$, then $N(s_0, \omega, \theta) \geq N(s_1, \omega^{-1}, \theta)$, $\alpha(s_0, \omega, \theta) \geq \alpha(s_1, \omega^{-1}, \theta)$ and $\delta(s_0, \omega^{-1}, \theta) \leq \delta(s_1, \omega, \theta)$.

Notice that $x_t(s_0, \omega, \theta)$ in $(C.2')$ refers to the first component vector of $s_t(s_0, \omega, \theta)$ as referred to in $(C.2)$.

THEOREM 3.4: *Let (A.1)-(A.4) be satisfied. Then, under either* $(C.1)'$ *or* $(C.2)'$ *for all* $x_0$*, and almost all* $(z_0, \omega, \tilde{\mathbf{s}})$ *the SME* $\{\hat{\theta}_{1T}(x_0, z_0, \omega, \tilde{\mathbf{s}}, \theta_0^2)\}_{T \geq 1}$ *converges to* $\theta_1^0$.

COROLLARY 3.5: *Suppose that for almost all* $(z_0, \omega, \tilde{\mathbf{s}}_0)$ *the estimator* $\{\hat{\theta}_{2T}(z_0, \omega, \tilde{\mathbf{s}})\}_{T \geq 1}$ *converges to* $\theta_2^0$*. Then, under the conditions of Theorem 3.4, for all* $x_0$*, and almost all* $(z_0, \omega, \tilde{\mathbf{s}})$ *the SME* $\{\hat{\theta}_{1T}(x_0, z_0, \omega, \tilde{\mathbf{s}}, \hat{\theta}_{2T}(z_0, \omega, \tilde{\mathbf{s}}))\}_{T \geq 1}$ *converges to* $\theta_2^0$.

### *3.3 Estimation of Numerical Approximations*

In most dynamical models the equilibrium solution $\varphi$ cannot be computed exactly. Hence, a typical situation is that the researcher can only perform functional evaluations of a numerical approximation $\varphi^n$. This approximate function $\varphi^n$ generates a new vector of parameters $\theta^n$ as a solution to optimization problem (2.1). More specifically,

$$\theta^n = \arg\min_{\theta \in \Theta} G(\int f(s)\mu_\theta^n(ds), \bar{f}) \tag{3.6}$$

where $\mu_\theta^n$ is an invariant distribution for the mapping $\varphi^n(\cdot, \cdot, \theta)$ for each $\theta$ in $\Theta$. The invariant distribution $\mu_\theta^n$ may not be unique, even though for each $\theta$ the original mapping $\varphi(\cdot, \cdot, \theta)$ is assumed to have a unique invariant distribution $\mu_\theta$. Also, the solution $\theta^n$ may not be unique. The idea is that certain economic assumptions may guarantee the existence of an invariant distribution $\mu_\theta$ for $\varphi(\cdot,\cdot,\theta)$ but uniqueness of the invariant distribution is not generally preserved under numerical approximations or under some other perturbations of the model. Hence, problem (3.6) may be understood as a minimization over all possible invariant distributions $\mu_\theta^n$. Then, it is of interest to know whether the set of solutions $\{\theta^n\}$ defined by (3.6) converge to the original solution $\theta^0$ defined by (2.1) as $\varphi^n$ approaches $\varphi$.

To substantiate this latter convergence property, Condition $(C.1)$ will be replaced by a related contractivity condition which is widely used in the literature on random contractions [cf., Norman (1972), Futia (1982), and Stenflo (2001)].

**(C.3)** For every $\theta$ there exists a constant $0 < \alpha < 1$ such that $\int \|\varphi(s', \varepsilon, \theta) - \varphi(s, \varepsilon, \theta)\| \, Q(d\varepsilon) \leq \alpha \|s' - s\|$ for all $s', s$ in $S$.

THEOREM 3.6: *Assume that the sequence of functions* $\{\varphi^n\}$ *converges to* $\varphi^n$. *Let* $\varphi^n$ *satisfy (A.2)-(A.3) for each* $n$. *Let* $\varphi$ *satisfy (A.1)-(A.4), and either* $(C.2)$ *or* $(C.3)$. *Then every sequence of optimal solutions* $\{\theta^n\}$ *defined by (3.6) must converge to* $\theta^0$ *defined by (2.1).*

REMARK: (a) The convergence of the sequence of functions $\{\varphi^n\}$ should be understood in the sup norm defined in Section 2. Observe that no contractivity conditions are imposed on the approximate functions $\{\varphi^n\}$. This is relevant for numerical approximations since these contractivity properties may not hold true for numerical interpolations.

