Accuracy of Simulations for Stochastic Dynamic Models*

Manuel S. Santos
Department of Economics
Arizona State University

Adrian Peralta-Alva
Department of Economics
University of Miami

September 14, 2003

This paper provides a general framework for the simulation of stochastic dynamic models. Our analysis rests upon a continuity property of invariant distributions and a generalized law of large numbers. We then establish that the simulated moments from numerical approximations converge to their exact values as the approximation errors of the computed solutions converge to zero. These asymptotic results are of further interest in the comparative study of dynamic solutions, model estimation, and derivation of error bounds for the simulated moments.

KEYWORDS: Stochastic Dynamic Model, Invariant Distribution, Numerical Solution, Approximation Error, Simulated Moments, Convergence.

1 Introduction

Stochastic dynamic models pervade many areas of economics, since time and uncertainty are basic considerations in most economic decisions. A typical situation in the analysis of these models is portrayed in Figure 1. A researcher is interested in the predictions or long-run behavior of a stochastic dynamic model. These predictions are often summarized by the moments and other statistics of its invariant distributions, and may form a basis for the calibration, estimation, and testing of the model. Most often the model

*We have benefitted form several discussions with participants at the seminars of Arizona State University, Universidad Carlos III of Madrid, the University of Minnesota, the Federal Reserve Bank of Minneapolis, and the SED meetings (Paris, June 2003). We are grateful to Jorge Aseff for several detailed comments to an earlier draft.

does not admit an analytical equilibrium solution, and hence it is approximated by numerical methods. Also, numerical simulations may be needed to compute the moments of an invariant distribution and for the study of further quantitative properties of the equilibrium solution. But in spite of the widespread use of numerical methods, there is not a well established theory for the simulation of stochastic dynamic models and most available results in probability theory cannot be readily applied to the analysis and computation of non-linear economic models.

This paper presents several results on accuracy properties of statistics derived from numerical simulations. Our analysis rests upon a continuity property of the correspondence of invariant distributions and a generalized law of large numbers. The continuity property of the correspondence of invariant distributions is shown to hold under general assumptions, and it implies that the moments of invariant distributions generated by a sufficiently good numerical approximation remain close to those generated by the original model. This result justifies the use of numerical methods to approximate quantitative properties of invariant distributions in stochastic dynamic models. Furthermore, under a certain contractivity property of the equilibrium solution, we provide an upper bound for the errors of the moments of invariant distributions generated by a numerical approximation. We establish that the errors are of the same order of magnitude as that of the computed solution. Some simple examples below illustrate that the constants bounding these orders of convergence are reasonably tight and can be explicitly computed.

The moments of an invariant distribution are generally hard to compute because of lack of information about the domain and form of the probability measure. Hence, laws of large numbers are ordinarily invoked to recover such statistics from numerical simulations. Available laws of large numbers for non-linear dynamical systems, however, are based upon some technical conditions [cf. Doob (1953, Ch. V)] that are difficult to check for both the economic model and the numerical approximation. We shall establish a generalized law of large numbers that builds upon some methods developed by Crauel (2002). This generalized law dispenses completely with the aforementioned technical conditions, and is valid for all initial values of the state variables.

Combining all these results we obtain that the average behavior of a typical sample path generated by a good enough numerical solution must be close to the corresponding expected value of some invariant distribution of the original model. Therefore, the moments computed from simulations of numerical solutions converge to the true moments as the approximation errors of the numerical solutions converge to zero. This proposition is at the foundations of statistical inference –and related empirical work– for non-linear dynamic models. It is worth pointing out that the main assumptions leading to these results are compactness of the domain and continuity of the equilibrium solution. These assumptions are usually derived from primitive conditions of economic

models. Under certain regularity properties the compactness of the domain may be weakened, but some examples below illustrate that the continuity of the equilibrium solution is indispensable.

A considerable part of the literature in computational economics has focussed on the performance of numerical algorithms [e.g., see Taylor and Uhlig (1990), Judd (1998), Santos (1999), and Aruoba, Fernandez-Villaverde and Rubio-Ramirez (2003)]. These algorithms are evaluated in some test cases or under some accuracy checks [e.g., den Haan and Marcet (1994) and Santos (2000)]. But given the complex dynamic behavior involved in random sample paths, error bounds on the computed solution are generally not operative to assess steady-state predictions of stochastic dynamic models. What it seems missing is a theory that links the approximation error of the numerical solution with the corresponding error in the statistics generated by this solution. In the absence of this theory, error bounds or accuracy checks on the numerical solution fall short of what it is required for the calibration, estimation and testing of an economic model. To accomplish these tasks a researcher usually defines a notion of distance in which a selected set of statistics is compared with its data counterparts. In the business cycle literature, for instance, this notion of distance is determined by second-order moments [e.g., Cooley and Prescott (1995)]. In the finance literature, measures of skewness and kurtosis are typically invoked. These exercises are only meaningful if the statistics computed from numerical approximations are sufficiently close to the true ones. We are not aware of any previous theoretical work concerned with accuracy properties of the simulated moments from numerical approximations.

Christiano and Eichenbaum (1992) develop a general framework for the calibration of dynamic models that takes account of the sampling error from time series data. These authors provide confidence intervals for the model's parameters, and consequently for the moments of the corresponding invariant distributions. Using the present results, some other sources of error may be integrated in this analysis. Rust (1994) presents several methods for the estimation and testing of structural dynamical systems in the presence of measurement and numerical errors and under some other misspecifications of the model. Such theories of estimation and testing will have to confront the issues addressed in this paper regarding sensitivity properties of invariant distributions.

Another application of our results is in the area of comparative statics analyses. Continuity properties of invariant distributions are a prerequisite to evaluate the effects of small changes in fiscal and monetary policies and in the model's parameters. Also, the continuity of the moments of the invariant distribution over a vector of parameters is assumed in proofs of consistency and asymptotic normality of simulation-based estimators [e.g., Duffie and Singleton (1993)]. It is then instructive to tie down this continuity condition to assumptions on the model's primitives. Moreover, building upon the results of this paper Santos (2003) proves various consistency properties of a simulated

estimator for a class of monotone Markov processes.

2 Stochastic Dynamics

Computable dynamic models are often characterized by equilibrium solutions that take the form of a Markov stochastic process or a stochastic difference equation. In most economic applications the solution system cannot be written explicitly, and hence the model is often simulated by numerical methods. Of course, the problem is that the stochastic dynamics may be substantially affected by the numerical approximation.

2.1 Random Dynamical Systems

In several economic models [e.g., Stokey, Lucas and Prescott (1989)] the equilibrium law of motion of the state variables can be specified by a dynamical system of the following form

$$\begin{aligned} z_{n+1} &= \Psi(z_n, \varepsilon_{n+1}) \\ k_{n+1} &= g(k_n, z_n, \varepsilon_{n+1}), \qquad\qquad n = 0, 1, 2, \ldots . \end{aligned} \tag{2.1}$$

Here, z may represent a vector made up of stochastic exogenous variables such as some indices of factor productivity or market prices. This random vector lies in a space Z, and it evolves according to a function Ψ and an *iid* shock ε in a set of "events" E governed by a probability measure Q. Vector k represents the endogenous state variables which may correspond to several types of capital stocks and measures of wealth. The evolution of k is determined by an equilibrium decision rule g taking values in a set K. Hence, $s = (k, z)$ is the vector of state variables that belongs to the set $S = K \times Z$. Let $(S, \mathbb{S})$ denote a measurable space.

For expository purposes, let us express (2.1) in the more compact form

$$s_{n+1} = \varphi(s_n, \varepsilon_{n+1}) \qquad\qquad n = 0, 1, 2, \ldots . \tag{2.2}$$

ASSUMPTION 1 *The set $S = K \times Z \subset R^l \times R^m$ is compact, and $\mathbb{S}$ is the Borel σ-field. $(E, \mathbb{E}, Q)$ is a probability space.*

ASSUMPTION 2 *Function $\varphi : S \times E \to S$ is bounded. Moreover, for each ε the mapping $\varphi(\cdot, \varepsilon)$ is continuous, and for each s the mapping $\varphi(s, \cdot)$ is measurable.*

It follows from a standard result in measure theory [cf., Futia (1982, Sect. 5)] that function φ is measurable on the product space $(S \times E, \mathbb{S} \times \mathbb{E})$. Therefore, this mapping gives rise to a transition function $P : S \times \mathbb{S} \to [0,1]$ that defines a time homogenous Markov process, such that $P(s,\cdot)$ is a probability on $\mathbb{S}$ for each given s in S and $P(\cdot, A)$ is a measurable function for each Borel set A in $\mathbb{S}$. The transition function P is defined as

$$P(s,A) = Q(\{\varepsilon | \varphi(s,\varepsilon) \in A\}) \tag{2.3}$$

for every $s \in \mathbb{S}$ and $A \in \mathbb{S}$. Function P characterizes the dynamics in the space of distributions on $\mathbb{S}$. Thus, for any initial probability measure μ_0 on $\mathbb{S}$, the evolution of future probabilities is determined by the recursion law

$$\mu_{n+1}(A) = \int P(s,A)\mu_n(ds) \tag{2.4}$$

for all $A \in \mathbb{S}$ and $n \geq 0$.

An invariant probability measure or invariant distribution μ^* is a fixed point of system (2.4). More precisely,

$$\mu^*(A) = \int P(s,A)\mu^*(ds) \tag{2.5}$$

for all $A \in \mathbb{S}$. Hence, an invariant distribution is a stationary solution of the original system (2.1). The analysis of invariant distributions seems then a very first step to investigate the dynamics of the system. Further, uniqueness of the invariant distribution is a highly desirable property, since the model may have sharper predictions.

2.2 Numerical Approximations

In many economic applications an explicit solution for the equilibrium function φ is not available. Then, the most one can hope for is to get a numerical approximation $\widehat{\varphi}$. Moreover, using some accuracy checks [cf., Santos (2000)] we may be able to bound the distance between functions φ and $\widehat{\varphi}$. Every numerical approximation $\widehat{\varphi}$ satisfying Assumptions 1-2 will give rise to a transition probability $\widehat{P}$ on $(S, \mathbb{S})$. But even if $\widehat{\varphi}$ is an arbitrarily good approximation of function φ, the following questions may come to the fore:

- How different are the dynamics under transition functions P and $\widehat{P}$?
- More specifically, do both functions generate the same number of invariant distributions?

- How close are the statistics defined over these invariant distributions?

As the following simple examples illustrate, without further regularity assumptions we cannot expect good stability properties.

EXAMPLE 2.1: Let g be the real-valued function depicted in Figure 2. This function has three interior steady states k_l, k_m, k_h. Consider now a downward vertical shift of the graph of this function so as to get the neighboring function $\widehat{g}$. Observe that steady states k_m and $\widehat{k}_m$ are unstable. Hence, any initial condition over the interval $(k_m, \widehat{k}_m)$ will converge to point k_h under g, but will converge to point $\widehat{k}_l$ under $\widehat{g}$. Consequently, a small perturbation on function g will not generally have good stability properties near the stationary solution k_m. From these functions we can now construct a random dynamical system under the following *iid* process. At each date $n = 0, 1, 2, \ldots$ let the system move by function g with probability 0.5, and by function $\widehat{g}$ with the remaining equal probability. One readily sees that the resulting random dynamical system has two ergodic invariant distributions [1] whose supports are the intervals $[\widehat{k}_l, k_l]$ and $[\widehat{k}_h, k_h]$. Points over the interval $[k_m, \widehat{k}_m]$ will leave this domain with probability one, and so this interval no longer will contain a stationary solution. As a matter of fact, Kolmogorov has been credited as one of the first to stress that adding a small stochastic perturbation to a deterministic dynamical system may lead to substantial discrepancies in the long-run dynamics of the system.

EXAMPLE 2.2: This example contains an analogous perturbation on the transition probability P. The state space S is a discrete set with three possible states, s_1, s_2, s_3. Transition probability P is defined by the following Markov matrix

$$\Pi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$

Each row i specifies the probability of moving from state s_i to any state in S, so that an element π_{ij} corresponds to the value $P(s_i, \{s_j\})$, for $i, j = 1, 2, 3$. Note that $\Pi^n = \Pi$ for all $n \geq 1$. Hence, $p = (1, 0, 0)$, and $p = (0, 1/2, 1/2)$ are invariant distributions under Π, and $\{s_1\}$ and $\{s_2, s_3\}$ are the ergodic sets. All other invariant distributions are convex combinations of these two probabilities.

