



Working Paper 03-47
Economics Series 16
October 2003

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624 98 75

SIMULATION-BASED ESTIMATION OF DYNAMIC MODELS WITH CONTINUOUS EQUILIBRIUM SOLUTIONS *

Manuel S. Santos¹

Abstract

This paper presents several results on consistency properties of simulation-based estimators for a general class of dynamic models with continuous laws of motion. The consistency of these estimators follows from a uniform convergence property of the model's sample paths over the vector of parameters. This convergence property is established under either contractivity or 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 M. Santos, Department of Economics, Arizona State University; E.mail: msantos@asu.edu

* I am grateful to Jorge Aseff, Seongman Moon and Adrian Peralta-Alva for some helpful remarks. I have also benefitted from several discussions with participants at the macro workshop of Universidad Carlos III de Madrid. This research was partially supported by the Ministerio de Ciencia y Tecnología, project no. SEC 2002-004318.

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 mechanisms of allocation of commodities in these economies and in the decision rules of economic agents in environments which may comprise time and uncertainty. 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 probability measures 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 further statistics of these distributions. Therefore, the process of estimation may entail the simulation of a parameterized family of models. Moreover, properties of these estimators such as consistency and asymptotic normality will 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, for a Markov process that is continuous in the state variables defined over a com-

compact space there always exists 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). As explained in detail below, I develop here some analytical methods that improve substantially the consistency results of these two papers. These results are proved under weaker assumptions, and are extended to constrained estimation (i.e., to cases in which some parameter values are known) and to the convergence of estimates from numerical approximations. A key step in the proof of consistency of these estimators is to establish the uniform convergence of the simulated moments –defined over the model’s sample paths– 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

¹For present purposes, a mapping $h : X \rightarrow X$ is contractive if $\|h(x) - h(x')\| < \|x - x'\|$, where $\|x\|$ is the max norm on X . A mapping $h : X \rightarrow X$ 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 $x' \geq x$.

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.

A broad conclusion of the present study is then that the two families of Markov processes investigated by 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), Bhattacharya and Majumdar (2003), Hopenhayn and Prescott (1992), and Stenflo (2001)] 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 aforementioned continuity assumption of the stochastic process in the state variables. Hence, an open issue is whether some families of models with non-continuous equilibrium solutions 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

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

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

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 ε_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 θ_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 θ_1 . Functions ξ and ψ 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 ξ governs the law of motion of the vector of endogenous variables x and ψ 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 θ^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 θ^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 errors vanish the sequence of estimated values $\{\hat{\theta}_T\}$ should converge to the optimal solution θ^0 .

Since a change in θ 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 full-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 θ . Likewise, standard non-linear least squares [e.g., Jennrich (1969)] and other generalized estimators [cf., Newey and McFadden (1994)] presuppose that functions ξ and ψ 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

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);

²Data-based estimation 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 subsequently these estimation methods are only valid under certain functional restrictions. The analysis may break down in the presence of *latent* variables or some private information not available to the econometrician. 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.

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 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 approximate 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 flexibility 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 con-

sistency for simulation-based estimators in which the uniform convergence of the simulated statistics must hold over a continuum of invariant distributions.

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 in which state variables are correlated over time.

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) and Santos (2003). It should be stressed that the methods of analysis developed below 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 θ_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 results in 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 monotone 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 each local majorizing and minorizing function in the presence of 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 convergence properties for standard numerical methods are only obtained under certain mathematical conditions. Consequently, one should expect that the conditions 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 put forward by Lee and Ingram (1991). 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 ε_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 Θ 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 $\varphi : S \times \mathcal{E} \times \Theta \rightarrow S$ is bounded.

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

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

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

Observe that (A.1)–(A.4) will all be satisfied if φ 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 θ there exists an invariant distribution μ_θ on S under mapping $\varphi(\cdot, \cdot, \theta)$. For a random contraction this invariant distribution μ_θ is unique [e.g., see Stenflo (2001)]. Also, some simple conditions guarantee the existence of a unique invariant distribution μ_θ 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 μ_θ corresponding to each parameter θ . 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

method 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 θ^0 .

(1) The *target function* $f : S \rightarrow R^p$ is assumed to be continuous. This function may represent p moments of an invariant distribution μ_θ 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} .

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

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

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 θ^0 . Moreover, for the analysis below there is no restriction of generality to consider that θ^0 is unique.

(3) The *estimation method* yields a sequence of estimated values $\{\hat{\theta}_T\}$ so as to approximate the solution θ^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 θ and initial condition $s_0 = (x_0, z_0)$ let $\{s_t(s_0, \omega, \theta)\}$ be the sequence generated by 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

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

The distance function G_T may depend on information available up to time T [e.g., see Duffie and Singleton (1993)]. The sequence of functions $\{G_T\}$ is assumed to converge uniformly to function G as $T \rightarrow \infty$. The rule $\tau(T)$ reflects that model's simulations may be of a different length than data samples, but it is required that $\tau(T) \rightarrow \infty$ as $T \rightarrow \infty$.

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 φ at any given point (s, ε, θ) ; later, the analysis will consider the more typical situation in which the researcher can only perform evaluations of a numerical approximation φ^n . The process $\{\tilde{s}_t\}$ is assumed to be stationary and ergodic. Also, as is typical in numerical simulation the postulated distribution of $\{\varepsilon_t\}$ is known, but no knowledge of the actual realization of the shock process $\{\varepsilon_t\}$ is required. This latter assumption is too strong since the sequence of shocks $\{\varepsilon_t\}$ may be unobservable, but this assumption 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 $\{\hat{\varepsilon}_t\}$ 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 γ is defined over the space of random shocks $\omega = \{\varepsilon_t\}$. (The construction of measure γ follows from standard arguments [e.g., see Stokey, Lucas and Prescott (1989, Ch. 8)].) Let $\lambda = \gamma \times \tilde{\gamma}$ represent the product measure.