(b) The main step in the proof of Theorem 3.6 is to establish the uniform convergence in the weak topology of the sequence of invariant distributions $\{\mu_\theta^n\}$ to $\mu_\theta$ in $\theta$ as $n$ goes to $\infty$.

(c) The above results on constrained estimation (Theorems 3.4 and Corollary 3.5) can also be extended to the present setting of estimation of numerical approximations. Also, for each $n$ one can define the SME $\{\hat{\theta}_T^n(s_0, \omega, \tilde{\mathbf{s}})\}_{T\geq 1}$ over all sample paths $s_t^n(s_0, \omega, \tilde{\mathbf{s}})$ generated by the approximate function $\varphi^n$. Then, combining Theorems 3.1 and 3.6 we get that generically $\hat{\theta}_T^n(s_0, \omega, \tilde{\mathbf{s}})$ and $\hat{\theta}_T(s_0, \omega, \tilde{\mathbf{s}})$ will be arbitrarily close provided that $n$ and $T$ are large enough.

## 4 RANDOM MONOTONE PROCESSES

This section studies analogous consistency properties of the SME under order-preserving conditions on the dynamics of system (2.1). These order preserving conditions are usually easier to verify, since they can be derived from primitive assumptions of economic models [cf., Hopenhayn and Prescott (1992) and Mirman, Morand and Reffet (2003)].

The analysis draws on earlier paper [Santos (2003)]. These earlier results are now extended using the following weaker assumptions: (i) $(A.1) - (A.4)$ replace a stronger continuity assumption on function $\varphi$, (ii) the moment function $f$ is only assumed to be continuous whereas previously this function was also assumed to be monotone, and (iii) a suitable law of large numbers from Santos and Peralta-Alva (2003) is invoked –rather than the familiar ergodic theorem– so that convergence holds for all initial conditions $s_0$ over $\lambda$-almost all $(\omega, \tilde{\mathbf{s}})$.

### *4.1 Consistency of the SME*

Assume that an order relation $\geq$ is defined on $S$. For concreteness, let $\geq$ be the Euclidean order. Hence, if $s = (\cdots, s_i, \cdots)$ and $s' = (\cdots, s'_i, \cdots)$ are two vectors in $S$, then $s \geq s'$ means that $s_i \geq s'_i$ for each coordinate $i$. A function $h : S \to S$ is called *order preserving* or *monotone increasing* if $h(s) \geq h(s')$ for all $s \geq s'$. Conversely, a function $h : S \to S$ is called *order reversing* or *monotone decreasing* if $h(s) \geq h(s')$ for all $s' \geq s$. All results in this section are stated for monotone increasing functions.

**(M)** For all $(\varepsilon, \theta)$ the mapping $\varphi(\cdot, \varepsilon, \theta) : S \to S$ is monotone increasing.

One should realize that no order preserving assumptions are made over the space of shocks $\mathcal{E}$ and over the parameter space $\Theta$.

THEOREM 4.1: *Under (A.1)-(A.4) and (M), for all $s_0$ and $\lambda$-almost all $(\omega, \tilde{\mathbf{s}})$ the SME $\{\hat{\theta}_T(s_0, \omega, \tilde{\mathbf{s}})\}_{T \geq 1}$ converges to $\theta^0$.*

The proof of Theorem 4.1 relies on a repeated application of a law of large numbers from Santos and Peralta-Alva (2003) to a countable collection of local majorizing and minorizing functions for the parameterized family of dynamical systems (1.1) over small neighborhoods of the parameter space $\Theta$. The orbits generated by these local bounding functions place upper an lower limits on the orbits generated by the individual dynamical systems (1.1) over these small neighborhoods of parameter values. Then, the uniform convergence of the simulated moments $\{\frac{1}{T}\sum_{t=1}^{T} f(s_t(s_0, \omega, \theta))\}$ to $E_\theta(f)$ in $\theta$ follows from a *sandwich* argument which is familiar in the theory of estimation under a fixed empirical process [e.g., see Jennrich

(1969), Pollard (1984), and van der Vaart and Wellner (2000)]. The extension of this familiar argument to a continuum family of stochastic processes involves the use of Condition $(M)$ and a continuity property on the set of invariant distributions from Santos and Peralta-Alva (2003).