[1] We say that a set $A \in \mathbb{S}$ is invariant if $P(s, A) = 1$ for all s in A. An invariant set A is called ergodic if there is no other invariant subset $B \subset A$. The support $\sigma(\mu^*)$ of a probability measure μ^* is the smallest closed set such that $\mu^*(\sigma(\mu^*)) = 1$. An invariant distribution μ^* is called ergodic if $\mu^*(A) = 0$ or $\mu^*(A) = 1$ for every invariant set A.

Let us now perturb Π slightly so that the new stochastic matrix is the following

$$\widehat{\Pi} = \begin{bmatrix} 1-2\alpha & \alpha & \alpha \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix} \text{ for } 0 < \alpha < 1/2.$$

As $n \to \infty$, the sequence of stochastic matrices $\{\widehat{\Pi}^n\}$ converges to

$$\begin{bmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$

Hence, $p = (0, 1/2, 1/2)$ is the only possible long-run distribution for the system. Moreover, $\{s_1\}$ is a transient state, and $\{s_2, s_3\}$ is the only ergodic set. Consequently, a small perturbation on a transition probability P may lead to a pronounced change in its invariant distributions. Indeed, small errors may propagate over time and alter the existing ergodic sets.

If we consider the correspondence from Π to the set of invariant distributions $\{p|p\Pi = p\}$, then the present example shows that such correspondence fails to be lower semicontinuous. (A similar result was obtained under the stochastic perturbation in Example 2.1.) A key step below is to establish that this correspondence is upper semicontinuous for Markov processes generated by a general class of random dynamical systems.

A further fundamental issue in stochastic dynamics is the convergence of the sequence of transition functions $\{\Pi^n\}$. This sequence may not always converge, and hence for certain distributions the system may not settle down to a stationary invariant probability. There are situations in which laws of large numbers fail to exist, and hence statistical inference may lose its full strength. For some important purposes it is sufficient to establish a weak notion of convergence for operators $A_N(\Pi) = \frac{1}{N}\sum_{n=1}^{N} \Pi^n$. For discrete state spaces, $\{A_N(\Pi)\}$ always converges to an invariant stochastic matrix, but further regularity conditions are required for more general state spaces [cf. Doob (1953, Ch. V)].

3 Main Results

This section contains our main analytical results that include an upper semicontinuity property of the invariant distributions correspondence and a generalized law of large numbers. These results entail that the moments computed from numerical simulations

converge to the moments of the model's invariant distributions as the approximation errors of the computed solutions converge to zero. Moreover, under a contractivity property on the equilibrium law of motion, we establish error bounds for the moments generated by numerical approximations.

3.1 Upper Semicontinuity of the Invariant Distributions Correspondence

We begin with some basic definitions in probability theory. Let $B(S)$ be the set of all bounded, $\mathbb{S}$-measurable, real-valued functions on S. Then, $B(S)$ is a Banach space when endowed with the norm $\|f\| = \sup_{s\in S} |f(s)|$. A transition function P on $(S, \mathbb{S})$ defines a Markov operator T from $B(S)$ to itself via the following integration operation

$$(Tf)(s) = \int f(s')P(s, ds') \quad \text{all } s \in S. \tag{3.1}$$

Each operator T can be associated with its adjoint T^* so that if we write $\langle f, \mu\rangle = \int f(s)\mu(ds)$, then $\langle Tf, \mu\rangle = \langle f, T^*\mu\rangle$ for all f and μ. Moreover, T^* is given by

$$(T^*\mu)(A) = \int P(s, A)\mu(ds) \quad \text{all } A \in \mathbb{S}. \tag{3.2}$$

The adjoint operator T^* maps the space of probability measures on $\mathbb{S}$ to itself.

Let $C(S)$ be the space of all continuous real-valued functions f on S. The weak topology is the coarsest topology such that every linear functional in the set $\{\mu \to \int f(s)\mu(ds), f \in C(S)\}$ is continuous. Then, a sequence $\{\mu_j\}$ of probability measures on $\mathbb{S}$ is said to converge weakly to a probability measure μ if $\int f(s)\mu_j(ds) \to_j \int f(s)\mu(ds)$ for every $f \in C(S)$. The weak topology is metrizable [e.g., see Billingsley (1968)]. Hence, every weakly convergent sequence $\{\mu_j\}$ of probability measures has a unique limit point.

THEOREM 3.1 *Under Assumptions 1-2, there exists a probability measure μ^* such that $\mu^* = T^*\mu^*$.*

The existence of an invariant probability measure μ^* can be established as follows. First, one can show [e.g., see Futia (1982, Prop. 5.6)] that T maps the space $C(S)$ of continuous functions into itself. Hence, operator T^* must be weakly continuous. Moreover, in the weak topology the set of all probability measures on $\mathbb{S}$ is compact, and it is obviously a convex set. Therefore, by the Markov-Kakutani fixed-point theorem there exists μ^* such that $\mu^* = T^*\mu^*$.

For a vector-valued function $\varphi = (\dots, \varphi^i, \dots)$, let $\|\varphi\| = \max_{1\le i\le l+m} \|\varphi^i\|$. Convergence of a sequence of functions $\{\varphi_j\}$ should be understood in the metric induced

by this norm. By Assumptions 1-2 each φ_j defines the associated triple (P_j, T_j, T_j^*); moreover, by Theorem 3.1 there always exists an invariant distribution $\mu_j^* = T_j^* \mu_j^*$.

THEOREM 3.2 *Let $\{\varphi_j\}$ be a sequence of functions converging to φ. Let $\{\mu_j^*\}$ be a sequence of probabilities on $\mathbb{S}$ such that $\mu_j^* = T_j^* \mu_j^*$ for each j. Then, under Assumptions 1-2 every weak limit point μ^* of $\{\mu_j^*\}$ is an invariant probability measure, $\mu^* = T^* \mu^*$.*

Observe that the theorem asserts the bilinear convergence of $T_j^* \mu_j^*$ to $T^* \mu^*$. This result is stronger than the standard notion of weak convergence, and entails that the invariant distributions correspondence is upper semicontinuous.[2] An early statement of the upper semicontinuity of this correspondence appears in Dubins and Freedman (1966, Th. 3.4). Further extensions can be found in Manuelli (1985) and Stokey, Lucas and Prescott (1989, Th. 12.3). All these authors validate the upper semicontinuity of invariant distributions under assumptions on function P, but these assumptions seem hard to check in applications. The practical importance of Theorem 3.2 is that it applies to numerical approximations of Markov processes generated by stochastic systems satisfying Assumptions 1-2.

COROLLARY 3.3 *Let f belong to $C(S)$. Then under the conditions of Theorem 3.2, for every $\eta > 0$ there exits J such that for each μ_j^* with $j \geq J$ there is μ^* with the property*

$$\left| \int f(s) \mu_j^*(ds) - \int f(s) \mu^*(ds) \right| < \eta. \tag{3.3}$$

This is an alternative formulation of the upper semicontinuity of the invariant distributions correspondence. Observe that function f may define a moment or statistic of an invariant probability measure. A related result can be stated in terms of the expectations operators for functions f in $C(S)$. Let

$$E^{max}(f) = \max_{\{\mu^* | \mu^* = T^* \mu^*\}} \int f(s) \mu^*(ds) \tag{3.4}$$

$$E^{min}(f) = \min_{\{\mu^* | \mu^* = T^* \mu^*\}} \int f(s) \mu^*(ds). \tag{3.5}$$

[2]The proofs of our main results are gathered in the Appendix. One can check that the proof of Theorem 3.2 goes through under the less demanding metric $\|\varphi - \varphi_j\|_L = \max_{s \in S} [\int \|\varphi(s, \varepsilon) - \varphi_j(s, \varepsilon)\| \, Q(d\varepsilon)]$.

REMARK 3.4 Note that the set of invariant distributions $\{\mu^* | \mu^* = T^*\mu^*\}$ is weakly compact and convex. Hence, every continuous linear functional $\mu^* \to \int f(s)\mu^*(ds)$ always attains a maximum and a minimum value over this set. Moreover, convexity entails that for every point $E(f)$ in the interval of values $[E^{min}(f), E^{max}(f)]$ there exists an invariant distribution μ^* such that $E(f) = \int f(s)\mu^*(ds)$.

As a straightforward consequence of Corollary 3.3, we obtain the following simple inequalities.

COROLLARY 3.5 *Let f belong to $C(S)$. Then under the conditions of Theorem 3.2, for every $\eta > 0$ there exists J such that*

$$E^{max}(f) + \eta \geq \int f(s)\mu_j^*(ds) \geq E^{min}(f) - \eta \tag{3.6}$$

for each μ_j^* with $j \geq J$.

It is worth mentioning the following two applications of Corollaries 3.3 and 3.5. First, for some computational methods [e.g., see Santos (1999)] one can obtain a sequence of numerical approximations $\{\varphi_j\}$ that converge to the exact solution φ. Then, (3.6) implies that the invariant distributions generated by these approximations will eventually be contained in an arbitrarily small weak neighborhood of the set of all invariant distributions generated by the original model. Second, most economic modelizations involve a family of solutions $\varphi(s, \varepsilon, \theta)$ parameterized by a vector θ in a space Θ. (Note that if φ is a continuous mapping on a compact domain, then this function is uniformly continuous.) As above for each θ one could define the functional mappings $E_\theta^{max}(f)$ and $E_\theta^{min}(f)$. Then, (3.6) implies that $E_\theta^{max}(f)$ is an upper semicontinuous function in θ, and $E_\theta^{min}(f)$ is a lower semicontinuous function in θ. If there exists but a unique invariant distribution μ_θ^* so that $E_\theta(f) = E_\theta^{max}(f) = E_\theta^{min}(f)$ for all θ, then the expectations operator $E_\theta(f)$ varies continuously with θ.

3.2 Error Bounds

In numerical applications it is often desirable to bound the size of the approximation error. Computations must stop in finite time, and hence error bounds can dictate efficient stopping criteria. In most theoretical work the size of the error is estimated from the convergence order of the numerical approximation and the constant bounding such a convergence order. It is then of paramount importance to relate these two components of the approximation error to primitive parameters of the model.

We shall now impose a contractivity condition on function φ. Then, we show that the error of the statistics generated by a numerical approximation is of the same order of magnitude as that of the computed solution.

CONDITION C: There exits a constant $0 < \alpha < 1$ such that $\int_E \|\varphi(s,\varepsilon) - \varphi(s',\varepsilon)\| \, Q(d\varepsilon) \leq \alpha \|s - s'\|$ for all pairs s, s'.

Condition C is familiar in 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, Condition C arises naturally in the one-sector Solow model [e.g., Schenk-Hoppé and Schmalfuss (2001)] and in concave dynamic programs [e.g., see Foley and Hellwig (1975) and Examples 4.2-4.3 below]. Stochastic contractivity properties are also encountered 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)].

It has been shown under various forms that a random contractive system has a unique invariant distribution μ^*. To present a formal version of this result –which will be needed for Theorem 3.7 below– let us now introduce some simple definitions. We say that a real-valued function f on S is Lipschitz with constant L if $\|f(s) - f(s')\| \leq L \|s - s'\|$ for all pairs s and s'. Let $Z^n(s_0)$ denote the random vector $\varphi(\varphi \cdots (\varphi(s_0, \varepsilon_1), \varepsilon_2) \cdots \varepsilon_n)$ for each initial value s_0, and let $\mu^n_{s_0}$ denote the distribution of $Z^n(s_0)$. Finally, let $d = diam(S)$.

THEOREM 3.6 [cf., Stenflo (2001)] *Let f be a Lipschitz function with constat L. Assume that φ satisfies Condition C. Then under Assumptions 1-2,*

(i) There exists a unique invariant distribution $\mu^* = T^*\mu^*$.

(ii) For all initial conditions s_0 and all n,

$$\left| \int f(s)\mu^*(ds) - \int f(s)\mu^n_{s_0}(ds) \right| \leq \frac{Ld\alpha^n}{1-\alpha}. \tag{3.7}$$

Consider now a numerical approximation $\widehat{\varphi}$. As before, this function defines an operator $\widehat{T}^*$ that has a fixed-point solution $\widehat{\mu}^* = \widehat{T}^*\widehat{\mu}^*$. We can now establish the following result.