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 λ -almost all $(\omega, \tilde{\mathbf{s}})$.*

REMARK: Because of the recursive structure embedded in the parameter space Θ , sometimes the value θ_2^0 may be known or may be estimated independently by a more efficient method. In those situations, for a fixed θ_2^0 one may consider a constrained version of optimization problem (2.2) over Θ_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). Several contractivity properties can be found in the literature on random dynamical systems [e.g., see Stenflo (2001)]. Our analysis will focus on two main alternative contractivity conditions. The consistency of the SME is first established for the whole vector of parameters θ . Then, these contractivity conditions will be relaxed for the consistency of the estimator in the first component vector θ_1 when the true value θ_2^0 is known. Further convergence results are likewise derived for estimates obtained from numerical approximations of function φ .

3.1 Consistency of the SME

The consistency of the SME rests on the dampening behavior of the dynamics imposed by each of the contractivity conditions. The first condition draws on some methods developed by Kifer (1986, Ch. 1) who proposed a notion of characteristic exponent in metric spaces. This notion seems appropriate for non-smooth functions. Let

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

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

(C.1) For every θ 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 φ with respect to s is on average a negative number. For fixed θ , 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 γ -almost all ω , for every vector θ and initial condition s_0 ,

- (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

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

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 $\delta(s_0, \omega^{-1}, \theta) \leq \delta(s_1, \omega^{-1}, \theta)$.

REMARK: For simplicity, this contractivity condition holds for a set of unit probability that it is common to all θ , but this set may be allowed to depend on θ along the lines of (C.1) above. 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

$$\|s_t(s, \omega, \theta) - s_t(s_0, \omega, \theta)\| \leq N \alpha^t \|s - s_0\| \text{ for some constants } N > 0 \text{ and } 0 < \alpha < 1. \text{ Note}$$

that these constants are allowed to depend on (s_0, ω, θ) . 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 and these bounds are locally

uniform in θ . Of course, if we neglect transitional dynamic behavior and the local uniformity property 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 λ -almost all (ω, \tilde{s}) the SME $\{\hat{\theta}_T(s_0, \omega, \tilde{s})\}_{T \geq 1}$ converges to θ^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 study is to dispense with some rather technical conditions invoked by these authors (cf., *op.cit.* Th. 2). The method of proof is based on some auxiliary results of independent interest, which we now pass to discuss.

The first lemma requires a canonical extension of the space of shocks [cf., Krengel (1985, Ch. 1)] in which t ranges from $-\infty$ to ∞ . Hence, for this result every sequence of shocks $\tilde{\omega}$ is of the form $\tilde{\omega} = (\cdots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \cdots)$. Let ϑ_t denote the shift operator defined as $\vartheta_t(\cdots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \cdots) = (\cdots, \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \cdots)$.

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 θ ,*

$$s_{t+1}^*(\tilde{\omega}, \theta) = \varphi(s_t^*(\tilde{\omega}, \theta), \varepsilon_{t+1}, \theta) \text{ and } s_t^*(\vartheta_t(\tilde{\omega}), \theta) = s_0^*(\tilde{\omega}, \theta) \text{ for all } t. \quad (3.3)$$

Function $s_t^(\cdot, \theta)$ is measurable for every t . Moreover, for each initial condition s_0 every sample path $s_t(s_0, \tilde{\omega}, \theta)$ converges uniformly to $s_t^*(\tilde{\omega}, \theta)$ in θ as t goes to ∞ for almost all $\tilde{\omega}$.*

This lemma is an 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 under (C.1) it suffices to analyze the convergence properties of the sequences $\frac{1}{\tau(T)} \sum_{t=1}^{\tau(T)} f(s_t^*(\tilde{\omega}, \theta))$ for all θ in Θ . 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 θ . This continuity property is established in the following result.

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

3.2 Constrained Estimation

In some applications it may be possible to get independent estimates of the true value θ_2^0 by more practical estimation methods. In those situations simulation-based estimation can be restricted to the first component vector θ_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

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

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

(C.1') For every θ_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) $E(c(z, \varepsilon)) < 0$

REMARK: The expectation $E(c(z, \varepsilon))$ in (ii) is taken with respect to the invariant distribution of vector (z, ε) . Since ε is an *iid* process, this invariant distribution is a product measure conformed by the invariant distributions of random vectors z and ε . Note that the invariant distribution of z is determined by θ_2^0 .

Regarding Condition (C.2) the following weakened version applies to the dynamics of the vector of endogenous state variables, x .

(C.2') For γ -almost all ω , for every vector θ and initial condition $s_0 = (x_0, z_0)$,

(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

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

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 $\delta(s_0, \omega^{-1}, \theta) \leq \delta(s_1, \omega, \theta)$.

REMARK: $x_t(s_0, \omega, \theta)$ in (C.2') refers to the first component vector of $s_t(s_0, \omega, \theta)$ as asserted 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 , almost all z_0 , and λ -almost all $(\omega, \tilde{\mathbf{s}})$ the constrained SME $\{\hat{\theta}_{1T}(x_0, z_0, \omega, \tilde{\mathbf{s}}, \theta_2^0)\}_{T \geq 1}$ converges to θ_1^0 .*

COROLLARY 3.5: *Suppose that for almost all z_0 and λ -almost all $(\omega, \tilde{\mathbf{s}}_0)$ the estimator $\{\hat{\theta}_{2T}(z_0, \omega, \tilde{\mathbf{s}})\}_{T \geq 1}$ converges to θ_2^0 . Then, under the conditions of Theorem 3.4, for all x_0 , almost all z_0 and λ -almost all $(\omega, \tilde{\mathbf{s}})$ the constrained 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 θ_2^0 .*

3.3 Estimation of Numerical Approximations

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

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

where μ_θ^n is an invariant distribution for the mapping $\varphi^n(\cdot, \cdot, \theta)$ for each θ in Θ . The invariant distribution μ_θ^n may not be unique, even though for each θ the original mapping $\varphi(\cdot, \cdot, \theta)$ is assumed to have a unique invariant distribution μ_θ . Also, the solution θ^n may not be unique but it is postulated that one such solution exists. The idea is that certain economic assumptions may guarantee the existence of an invariant distribution μ_θ for $\varphi(\cdot, \cdot, \theta)$. Uniqueness of the invariant distribution, however, 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 μ_θ^n . Then, it is of interest to know whether the set of solutions $\{\theta^n\}$ defined by (3.6) converge to the original solution θ^0 defined by (2.1) as φ^n approaches φ .

