Discussion Paper Series

# RIEB

Kobe University

DP2014-38

# Seeking Ergodicity in Dynamic Economies

Takashi KAMIHIGASHI
John STACHURSKI

November 21, 2014

Research Institute for Economics and Business Administration
Kobe University
2-1 Rokkodai, Nada, Kobe 657-8501 JAPAN

# Seeking Ergodicity in Dynamic Economies[1]

Takashi Kamihigashi[a,b] and John Stachurski[c]

[a]Research Institute for Economics and Business Administration, Kobe University

[b]IPAG Business School, Paris, France

[c]Research School of Economics, Australian National University

November 19, 2014

ABSTRACT. In estimation and calibration studies the convergence of time series sample averages plays a central role. At the same time, a significant number of economic models do not satisfy the classical ergodicity conditions. Motivated by existing work on asymptotics of stochastic economic models, we develop a new set of results on limits of sample moments and other sample averages using an order-theoretic approach. Our results include a condition that is necessary and sufficient for convergence over a broad class of moment functions. We discuss implications, sufficient conditions and a range of economic applications.

*JEL Classifications:* C62, C63
*Keywords:* Ergodicity, monotonicity, calibration

## 1. INTRODUCTION

It has frequently been observed that many of the economic models used for quantitative studies fail to satisfy standard stability conditions from the classical Markov process literature (see, e.g., Stokey and Lucas (1989), chapter 12). This situation has spurred the growth of an alternative approach to treating asymptotics of economic models based around order theoretic notions. Well known contributions include Razin and Yahav (1979), Stokey and Lucas (1989), Bhattacharya and Lee

[1]This paper has benefited from helpful comments by David Backus, Anmol Bhandari, Tom Sargent and Venky Venkateswaran, as well as participants at the NYU Stern School of Business macroeconomics seminar. We acknowledge financial support from the Japan Society for the Promotion of Science and Australian Research Council Discovery Grant DP120100321.

*Email addresses:* tkamihig@rieb.kobe-u.ac.jp, john.stachurski@anu.edu.au

(1988), Hopenhayn and Prescott (1992) and Bhattacharya and Majumdar (2001). More recent work can be found in Zhang (2007), Szeidl (2013) and Kamihigashi and Stachurski (2014). Applications of these ideas can be found throughout economics. Representative examples include Huggett (1993), De Hek (1999), Aghion and Bolton (1997), Piketty (1997), Owen and Weil (1998), Cabrales and Hopenhayn (1997), Cooley and Quadrini (2001) and Morand and Reffett (2007).

Almost all of this literature has focused on distributional properties, by which we mean existence, uniqueness and stability of stationary (or invariant) distributions. The objective of this paper is to complement these distributional results by extending the order theoretic analysis of economic dynamics to the problem of sample path properties. In particular, we seek conditions suitable for economic modeling under which time series averages converge to their population counterparts, in the sense that any time series $\{X_t\}$ generated by the model in question satisfies

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} h(X_t) = \mathbb{E}\, h(X_t) \tag{1}$$

with probability one for some suitably large class of "moment" functions $h$. Here $\mathbb{E}\, h(X_t)$ denotes expectation with respect to the stationary distribution of the model.

Such results are fundamental to quantitative analysis. They support a great variety of computations and theoretical results, from consistency of estimators to simulation of stationary equilibria, calibration, and simulation-based time series estimation (e.g., Hansen (1982); Santos and Peralta-Alva (2005); Duffie and Singleton (1993)). Even Bayesian results that make no direct appeal to asymptotics often require Markov chain Monte Carlo for actual computation, and this in turn requires convergence of time series averages (see, e.g., Geweke (2005)).