THEOREM 3.7 *Let f be a Lipschitz function with constant L. Let $\|\widehat{\varphi} - \varphi\| < \delta$ for some $\delta > 0$. Assume that φ satisfies Condition C. Then under Assumptions 1-2,*

$$\left| \int f(s)\mu^*(ds) - \int f(s)\widehat{\mu}^*(ds) \right| \leq \frac{L\delta}{1-\alpha} \tag{3.8}$$

where μ^* is the unique invariant distribution under φ, and $\widehat{\mu}^*$ is any invariant distribution under $\widehat{\varphi}$.

Note that function $\widehat{\varphi}$ may not satisfy Condition C, and hence it may contain multiple Markovian invariant distributions $\widehat{\mu}^*$. Stenflo (2001) proves a related result in which the approximate function $\widehat{\varphi}$ is required to satisfy Condition C. This contractivity condition, however, is rather restrictive for our purposes, since it may not be preserved for polynomial interpolations and related high-order approximation schemes under which function $\widehat{\varphi}$ may have been calculated.

3.3 A Generalized Law of Large Numbers

In applied work, laws of large numbers are often invoked to compute statistics of invariant distributions. For stochastic dynamical systems, usual derivations of laws of large numbers proceed along the lines of the ergodic theorem [e.g., see Krengel (1985)]. But to apply the ergodic theorem the initial value of the state variable s_0 must lie inside an ergodic set. A certain technical condition –known as Hypothesis D from Doob (1953)– ensures that for every initial value s_0 the dynamical system will enter one of its ergodic sets almost surely. Then, using this hypothesis standard ergodic theorems can be extended to incorporate all initial values s_0 outside these ergodic sets. Hypothesis D, however, is usually difficult to verify in economic applications [e.g., see Stokey, Lucas and Prescott (1989, Ch. 11)].

Under conditions similar to Assumptions 1-2 above, Breiman (1960) proves a law of large numbers that is valid for all initial values s_0. This author dispenses with Hypothesis D, but requires the Markov process to have a unique invariant distribution. Uniqueness of the invariant distribution seems, however, a rather limiting restriction for numerical approximations. Primitive conditions may be imposed on the original model that guarantee the existence of a unique invariant distribution, but these conditions may not be preserved by the discretization procedure leading to the numerical approximation. Indeed, uniqueness of the invariant distribution is not robust to continuous perturbations of the model. This is illustrated in Figure 3 that portrays a deterministic policy function with a unique stationary point. The dotted line depicts a close approximation that contains a continuum of stationary points.

Our goal is then to derive a law of large numbers that holds true for all initial values s_0 but without imposing the technical Hypothesis D assumed in Doob (1953) or the existence of a unique invariant distribution assumed in Breiman (1960). Our generalized law of large numbers is a strengthening of some analytical methods developed by Crauel (1991, 2002). These methods cannot be directly applied to Markovian invariant distributions, since function φ may contain some other invariant distributions. The method of proof relies on an argument related to the subadditive ergodic theorem of Kingman (1968) and on certain properties of conditional expectations for non-Markovian invariant distributions as spelled out in Arnold (1998). Section 5 below provides a further

discussion of these results along with Hypothesis D.

In preparation for our analysis, we define a new probability space comprising all infinite sequences $\{\varepsilon_n\}$. Let $\Omega = E^\infty$ be the countably infinite cartesian product of copies of E. Let $\mathbb{F}$ be the σ-field in E^∞ generated by the collection of all cylinders $A_1 \times A_2 \times ... \times A_n \times E \times E \times E \times ...$ where $A_i \in \mathbb{E}$ for $i = 1, ..., n$. A probability measure λ can be constructed over these finite-dimensional sets $A_1 \times A_2 \times ... \times A_n \times E \times E \times E \times ...$ as the product of probabilities $Q(A_i)$. That is, let

$$\lambda\{\omega : \varepsilon_1 \in A_1, \varepsilon_2 \in A_2, ..., \varepsilon_n \in A_n\} = \prod_{i=1}^{n} Q(A_i) \tag{3.9}$$

for $\omega = (\varepsilon_1, \varepsilon_2, ...)$, and $A_i \in \mathbb{E}$ for $1 \leq i \leq n$. This measure λ has a unique extension on $\mathbb{F}$. Hence, let $(\Omega, \mathbb{F}, \lambda)$ denote the probability space. Finally, for every initial value s_0 and sequence of shocks $\omega = \{\varepsilon_n\}$, let $\{s_n(s_0, \omega)\}$ be the sample path generated by function φ, so that $s_{n+1}(s_0, \omega) = \varphi(s_n(s_0, \omega), \varepsilon_{n+1})$ for all $n \geq 1$ and $s_1(s_0, \omega) = \varphi(s_0, \varepsilon_1)$.

THEOREM 3.8 *Let f belong to $C(S)$. Then, under Assumptions 1-2 for all s_0 and for λ-almost all ω we have*

$$(i) \qquad \limsup_N \frac{1}{N} \sum_{n=1}^{N} f(s_n(s_0, \omega)) \leq E^{max}(f) \tag{3.10}$$

$$(ii) \qquad \liminf_N \frac{1}{N} \sum_{n=1}^{N} f(s_n(s_0, \omega)) \geq E^{min}(f). \tag{3.11}$$

TECHNICAL REMARKS: *(a)* Example 4.1 below illustrates that inequalities (3.10)-(3.11) may fail to hold in the absence of the continuity of $\varphi(s, \varepsilon)$ in s. Moreover, the proof of Theorem 3.8 shows that there are invariant probability measures for which these bounds are tight in the sense that both (3.10) and (3.11) must hold with equality for some (s_0, ω).

(b) Theorem 3.8 is a sharpened version of Crauel (2002, Prop. 6.21). Our proof has to deal with a further added technicality, since mapping φ and transition function P may not have the same invariant distributions. Indeed, φ may generate some non-Markovian probability measures μ on $\mathbb{S} \times \mathbb{F}$ [e.g., see Arnold (1998, p. 56)]. Our operators $E^{max}(f)$ and $E^{min}(f)$ in (3.10)-(3.11) are defined in (3.4)-(3.5) over the set of probability measures μ^* which are fixed points of P, whereas Crauel proves the result over the set of all invariant probability measures of φ. Therefore, our bounds (3.10)-(3.11) are tighter, since the operators $E^{max}(f)$ and $E^{min}(f)$ are defined over a smaller

set of invariant distributions, but the bounds in Crauel hold for functions f defined over the set $S \times \Omega$.

(c) It should be stressed that (3.10)-(3.11) hold for all s_0 over λ-almost all ω. As a matter of fact, the proof of Theorem 3.8 shows that (3.10)-(3.11) hold *uniformly* in s_0. More precisely, inequality (3.10) can be replaced by the stronger condition $\limsup_N [\max_{s_0 \in S}(\frac{1}{N}\sum_{n=1}^{N} f(s_n(s_0,\omega)))] \leq E^{max}(f)$, and inequality (3.11) can be replaced by $\liminf_N [\min_{s_0 \in S}(\frac{1}{N}\sum_{n=1}^{N} f(s_n(s_0,\omega)))] \geq E^{min}(f)$.

(d) Observe that Theorem 3.8 does not guarantee convergence of a typical sequence $\{\frac{1}{N}\sum_{n=1}^{N} f(s_n(s_0,\omega))\}$. Note, however, that by the convexity of the set of invariant distributions μ^* (see Remark 3.4) for every limit point $E(f)$ of the sequence $\{\frac{1}{N}\sum_{n=1}^{N} f(s_n(s_0,w))\}$ there exists an invariant distribution μ^* such that $E(f) = \int f(s)\mu^*(ds)$.

If there exists a unique invariant distribution μ^*, then $E^{max}(f) = E^{min}(f)$. Hence, both limits in (3.10)-(3.11) are the same and correspond to the unique expected value $E(f)$. Therefore, as a special case of Theorem 3.8 we obtain a standard formulation of the law of large numbers for systems with a unique invariant distribution.

COROLLARY 3.9 [c.f., Breiman (1960)] *Assume that there exits a unique invariant distribution $\mu^* = T^*\mu^*$. Then under the conditions of Theorem 3.8,*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(s_n(s_0,\omega)) = E(f) \tag{3.12}$$

for all s_0 and for λ-almost all ω.

3.4 Accuracy of Numerical Simulations

The foregoing results are now applied to the problem of original interest represented in Figure 1, which can be restated as follows. A researcher is concerned with the predictions of a stochastic dynamic model whose equilibrium solution may be specified by a function φ. Function φ may not have an analytical representation. The model is then solved by numerical methods. However, most often the invariant distributions generated by a numerical approximation cannot be computed directly. Hence, numerical methods are again brought up into the analysis. This time in connection with some law of large numbers. As is typical in the simulation of stochastic models we suppose that the researcher can draw sequences $\{\widehat{\varepsilon}_n\}$ from a generating process that can mimic the distribution of the shock process $\{\varepsilon_n\}$. A probability measure λ is defined over all sequences $\omega = (\varepsilon_1, \varepsilon_2, ...)$. Once a numerical approximation φ_j is available, it is generally not so costly to generate sample paths $\{s_{jn}(s_0,\omega)\}$ defined recursively as

$s_{jn+1}(s_0,\omega) = \varphi_j(s_{jn}(s_0,\omega),\varepsilon_{n+1})$ for every $n \geq 0$ for fixed s_0 and ω. Averaging over these sample paths we get sequences of simulated statistics $\{\frac{1}{N}\sum_{n=1}^{N} f(s_{jn}(s_0,\omega))\}$ as defined by the real-valued function f. Our goal is to establish that the limit points of the sequence of simulated statistics $\{\frac{1}{N}\sum_{n=1}^{N} f(s_{jn}(s_0,\omega))\}$ from a good enough numerical approximation φ_j are close to the exact values $E(f) = \int f(s)\mu^*(ds)$ of invariant distributions μ^* of the solution function φ.

THEOREM 3.10 *Under the conditions of Theorem 3.2, for every $\eta > 0$ there exists J such that for each $j \geq J$ we can find a function $N_j(w)$ so that*

$$E^{max}(f) + 2\eta \geq \frac{1}{N}\sum_{n=1}^{N} f(s_{jn}(s_0,\omega)) \geq E^{min}(f) - 2\eta \tag{3.13}$$

for all s_0 and λ-almost all ω, and for all $N \geq N_j(\omega)$.

Theorem 3.10 is a simple consequence of Corollary 3.5 and Theorem 3.8. Note that by the technical remark *(c)* after Theorem 3.8 the convergence in (3.10)-(3.11) is uniform in s_0. Hence, (3.13) is valid for λ-almost all ω independently of s_0. Also, we should stress that by the convexity of the set of invariant distributions (see Remark 3.4) for every function f in $C(S)$ the whole interval of values $[E^{min}(f), E^{max}(f)]$ can be generated by the linear functionals $\mu^* \to \int f(s)\mu^*(ds)$ over the set of invariant distributions $\{\mu^* | \mu^* = T^*\mu^*\}$. Therefore, every limit point of the sequence $\{\frac{1}{N}\sum_{n=1}^{N} f(s_{jn}(s_0,\omega))\}$ in (3.13) must be close to some expected value $E(f) = \int f(s)\mu^*(ds)$ for an invariant distribution μ^*. Of course, if there exists a unique invariant distribution $\mu^* = T^*\mu^*$ we then have

COROLLARY 3.11 *Assume that there exits a unique invariant distribution $\mu^* = T^*\mu^*$. Then under the conditions of Theorem 3.10,*

$$|E(f) - \frac{1}{N}\sum_{n=1}^{N} f(s_{jn}(s_0,\omega))| \leq 2\eta \tag{3.14}$$

for all s_0 and λ-almost all ω, and for all $N \geq N_j(\omega)$ with $j \geq J$.

Regarding this result, notice that each approximating function φ_j may contain multiple invariant distributions μ_j^*. In fact, this corollary is a good application of Theorem 3.8, since it would be quite restrictive to assume Hypothesis D or the existence of a unique invariant distribution for every approximating function φ_j.

Finally, let us consider the numerical approximation $\widehat{\varphi}$ of Theorem 3.7. For given s_0 and ω, let $\widehat{s}_{n+1}(s_0,\omega) = \widehat{\varphi}(\widehat{s}_n(s_0,\omega),\varepsilon_{n+1})$ for all $n \geq 0$.

COROLLARY 3.12 *Under the conditions of Theorem 3.7, for every $\eta > 0$ there exists a function $\widehat{N}(\omega)$ such that for all $N \geq \widehat{N}(\omega)$,*

$$|E(f) - \frac{1}{N}\sum_{n=1}^{N} f(\widehat{s}_n(s_0, \omega))| \leq \frac{L\delta}{1-\gamma} + \eta \tag{3.15}$$

for all s_0 and λ-almost all ω.