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 θ 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 φ . Let φ^n satisfy (A.2)-(A.3) for each n . Let φ 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 θ^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. Such convergence property is attained for some numerical approximations [e.g, see Santos (1999)]. 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 μ_θ in θ as n goes to ∞ .

(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 $\{s_t^n(s_0, \omega, \theta)\}$ generated by the approximate function φ^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 Reffett (2003)].

The analysis draws on an earlier paper [Santos (2003)]. These earlier results will be extended using the following weaker assumptions: (i) (A.1) – (A.4) replace a stronger continuity assumption on function φ , (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 λ -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 $h : S \rightarrow S$ is called *order preserving* or *monotone increasing* if $h(s) \geq h(s')$ for all $s \geq s'$. Conversely, a function $h : S \rightarrow S$ is called *order reversing* or *monotone decreasing* if $h(s) \leq h(s')$ for all $s' \geq s$. The analysis will focuss on monotone increasing dynamics.

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

Note that no order preserving assumptions are made over the space of shocks \mathcal{E} and over the parameter space Θ .

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

As in the preceding section a key step in the method of proof of this consistency result is to establish the uniform convergence of the series $\{\frac{1}{T} \sum_{t=1}^T f(s_t(s_0, \omega, \theta))\}$ to $E_\theta(f)$ in θ , for each fixed s_0 and for almost all ω . A law of large numbers for continuous Markov processes [cf. Breiman (1960)] implies that for each θ the sequence $\{\frac{1}{T} \sum_{t=1}^T f(s_t(s_0, \omega, \theta))\}$ converges almost surely to $E_\theta(f)$ for all s_0 . This pointwise property over the parameter space Θ is, however, too weak for present purposes. Then, the proof of Theorem 4.1 relies on a judicious construction of 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 Θ . The orbits generated by these local bounding functions place upper and 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 θ 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 builds upon Condition (M), a continuity property on the set of invariant distributions and a generalized law of large numbers for Markov processes. These last two results are taken from Santos and Peralta-Alva (2003).

Here is a detailed description of the main elements conforming the proof of Theorem 4.1. For each given (ε, θ) and constant $\kappa > 0$ define the *majorizing* function

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

and the *minorizing* function,

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

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

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

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

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

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

(iv) For each θ the mapping $\varphi^{sup}(\cdot, \cdot, \theta, \kappa)$ converges uniformly to $\varphi(\cdot, \cdot, \theta)$ as κ goes to 0.

This lemma is a straightforward consequence of definition (4.1) and Assumptions (A.1) – (A.4). Observe that

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

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

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

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

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

Therefore, proceeding by induction

$$s_t^{sup} \geq s_t \geq s_t^{inf} \quad (4.6)$$

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 θ 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 (θ, κ) . To carry out this argument the following auxiliary results are invoked: (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.

Let $\mu_{\theta, \kappa}^{sup}$ be an invariant distribution under function $\varphi^{sup}(s, \varepsilon, \theta, \kappa)$ and μ_θ be the unique invariant distribution for function $\varphi(s, \varepsilon, \theta)$. We do rule out the possibility that function $\varphi^{sup}(s, \varepsilon, \theta, \kappa)$ may contain multiple invariant distributions. Let $\Delta_{\theta, \kappa}^{sup}$ be the set of all the invariant distributions $\mu_{\theta, \kappa}^{sup}$ under $\varphi^{sup}(s, \varepsilon, \theta, \kappa)$. Note that $\Delta_{\theta, \kappa}^{sup}$ is a compact convex set in the weak topology of measures. Then, the linear functional $\mu_{\theta, \kappa}^{sup} \rightarrow \int f(s) \mu_{\theta, \kappa}^{sup}(ds)$ attains a maximum and a minimum over all $\mu_{\theta, \kappa}^{sup}$ in $\Delta_{\theta, \kappa}^{sup}$. Also, for every sequence of shocks $\omega = \{\varepsilon_n\}$ and initial condition s_0 , let $s_{t+1}^{sup}(s_0, \omega, \theta, \kappa) = \varphi^{sup}(s_t^{sup}(s_0, \omega, \theta, \kappa), \varepsilon_{t+1}, \theta, \kappa)$ for all $t \geq 1$.

LEMMA 4.3 [SANTOS AND PERALTA-ALVA (2003, TH. 3.2)]: *Let (A.1) – (A.4) be satisfied. Then, every sequence of probability measures $\{\mu_{\theta, \kappa}^{sup}\}$ converges to $\{\mu_\theta\}$ as κ goes to 0.*

This result follows from Lemma 4.2 and 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 [SANTOS AND PERALTA-ALVA (2003, TH. 3.8)]: *Let (A.1) – (A.4) be satisfied. Then, for all s_0 and almost all ω ,*

$$\limsup \frac{1}{T} \sum_{t=1}^T f(s_t^{sup}(s_0, \omega, \theta, \kappa)) \leq \max_{\mu_{\theta, \kappa}^{sup} \in \Delta_{\theta, \kappa}^{sup}} \int f(s) \mu_{\theta, \kappa}^{sup}(ds) \quad (4.7)$$

$$\liminf \frac{1}{T} \sum_{t=1}^T f(s_t^{sup}(s_0, \omega, \theta, \kappa)) \geq \min_{\mu_{\theta, \kappa}^{sup} \in \Delta_{\theta, \kappa}^{sup}} \int f(s) \mu_{\theta, \kappa}^{sup}(ds). \quad (4.8)$$

This result places upper and lower bounds for the average behavior of a typical orbit in the presence of multiple invariant distributions. If there is a unique invariant distribution $\mu_{\theta, \kappa}^{sup}$,

then Lemma 4.4 reduces to the law of large numbers of Breiman (1960).