While convergence in the sense of (1) is usually automatic in cross sectional models as a result of the law of large numbers for independent random variables,[2] convergence for dynamic models is more subtle. In a recursive time series setting, perhaps the most famous general result is the classical Markov ergodic theorem. For a Markov process $\{X_t\}$ with stationary distribution $\pi$, the theorem gives necessary

[2]There are some obvious exceptions. See, for example, Brock and Durlauf (2001) or Nirei (2006).

and sufficient conditions under which (1) holds almost surely for any function $h$ such that the expectation is finite, and any initial condition $X_0$.[3]

Although this is a powerful and important result, the conditions of the theorem fail to hold for many well known economic models. For example, it cannot be established under the stated assumptions for the capital and income processes in the canonical stochastic optimal growth model of Brock and Mirman (1972). The same is true for various extensions, including the multi-sector version in §10.3 of Stokey and Lucas (1989), the correlated shock version in Hopenhayn and Prescott (1992) and the distorted version in Greenwood and Huffman (1995). Similar issues arise with models from economic development, monetary economics, industrial organization and so on.

Thus the situation for time series averages is essentially analogous to that for distributional results discussed above: classical Markov process theory delivers very strong forms of convergence but at the same time its conditions are too strict for many economic models. In fact the conditions of the classical Markov ergodic theorem are stricter than irreducibility (Meyn and Tweedie, 2009, p. 82 and theorem 17.1.7), and the prevalence of economic models that fail irreducibility is the major motivating factor behind the development of the order theoretic approach to economic dynamics by Razin and Yahav (1979), Bhattacharya and Lee (1988), Stokey and Lucas (1989), Hopenhayn and Prescott (1992) and others.

Drawing on the work of these authors, we investigate sample path properties using an order theoretic approach, in a setting where irreducibility is not required. There are in fact some existing results along these lines. In particular, Bhattacharya and Lee (1988) and Bhattacharya and Majumdar (2001) provide important results showing that a version of (1) holds for monotone functions of the state under a

[3]See, for example, Meyn and Tweedie (2009), theorem 17.1.7. Note that some versions of the ergodic theorem require that $X_0$ is drawn from the stationary distribution itself, and that this distribution is extremal in the set of stationary distributions of the model (see, e.g., Breiman (1992)). In the Markov ergodic theorem considered here, the initial condition is irrelevant. This can be helpful in applications, since it is not necessary to check whether a stationary distribution is extremal or otherwise, and since it means that we can compute stationary outcomes by simulation, starting the process from an arbitrary initial position and allowing for sufficient "burn in" (as in, e.g., Markov chain Monte Carlo).

so-called "splitting" condition, which is a type of strong mixing condition based on order, and closely related to the conditions of Hopenhayn and Prescott (1992).

Despite the usefulness of these sample path results based on splitting, they still fail to cover many moment convergence problems studied by economists. One issue is that we would ideally like to know whether the convergence in (1) holds for continuous functions as well, since the moment conditions tested by economists are sometimes continuous but not monotone. A more important issue is that the splitting conditions themselves can also fail under standard assumptions.

Part of the difficulty is that splitting conditions require a uniform mixing rate from anywhere in the state space that is problematic in settings where the state space is unbounded. Although unboundedness can sometimes be circumvented by assuming bounded shocks, there is a cost in terms of loss of information. For example, Rossi-Hansberg and Wright (2007) study among other things the tail properties of the firm size distribution under different levels of aggregation. In order to exploit stability results from Stokey and Lucas (1989), they compactify their state space (Rossi-Hansberg and Wright, 2007, proposition 5). But compactification clearly discards information about tail properties.

In addition, for some models the state space cannot be compactified at all, even when shocks are assumed to be bounded. One example is Benhabib et al. (2011), where wealth is affected by multiplicative shocks. Even if these shocks are bounded above, a sufficiently long sequence of positive multiplicative shocks can drive wealth above any given threshold. In other words, the state variable exceeds any finite threshold with positive probability. Thus the state space cannot be compact.