Finally, we would like to emphasize that all these results have been presented in a framework of numerical simulation, but they may be recasted under alternative approximations that can bear on issues related to convergence of parameter values, convergence of the distributions for the stochastic shock ε, or stochastic perturbations of the model. See Santos (2003) for an application of Theorems 3.2 and 3.8 to simulation-based estimation.

4 Examples

This section contains three illustrative examples. The first example highlights the role of the continuity condition of Assumption 2 in our main results. The remaining two examples are concerned with the derivation of error bounds in a simple model of economic growth. For many economic models Condition C is quite restrictive, but it has a nice application in the macroeconomics literature. Hence, the second example is concerned with the analysis of error bounds in the case that the model contains a closed form solution, and the third example extends this analysis to the more general case.

EXAMPLE 4.1: As one can observe from the proofs of Theorems 3.1-3.2 the continuity of $\varphi(s, \varepsilon)$ in s (Assumption 2) plays a fundamental role for the existence and upper semi-continuity of the invariant distributions correspondence. The following two illustrations enhance the further role of this continuity condition for the law of large numbers of Theorem 3.8. The failure of the law of large numbers entails that the invariant distributions no longer reflect the average behavior of a typical sample path.

Figure 4 depicts two real-valued functions g and $\widehat{g}$, which are discontinuous from the left at point $s = 1$. Moreover, $g(s) > \widehat{g}(s) > s$ for all $s < 1$, and $\lim_{s \to 1^-} g(s) = \lim_{s \to 1^-} \widehat{g}(s) = 1$; also, $s > g(s) > \widehat{g}(s)$ for all $s \geq 1$. Using the same randomization device of Example 2.1 we can then obtain a stochastic Markov process. But the resulting stochastic Markov process from the random application of these two functions fails to have an invariant distribution, since functions g and $\widehat{g}$ display a simple discontinuity at point $s = 1$. Furthermore, for every sample path $\{s_n\}$ generated by this random process

the partial sums $\{\frac{1}{N}\sum_{n=1}^{N} s_n\}$ converge to 1. This limit point cannot be the mean or expected value of an invariant distribution, since such distribution does not exist.

Now, following Figure 5, let us change the value of the functions g and $\widehat{g}$ so that $g(s) > \widehat{g}(s) > s$ for all $s \geq 1$. Then, the resulting random dynamical system generated by these new functions has a unique invariant distribution whose support is the interval $[1.9, 2]$. Note that as in Figure 4, for every initial value $s_0 < 1$ all the partial sums $\{\frac{1}{N}\sum_{n=1}^{N} s_n\}$ of the sample paths $\{s_n\}$ generated by the random dynamical system converge to point 1. Again, this limit point is not the mean value of an invariant distribution, since the support of the existing invariant distribution is the interval $[1.9, 2]$. Hence, this limit point is off the feasible bounds asserted in Theorem 3.8, and the failure of this theorem should be linked to the non-continuity of the stochastic process.

EXAMPLE 4.2: In this example we consider a simple version of the stochastic growth model of Brock and Mirman (1972):

$$\max E \sum_{t=0}^{\infty} \beta^t \log c_t \tag{P}$$

$$\text{s. t. } k_{t+1} = A\varepsilon_{t+1} k_t^{\alpha} + (1-\pi)k_t - c_t \tag{4.1}$$

$$k_0 \text{ fixed, } k_t \geq 0,\ t = 0, 1, \ldots$$

$$0 < \beta < 1,\ A > 0,\ 0 < \alpha < 1,\ 0 < \pi < 1$$

where E is the expectations operator, and $\{\varepsilon_t\}$ is an *iid* process drawn from a log-normal distribution with mean 0 and variance σ_{ε}^2.

The shock ε_{t+1} is realized at the beginning of each date for $t = 0, 1, \ldots$. Then, the decision problem is to allocate at each t the amounts of consumption c_t and capital for the next period k_{t+1} so as to satisfy feasibility constraint (4.1). As is well known, the set of optimal solutions for problem (P) has a Markovian representation. Hence, every optimal path $\{k_t\}$ can be generated by a stochastic difference equation $k_{t+1} = g(k_t, \varepsilon_{t+1})$ where g is a continuous function. Moreover, for $\pi = 1$ the optimal policy $k_{t+1} = g(k_{t+1}, \varepsilon_{t+1})$ takes the form

$$k_{t+1} = \alpha\beta A\varepsilon_{t+1} k_t^{\alpha}, \qquad t = 0, 1, \ldots \tag{4.2}$$

Now, taking logs in (4.2) we get

$$\log k_{t+1} = \log(\alpha\beta A) + \log \varepsilon_{t+1} + \alpha \log k_t. \tag{4.3}$$

Clearly, this transformed system satisfies Condition C. Moreover, it follows from this equation that k is log-normally distributed, and

$$E(\log k) = \frac{\log(\alpha\beta A)}{1-\alpha} \tag{4.4a}$$

$$Var(\log k) = \frac{\sigma_\varepsilon^2}{1-\alpha^2}. \tag{4.4b}$$

We now introduce the following simple perturbations to function g:

(a) A simple translation g_δ of the exact policy, $g_\delta(k_t, \varepsilon_{t+1}) = g(k_t, \varepsilon_{t+1}) + \delta$ with $\delta > 0$; hence, $k_{t+1} = g_\delta(k_t, \varepsilon_{t+1}) = \alpha\beta A\varepsilon_{t+1}k_t^\alpha + \delta$.

(b) A linear approximation g_L of the exact policy g around the deterministic steady state $k^* = g(k, 1)$; hence, $k_{t+1} = g_L(k_t, \varepsilon_{t+1}) = k^* + \alpha^2\beta A\varepsilon_{t+1}k_t^{*\alpha-1}(k_t - k^*)$ for $k^* = \sqrt[1-\alpha]{\alpha\beta A}$.

For the numerical experiments below we consider the following baseline parameterization,

$$\beta = 0.95, A = 10, \alpha = 0.34, \pi = 1, \sigma_\varepsilon = 0.008. \tag{4.5}$$

Whenever some of these parameter values are changed, normalizing constant A will be adjusted so that the deterministic steady state $k^* = g(k^*, 1)$ always remains at the benchmark value $k^* = 5.909$. Also, in order to confine the analysis to a compact domain of capitals, $[k_l, k_h]$, we restrict the set of possible values for ε to the interval $[e^{-4\sigma_\varepsilon}, e^{4\sigma_\varepsilon}]$. Accordingly, the density function of ε is rescaled so as to have a unit mass over this interval. The bounds k_l and k_h are computed from (4.2) as the fixed points $k_l = g(k_l, e^{-4\sigma_\varepsilon})$ and $k_h = g(k_h, e^{4\sigma_\varepsilon})$.

Error Bounds

In what follows, k refers to the values for the capital stock generated by the policy function (4.2), k_δ refers to those generated by function g_δ, and k_L refers to those generated by function g_L. For the computation of the sample moments we use a sequence of pseudo-random numbers for ε of length $N = 300,000$. The length of this path seems larger than what it is necessary in all cases.

According to Theorem 3.7 the approximation error of the sample statistics from a numerical approximation $\widehat{g}$ is up to a constant determined by $\frac{\|g-\widehat{g}\|}{1-\alpha}$, where $\|g - \widehat{g}\|$ is the approximation error of the numerical solution and α is the modulus of the random

contraction in Condition C. Note that by (4.3) above the modulus of contraction applies to the logarithmic transformation of the function. Hence, $\|\log(g) - \log(\widehat{g})\|$ seems the proper notion of distance for our calculations.

Table 1 refers to the perturbed solution $\widehat{g} = g_\delta$ for $\delta = 0.01$. This table illustrates how the error of the first and second order moments varies with changes in α. The second column considers the ratio $\frac{\|\log(g)-\log(\widehat{g})\|}{1-\alpha}$, and the third and fourth columns report the differences of simulated values generated under these functions for the mean $E(\log k)$ and the variance $Var(\log k)$. (The expression xE-y means x times 10^{-y}.) In the case of $E(\log k)$ the derivative of this function is always equal to 1, and so the Lipschitz constant in Theorem 3.7 is equal to 1. Therefore, the theory predicts that the entries in the second column should dominate those of the third column. As we can see in this table, the actual errors in the third column are always over one half of the theoretical estimates of the second column. Hence, for the mean values these upper error estimates are reasonably tight, and the observed values move consistently with those predicted by the theory. Regarding the fourth column of this table, after some simple calculations we obtain an upper bound equal to 0.2 for the Lipschitz constant of $Var(\log k)$ in Theorem 3.7. The corresponding errors reported in the fourth column, however, go down by three orders of magnitude. The fact that our upper estimates here are not so tight is in a way expected. Our theory provides upper estimates based upon a Lipschitz constant of the moment function, but the derivative of $Var(\log k)$ varies sharply with $\log k$. For instance, at the mean value the derivative of this function is equal to zero.

The same calculations were replicated for the linear approximation g_L. These numbers are reported in Table 2. Note that as α goes from 0.34 to 0.17 our theory predicts a decrease in the error of the mean by a factor of 3.9538, whereas the observed error went down by a factor of 10. Also an increase in α from 0.34 to 0.68 should lead to an increment in the error of the mean by a factor of 14.43, but the observed error just went up by a factor of 1.9. Therefore, our error bounds do not work as well for the linear approximation g_L. Again, this is to be expected since our theory is concerned with worst-case error bounds that may just be tight for uniform numerical approximations. For the linear approximation the maximum distance between functions g and g_L is attained at the tails, but a typical sample path would be fluctuating near the mean value. A further illustration of this point appears in Table 3. Here, we let $\sigma_\varepsilon = 0.08$ and compare the errors of the moments generated by the approximations g_δ and g_L. Although the error of the policy function g_L is about 150 times larger than that of g_δ, the errors of the moments generated by g_L are only about 10 times larger.

In conclusion, the upper bounds in Theorem 3.7 are the best possible. But these estimates cannot be expected to be tight in all circumstances. These estimates should be operative if the distance between the numerical approximation and the true solution is constant over all values of state variable k and if the Lipschitz constant of the moment

function is a tight upper bound for its derivative.

EXAMPLE 4.3: We now extend the analysis of the previous example for depreciation factors $0 < \pi < 1$. In this case, the model does not possess an analytical solution. Hence, we propose a simple operational way to estimate numerically the modulus of the random contraction in Condition C.

It should be clear that the model has a unique ergodic set, and for small values of σ_ε Condition C has to be satisfied. To bound the ergodic set we consider the deterministic steady states k_l and k_h corresponding to the worst realization $\varepsilon_t = e^{-4\sigma_\varepsilon}$ and the best realization $\varepsilon_t = e^{4\sigma_\varepsilon}$ for all $t \geq 1$. Thus, let $\zeta(k, \varepsilon)$ be the policy function of the deterministic model in which $\varepsilon_t = \varepsilon$ for all $t \geq 1$. Then,

$$k_l = \zeta(k_l, e^{-4\sigma_\varepsilon}) \tag{4.5a}$$

$$k_h = \zeta(k_h, e^{4\sigma_\varepsilon}). \tag{4.5b}$$

These steady states k_l and k_h can be calculated from the associated first-order conditions. This interval of values $[k_l, k_h]$ contains the original ergodic set.