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 production 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, ε, θ) the mapping $\xi(\cdot, z, \varepsilon, \theta) : X \rightarrow X$ is monotone increasing.

THEOREM 4.5: Under (A.1)-(A.4) and (M'), for all s_0 and λ -almost all $(\omega, \tilde{\mathbf{s}})$ the constrained SME $\{\hat{\theta}_{1T}(s_0, \omega, \tilde{\mathbf{s}}, \theta_2^0)\}_{T \geq 1}$ converges to θ_1^0 .

COROLLARY 4.6: Suppose that for almost all z_0 and λ -almost all $(\omega, \tilde{\mathbf{s}})$ the estimator $\{\hat{\theta}_{2T}(z_0, \omega, \tilde{\mathbf{s}})\}_{T \geq 1}$ converges to θ_2^0 . Then, under the conditions of Theorem 4.5, for all x_0 , almost all z_0 and λ -almost all $(\omega, \tilde{\mathbf{s}})$ the constrained 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 θ_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 μ_θ^n for every θ in Θ . Let θ^n solve optimization problem (3.6).

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

REMARK: Note that in this result every approximate function φ^n is also required to satisfy Condition (M). This is a main restriction as compared to Theorem 3.6 in which every function φ^n is not assumed to satisfy any of the contractivity conditions (C.2) and (C.3).

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:

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

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

$$z_1 = \psi(z_0, \varepsilon_1, \rho)$$

$$x_0 \text{ and } z_0 \text{ given, } 0 < \beta < 1, 0 < \pi < 1,$$

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 ε_1 takes place next period. Total production of the aggregate good $y_0 = z_0 f(x_0, \phi)$ 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 π . 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 σ and ϕ characterize the utility function $u(\cdot, \sigma)$ and the production function $f(\cdot, \phi)$ 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 $f(\cdot, \phi)$ 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 ε follows an *iid* process, and for each ρ function $\psi(z, \varepsilon, \rho)$ is assumed to contain a unique invariant Markovian distribution. The parameter region Θ is conformed by vectors of the form $\theta = (\beta, \sigma, \phi, \pi, \rho)$.

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, θ) 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 ξ is jointly continuous in all arguments, and characterizes the dynamics of optimal paths.

Monotonicity properties of the policy function ξ 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 $f(\cdot, \phi)$ imply that for each given z the mapping $\xi(\cdot, z, \theta)$ is monotone increasing. The monotonicity of ξ 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 $f(\cdot, \phi)$ 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 $\psi(\cdot, \rho)$ 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 $y = zf(x, \phi)$, 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 ξ 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 4.5 on constrained estimation seems quite pertinent for economic applications.

In multidimensional models, to preserve the above monotonicity properties the concavity of functions $u(\cdot, \sigma)$ and $f(\cdot, \phi)$ 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 ξ with respect to the first component variable x .

This procedure yields the value of the maximum characteristic exponent of the dynamical

³Under 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.

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 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.
- Bhattacharya, R., Majumdar, M., 2003, *Random dynamical systems: A review*, forthcoming in *Economic Theory*.
- 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, *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.
- Dubins, L. E., Freedman, D., 1966, Invariant probabilities for certain Markov processes, *Annals of Mathematical Statistics* 37, 837-858.
- 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 Structural 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., 1996, 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.-S., 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., 2003, 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).
- Rudin, W., 1974, Real and complex analysis (McGraw-Hill, New York).
- Rust, J., 1994, Structural estimation of Markov decision processes, in R. F. Engle and D. L. McFadden, eds., *Handbook of Econometrics*, vol. 4, (Elsevier, Amsterdam) pp. 3081-3143.
- 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) pp. 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 simulations for stochastic dynamic models, 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.

Appendix

PROOF OF THEOREM 3.1: (a) Assume that condition (C.1) is satisfied. The proof of this part requires to establish Lemmas 3.2 and 3.3.

PROOF OF LEMMA 3.2 (SKETCH): The existence of a unique fixed-point solution $s_t^*(s_0, \tilde{\omega})$, for $-\infty < t < \infty$, can be proved along the lines of Schmalfuss (1996, Th. 2.2). The uniform convergence of the sample paths $s_t(s_0, \tilde{\omega}, \theta)$ to $s_t^*(\tilde{\omega}, \theta)$ in θ as t goes to ∞ (for all $\tilde{\omega}$ in a set of full measure) can be proved along the lines of Duffie and Singleton (1993), since the space S is compact and the bounds asserted in Condition (C1) are locally uniform.