Here are the main elements of the proof of Theorem 4.1. For each given $(\varepsilon, \theta)$ and constant $\kappa > 0$ define the majorizing function

$$\varphi^{sup}(s, \varepsilon, \theta, \kappa) = \sup_{\theta'} \varphi(s, \varepsilon, \theta') \qquad s.\ t.\ \{\theta' : \| \theta' - \theta \| < \kappa\} \tag{4.1}$$

and the minorizing function,

$$\varphi^{inf}(s, \varepsilon, \theta, \kappa) = \inf_{\theta'} \varphi(s, \varepsilon, \theta') \qquad s.\ t.\ \{\theta' : \| \theta' - \theta \| < \kappa\}. \tag{4.2}$$

In these definitions the $sup$ and $inf$ are taken coordinate by coordinate. The following simple results are stated for function $\varphi^{sup}$, but analogous results hold for function $\varphi^{inf}$.

LEMMA 4.2: *Let (A.1)-(A.4) be satisfied. Then,*

**(i)** *For each $(s, \theta, \kappa)$ the mapping $\varphi^{sup}(s, \cdot, \theta, \kappa) : \mathcal{E} \to S$ is measurable.*

**(ii)** *For each $\varepsilon$ the mapping $\varphi^{sup}(\cdot, \varepsilon, \cdot, \cdot) : S \times \Theta \times R_+ \to S$ is continuous.*

**(iii)** *For each $(s, \theta, \kappa)$ the mapping $\varphi^{sup}(\cdot, \varepsilon, \theta, \kappa) : S \to S$ is monotone.*

This lemma is a straightforward consequence of the definition (4.1) and the above assumptions. Observe that

$$\varphi^{sup}(\cdot, \varepsilon, \theta, \kappa) \geq \varphi(\cdot, \varepsilon, \theta') \geq \varphi^{inf}(\cdot, \varepsilon, \theta, \kappa) \tag{4.3}$$

for every $\theta'$ such that $\parallel \theta' - \theta \parallel < \kappa$. Hence,

$$s_1^{sup} = \varphi^{sup}(s_0, \varepsilon_1, \theta, \kappa) \geq s_1 = \varphi(s_0, \varepsilon_1, \theta') \geq s_1^{inf} = \varphi^{inf}(s_0, \varepsilon_1, \theta, \kappa) \tag{4.4}$$

for all $s_0$. Now, by (4.3)-(4.4) and $(M)$,

$$s_2^{sup} = \varphi^{sup}(s_1^{sup}, \varepsilon_2, \theta, \kappa) \geq s_2 = \varphi(s_1, \varepsilon_2, \theta') \geq s_2^{sup} = \varphi^{sup}(s_1^{sup}, \varepsilon_2, \theta, \kappa). \tag{4.5}$$

Therefore, proceeding by induction

$$s_t^{sup} \geq s_t \geq s_t^{sup} \tag{4.6}$$

for all $t \geq 1$.

This order-preserving property of the dynamics reduces the proof of uniform convergence of the simulated moments $\{\frac{1}{T}\sum_{t=1}^{T} f(s_t(s_0, \omega, \theta))\}$ to $E_\theta(f)$ in $\theta$ to a sandwich argument over a countable sequence of functions $\varphi^{sup}(s_0, \varepsilon_1, \theta, \kappa)$ and $\varphi^{inf}(s_0, \varepsilon_1, \theta, \kappa)$ for selected $(\theta, \kappa)$. To carry out this argument the following auxiliary results are needed: (i) A continuity property on the set of invariant distributions, and (ii) a law of large numbers for the bounding functions $\varphi^{sup}(s_0, \varepsilon_1, \theta, \kappa)$ and $\varphi^{inf}(s_0, \varepsilon_1, \theta, \kappa)$ for all arbitrary initial conditions $s_0$, even if these functions contain multiple invariant distributions. Both results are established in an earlier paper [Santos and Peralta-Alva (2003)].