Now, as above let $g(k, \varepsilon)$ be the policy function for Problem (P). By the mean-value theorem,

$$g(k, \varepsilon) - g(k', \varepsilon) = D_1 g(\widetilde{k}, \varepsilon) \cdot (k - k') \tag{4.6}$$

where $D_1 g(\widetilde{k}, \varepsilon)$ is the partial derivative of g with respect to k evaluated at $(\widetilde{k}, \varepsilon)$, for some point $\widetilde{k}$ in the segment (k, k'). Let

$$\begin{aligned} D_1 g(\widehat{k}, \widehat{\varepsilon}) &= \max D_1 g(k, \varepsilon) \\ \text{s. t.} \quad & \{k \in [k_l, k_h], \varepsilon \in [e^{4\sigma_\varepsilon}, e^{-4\sigma_\varepsilon}]\} \end{aligned} \tag{4.7}$$

Then, for any k in $[k_l, k_h]$ it follows from (4.6)-(4.7) that

$$\int |g(k, \varepsilon) - g(k', \varepsilon)| Q(d\varepsilon) \leq |D_1 g(\widehat{k}, \widehat{\varepsilon})||k - k'|. \tag{4.8}$$

Moreover, as already pointed out, $|D_1 g(\widehat{k}, \widehat{\varepsilon})| < 1$ for σ_ε small enough. To get an estimate for $D_1 g(\widehat{k}, \widehat{\varepsilon})$ we just calculate the maximum value of the derivative $D_1 \zeta(k, \varepsilon)$ over the set of points $\{(k_\varepsilon, \varepsilon) | k_\varepsilon = \zeta(k_\varepsilon, \varepsilon), \varepsilon \text{ in } [e^{-4\sigma_\varepsilon}, e^{4\sigma_\varepsilon}]\}$. This is the set of deterministic steady states $k_\varepsilon = \zeta(k_\varepsilon, \varepsilon)$ for all ε in $[e^{-4\sigma_\varepsilon}, e^{4\sigma_\varepsilon}]\}$. For each point the derivative $D_1 \zeta(k_\varepsilon, \varepsilon)$ can be calculated as the smallest eigenvalue of the linearization of the Euler equation evaluated at the steady state solution for the deterministic growth model in

which $\varepsilon_t = \varepsilon$ for all t. This computational procedure is relatively easy to implement, and should provide a good approximation $\widehat{\alpha}$ of the true value $D_1 g(\widehat{k}, \widehat{\varepsilon})$.

We now carry out several numerical experiments to evaluate the error of the sample moments using the baseline parameter values (4.5) with $\pi = 0.1$. As before, the scale parameter A is adjusted so as to keep the stationary solution $k^* = 5.909$ for $k^* = \zeta(k^*, \varepsilon)$ and $\varepsilon = 1$. For this parameterization we get $\widehat{\alpha} = 0.89$ as our estimate for the modulus of contraction $D_1 g(\widehat{k}, \widehat{\varepsilon})$. Since the model does not have an explicit solution, we use a PEA computational algorithm with Chebyshev polynomial interpolation and collocation along the lines of Christiano and Fisher (2000). Our finest grid uses 8 collocation points over $[k_l, k_h]$ and 5 collocation points over $[e^{-4\sigma_\varepsilon}, e^{4\sigma_\varepsilon}]$. Let $g_{PEA(8,5)}$ denote the computed policy function for the collocation pair (8, 5). The Euler equation residuals generated by this policy function are of order at most 10^{-9}, which seems a very good approximation to the true solution [cf. Santos (2000)].

Order of Convergence

Table 4 compares the simulated first and second order moments generated by numerical solution $g_{PEA(8,5)}$ with those generated by other solutions obtained from coarser grids. The column of the Euler residuals lists the maximum value for the Euler equation residuals generated by each of these policies. This value is of the same order of magnitude as the approximation error of the numerical solution. As one can see, the errors of the moments vary quite evenly with the Euler residuals, especially for the mean values. Hence, as predicted by our theory the error of the moments is of the same order of magnitude as the approximation error of the numerical solution.

To see more neatly the order of convergence of the moments, we also analyzed the following perturbation of the computed policy

$$\widehat{g}_\delta = g_{PEA(8,5)} + \delta \tag{4.9}$$

for $\delta > 0$. Figures 6 and 7 plot the simulated values for the ratios $\frac{|Ek_{PEA(8,5)} - E\widehat{k}_\delta|}{\delta}$ and $\frac{|Vark_{PEA(8,5)} - Var\widehat{k}_\delta|}{\delta}$ against δ, where $k_{PEA(8,5)}$ refers to the random process generated by the solution $g_{PEA(8,5)}$ and $\widehat{k}_\delta$ refers to the random process generated by function $\widehat{g}_\delta$ in (4.9). These figures cast no doubt that these ratios settle down as δ goes to 0. Therefore, for this type of perturbation the errors of the moments have the same convergence order as the approximation error of the numerical solution.

5 Hypothesis D

In his classical treatise, Doob (1953, Ch. V) proves a law of large numbers for Markov processes under a technical condition called Hypothesis D. In this section we present various useful regularity properties of invariant distributions implied by Hypothesis D. We also point out certain differences with respect to our approach, and analyze some important environments in which condition D is satisfied.

Let P^n be the n-step transition function. That is, $P^n(s,A) = \int P^{n-1}(t,A)P(s,dt)$ for $n \geq 2$ and $P^1(s,A) = P(s,A)$. Note that $P^n(s,A) = (T^{*n}\delta_s)(A)$, where δ_s is the Dirac measure at s.

HYPOTHESIS D: There exists a finite-valued measure γ on $\mathbb{S}$, an integer n, and a positive constant η, such that for all s,

$$P^n(s,A) \leq 1-\eta \quad \text{if } \gamma(A) \leq \eta. \tag{5.1}$$

THEOREM 5.1 [cf. Doob (1953, Ch. V)] *Suppose that Hypothesis D is satisfied. Then*

(i) There exists a finite number M of ergodic invariant probabilities μ_m^*, $m = 1,...,M$. Every other invariant probability μ^* is a convex combination of the ergodic invariant probabilities μ_m^*.

(ii) For every probability μ_0 on $\mathbb{S}$, the sequence $\{\frac{1}{N}\sum_{n=1}^N T^{*n}\mu_0\}$ converges to some invariant probability measure μ^*. The limit of each sequence $\{\frac{1}{N}\sum_{n=1}^N T^{*n}\mu_0\}$ may depend on μ_0, but the convergence is uniform to the set of invariant probability measures μ^*.

(iii) Let f be an element in $B(S)$. Then, for almost all s_0 and λ-almost all ω,

$$\lim_{N\longrightarrow\infty} \frac{1}{N}\sum_{n=1}^{N} f(s_n(s_0,\omega)) = \int f(s)\mu_m^*(ds) \tag{5.2}$$

for some ergodic invariant probability μ_m^*, $m = 1,...,M$.

Let us now explain these results. Hypothesis D is a generalization of the so called *Doeblin's condition*:

$$P(s,A) \geq \gamma(A) \tag{5.3}$$

for all $s \in S$ and $A \in \mathbb{S}$, for a finite-valued measure γ on $\mathbb{S}$. Note that under (5.3) the transition function P can have but a unique invariant distribution μ^*, whereas under

Hypothesis D there can be multiple invariant distributions. Both (5.1) and (5.3) are usually hard to check in applications. Some basic cases satisfying Hypothesis D are discussed below; for further economic applications making use of these conditions, see Stokey, Lucas and Prescott (1989, Ch. 11), Easley and Spulberg (1981), and Laitner (1988).

Since the supports E_m and E_n of two ergodic invariant probabilities μ_m^* and μ_n^* must be disjoint, part (i) of Theorem 5.1 is just a straightforward consequence of Hypothesis D. Indeed, (5.1) implies that if E_m is the support of an invariant probability μ_m^*, then $\gamma(E_m) > \eta$. Hence, the number M must be bounded by $\frac{\gamma(S)}{\eta}$. Note that if γ is an absolutely continuous measure, then the support E_m of every invariant probability μ_m^* must have a non-empty interior. Part (ii) of the theorem is another technical consequence of Hypothesis D. The convergence of $\{\frac{1}{N}\sum_{n=1}^{N} T^{*n}\mu_0\}$ holds in a strong sense. This convergence property implies that for every s_0 a generic sample path $\{s_n(s_0,\omega)\}$ will reach the center of some ergodic set E_m, for λ-almost all ω. The law of large numbers in (5.2) can now be easily deduced. Indeed, the ergodic theorem [e.g., Krengel (1985)] can be applied to each individual set E_m, for $m = 1, ..., M$; hence, for μ_m^*-almost all s_0 and λ-almost all ω the series $\{\frac{1}{N}\sum_{n=1}^{N} f(s_n(s_0,\omega))\}$ will converge to $\int f(s)\mu_m^*(ds)$. Moreover, every other point s_0 outside the set $\bigcup_{m=1}^{M} E_m$ will eventually fall into one of the ergodic sets E_m with probability one. Therefore, for every initial condition s_0 there is only a finite set of possible limit points $\int f(s)\mu_m^*(ds)$ in (5.2) corresponding to each of the ergodic invariant measures μ_m^*.

Hypothesis D is actually too strong for the derivation of the law of large numbers in (5.2). This hypothesis may be replaced by some weaker, local irreducibility conditions [e.g., see Meyn and Tweedie (1993) and Revuz (1975)]. These irreducibility conditions, however, are also hard to verify in economic applications. In contrast, the assumptions underlying our generalized law of large numbers in Theorem 3.8 hold true from regular primitive assumptions on economic models.

Under the conditions contemplated in Theorem 3.8 there could be an infinite number of ergodic invariant distributions μ^*; moreover, the sequence of distributions $\{\frac{1}{N}\sum_{n=1}^{N} T^{*n}\mu\}$ may not have a well defined limit in the weak topology of measures. Hence, there is no guarantee that a typical sequence $\{\frac{1}{N}\sum_{n=1}^{N} f(s_n(s_0,\omega))\}$ will converge. But the existence of a well defined limit as in (5.2) will at most improve marginally the content of our final convergence result in Theorem 3.10. The inequalities in Theorem 3.10 stem from the inequalities of both Corollary 3.5 and Theorem 3.8. Hence, unless one could control for sudden explosions of the set of invariant distributions as considered in Corollary 3.5, the bounds in Theorem 3.10 cannot be tightened. Moreover, Theorem 3.8 establishes tight upper and lower bounds for the set of limit points of all these sequences, and by the convexity of the set of invariant distributions (see Remark 3.4)

every limit point of the sequence $\{\frac{1}{N}\sum_{n=1}^{N} f(s_{jn}(s_0,\omega))\}$ in (3.13) must be arbitrarily close to some expected value $\int f(s)\mu^*(ds)$ for an invariant probability measure μ^*.

Although Hypothesis D may be hard to check in particular economic applications, we now show that this condition holds generically provided that one can enlarge adequately the noise space. Hence, this condition is satisfied by the transition probabilities P of an open and dense set of mappings φ over an enlarged space of random shocks. First, notice that if the transition probability P of an equilibrium function φ satisfies Hypothesis D, then this property will be preserved by the transition probability $\widehat{P}$ of a sufficiently good approximation $\widehat{\varphi}$ of the mapping φ [e.g., see Futia (1982, Sect. 4.3)]. Second, Doob (1953, p. 192) illustrates that if the transition probability $P(s,A)$ can be written as

$$P(s,A) = \int_A p(s,t)\gamma(dt) \quad \text{for all } s \in S \text{ and } A \in \mathbb{S} \tag{5.4}$$

where $p(s,t)$ is a bounded measurable function, then P satisfies Hypothesis D. Now, the following example considers an enlargement of the noise space such that the resulting transition probability P_a in this construction admits a representation of the form (5.4). Hence, the addition of an independent shock $\widetilde{\varepsilon}$ with a continuous positive density function over the whole space S will give rise to a transition probability P_a satisfying Hypothesis D. Further, in this example the sequence of mappings φ_a converge to φ as $a \to 0$. Therefore, Hypothesis D is satisfied by the probabilities P_a of a dense set of mappings φ_a over the enlarged noise space. Similar constructions are often encountered in the econometrics literature. See Rust (1994) for several possible interpretations of this additional error term in the estimation and testing of structural dynamic models. This enlargement of the shock process is not always justifiable on economic grounds.

EXAMPLE 5.1: We now add a non-singular random vector $\tilde{\varepsilon}$ to the original dynamical system φ so as to guarantee the existence of a finite number of ergodic invariant measures whose supports have non-empty interiors. Moreover, under this artificial enlargement of the noise space a law of large numbers along the lines of Theorem 5.1 can be established without the continuity assumption discussed in Example 4.1.

In order to perturb mapping φ, let us assume that the following regularity condition is satisfied

$$int(S) \neq \emptyset \text{ and } (clos[\varphi(S \times E)]) \bigcap (\partial S) = \emptyset \tag{5.5}$$

where ∂ denotes the boundary of S. The random vector $\tilde{\varepsilon}$ is now introduced in a rather formal way. Let $(G, \mathbb{G}, \nu)$ be a probability space, and let $\tilde{\varepsilon} : G \to S$ be a measurable function. Suppose that ν is such that the random vector $\tilde{\varepsilon}$ is uniformly distributed.