PROOF OF LEMMA 3.3 (SKETCH): This proof is a simple consequence of the following two facts which are derived from our basic assumptions: (i) The set of fixed point solutions is an upper semicontinuous correspondence over Θ , and (ii) by Lemma 3.2 this correspondence is univalued, and hence the mapping $s_t^*(\tilde{\omega}, \theta)$ must be a continuous function of θ . More formally, consider a sequence $\{\theta_i\}$ converging to some vector θ . For each t , let $\bar{s}_t^*(\tilde{\omega}, \theta)$ be the limsup of the sequence of functions $\{s_t^*(\tilde{\omega}, \theta_i)\}$, and let $\underline{s}_t^*(\tilde{\omega}, \theta)$ be the liminf of $\{s_t^*(\tilde{\omega}, \theta_i)\}$. Both mappings $\bar{s}_t^*(\tilde{\omega}, \theta)$ and $\underline{s}_t^*(\tilde{\omega}, \theta)$ are measurable function of $\tilde{\omega}$ [e.g., see Rudin (1974, p. 15)]. Moreover, by (A.2) and (A.4) both $\bar{s}_t^*(\tilde{\omega}, \theta)$ and $\underline{s}_t^*(\tilde{\omega}, \theta)$ must satisfy (3.3) for all $-\infty < t < \infty$. But the uniqueness of the fixed-point solution (Lemma 3.2) entails that $\bar{s}_t^*(\tilde{\omega}, \theta) = \underline{s}_t^*(\tilde{\omega}, \theta)$ for all t . Therefore, $s_t^*(\tilde{\omega}, \theta)$ is a continuous function of θ for almost all $\tilde{\omega}$.

By virtue of Lemma 3.2, for the purposes of this part of the proof of Theorem 3.1 it

suffices to show the uniform convergence of the sequence $\frac{1}{\tau(T)} \sum_{t=1}^{\tau(T)} f(s_t^*(\tilde{\omega}, \theta))$ to $E_\theta(f) = \int f(s) \mu_\theta(ds)$ in the vector of parameters θ , as t goes to ∞ . This is a standard result in econometrics [e.g., Jennrich (1969, Th. 2)], which follows from the continuity of $s_t^*(\tilde{\omega}, \theta)$ in θ and the measurability of $s_t^*(\tilde{\omega}, \theta)$ in $\tilde{\omega}$ for each t , and the fact that $\tilde{\omega}$ is generated by an *iid* process.

Finally, the convergence of the SME $\{\hat{\theta}_T(s_0, \omega, \tilde{s})\}$ to θ^0 for all s_0 and λ -almost all (ω, \tilde{s}) is a simple consequence of the following results: (a) The assumed convergence of the sequence $\frac{1}{T} \sum_{t=0}^T f(\tilde{s}_t)$ to \bar{f} as T goes to ∞ for $\tilde{\gamma}$ -almost all \tilde{s} , (b) the uniform convergence proved above of the sequence $\frac{1}{\tau(T)} \sum_{t=0}^T f(s_t(s_0, \tilde{\omega}, \theta))$ to $E_\theta(f)$ in θ , as T goes to ∞ , for all s_0 and γ -almost all ω , (c) the uniform convergence of $\{G_T\}$ to G , (d) the continuity of function G and the continuity of $E_\theta(f)$ in θ , and (e) the compactness of the set Θ , and the uniqueness of the maximizer $\theta^0 = (\theta_1^0, \theta_2^0)$ in (2.1).

(b) Assume that Condition (C.2) is satisfied. The proof of this part is omitted as it builds along similar arguments to those of Theorem 3.6 below. The following basic result will be used in that proof.

LEMMA 6.1: Let $h : S \rightarrow S$ and $g : S \rightarrow S$. For fixed s_0 , let $s_t^h(s_0) = h(s_{t-1}^h(s_0))$ and $s_t^g(s_0) = g(s_{t-1}^g(s_0))$ for all $t \geq 1$. Suppose that $\|h - g\| < \delta$ for some $\delta > 0$. Assume that there exists a constant $0 < \alpha < 1$ such that $\|h(s) - h(s')\| < \alpha \|s - s'\|$ for all s, s' in S . Then,

$$\|s_t^h(s_0) - s_t^g(s_0)\| \leq \frac{\delta}{1 - \alpha} \text{ for all } t \geq 1. \quad (6.1)$$

PROOF: The proof is by an inductive argument. Suppose that $\|s_{t-1}^h(s_0) - s_{t-1}^g(s_0)\| \leq \frac{\delta}{1-\alpha}$.

Then,

$$\begin{aligned} \|s_t^h(s_0) - s_t^g(s_0)\| &\leq \|h(s_{t-1}^h(s_0)) - g(s_{t-1}^g(s_0))\| \leq \\ &\|h(s_{t-1}^h(s_0)) - h(s_{t-1}^g(s_0))\| + \|h(s_{t-1}^g(s_0)) - g(s_{t-1}^g(s_0))\| \leq \\ &\alpha\|s_{t-1}^h(s_0) - s_{t-1}^g(s_0)\| + \delta \leq \alpha\frac{\delta}{1-\alpha} + \delta = \frac{\delta}{1-\alpha}. \end{aligned} \quad (6.2)$$

The first inequality comes from the triangle inequality, and the remaining inequalities follow from the postulated conditions. The result is thus established.

PROOF OF THEOREM 3.4: This proof is omitted, as it follows from the same arguments as in Theorem 3.1.

PROOF OF THEOREM 3.6: (a) Assume that Condition (C.2) is satisfied. By (C.2) for every θ and s_0 there are constants $N(s_0, \omega, \theta)$, $\alpha(s_0, \omega, \theta)$ and $\delta(s_0, \omega, \theta)$ such that (3.2) holds for all s in $B_{(s_0, \omega, \theta)}(s_0)$ and all $t \geq 1$, for almost all ω . Note that by a basic result in measure theory, if $\{A_n\}$ is a sequence of sets such that $A_n \subset A_{n+1}$ for every n and $\bigcup_{n=1}^{\infty} A_n = A$ then $\lim_n \gamma(A_n) = \gamma(A)$. Using this basic result it follows that for large enough constants $\tilde{N}(s_0, \theta) > 0$ and $0 < \tilde{\alpha}(s_0, \theta) < 1$ and a small enough constant $\tilde{\delta}(s_0, \theta)$ there is a set $\Omega_{(s_0, \theta)}$ with $\gamma(\Omega_{(s_0, \theta)}) > 0$ such that the relation in (3.2) holds under constants $\tilde{N}(s_0, \theta) > 0$, $0 < \tilde{\alpha}(s_0, \theta) < 1$ and $\tilde{\delta}(s_0, \theta) > 0$ for all ω in $\Omega_{(s_0, \theta)}$. Moreover, for present purposes there is no restriction of generality to assume that constant $\tilde{N}(s_0, \theta) > 0$ is less than one.