Let $\mu_{\theta,\kappa}^{sup}$ be an invariant distribution under function $\varphi^{sup}(s, \varepsilon, \theta, \kappa)$ and $\mu_\theta$ be the unique invariant distribution for function $\varphi(s, \varepsilon, \theta)$. Function $\varphi^{sup}(s, \varepsilon, \theta, \kappa)$ may contain multiple invariant distributions. Let $\Delta^{sup}_{\theta,\kappa}$ be the set of all the invariant distributions $\mu^{sup}_{\theta,\kappa}$ under $\varphi^{sup}(s,\varepsilon,\theta,\kappa)$. Note that $\Delta^{sup}_{\theta,\kappa}$ is a compact convex set in the weak topology of measures. Also, for every sequence of shocks $\omega = \{\varepsilon_n\}$ and initial condition $s_0$, let $s^{sup}_{t+1}(s_0,\omega,\theta,\kappa) = \varphi^{sup}(s^{sup}_t(s_0,\omega,\theta,\kappa),\varepsilon_{t+1},\theta,\kappa)$ for all $t \geq 1$.

LEMMA 4.3: *Every sequence of probability measures $\{\mu^{sup}_{\theta,\kappa}\}$ converges to $\{\mu_\theta\}$ as $\kappa$ goes to zero.*

This lemma follows from the upper semicontinuity of the correspondence of invariant distributions. The next result shows that these invariant distributions bound the range of variation of the average behavior of a typical sample path.

LEMMA 4.4: *For all $s_0$ and almost all $\omega$,*

$$\limsup \frac{1}{N}\Sigma_{n=1}^{N} f(s^{sup}_n(s_0,\omega,\theta,\kappa)) \leq \max_{\mu^{sup}_{\theta,\kappa}\in\Delta^{sup}} \int f(s)\mu^{sup}_{\theta,\kappa}(ds) \tag{4.7}$$

$$\liminf \frac{1}{N}\Sigma_{n=1}^{N} f(s^{sup}_n(s_0,\omega,\theta,\kappa)) \geq \min_{\mu^{sup}_{\theta,\kappa}\in\Delta^{sup}} \int f(s)\mu^{sup}_{\theta,\kappa}(ds). \tag{4.8}$$

If there is a unique invariant distribution $\{\mu^{sup}_{\theta,\kappa}\}$, then Lemma 4.4 reduces to the law of large numbers by Breiman (1960). Hence, the lemma places upper and lower bounds for the average behavior of a typical orbit in the presence of multiple invariant distributions.

### *4.2 Constrained Estimation*

As discussed in Section 5 below, the monotonicity of $\xi(x,z,\varepsilon,\theta)$ in $x$ is usually derived from monotonicity and concavity assumptions on the utility and functions. The monotonicity of $\xi(x,z,\varepsilon,\theta)$ in $z$ is, however, a more delicate assumption, since after the occurrence of a good realization $z$ the associated income effects may reverse the order-preserving property of the dynamics. Hence, the following milder monotonicity condition should be useful in applications.

**(M′)** For each vector $(z, \varepsilon, \theta)$ the mapping $\xi(\cdot, z, \varepsilon, \theta) : X \to X$ is monotone increasing.

THEOREM 4.5: *Under (A.1)-(A.4) and* $(M')$*, for all* $s_0$ *and* $\lambda$*-almost all* $(\omega, \tilde{\mathbf{s}})$ *the SME* $\{\hat{\theta}_{1T}(s_0, \omega, \tilde{\mathbf{s}}, \theta_2^0)\}_{T\geq 1}$ *converges to* $\theta_1^0$*.*

COROLLARY 4.6: *Suppose that for almost all* $(z_0, \omega, \tilde{\mathbf{s}}_0)$ *the estimator* $\{\hat{\theta}_{2T}(z_0, \omega, \tilde{\mathbf{s}})\}_{T\geq 1}$ *converges to* $\theta_2^0$*. Then, under the conditions of Theorem 4.5, for all* $x_0$*, and almost all* $(z_0, \omega, \tilde{\mathbf{s}})$ *the SME* $\{\hat{\theta}_{1T}(x_0, z_0, \omega, \tilde{\mathbf{s}}, \hat{\theta}_{2T}(z_0, \omega, \tilde{\mathbf{s}}))\}_{T\geq 1}$ *converges to* $\theta_2^0$*.*