Specifically, if γ denotes the Lebesgue measure, suppose that a probability $\tilde{\gamma}$ is defined on $\mathbb{S}$ such that for every $A \in \mathbb{S}$,

$$\tilde{\gamma}(A) = \nu(\tilde{\varepsilon}^{-1}(A)) = \frac{\gamma(A)}{\gamma(S)}. \tag{5.6}$$

Let

$$\varphi_a(s, \varepsilon, \tilde{\varepsilon}) = \varphi(s, \varepsilon) + a\tilde{\varepsilon} \tag{5.7}$$

where a is a multiplicative constant. Notice that by (5.5) the mapping $\varphi_a : S \times E \times G \to S$ is well defined for a small enough. Then, let $\{(\varepsilon_t, \tilde{\varepsilon}_t)\}$ be the bivariate *iid* process generated under the product measure $Q \times \tilde{\gamma}$. As in (2.3), let

$$P_a(s, A) = Q \times \tilde{\gamma}(\{(\varepsilon, \tilde{\varepsilon}) | \varphi_a(s, \varepsilon, \tilde{\varepsilon}) \in A\}). \tag{5.8}$$

Now, since the shock $\tilde{\varepsilon}$ has a continuous density and it is independent of ε it follows that transition function P_a can be written as in (5.4) for all $s \in S$ and $A \in \mathbb{S}$. Hence, P_a satisfies Hypothesis D. Then, by Theorem 5.1, there is a finite number M_a of ergodic invariant probabilities μ^*_{am} for P_a with non-overlapping supports $\sigma(\mu^*_{am})$. Any other invariant probability μ^*_a for P_a must be a convex combination of the ergodic invariant probabilities μ^*_{am}, for $m = 1, ..., M_a$. The support $\sigma(\mu^*_a)$ of each invariant probability μ^*_a must have non-empty interior, since random vector $\tilde{\varepsilon}$ is uniformly distributed over S, and independent of ε. Moreover, letting $a \to 0$, by Theorem 3.2 these ergodic invariant probabilities $\{\mu^*_{am}\}$ must approach the set of invariant probability measures $\{\mu^* | \mu^* = T^*\mu^*\}$.

Finally, it should be clear that Hypothesis D does not require the vector of shocks ε to be of at least the same dimension as the vector of state variables s, since (5.1) may be satisfied for some n-step transition probability. Here we provide another application in which condition (5.1) can be verified by invoking an inverse function theorem argument over a sequence of shocks ε.

EXAMPLE 5.2: Consider the following dynamical system

$$\begin{aligned} k_{t+1} &= g(k_t, z_t) \\ z_{t+1} &= \Psi(z_t, \varepsilon_{t+1}) \end{aligned} \tag{5.9}$$

where k, z are non-negative numbers that lie in the compact set $K \times Z$ and ε_t is an *iid* random variable with an absolutely continuous distribution over some interval $[\underline{\varepsilon}, \overline{\varepsilon}]$. Now, assume that functions g and Ψ are continuously differentiable, and that each of

their partial derivatives is always different from zero. Then, we claim that the 2-step transition probability P^2 satisfies (5.1). Indeed, we may write (5.9) as

$$\begin{aligned} k_{t+1} &= g(g(k_{t-1}, z_{t-1}), \Psi(z_{t-1}, \varepsilon_t)) \\ z_{t+1} &= \Psi(\Psi(z_{t-1}, \varepsilon_t), \varepsilon_{t+1}). \end{aligned} \tag{5.10}$$

Now, differentiating (5.10) with respect to $(\varepsilon_t, \varepsilon_{t+1})$, one can check that the matrix of partial derivatives is non-singular. Hence, by an inverse function argument one can see that for each (k_{t-1}, z_{t-1}) every small neighborhood of $(\varepsilon_t, \varepsilon_{t+1})$ is mapped into a neighborhood of some (k_{t+1}, z_{t+1}). Furthermore, by a compactness argument there are positive constants δ and δ' such that for any given (k_{t-1}, z_{t-1}) and all $(\varepsilon_t, \varepsilon_{t+1})$ every ball $B_\delta(\varepsilon_t, \varepsilon_{t+1})$ with center $(\varepsilon_t, \varepsilon_{t+1})$ and radius δ will be mapped by (5.10) onto a neighborhood containing a ball $B_{\delta'}(k_{t+1}, z_{t+1})$ with center (k_{t+1}, z_{t+1}) and radius δ'. From this injective uniform lower bound, it is now easy to show that Hypothesis D must be satisfied for the 2-step transition function P^2.

To summarize, under Hypothesis D the ergodic theorem can be extended to encompass all initial conditions s_0 outside the ergodic sets. This is a very useful result since in general it is not possible to locate these ergodic sets. In economic models, however, Hypothesis D is hard to verify from primitive conditions. Hence, our law of large numbers in Theorem 3.8 has wider applicability. Examples 5.1 and 5.2 illustrate nevertheless two important applications covered by Hypothesis D that often arise in the economics literature.

6 Concluding Remarks

As argued by Lucas (1980), theoretical economics should provide fully articulated, artificial economic systems that can be simulated and contrasted with available data sets. For reasons of mathematical tractability and further considerations, most economic models are not conceived as computer programs that can be realized in a finite number of instructions. These models are characterized by non-linear functions and correspondences, and are discretized by numerical methods. Therefore, typically a researcher simulates a numerical approximation in order to learn about the behavior of an economic model.

In this paper we have delved into the theoretical foundations of numerical simulation for stochastic dynamic models. One issue of major concern in this investigation is that some of the available theory is grounded on assumptions which are not readily verifiable for economic models. More specifically, existing results on continuity properties of the correspondence of invariant distributions are not suitable for numerical approximations as they are formulated under conditions on the transition probability P; error bounds

concerning perturbations of stochastic contractions require that the approximating function $\widehat{\varphi}$ must also satisfy the contractivity condition; and laws of large numbers build along the lines of Hypothesis D or the assumption of a unique invariant distribution. All these conditions are difficult to check in economic models, and may not be preserved for numerical approximations or under further stochastic perturbations of the original system.

Our analysis builds on the assumptions of a compact domain and continuity of the equilibrium solution. These hypotheses are usually validated from primitive conditions of economic models, and hold true for most numerical approximation schemes. The assumption of a compact domain can be weakened [e.g., see Billingsley (1968) and Futia (1982)]. The continuity of the equilibrium solution is a more delicate assumption,[3] and can only be dispensed at the cost of some other specific conditions [e.g., Hopenhayn and Prescott (1992)]. As discussed in Example 4.1 above, the continuity of the equilibrium function plays an essential role for the existence of an invariant distribution, the upper semicontinuity of the correspondence of invariant distributions, and our generalized version of the law of large numbers. Therefore, a main message of the present paper is that the continuity of the equilibrium function is of fundamental importance to guarantee convergence of the simulated moments from numerical solutions to the moments of the model's invariant distributions as the approximation errors of these solutions converge to zero. These asymptotic results are a first step to develop theories of comparative analyses and error bounds for approximate solutions in stochastic dynamic models. Error bounds for the moments of the invariant distributions were obtained under the assumption of a random contraction. This assumption may be difficult to check in some economic applications, but it seems plausible to establish related error bounds for differentiable perturbations of the model. However, under the postulated continuity property of the dynamical system our asymptotic results appear to be optimal.

Continuity properties of invariant distributions are usually assumed in estimation-based simulation and testing of dynamic economic models in which one generally considers a family of models $\varphi(s, \varepsilon, \theta)$ indexed by a vector of parameters θ in a space Θ. Then, sequences of simulated moments $\{\frac{1}{N}\sum_{n=1}^{N} f(s_n(s_0, \omega, \theta))\}$ are generated so as to approximate the expectations operator $E_\theta(f)$. Hence, an important property is the continuity of operator $E_\theta(f)$ in θ. This continuity property follows from our results above (cf., Corollary 3.5), and plays a central role in proofs of consistency of simulated estimators [e.g., Duffie and Singleton (1993) and Santos (2003)] to establish the uniform convergence of the sequence of simulated moments $\{\frac{1}{N}\sum_{n=1}^{N} f(s_n(s_0, \omega, \theta))\}$ to $E_\theta(f)$ in θ.

[3]Continuity of the equilibrium function may cease to hold in non-convex optimization problems or in competitive convex economies with incomplete markets [Krebs (2002)] or with taxes, externalities, and money [Santos (2002)].

Finally, we would like to conclude with a brief discussion of an important and controversial issue regarding the simulation of stochastic dynamic models. It seems to be a common practice in macroeconomics to compute the distribution of the simulated moments from a large number of sample paths of length equal to that of the data sample. Thus, if the length of the data sample is N the simulation exercise would proceed by producing a large number of the model's sample paths of the same length N. Then, the simulated moments computed over each of these sample paths are compared with those of the data sample. Our analysis, however, seems to suggest that a proper way to simulate a dynamic model would be produce one single sample path of arbitrarily large length. This is because by the law of large numbers the simulated moments would be approaching generically the true moments as the length of the sample path gets large. Of course, note that once a model has been specified in our framework the only uncertainties in the computation of the moments of an invariant distribution stem from the approximation and truncation errors of the numerical solution. Therefore, our analysis offers no justification for replicating a large number of the model's sample paths –each of the same length N as that of the data sample. These sample paths may be heavily influenced by the choice of the initial values s_0.

7 Appendix

Proof of Theorem 3.2: For an associated triple (φ, T, T^*) and a probability μ, let $\varphi \cdot \mu$ stand for $T^*\mu$. Then, following Dubins and Freedman (1966, p. 239) the theorem will be established if we can show the continuity of the evaluation map $ev(\varphi, \mu) = \varphi \cdot \mu$. Recall that the space of mappings φ is endowed with the metric topology induced by the norm $\|\varphi\|$ and the space of probability measures is endowed with the topology of weak convergence. Note that the topology of weak convergence is metrizable [cf. Billingsley (1968)]. As is well known [e.g., see Shiryaev (1996)], the following metric is compatible with this topology:

$$d(\mu, \nu) = \sup_{f \in \mathcal{A}} \{ | \int f(s)\mu(ds) - \int f(s)\nu(ds) | \} \tag{7.1}$$

where $\mathcal{A}$ is the space of Lipschitz functions on S with constant $L \leq 1$ and such that $\|f\| \leq 1$.

Let f belong to $\mathcal{A}$. Then, for any two mappings φ and $\widehat{\varphi}$, and any two measures μ

and ν, we have

$$\begin{aligned}
&|\int f(s)[\varphi \cdot \mu(ds)] - \int f(s)[\widehat{\varphi} \cdot \nu(ds)]| \\
&= |\int [\int f(\varphi(s,\varepsilon))Q(d\varepsilon)]\mu(ds) - \int [\int f(\widehat{\varphi}(s,\varepsilon))Q(d\varepsilon)]\nu(ds)| \\
&\leq |\int [\int f(\varphi(s,\varepsilon))Q(d\varepsilon)]\mu(ds) - \int [\int f(\varphi(s,\varepsilon))Q(d\varepsilon)]\nu(ds)| \\
&+|\int [\int f(\varphi(s,\varepsilon))Q(d\varepsilon)]\nu(ds) - \int [\int f(\widehat{\varphi}(s,\varepsilon))Q(d\varepsilon)]\nu(ds)| \\
&\leq |\int [\int f(\varphi(s,\varepsilon))Q(d\varepsilon)][\mu(ds) - \nu(ds)] + \|\varphi - \widehat{\varphi}\|.
\end{aligned}$$

Note that the first inequality comes from the triangle inequality, and after some simple arrangements the second inequality comes from the definition of the norm $\|\varphi - \widehat{\varphi}\|$ for f in $\mathcal{A}$.

Then, by (7.1) the theorem will be established if we can show that for every arbitrary $\eta > 0$ there exists a weak neighborhood $V(\mu)$ of μ such that for all ν in $V(\mu)$ and all f in $\mathcal{A}$,

$$|\int [\int f(\varphi(s,\varepsilon))Q(d\varepsilon)][\mu(ds) - \nu(ds)]| < \eta. \tag{7.2}$$

By the Arzela-Ascoli theorem, the set $\mathcal{A}$ is compact. Hence, we can find a finite set of elements $\{f^j\}$ such that for every f in $\mathcal{A}$ there exists an element f^j so that $\|f - f^j\| < \frac{\eta}{3}$. Also, by Assumption 2 the mapping $\int f(\varphi(s,\varepsilon))Q(d\varepsilon)$ is continuous in s. Hence, for every f^j there exists a weak neighborhood $V_j(\mu)$ such that for all ν in $V_j(\mu)$,

$$|\int [\int f^j(\varphi(s,\varepsilon))Q(d\varepsilon)][\mu(ds) - \nu(ds)]| < \frac{\eta}{3}.$$

Therefore, (7.2) must hold for all f with $\|f - f^j\| < \frac{\eta}{3}$. Finally, let $V(\mu) = \bigcap_j V_j(\mu)$. Then, (7.2) must hold for every ν in $V(\mu)$ and all f in $\mathcal{A}$. The proof is complete.