Theorem 3.6 will be established if we can show the uniform convergence of $\int f(s)\mu_{\theta}^n(ds)$ to $\int f(s)\mu_{\theta}(ds)$ over the set Θ as n goes to ∞ . Let us first prove the result for a Lipschitz function \tilde{f} with Lipschitz constant L . Hence, we need to show that for every $\eta > 0$ there

exists χ such that

$$|\int \tilde{f}(s)\mu_\theta^n(ds) - \int \tilde{f}(s)\mu_\theta(ds)| < \eta \quad (6.3)$$

for all θ in Θ and $n \geq \chi$.

Since Θ is a compact set, this set can be covered by a finite collection of neighborhoods $V(\theta_i)$ such that for all θ in $V(\theta_i)$,

$$\|\varphi(s, \varepsilon, \theta) - \varphi(s, \varepsilon, \theta_i)\| < \frac{\eta(1 - \tilde{\alpha}_i)}{4L} \quad (6.4)$$

for all (s, ε) , where $0 < \tilde{\alpha}_i < 1$ is equal to $\tilde{\alpha}_i = \tilde{\alpha}(s_0, \theta_i)$ as defined above, for an initial condition s_0 and the vector of parameters θ_i , for $i = 1, \dots, I$. Further, we can require that $\frac{\eta(1 - \tilde{\alpha}_i)}{L} < \tilde{\delta}_i$ for $\tilde{\delta}_i = \tilde{\delta}(s_0, \theta_i)$.

Also, as the sequence of functions $\{\varphi^n\}$ converges to φ , we can pick χ_i such that for all $n \geq \chi_i$

$$\|\varphi^n(s, \varepsilon, \theta) - \varphi(s, \varepsilon, \theta)\| < \frac{\eta(1 - \tilde{\alpha}_i)}{4L} \quad (6.5)$$

for all (s, ε) and all θ in $V(\theta_i)$. Now, by the triangle inequality it follows from (6.4)-(6.5) that

$$\|\varphi^n(s, \varepsilon, \theta) - \varphi(s, \varepsilon, \theta_i)\| < \frac{\eta(1 - \tilde{\alpha}_i)}{2L} \quad (6.6)$$

for all $n \geq \chi_i$, and all (s, ε) and θ in $V(\theta_i)$. Since constants \tilde{N} , $\tilde{\alpha}_i$ and $\tilde{\delta}_i$ hold for all ω in $\Omega(s_0, \theta)$ and $\gamma(\Omega(s_0, \theta)) > 0$, Lemma 4.4 and (6.1) and (6.6) imply that

$$|\int \tilde{f}(s)\mu_\theta^n(ds) - \int \tilde{f}(s)\mu_{\theta_i}(ds)| < \eta/2 \quad (6.7)$$

for all θ in $V(\theta_i)$. Moreover, by (6.4) the same argument establishes that

$$|\int \tilde{f}(s)\mu_\theta(ds) - \int \tilde{f}(s)\mu_{\theta_i}(ds)| < \eta/4 \quad (6.8)$$

for all θ in $V(\theta_i)$. Let $\chi = \max\{\chi_i\}$. Then, (6.3) is a simple consequence of (6.7)-(6.8).

Now, it remains to prove (6.3) for every continuous function f . Since the Lipschitz functions are dense in the set of continuous functions, for every $\eta > 0$ there exists a Lipschitz function such that $\|f - \tilde{f}\| < \eta$. Then, the above argument is easily extended to a continuous function f since $\|f - \tilde{f}\| < \eta$ implies that $|\int f(s)\mu_\theta(ds) - \int \tilde{f}(s)\mu_\theta(ds)| < \eta$ and $|\int f(s)\mu_\theta^n(ds) - \int \tilde{f}(s)\mu_\theta^n(ds)| < \eta$ for all probability measures μ_θ and μ_θ^n .

(b) Assume that Condition (C.3) is satisfied. The proof hinges upon the following result from Santos and Peralta-Alva (2003, Th. 3.7)

THEOREM 6.2: Let \tilde{f} be a Lipschitz function with constant L . For a given θ , let $\tilde{\varphi} : S \times \mathcal{E} \rightarrow S$ be a measurable function such that $\|\tilde{\varphi}(s, \varepsilon) - \varphi(s, \varepsilon, \theta)\| < \delta$ for all (s, ε) and $\delta > 0$. Assume that φ satisfies Condition (C.3). Then, under Assumptions (A.1) – (A.3),

$$|\int \tilde{f}(s)\tilde{\mu}(ds) - \int \tilde{f}(s)\mu_\theta(ds)| < \frac{L\delta}{1 - \alpha} \quad (6.9)$$

for every invariant distribution $\tilde{\mu}$ under $\tilde{\varphi}$ and for the invariant distribution μ_θ under the mapping $\varphi(\cdot, \cdot, \theta)$.

As in part (a) of this theorem, it suffices to establish (6.3) for a Lipschitz function \tilde{f} with constant L . As before, the set Θ can be covered by a finite number of neighborhoods $V(\theta_i)$, for $i = 1, \dots, I$, such that

$$\|\varphi(s, \varepsilon, \theta) - \varphi(s, \varepsilon, \theta_i)\| < \frac{\eta(1 - \alpha_i)}{4L} \quad (6.10)$$

for all θ in $V(\theta_i)$, where α_i is the modulus of contraction in (C.3) corresponding to the vector of parameters θ_i . Moreover, the postulated uniform convergence of the sequence of functions $\{\varphi^n\}$ implies that (6.5) and (6.6) above will be satisfied for n large enough. Then, by Theorem 6.2 the above inequalities (6.7)-(6.8) must hold in this case, and these inequalities yield the desired result in (6.3).