In this paper we provide new results that deliver convergence of sample averages in the sense of (1) over a large class of moment functions in a large range of settings. Our conditions are weaker than the conditions in the literature listed above. In particular our conditions for convergence over the class of monotone functions are both necessary and sufficient (see theorem 3.1), and in this sense cannot be improved upon. The generality of the conditions means that they are straightforward to apply in both bounded and unbounded state spaces. We provide sufficient conditions making them relatively easy to check in applications, as well as a number of examples as to how this can be done.

We also show that, under mild additional restrictions on the state space that are satisfied in all standard economic applications, convergence in the form of (1) extends from monotone functions to continuous functions of the state vector. Under the same conditions, we show that the empirical distribution function computed from any sample path converges to the true stationary distribution with probability one. This result can be used to justify computation of the stationary distribution by simulation or estimation methods using the empirical distribution.

Finally, our main theorem is "parametric" in the partial order, in the sense that varying the partial order changes the definition of monotonicity, and hence the conditions and implications of the theorem. Using the standard partial order for vectors gives weak conditions for convergence of sample averages. This is the most practical use case, and the focus of our applications. However, from a theoretical point of view it is notable that with a different choice of partial order the main theorem includes the classical Markov ergodic theorem as a special case.

The remainder of this paper is structured as follows. Section 2 gives some preliminary definitions and results. Sections 3 and 4 present our results on ergodicity and discuss their implications. Section 5 provides sufficient conditions for the form of ergodicity considered in the paper. Section 6 treats application and section 7 concludes. All proofs are deferred to section 8.

## 2. Preliminaries

In this paper, as in Hansen and Sargent (2010), an economic model is a probability distribution on a sequence space. Our main interest is in identifying suitable conditions under which these distributions pick out time series with sample averages that converge to stationary expectations, in a sense to be made precise. In what follows, the sequence space is $S^\infty = S \times S \times \cdots$, where $S$ is called the state space. Elements of $S$ summarize the state of the economy at any point in time, while elements of $S^\infty$ are called time series. A typical probability distribution on $S^\infty$ is denoted by $\mathbb{P}^Q_x$. In this first section, we describe how this distribution is constructed from objects $Q$ and $x$, where $Q$ is a primitive representing the first order transition probabilities induced by preferences, technology and other economic considerations, and $x$ is an initial condition.[4]

2.1. **Model Primitives.** Let $S$ be a separable and completely metrizable topological space and let $\preceq$ be a closed partial order on $S$.[5] Let $\mathscr{B}$ be the Borel sets and let $\mathscr{P}$ be the set of probability measures on $(S, \mathscr{B})$. A function $h \colon S \to \mathbb{R}$ is called *increasing* if $x \preceq x'$ implies $h(x) \leq h(x')$, and *decreasing* if $-h$ is increasing. A subset $B$ of $S$ is called increasing if $x \in B$ and $x \preceq y$ implies $y \in B$; and decreasing if $x \in B$ and $y \preceq x$ implies $y \in B$.

Throughout the paper, we consider models that are time-homogeneous and Markovian. The dynamics of any such model can be summarized by a *stochastic kernel* $Q$, which is a function $Q \colon S \times \mathscr{B} \to [0, 1]$ such that

1. $Q(x, \cdot) \in \mathscr{P}$ for each $x \in S$, and
2. $Q(\cdot, B)$ is measurable for each $B \in \mathscr{B}$.

In the applications treated below, $Q(x, B)$ represents the probability that the state of the economy transitions from point $x \in S$ into set $B \in \mathscr{B}$ over one unit of time. A distribution $\pi \in \mathscr{P}$ is called *stationary* for $Q$ if

$$\int Q(x, B)\pi(dx) = \pi(B), \qquad \forall\, B \in \mathscr{B}.$$

In essence this means that if the current state $X_t$ is drawn from $\pi$ and then $X_{t+1}$ is drawn from $Q(X_t, \cdot)$, the distribution of $X_{t+1}$ will again be $\pi$. As in many other studies (e.g., Brock and Mirman (1972), Stokey and Lucas (1989), Duffie et al. (1994), etc.), a stationary probability is understood here as representing an equilibrium distribution for a stochastic economic model with dynamics given by $Q$.