### *4.3 Estimation of Numerical Approximations*

As in the preceding section, let us now consider a sequence of approximate functions $\{\varphi^n\}$. As before, assume that for each approximate mapping $\varphi^n(\cdot, \cdot, \theta)$ there exists an invariant distribution $\mu_\theta^n$ for every $\theta$ in $\Theta$, and let $\theta^n$ solve optimization problem (3.6).

THEOREM 4.7: *Assume that the sequence of functions* $\{\varphi^n\}$ *converges to* $\varphi$*. Then, under (A.1)-(A.4) and* $(M)$ *every sequence of optimal solutions* $\theta^n$ *defined by (3.6) must converge to the original solution* $\theta^0$ *defined by (2.1).*

REMARK: Note that in this result every approximate function $\varphi^n$ is also required to satisfy Condition $(M)$.

## 5 THE ONE-SECTOR GROWTH MODEL

This section contains a discussion of the above assumptions in the context of the one-sector stochastic growth model with correlated shocks. In this version of the model Condition ($M'$) holds under regular standard assumptions of the utility and production functions, but Conditions ($M$) and ($C1$) $-$ ($C2$) require further specific restrictions. Formally, the model is summarized by the following dynamic optimization program:

$$W(x_0, z_0, \theta) = \max_{\{c_0, x_1\}} u(c_0) + \beta EW(x_1, z_1, \theta) \tag{5.1}$$

$$\text{s. t. } x_1 + c_0 = z_0 f(x_0, \alpha) + (1 - \delta)x_0$$

$$z_1 = \psi(z_0, \varepsilon_1, \rho)$$
$$x_0 \text{ and } z_0 \text{ given, } 0 < \beta < 1,\ 0 < \delta < 1,$$

where $E$ denotes the expectations operator. The vector of state variables $s_0 = (x_0, z_0)$ is known at time $t = 0$, and the realization of the exogenous stochastic perturbation $\varepsilon_1$ takes place next period. Total production of the aggregate good $y_0 = z_0 f(x_0, \alpha)$ depends on the exogenous level of productivity $z_0$ and the amount of initial capital $x_0$. Capital $x_0$ can also be consumed, and it is subject to a depreciation factor $\delta$. The optimization problem is to choose the amounts of consumption $c_0$ and capital for the next period $x_1$ so as to attain a maximum value for the discounted objective in (5.1). Parameters $\sigma$ and $\alpha$ characterize the utility function $u(\cdot, \sigma)$ and the production function $f(\cdot, \alpha)$ respectively. Standard regular conditions are that functions $u : R_+ \times R \rightarrow R$ and $f : R_+ \times R \rightarrow R$ are bounded and continuous, and $u(\cdot, \sigma)$ and $f(\cdot, \alpha)$ are monotone increasing and strictly concave. Also, it is typical to assume that function $\psi : R_+ \times R_+ \times R \rightarrow R_+$ is bounded and continuous. The shock $\varepsilon$ follows an *iid* process, and function $\psi(z, \varepsilon, \rho)$ is assumed to contain a unique invariant Markovian distribution. The parameter region $\Theta$ is conformed by vectors of the form $\theta = (\beta, \sigma, \alpha, \delta, \rho)$.

Equation (5.1) is Bellman's equation of dynamic programming, and the value function $W$ is the unique fixed-point solution of this functional equation. Function $W$ is bounded and continuous. Moreover, for each $(z_0, \theta)$ the mapping $W(\cdot, z_0, \theta)$ is monotone increasing and strictly concave. The optimal solution to (5.1) is attained at unique $x_1$ given by the policy function $x_1 = \xi(x_0, z_0, \theta)$. Function $\xi$ is jointly continuous in all arguments, and characterizes the dynamics of optimal paths.