PROOF OF THEOREM 4.1: In view of our preceding arguments we shall only focuss on the uniform convergence of the simulated sequence $\{\frac{1}{T}\sum_{t=1}^T f(s_t(s_0, \omega, \theta))\}$ on the space Θ . The uniform convergence of $\{\frac{1}{T}\sum_{t=1}^T f(s_t(s_0, \omega, \theta))\}$ is first shown for some simple functions, and then for every continuous function f on S .

(1) *Uniform convergence of $\{\frac{1}{T}\sum_{t=1}^T \|s_t(s_0, \omega, \theta)\|\}$ to $E_\theta(\|s\|)$ over the space Θ for all s_0 and γ -almost all ω .*

Since a countable union of sets of measure zero has also measure zero, we only need to prove that for a fixed rational number $\eta > 0$ there is $T(\omega)$ such that for all θ in Θ and all $T \geq T(\omega)$,

$$|\frac{1}{T}\sum_{t=1}^T \|s_t(s_0, \omega, \theta)\| - \int \|s\| \mu_\theta(ds)| < \eta \quad (6.11)$$

for all s_0 and γ -almost all ω .

For $\eta > 0$, by Lemmas 4.2 and 4.3 the compact set Θ can be covered by a finite number of balls $B_{\kappa_i}(\theta_i)$ with center θ_i and radius κ_i , for $i = 1, 2, \dots, I$ such that

$$|\int \|s\| \mu_{\theta_i, \kappa_i}^{sup}(ds) - \int \|s\| \mu_{\theta_i, \kappa_i}^{inf}(ds)| < \frac{\eta}{3} \quad (6.12)$$

for all $\mu_{\theta_i, \kappa_i}^{sup}$ in $\Delta_{\theta_i, \kappa_i}^{sup}$ and all $\mu_{\theta_i, \kappa_i}^{inf}$ in $\Delta_{\theta_i, \kappa_i}^{inf}$. By the same arguments as in (4.3)-(4.6) above and the monotonicity of the max norm, for all θ in $B_{\kappa_i}(\theta_i)$ the following

inequalities must hold true:

$$\frac{1}{T} \sum_{t=1}^T \|s_t^{sup}(s_0, \omega, \theta_i, \kappa_i)\| \geq \frac{1}{T} \sum_{t=1}^T \|s_t(s_0, \omega, \theta)\| \geq \frac{1}{T} \sum_{t=1}^T \|s_t^{inf}(s_0, \omega, \theta_i, \kappa_i)\| \quad (6.13)$$

for all $t \geq 1$, and

$$\int \|s\| \mu_{\theta_i, \kappa_i}^{sup}(ds) \geq \int \|s\| \mu_{\theta}(ds) \geq \int \|s\| \mu_{\theta_i, \kappa_i}^{inf}(ds). \quad (6.14)$$

for every pair of invariant distributions $(\mu_{\theta_i, \kappa_i}^{sup}, \mu_{\theta_i, \kappa_i}^{inf})$. Moreover, by Lemma 4.4 for all s_0 and γ -almost all ω there exists $T_i(\omega)$ such that for all $T \geq T_i(\omega)$

$$\frac{1}{T} \sum_{t=1}^T \|s_t^{sup}(s_0, \omega, \theta_i, \kappa_i)\| \leq \max_{\mu_{\theta_i, \kappa_i}^{sup} \in \Delta_{\theta_i, \kappa_i}^{sup}} \int \|s\| \mu_{\theta_i, \kappa_i}^{sup}(ds) + \frac{\eta}{6} \quad (6.15)$$

and

$$\frac{1}{T} \sum_{t=1}^T \|s_t^{inf}(s_0, \omega, \theta_i, \kappa_i)\| \geq \min_{\mu_{\theta_i, \kappa_i}^{inf} \in \Delta_{\theta_i, \kappa_i}^{inf}} \int \|s\| \mu_{\theta_i, \kappa_i}^{inf}(ds) - \frac{\eta}{6} \quad (6.16)$$

Let $T(\omega) = \max\{T_i(\omega)\}$, for $i = 1, \dots, I$. Then, for all s_0 and γ -almost all ω it follows from (6-12)-(6-16) that (6.11) must hold true for all θ and all $T \geq T(\omega)$.

- (2) *Uniform convergence of $\{\frac{1}{T} \sum_{t=1}^T \tilde{f}(s_t(s_0, \omega, \theta))\}$ to $E_{\theta}(\tilde{f})$ over the space Θ for all s_0 and γ -almost all ω for every Lipschitz function \tilde{f} .*

For this part assume that the norm is the *sum* norm $\|s\|_{sum} = \sum_i |s_i|$ for every vector $s = (\dots, s_i, \dots)$. Let \tilde{f} be a Lipschitz function with constant L . Then

$$\begin{aligned}
& \left| \frac{1}{T} \sum_{t=1}^T \tilde{f}(s_t^{sup}(s_0, \omega, \theta_i, \kappa_i)) - \frac{1}{T} \sum_{t=1}^T \tilde{f}(s_t^{inf}(s_0, \omega, \theta_i, \kappa_i)) \right| \leq \\
& \frac{1}{T} \sum_{t=1}^T |\tilde{f}(s_t^{sup}(s_0, \omega, \theta_i, \kappa_i)) - \tilde{f}(s_t^{inf}(s_0, \omega, \theta_i, \kappa_i))| \leq \\
& L \frac{1}{T} \sum_{t=1}^T \|s_t^{sup}(s_0, \omega, \theta_i, \kappa_i) - s_t^{inf}(s_0, \omega, \theta_i, \kappa_i)\|_{sum} = \\
& L \frac{1}{T} \sum_{t=1}^T (\|s_t^{sup}(s_0, \omega, \theta_i, \kappa_i)\|_{sum} - \|s_t^{inf}(s_0, \omega, \theta_i, \kappa_i)\|_{sum}) = \\
& L \left(\frac{1}{T} \sum_{t=1}^T \|s_t^{sup}(s_0, \omega, \theta_i, \kappa_i)\|_{sum} - \frac{1}{T} \sum_{t=1}^T \|s_t^{inf}(s_0, \omega, \theta_i, \kappa_i)\|_{sum} \right).
\end{aligned}$$