A stochastic kernel $Q$ is called *increasing* if $(Qh)(x) := \int h(y)Q(x, dy)$ is increasing in $x$ whenever $h \colon S \to \mathbb{R}$ is measurable, bounded and increasing. This condition

[4] Our assumptions and results are always stated in terms of first order models. This costs no generality, since greater lag lengths can be reformulated into the first order framework by suitable redefinition of state variables.

[5] In particular, $\preceq$ is reflexive ($x \preceq x$ for all $x \in S$), transitive ($x \preceq y$ and $y \preceq z$ implies $x \preceq z$) and antisymmetric ($x \preceq y$ and $y \preceq x$ implies $x = y$), and its graph is closed in the product space $S \times S$. Almost all economic settings of interest to us have these properties.

is typically satisfied in models where, holding all shocks fixed, increases in the current state shift up the future state (see, e.g., Stokey and Lucas (1989)).

2.2. **Markov Processes.** Let $S^\infty := S \times S \times \cdots$, and let $\mathscr{B}^\infty$ be the product $\sigma$-algebra. It is well known (see, e.g., Stokey and Lucas (1989), p. 222) that to each stochastic kernel $Q$ on $S$ and distribution $\mu \in \mathscr{P}$, we can associate a unique probability measure $\mathbb{P}^Q_\mu$ on the sequence space $(S^\infty, \mathscr{B}^\infty)$, which is uniquely defined by the expression

$$\mathbb{P}^Q_\mu(B_0 \times \cdots \times B_n \times S \times S \times \cdots) = \int_{B_0} \mu(dx_0) \int_{B_1} Q(x_0, dx_1) \cdots \int_{B_{n-1}} Q(x_{n-2}, dx_{n-1}) \int_{B_n} Q(x_{n-1}, dx_n) \quad (2)$$

for any finite collection $\{B_i\}_{i=0}^n \subset \mathscr{B}$.[6] In essence, $\mathbb{P}^Q_\mu$ is the joint distribution of the Markov process $\{X_t\}$ defined by drawing $X_0$ from $\mu$ and then, recursively, $X_{t+1}$ from $Q(X_t, \cdot)$. If $\mu = \delta_x$ then we simply write $\mathbb{P}^Q_x$.

We are interested in the properties of time series generated by models of this form. In studying these properties, it is helpful to have a canonical Markov process $\{X_t\}$ with which to state our results. To this end, recall that if $(E, \mathscr{E}, \mathbb{P})$ is any probability space and $X$ is the identity map $X(\omega) = \omega$, then $X$ is an $E$-valued random element with distribution $\mathbb{P}$. Following this construction, we take $(S^\infty, \mathscr{B}^\infty, \mathbb{P}^Q_\mu)$ as our probability space unless otherwise stated, and $\{X_t\}$ is just the identity map. This gives a generic Markov process generated by $Q$ and having initial condition $\mu$.

## 3. Ergodicity

In this section we first state the classical Markov ergodic theorem and then present an extension that depends on our partial order $\preceq$.

3.1. **Conditions for Ergodicity.** We begin by reviewing the classical Markov ergodic theorem. Recall that a bounded measurable function $h\colon S \to \mathbb{R}$ is called *invariant* for $Q$ if

$$\int h(y) Q(x, dy) = h(x) \quad (3)$$

[6] As is conventional in ergodic theory, the integrals in (2) are computed from right to left, with the integrand written to the right of the integrating measure.

for all $x \in S$. A stochastic kernel $Q$ on $S$ is said to be *ergodic* if the only bounded invariant functions are the constant functions. The classical Markov ergodic theorem states that, for any stochastic kernel $Q$ with stationary distribution $\pi$, the kernel $Q$ is ergodic if and only if

$$\forall\, x \in S,\ \forall\, \pi\text{-integrable } h, \quad \mathbb{P}_x^Q \left\{ \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} h(X_t) = \int h \, d\pi \right\} = 1.$$

Here "$\pi$-integrable" means that $h \colon S \to \mathbb{R}$ is measurable, and $\int |h| \, d\pi < \infty$, and we maintain this definition throughout. See, for example, proposition 17.1.4 and theorem 17.1.7 of Meyn and Tweedie (2009).