Proof of Theorem 3.7: For an initial point s_0, let $Z^n(s_0)$ denote the random vector

$\varphi(\varphi\cdots(\varphi(s_0,\varepsilon_1),\varepsilon_2)\cdots\varepsilon_n)$ and let $\widehat{Z}^n(s_0)$ denote $\widehat{\varphi}(\widehat{\varphi}\cdots(\widehat{\varphi}(s_0,\varepsilon_1),\varepsilon_2)\cdots\varepsilon_n)$. Then,

$$\begin{aligned}
|E[f(Z^n(s_0))] - E[f(\widehat{Z}^n(s_0))]| = \\
|E[f(\varphi(Z^{n-1}(s_0),\varepsilon_n))] - E[f(\widehat{\varphi}(\widehat{Z}^{n-1}(s_0),\varepsilon_n))]| \leq \\
|E[f(\varphi(Z^{n-1}(s_0),\varepsilon_n))] - E[f(\varphi(\widehat{Z}^{n-1}(s_0),\varepsilon_n))]| + \\
|E[f(\varphi(\widehat{Z}^{n-1}(s_0),\varepsilon_n))] - E[f(\widehat{\varphi}(\widehat{Z}^{n-1}(s_0),\varepsilon_n))]| = \\
|E[E[f(\varphi(Z^{n-1}(s_0),\varepsilon_n)) - f(\varphi(\widehat{Z}^{n-1}(s_0),\varepsilon_n))]|\varepsilon_n]| + \\
|E[f(\varphi(\widehat{Z}^{n-1}(s_0),\varepsilon_n))] - E[f(\widehat{\varphi}(\widehat{Z}^{n-1}(s_0),\varepsilon_n))]| \leq \\
L\alpha E\|Z^{n-1}(s_0) - \widehat{Z}^{n-1}(s_0)\| + L\delta.
\end{aligned}$$

Observe that the first inequality comes from the triangle inequality. The second equality results by first conditioning on ε_n, and then by an application of the law of iterated expectations. And the last inequality follows from the assumptions of the theorem. Now, by a similar argument we get

$$L\alpha E\|Z^{n-1}(s_0) - \widehat{Z}^{n-1}(s_0)\| \leq$$

$$L\alpha^2 E\|Z^{n-2}(s_0) - \widehat{Z}^{n-2}(s_0)\| + L\alpha\delta.$$

Hence, combining these inequalities and proceeding inductively it follows that

$$|E[f(Z^n(s_0))] - E[f(\widehat{Z}^n(s_0))]| \leq \frac{L\delta}{1-\alpha} \text{ for all } n \geq 1. \tag{7.3}$$

Now, consider an invariant distribution $\widehat{\mu}^*$ of mapping $\widehat{\varphi}$. Then,

$$|\int E[f(Z^n(s_0))]\widehat{\mu}^*(ds_0) - \int f(s)\widehat{\mu}^*(ds_0)| =$$

$$|\int E[f(Z^n(s_0))]\widehat{\mu}^*(ds_0) - \int E[f(\widehat{Z}^n(s_0))]\widehat{\mu}^*(ds_0)| \leq \frac{L\delta}{1-\alpha} \text{ for all } n \geq 1. \tag{7.4}$$

Observe that the equality comes from the fact that $\widehat{\mu}^*$ is an invariant distribution under φ; and the inequality in (7.4) is a consequence of (7.3). As this inequality holds true for all $n \geq 1$, and $Ef(Z^n(s_0)) = \int f(s)\mu^n_{s_0}(ds)$, by Theorem 3.6 we must have

$$|\int f(s)\mu^*(ds) - \int f(s)\widehat{\mu}^*(ds)| \leq \frac{L\delta}{1-\alpha}. \tag{7.5}$$

Therefore, (7.5) must be satisfied for every invariant distribution $\widehat{\mu}^*$ under $\widehat{\varphi}$. This concludes the argument.

Proof of Theorem 3.8: For the proof of this theorem, it is convenient to let time n range from $-\infty$ to ∞ so as to consider sequences of the form $(\ldots, \varepsilon_{-n}, \ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n, \ldots)$. Using the construction in (3.9), we can then define a probability space $(\widehat{\Omega}, \widehat{\mathbb{F}}, \widehat{\lambda})$ over these doubly infinite sequences. Also, we shall view $\mathbb{F}$ as σ-subfield of $\widehat{\mathbb{F}}$. For each integer J we define the J-shift operator ϑ_J: $\widehat{\Omega} \rightarrow \widehat{\Omega}$, as $\vartheta_J(\ldots, \varepsilon_{-n}, \ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n, \ldots) = (\ldots, \varepsilon_{-n+J}, \ldots, \varepsilon_{-1+J}, \varepsilon_J, \varepsilon_{1+J}, \ldots, \varepsilon_{n+J}, \ldots)$. Note that the mapping ϑ_J is bijective and measurable. Hence, $(\widehat{\Omega}, \widehat{F}, \widehat{\lambda}, \vartheta_J)$ is an ergodic system.

Let

$$F_N(\omega) = max_{s_0 \in S}(\sum_{n=1}^{N} f(s_n(s_0, \omega))). \tag{7.6}$$

Then, the following inequality must be satisfied,

$$F_{N+J}(\omega) \leq F_N(\vartheta_J(\omega)) + F_J(\omega)$$

for all positive integers N and J. Hence, by the subadditive ergodic theorem of Kingman (1968) there exists a constant F such that for λ-almost all ω,

$$\lim_{N \rightarrow \infty} \frac{F_N(\omega)}{N} = F. \tag{7.7}$$

Under these conditions, Crauel (1991, and 2002 pp. 96-97) shows existence of a φ-invariant probability measure $\widehat{\nu}$ on the product space $\mathbb{S} \times \widehat{\mathbb{F}}$ such that

$$\int f(s)\widehat{\nu}(ds, d\omega) = \int [\int f(s)\widehat{\nu}_\omega(ds)]\widehat{\lambda}(d\omega) = F \tag{7.8}$$

where $\widehat{\nu}_\omega(ds)\widehat{\lambda}(d\omega)$ denotes the disintegration of $\nu(ds, d\omega)$. Furthermore, by (7.7) and an iterated application of the ergodic theorem to the measure $\widehat{\nu}$ in (7.8) we get that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^{N} f(s_n(s_0, \omega)) = F \tag{7.9}$$

for $\widehat{\nu}$-almost all (s_0, ω).

For our purposes, the problem with Crauel's argument is that the φ-invariant probability measure $\widehat{\nu}$ in (7.8) may not be P-invariant [cf., Arnold (1998, p. 56)]. Hence, to complete the argument we need to show that there exists a φ-invariant probability measure satisfying (7.8) that can be expressed as a product measure.

Let $\nu^* = E(\widehat{\nu}|\mathbb{S} \times \mathbb{F})$ be the conditional expectations of $\widehat{\nu}$ over our original product space $\mathbb{S} \times \mathbb{F}$. Then, as illustrated in Arnold (1998, p. 38) this measure ν^* is φ-invariant. Furthermore, since $\{\varepsilon_n\}$ is an *iid* process the measure ν^* can be expressed as $\nu^* = \mu^* \times \lambda$ where μ^* is a P-invariant probability measure (*viz. op. cit.* Corollary 1.7.6.). Moreover, for any continuous function f on S,

$$\int_{S\times\Omega} f(s)\widehat{\nu}(ds, d\omega) = \int_{S\times\Omega} f(s)\nu^*(ds, d\omega) = \int_S f(s)\mu^*(ds).$$

Hence, (7.7)-(7.8) entail that $E^{max}(f) = F$. Therefore, (3.10) in Theorem 3.8 follows now from (7.6)-(7.7). After minor changes, the same argument will prove (3.11). The theorem is thus established.

References

Arnold, L. (1998), *Random Dynamical Systems*, Springer-Verlag, Berlin.

Aruoba, S. B., Fernandez-Villaverde, J. and J. F. Rubio-Ramirez (2003), "Comparing solution methods for dynamic equilibrium economies," manuscript, Federal Reserve Bank of Atlanta.

Bhattacharya, R. and M. Majumdar (2003), "Random dynamical systems: A review," forthcoming in *Economic Theory*.

Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley & Sons, New York.

Breiman, L. (1960), "The strong law of large numbers for a class of Markov chains," *Annals of Mathematical Statistics*, 31: 801-803.

Brock, W. A. and L. J. Mirman (1972), "Optimal economic growth and uncertainty: The discounted case," *Journal of Economic Theory*, 4: 479-513.

Christiano, L. J. and M. Eichenbaum (1992), "Current real business-cycle theories and aggregate labor-market fluctuations," *American Economic Review*, 82: 430-450.

Christiano, L. J. and J. D. M. Fisher (2000), "Algorithms for solving dynamic models with occasionally binding constraints," *Journal of Economic Dynamics and Control*, 24: 1179-1232.

Cooley, T. F. and E. C. Prescott (1995), "Economic growth and business cycles," in Cooley, ed., *Frontiers of Business Cycles Research*, Princeton University Press, Princeton, NJ.

Crauel, H. (1991), "Lyapunov exponents of random dynamical systems on Grassmanians," in L. Arnold, H. Crauel and J.-P. Eckmann, eds., *Lyapunov Exponents. Proceedings, Oberwolfach 1990*, Springer Lecture Notes in Mathematics vol. 1486, Springer-Verlag, New York.

Crauel, H. (2002), *Random Probability Measures on Polish Spaces*, Taylor & Francis, New York.

Den Haan, W. and A. Marcet (1994), "Accuracy in simulations," *Review of Economic Studies*, 61: 3-17.

Doob, J. L. (1953), *Stochastic Processes*, John Wiley & Sons, New York.

Dubins, L. E. and D. Freedman (1966), "Invariant probability measures for certain Markov processes," *Annals of Mathematical Statistics*, 37: 837-848.

Duffie, D. and K. Singleton (1993), "Simulated moments estimation of Markov models of asset prices," *Econometrica*, 61: 929-952.

Easley, D. and D. F. Spulberg (1981), "Stochastic equilibrium and optimality with rolling plans," *International Economic Review*, 22: 79-103.

Ellison G. and D. Fudenberg (1993), "Rules of thumb for social learning," *Journal of Political Economy*, 101: 612-643.

Foley, D. and M. Hellwig (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.

Hopenhayn, H. A. and E. C. Prescott (1992), "Stochastic monotonicity and stationary distributions for dynamic economics," *Econometrica*, 60: 1387-1406.

Judd, K. L. (1998), *Numerical Methods in Economics*, MIT Press, London.

Kingman, J. F. C. (1968), "The ergodic theory of subadditive stochastic processes," *Journal of the Royal Statistical Society: Series B*, 30: 499-510.

Krebs, T. (2002), "Non-existence of recursive equilibria on compact state spaces when markets are incomplete," manuscript, Brown University.

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

Laitner, J. (1988), "Bequests, gifts, and social security," *Review of Economic Studies*, 55: 275-299.

Lucas, R. E. (1980), "Methods and problems in business cycle theory," *Journal of Money, Credit and Banking*, 12: 696-715.

Manuelli, R. (1985), "A note on the behavior of solutions to dynamic stochastic models," mimeo, University of Wisconsin, Madison.

Meyn, S. P. and R. L. Tweedie (1993), *Markov Chains and Stochastic Stability*, Springer, New York.

Norman, F. M. (1972), *Markov Processes and Learning Models*, Academic Press, New York.

Revuz, D. (1975), *Markov Chains*, North-Holland, Amsterdam.

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.

Sanghvi, A. P. and M. J. Sobel (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. 1, Elsevier, Amsterdam.

Santos, M. S. (2000), "Accuracy of numerical solutions using the Euler equation residuals," *Econometrica*, 68: 1377-1402.

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," forthcoming in *Journal of Computational and Applied Mathematics*.

Schenk-Hoppé, K. R. and B. Schmalfuss (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.

Shiryaev, A. N. (1996), *Probability*, Springer-Verlag, New York.

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. and E. C. Prescott (1989), *Recursive Methods in Economic Dynamics*, Harvard University Press, Cambridge, MA.

Taylor, J. B. and H. Uhlig (1990), "Solving nonlinear stochastic growth models: A comparison of alternative solution methods," *Journal of Business and Economic Statistics*, 8: 1-18.