It should be stressed that the first equality is key in the method of proof. This equality follows from the definition of the norm $\|\cdot\|_{sum}$ and the order-preserving property of the orbits since $s_t^{sup} \geq s_t^{inf}$ for all $t \geq 1$.

Using these inequalities the uniform convergence argument in part (a) can be readily extended for any arbitrary Lipschitz function \tilde{f} . Hence, by a suitable change of the estimates in (6.12) and (6.15)-(6.16), the inequality in (6.11) will read as follows

$$\left| \frac{1}{T} \sum_{t=1}^T \tilde{f}(s_t(s_0, \omega, \theta)) - \int \tilde{f}(s) \mu_\theta(ds) \right| < \eta. \quad (6.17)$$

- (3) *Uniform convergence of the sequence $\{\frac{1}{T} \sum_{t=1}^T f(s_t(s_0, \omega, \theta))\}$ to $E_\theta(f)$ over the space Θ for all s_0 and γ -almost all w for every continuous function f .*

This further extension follows from the fact that the set of Lipschitz functions is dense in the space of continuous functions.

PROOF OF THEOREM 4.5: The proof of this theorem amounts to a simple extension of the method of proof of Theorem 4.1. Hence, the proof is omitted.

PROOF OF THEOREM 4.7: In view of the proof of Theorem 4.1, there is no restriction of generality to assume that function f is continuous and monotone on S . Also, as Θ is a compact set, for every $\eta > 0$ this set can be covered by a finite number of balls $B_{\kappa_i}(\theta_i)$ such that the invariant distributions of functions $\varphi^{sup}(s_0, \omega, \theta_i, \kappa_i)$ and $\varphi^{inf}(s_0, \omega, \theta_i, \kappa_i)$ satisfy the following property

$$| \int f(s) \mu_{\theta_i, \kappa_i}^{sup}(ds) - \int f(s) \mu_{\theta_i, \kappa_i}^{inf}(ds) | < \frac{\eta}{2}. \quad (6.18)$$

Moreover, over these balls $B_{\kappa_i}(\theta_i)$ it follows that

$$\int f(s) \mu_{\theta_i, \kappa_i}^{sup}(ds) \geq \int f(s) \mu_{\theta}(ds) \geq \int f(s) \mu_{\theta_i, \kappa_i}^{inf}(ds) \quad (6.19)$$

and

$$\int f(s) \mu_{\theta_i, \kappa_i}^{n, sup}(ds) \geq \int f(s) \mu_{\theta}^n(ds) \geq \int f(s) \mu_{\theta_i, \kappa_i}^{n, inf}(ds) \quad (6.20)$$

where $\mu_{\theta_i, \kappa_i}^{n, sup}$, μ_{θ}^n , $\mu_{\theta_i, \kappa_i}^{n, inf}$ are invariant distributions for $\varphi^{n, sup}(s_0, \omega, \theta_i, \kappa_i)$, $\varphi^n(s_0, \omega, \theta)$, and $\varphi^{n, inf}(s_0, \omega, \theta_i, \kappa_i)$, respectively, and mappings $\varphi^{n, sup}(s_0, \omega, \theta_i, \kappa_i)$ and $\varphi^{n, inf}(s_0, \omega, \theta_i, \kappa_i)$ are defined from $\varphi^n(s_0, \omega, \theta)$ under (4.1)-(4.2). Since φ^n converges to φ in the sup norm, the corresponding sequences of functions $\{\varphi^{n, sup}(s_0, \omega, \theta_i, \kappa_i)\}$ and $\{\varphi^{n, inf}(s_0, \omega, \theta_i, \kappa_i)\}$ must converge to $\varphi^{sup}(s_0, \omega, \theta_i, \kappa_i)$ and $\varphi^{inf}(s_0, \omega, \theta_i, \kappa_i)$, respectively, in the sup norm. Therefore, by a result analogous to Lemma 4.3 there exists χ_i such that for all $n \geq \chi_i$,

$$| \int f(s) \mu_{\theta_i, \kappa_i}^{n, sup}(ds) - \int f(s) \mu_{\theta_i, \kappa_i}^{sup}(ds) | < \frac{\eta}{2} \quad (6.21)$$

and

$$| \int f(s) \mu_{\theta_i, \kappa_i}^{n, inf}(ds) - \int f(s) \mu_{\theta_i, \kappa_i}^{inf}(ds) | < \frac{\eta}{2}. \quad (6.22)$$

Let $\chi = \max\{\chi_i\}$. Then, (6.18)-(6.22) taken together imply that for all $n \geq \chi$ and all θ in Θ ,

$$|\int f(s)\mu_\theta^n(ds) - \int f(s)\mu_\theta(ds)| < \eta. \quad (6.23)$$

Since (6.23) holds true for an arbitrary $\eta > 0$, it follows that $\int f(s)\mu_\theta^n(ds)$ converges uniformly to $\int f(s)\mu_\theta(ds)$ in Θ . Therefore, every sequence of solutions $\{\theta^n\}$ defined by (3.6) must converge to the solution θ^0 defined by (2.1).