As stated in the introduction, the conditions of this theorem are too strict for many standard economic models. We give several examples of how classical ergodicity can fail in section 6. Our next step is to provide a class of ergodicity results that are "parameterized" by the order $\preceq$ on $S$. By choosing the right order we can include many standard models. In the statement of our theorem, a stochastic kernel $Q$ is called *monotone ergodic* if the only increasing bounded invariant functions are the constant functions.

**Theorem 3.1.** *For any increasing stochastic kernel $Q$ with stationary distribution $\pi$, the following conditions are equivalent:*

(i) *$Q$ is monotone ergodic.*

(ii) *For every $x \in S$ and increasing $\pi$-integrable function $h$,*

$$\mathbb{P}_x^Q \left\{ \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} h(X_t) = \int h \, d\pi \right\} = 1.$$

Theorem 3.1 is in fact a generalization of the classical Markov ergodic theorem, as can be seen by setting $\preceq$ to equality, in the sense that $x \preceq y$ if and only if $x = y$. For this choice of $\preceq$, it's easily verified that every function from $S$ to $\mathbb{R}$ is increasing. As a consequence, the definitions of monotone ergodicity and ergodicity are identical, and every stochastic kernel on $S$ is increasing. In such a setting, the results of theorem 3.1 reduce to the classical case.

To see that theorem 3.1 is strictly more general, observe that for partial orders other than equality, the family of increasing functions is a strict subset of the family of all functions. When such a partial order is chosen, monotone ergodicity is strictly weaker than ergodicity. This allows us to capture the asymptotics of additional models that do not satisfy the classical conditions—provided that their stochastic kernels satisfy the requisite monotonicity. As discussed in the introduction, this is useful for a number of workhorse applications. Concrete examples are given in section 6.

One apparent concern with theorem 3.1 is that if $\preceq$ is a standard partial order such as the usual order $\leq$ on $\mathbb{R}$, then the set of increasing functions referred to in part (ii) of theorem 3.1 may be too small to be useful. For example, we might care about convergence of the second moment $h(x) = x^2$. This function is not monotone. Fortunately, it turns out that the convergence in theorem 3.1 extends to a larger set of functions, without additional assumptions. For example, let $Q$ be a fixed stochastic kernel with stationary distribution $\pi$. Let $\mathscr{L}$ denote the linear span of the set of increasing $\pi$-integrable functions.[7]

**Corollary 3.1.** *If $Q$ is increasing and monotone ergodic with stationary distribution $\pi$, then for all $\mu \in \mathscr{P}$ and all $h \in \mathscr{L}$ we have*

$$\mathbb{P}^Q_\mu \left\{ \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} h(X_t) = \int h \, d\pi \right\} = 1. \tag{4}$$

Convergence over all $h \in \mathscr{L}$ is sufficient for many applications. For example, if $S = \mathbb{R}$ and $\pi$ has $m$ finite moments, then the moment functions $h(x) = x^i$ lie in $\mathscr{L}$ for $i \leq m$, as does any polynomial of order $m$ or less.[8]

As suggested by the statement of theorem 3.1, monotone ergodicity is not sufficient to yield existence of a stationary distribution $\pi$. (We provide an existence result along with other results that imply monotone ergodicity in section 5.) However, the conditions of theorem 3.1 are sufficient for uniqueness:

[7] In other words, $\mathscr{L}$ is the set of all $h \colon S \to \mathbb{R}$ such that $h = \alpha_1 h_1 + \cdots + \alpha_k h_k$ for some scalars $\{\alpha_i\}_{i=1}^k$ and increasing measurable $\{h_i\}_{i=1}^k$ with $\int |h_i| d\pi < \infty$. Equivalently, $\mathscr{L}$ is all $h$ such that $h = f - g$ for increasing $\pi$-integrable $f$ and $g$.