Monotonicity properties of the policy function $\xi$ in $x$ and $z$ have been amply documented. For instance, Donaldson and Mehra (1983) illustrate that the strict concavity of functions $u(\cdot, \sigma)$ and $f(\cdot, \alpha)$ imply that for each given $z$ the mapping $\xi(\cdot, z, \theta)$ is monotone increasing. The monotonicity of $\xi$ jointly in $(x, z)$, however, requires some further limiting restrictions. The logic underlying these results is quite simple. After an increase in $x_0$ it becomes optimal to spread out the gain in consumption over time. Indeed, the concavity of functions $u(\cdot, \sigma)$ and $f(\cdot, \alpha)$ entails that the marginal utility of consumption and marginal productivity of capital are monotone decreasing. Hence, after an increase in $x_0$ both $c_0$ and $x_1$ should go up. But this argument does not extend to changes in $z_0$. Thus, if function $\psi(\cdot, \sigma)$ is monotone increasing, then a higher $z_0$ signals higher values for $z$ in the future. The expectations of future gains in $z$ may stimulate $c_0$ to a level such that $x_1$ may actually go down after the increase in $z_0$. Of course, if $z$ is modelled as an *iid* process [e.g., Brock and Mirman (1972)], then expectations about future income effects vanish, and so $\xi(\cdot,\cdot,\theta)$ must be jointly monotone. Indeed, if $z$ follows an *iid* process then the only state variable is $y = zf(x,\alpha)$, and increases in $x_0$ and $z_0$ must have the same qualitative effects.

Therefore, under standard regular assumptions, for correlated values of $z$ the mapping $\xi(\cdot,z,\theta)$ is monotone increasing. But the joint monotonicity of $\xi$ in $(x,z)$ is a much more restrictive condition, and requires some additional joint assumptions on the utility and production functions and on the evolution of the exogenous shock [cf., Donaldson and Mehra (1983)]. Condition $(M')$ is then easier to check and is much weaker than Condition $(M)$, and so Theorem 4.2 seems relevant for economic applications.

In multidimensional models, to preserve the above monotonicity properties the concavity of functions $u(\cdot,\sigma)$ and $f(\cdot,\alpha)$ needs to be strengthened. For the monotonicity of the mapping $\xi(\cdot,z,\theta)$ some key properties are that the objective must be supermodular and the feasible correspondence must be increasing in $x$. Supermodularity implies some form of complementarity among the various goods, or that the cross-partial derivatives must be non-negative. For recent developments in this area and further economic applications with monotone laws of motion, see Hopenhayn and Prescott (1994) and Mirman, Morand and Reffett (2003).

Monotonicity properties play a fundamental role in competitive-markets economies with distortions such as taxes, externalities, and money. In the presence of distortions, a Markov equilibrium may fail to exist. The monotonicity of $\xi(\cdot,z,\theta)$ has been the most effective tool to establish the existence of a Markov equilibrium for these economies [e.g., see Bizer and Judd (1989), Coleman (1991), Datta, Mirman and Reffett (2002), and Greenwood and Huffman (1995)]. Moreover, Santos (2002) provides some examples of non-existence of a Markov equilibrium in simple models with taxes and externalities in which this monotonicity property does not hold.

Conditions $(C.1)-(C.2)$ are much harder to verify in the above stochastic growth model. If the policy function is known and it is a differentiable function, an operational way to find if the system is a random contraction is to produce a large sample path to locate the ergodic set.[3] Then one can evaluate the derivatives of the policy function over the ergodic region. As a matter of fact, one could appeal directly to the multiplicative ergodic theorem [cf. Arnold (1998)] and get the expected value of the function $log(\|D_1\xi(x,z,\theta)\|)$, where $D_1\xi(x,z,\theta)$ denotes the derivative of $\xi$ with respect to the the first component variable $x$. This procedure yields the value of the maximum characteristic exponent of the dynamical system, and this exponent must be less than zero for Condition $(C.1')$ to be satisfied.

Also, the policy function $\xi(x,z,\theta)$ is contractive in $x$ if the domain of variation of the exogenous variable $z$ is small enough. Broadly speaking the argument goes as follows. If the utility and production functions are strongly concave and continuously differentiable, then under mild regularity conditions the deterministic version of the above growth model has a unique interior steady state in which the derivative of the policy function is less than one. Hence, by a continuity argument for a small stochastic differentiable perturbation of

[3]Under certain regularity conditions, by the law of large numbers the empirical measure generated by a typical sample path must converge weakly to the model's invariant distribution, assuming that such distribution is unique. Hence, the closure of a typical sample path must contain the ergodic set.

the model such derivative will be less than one over the corresponding ergodic set.