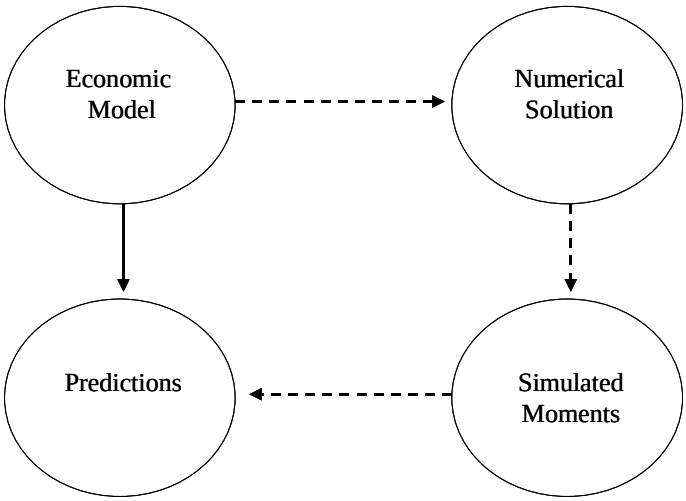

Figure 1: A framework for numerical simulation

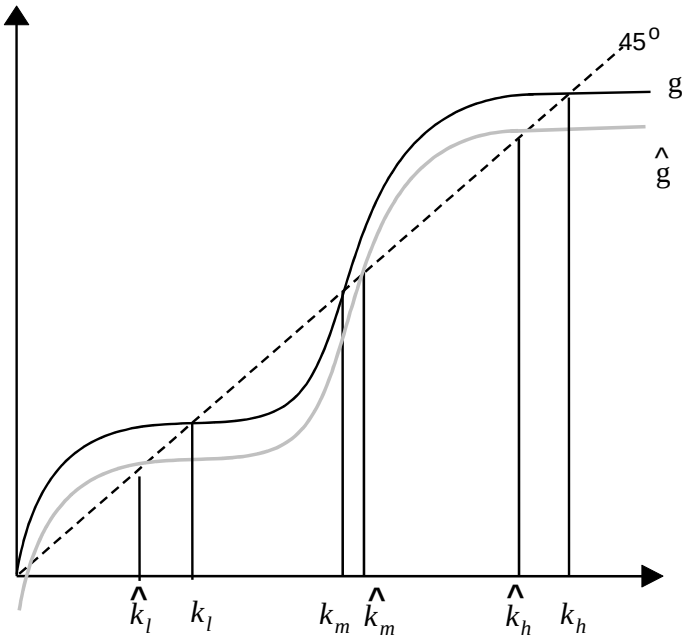

Figure 2: A numerical approximation

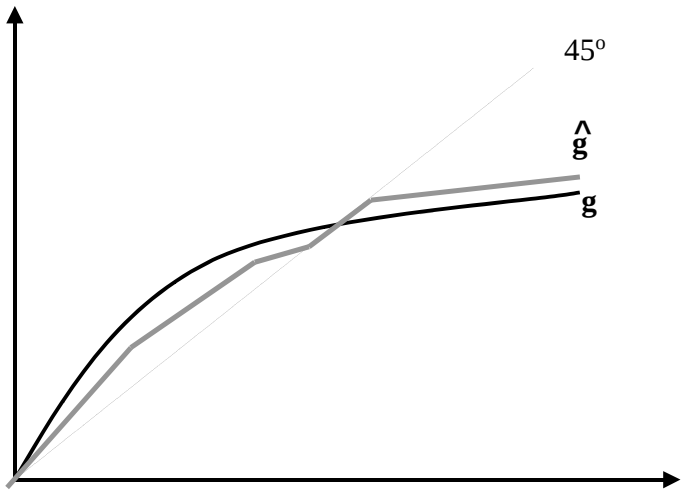

Figure 3: Multiplicity of stationary solutions for the numerical approximation

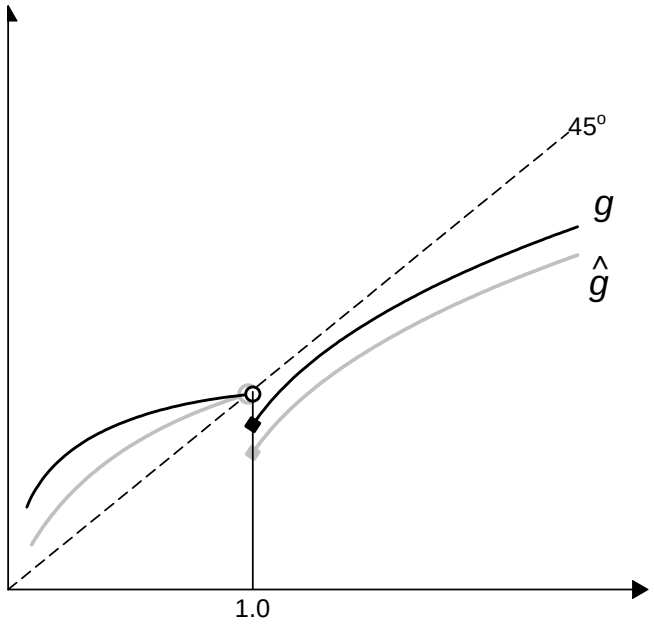

Figure 4: Non-existence of an invariant distribution

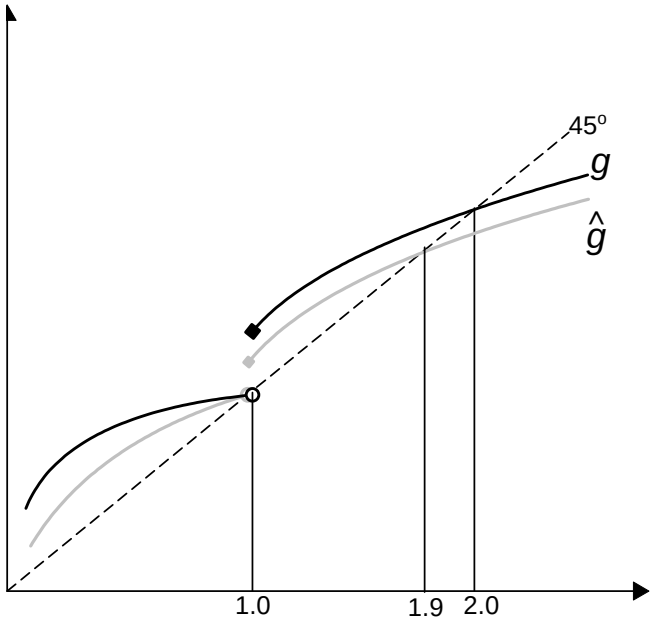

Figure 5: Failure of the law of large numbers

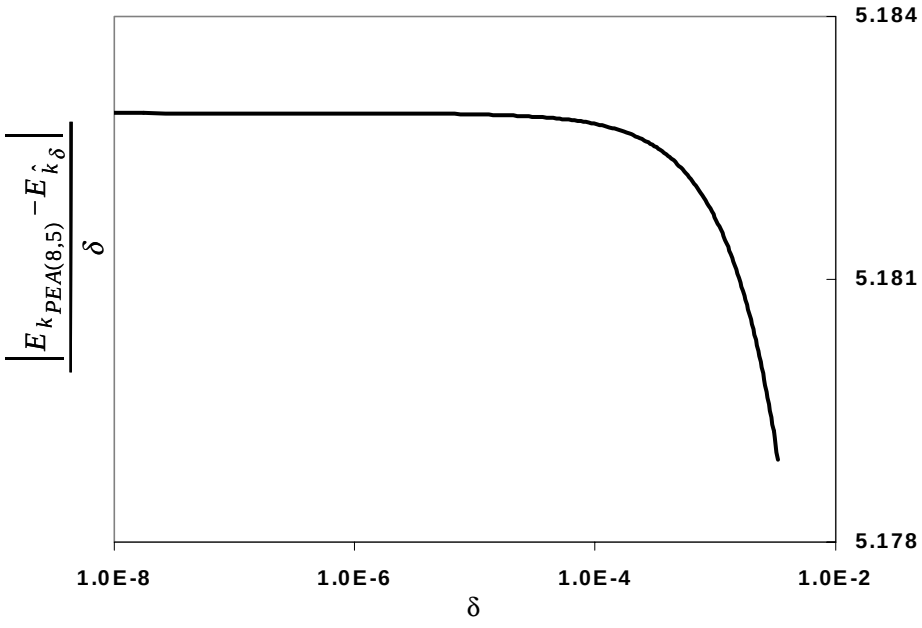

Figure 6: Convergence of the simulated mean

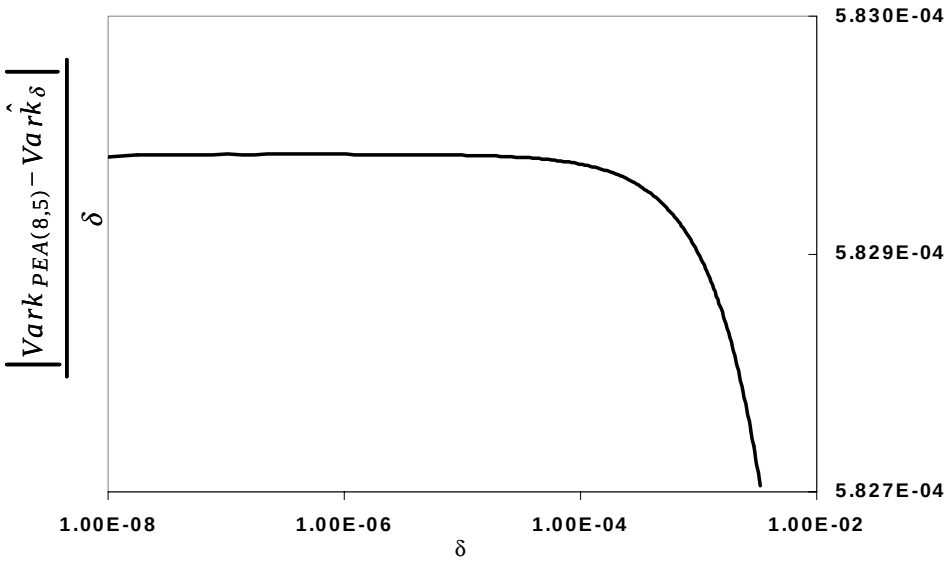

Figure 7: Convergence of the simulated variance

	$\frac{\\|\log g - \log g_\delta\\|}{1-\alpha}$	$\|E(\log k) - E(\log k_\delta)\|$	$\|Var(\log k) - Var(\log k_\delta)\|$
$\alpha = 0.34$	4.7730E-3	2.5716E-3	7.9716E-6
$\alpha = 0.17$	3.7198E-3	2.0460E-3	1.6230E-6
$\alpha = 0.68$	1.9134E-2	5.2894E-3	5.4532E-5

Table 1: Approximation errors for the simulated moments generated by function g_δ, $\delta = 0.01$

	$\frac{\\|\log g - \log g_L\\|}{1-\alpha}$	$\|E(\log k) - E(\log k_L)\|$	$\|Var(\log k) - Var(\log k_L)\|$
$\alpha = 0.34$	1.2145E-3	3.6033E-4	1.0159E-8
$\alpha = 0.17$	7.6283E-4	3.2839E-5	8.5709E-9
$\alpha = 0.68$	8.6107E-3	5.9152E-5	2.0313E-8

Table 2: Approximation errors for the simulated moments generated by function g_L

	$\|\|\log g - \log \hat{g}\|\|$	$\|E(\log k) - E(\log \widehat{k})\|$	$\|Var(\log k) - Var(\log \widehat{k})\|$
$\widehat{k} = k_\delta$	2.7444E-3	4.4762E-4	7.2232E-6
$\widehat{k} = k_L$	3.5350E-1	3.4643E-3	1.2847E-4

Table 3: Approximation errors for the simulated moments generated by functions g_δ and g_L

	Euler Residual	$\left\|E(k_{PEA(8,5)})\text{-}E(k_{PEA})\right\|$	$\|Var(k_{PEA(8,5)})\text{-}Var(k_{PEA})\|$
$k_{PEA(2,2)}$	4.0524E-2	2.7873E-2	9.6762E-6
$k_{PEA(4,4)}$	4.1722E-4	1.2821E-3	8.1683E-7
$k_{PEA(5,5)}$	3.2863E-5	3.1803E-5	9.4821E-8

Table 4: Approximation errors for the simulated moments generated by functions g_{PEA}