Random contractive systems are familiar from the literature on Markov chains [e.g., see Norman (1972) for an early analysis and applications, and Stenflo (2001) for a recent update of the literature]. Related contractivity conditions are studied in Dubins and Freedman (1966), Schmalfuss (1996) and Bhattacharya and Majumdar (2003). In the macroeconomics literature, Conditions $(C1)-(C2)$ arise naturally in the one-sector Solow model [e.g., Schenk-Hoppé and Schmalfuss (2001)] and in some concave dynamic programs [e.g., see Foley and Hellwig (1975) and Examples 4.2-4.3 in Santos and Peralta-Alva (2003)]. Stochastic contractivity properties are also found in learning models [e.g., Schmalensee (1975), and Ellison and Fudenberg (1993)] and in certain types of stochastic games [e.g., Sanghvi and Sobel (1976)].

## References

Ait-Sahalia, Y., 1996, Nonparametric pricing of interest rate derivative securities, Econometrica 64, 527-560.

Araujo, A., Scheinkman, J. A., 1977, Smoothness, comparative dynamics and the turnpike property, Econometrica 45, 601-620.

Arnold, L., 1998, Random dynamical systems (Springer-Verlag, Berlin).

Bhattacharya, R., Lee, O., 1988, Asymptotics of a class of Markov processes which are not in general irreducible, Annals of Probability 16, 1333-1347.

Billingsley, P., Topsoe, F., 1967, Uniformity in weak Convergence, Zeitschrift Fur Wahrscheinlichkeitstheorie Und Verwandte Gebiete 7, 1-16.

Bizer, D., Judd, K. L., 1989, Taxation under uncertainty, American Economic Review 79, 331-336.

Brock, W. A. , Mirman, L. J., 1972, Optimal economic growth under uncertainty: The discounted case, Journal of Economic Theory 4, 497-513.

Burke, J. L., 1990, A benchmark for comparative dynamics and determinacy in overlapping generations economies, Journal of Economic Theory 52, 268-303.

Coleman, W. J., 1991, Equilibrium in a production economy with an income tax, Econometrica 59, 1091-1104.

Datta, M., Mirman, L. J., Reffett, K. L., 2002, Existence and uniqueness of equilibrium in distorted dynamic economies with capital and labor, Journal of Economic Theory 103, 377-410.

Dehardt, J., 1971, Generalizations of the Glivenko-Cantelli theorem, The Annals of Mathematical Statistics 42, 2050-2055.

Donaldson, J. B., Mehra, R., 1983, Stochastic growth with correlated production shocks, Journal of Economic Theory 29, 282-312.

Duffie, D., Singleton, K. J., 1993 , Simulated moments estimation of Markov models of asset prices, Econometrica 61, 929-952.

Ellison, G., Fudenberg, D., 1993, Rules of thumb for social learning, Journal of Political Economy 101, 612-643.

Feinberg, S., Keane, M., 2002, A Structual model of the production location decisions of MNCs, manuscript, Yale University.

Foley, D., Hellwig, M., 1975, Asset management with trading uncertainty, Review of Economic Studies 42, 327-346.

Futia, C., 1982, Invariant distributions and the limiting behavior of Markovian economic models, Econometrica 50, 377-408.

Gallant, A. R., Tauchen, G., 2002, Which moments to match? Econometric Theory 12, 657-681.

Gaspar, J., Judd, K. L., 1997, Solving large-scale rational expectations models, Macroeconomic Dynamics 1, 45-75.

Gourieroux, C., Monfort, A., Renault, E., 1993, Indirect inference, Journal of Applied Econometrics 8, S85-S118.

Gourinchas, P.-O., Parker, J. A., 2002, Consumption over the Life Cycle, Econometrica 70, 47-89.

Greenwood, J., Huffman, G., 1995, On the existence of non-optimal equilibria in dynamic stochastic economies, Journal of Economic Theory 65, 611-623.

Hall, G., Rust, J., 2002, Econometric methods for endogenously sampled time series: The case of commodity speculation in the steel market, manuscript, Yale University.

Hansen L. P., Scheinkman, J. A., 1995, Back to the future: generating moments implications for continuous time Markov processes, Econometrica 63, 767-804.

Hansen, L. P., Singleton, K. J., 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, Econometrica 50, 1269-1286.

Hopenhayn H. A., Prescott, E. C., 1992, Stochastic monotonicity and stationary distributions for dynamic economies, Econometrica 60, 1387-1406.

Jennrich, R. I., 1969, Asymptotic properties of non-linear least squares estimators, Annals of Mathematical Statistics 40, 633-643.

Kifer, Y., 1986, Ergodic theory of random transformations (Birkhauser, Boston).

Krengel, U., 1985, Ergodic theorems (Walter de Gruyter, Berlin).

Krusell, P., Smith, T., 1998, Income and wealth heterogeneity in the macroeconomy, Journal of Political Economy 106, 867-896.

Lee, B., Ingram, B., 1991, Simulation estimation of time series models, Journal of Econometrics 47, 197-205.

Mariano, R. S., Schuermann, T., Weeks, M., 1999, Simulation-based inference in econometrics: Theory and applications, (Cambridge University Press, Cambridge).

McFadden, D. L., 1989, A method of simulated moments for estimation of discrete response models without numerical integration, Econometrica 57, 995-1026.

Mirman, L. J., Morand, O. F., Reffett, K. L. 2002, A quantitative approach to Markovian equilibrium in infinite horizon economies with capital, manuscript, Arizona State University.

Newey, W. K., McFadden, D., 1994, Large sample estimation and hypothesis testing, in: R. F. Engle and D. L. McFadden, eds., Handbook of Econometrics, Vol. 4 (Elsevier, Amsterdam) pp. 2113-2245.

Norman, F. M., 1972, Markov processes and learning models (Academic Press, New York).

Ortigueira, S., Santos, M. S., equilibrium dynamics in a two-sector model with taxes, Journal of Economic Theory 105, 99-119.

Pakes, A., Pollard, D., 1989, Simulation and the asymptotics of optimization estimators, Econometrica 57, 1027-105.

Pollard, D., 1984, Convergence of stochastic processes (Springer-Verlag, Berlin).

Sanghvi, A. P., Sobel, M. J., 1976, Bayesian games as stochastic processes, International Journal of Game Theory 5, 1-22.

Santos, M. S., 1999, Numerical solution of dynamic economic models, in: J. B. Taylor and M. Woodford, eds., Handbook of Macroeconomics, Vol. 1A (Elsevier, Amsterdam) 311-386.

Santos, M. S., 2002, On non-existence of Markov equilibria in competitive-market economies, Journal of Economic Theory 105 , 73-98.

Santos, M. S., 2003, Estimation by simulation of monotone dynamical systems, Journal of Computational and Applied Mathematics 158, 145-156.

Santos, M. S., Bona, J. L., 1989, On the structure of the equilibrium price set of overlapping-generations economies, Journal of Mathematical Economics 18, 209-230.

Santos, M. S., Peralta-Alva, A., 2003, Accuracy of the statistics from numerical simulations, manuscript, Arizona State University.

Schenk-Hoppé, K.R., Schmalfuss, B., 2001, Random fixed points in a stochastic Solow growth model, Journal of Mathematical Economics 36, 19-30.

Schmalensee, R., 1975, Alternative models of bandit selection, Journal of Economic Theory 10, 333-342

Schmalfuss, B., 1996, A random fixed-point theorem based upon Lyapunov exponents, Random and Computational Dynamics 4, 257-268.

Stenflo, O., 2001, Ergodic theorems for Markov chains represented by iterated function systems, Bulletin of the Polish Academy of Sciences: Mathematics 49, 27-43.

Stokey, N. L., Lucas, R. E., Prescott, E. C., 1989, Recursive methods in economic dynamics (Harvard University Press, Cambridge).

van der Vaart, A. W., Wellner, J., 2000, Weak convergence and empirical processes (Springer-Verlag, New York).

Williams, N., 2002, Small noise asymptotics for stochastic growth models, manuscript, Princeton University.