[8] Let $h(x) = x^i$ with $i \leq m$. If $i$ is odd, then $h$ is increasing. If not then write $h$ as $h = -f + g$, where $f(x) := -x^i \mathbb{1}\{x < 0\}$ and $g(x) := x^i \mathbb{1}\{x \geq 0\}$. Both $f$ and $g$ are increasing and hence $h \in \mathscr{L}$. Finally, if $p(x) = \sum_{i=1}^m a^i x^i$ then $p \in \mathscr{L}$ because $\mathscr{L}$ is closed under linear combinations.

**Proposition 3.1.** *If $Q$ is increasing and monotone ergodic, then $Q$ has at most one stationary distribution.*

3.2. **Continuous Functions and Empirical Distributions.** It is in fact possible to extend the convergence results beyond $\mathscr{L}$ in many situations. In this section we show that if $S$ is compact and $\preceq$ is suitably regular, then the convergence in corollary 3.1 extends to all continuous functions too. Moreover, if $S$ is not compact, then the same is true for any continuous bounded function. In fact we prove a considerably stronger result, related to convergence of the *empirical distribution* $\pi_n$, which is, as usual, defined by

$$\int h \, d\pi_n := \frac{1}{n} \sum_{t=1}^{n} h(X_t) \quad \text{for measurable } h \colon S \to \mathbb{R}.$$

The empirical distribution is a natural candidate for estimating $\pi$, and forms a standard tool for econometric analysis and calibration. We wish to know when $\pi_n \xrightarrow{w} \pi$ with probability one, where $\xrightarrow{w}$ represents the usual probabilist's notion of weak convergence (i.e., $\int h \, d\pi_n \to \int h \, d\pi$ for all continuous bounded $h$).[9]

**Assumption 3.1.** $(S, \preceq)$ is a normally ordered[10] and has the property that $K \subset S$ is compact if and only if it is closed and order bounded (i.e., there exist points $a$ and $b$ in $S$ with $a \preceq x \preceq b$ for all $x \in K$). Moreover, there exists a countable subset $A$ of $S$ such that, given any $x \in S$ and neighborhood $U$ of $x$, there are $a, a' \in A$ such that $a, a' \in U$ and $a \preceq x \preceq a'$.

Assumption 3.1 is satisfied for almost all state spaces used in economic applications, such as when $S = \mathbb{R}^m$ with its usual pointwise order $\leq$, or more generally, when $S$ is an open or closed interval or cone in $\mathbb{R}^m$ with the usual pointwise order.

[9]The statement $\int h \, d\pi_n \to \int h \, d\pi$ for all continuous bounded $h$ with probability one is much stronger than $\int h \, d\pi_n \to \int h \, d\pi$ with probability one for all continuous bounded $h$. The reason is that, even when the latter holds, the probability one set on which convergence obtains depends on $h$, and the set of continuous bounded functions on $S$ is uncountable.

[10]$S$ is called *normally ordered* if, given any disjoint pair of closed sets $I, D \subset S$ such that $I$ is increasing and $D$ is decreasing, there exists an increasing continuous bounded $h \colon S \to \mathbb{R}$ such that $h(x) = 0$ for all $x \in D$ and $h(x) = 1$ for all $x \in I$.

**Theorem 3.2.** *Let the state space $S$ satisfy assumption 3.1. In this setting, if $Q$ is increasing and monotone ergodic with stationary distribution $\pi$, then, for any $x \in S$,*

$$\mathbb{P}_x^Q \left\{ \lim_{n\to\infty} \int h \, d\pi_n = \int h \, d\pi, \quad \forall \text{ continuous bounded } h \colon S \to \mathbb{R} \right\} = 1.$$

In particular,

(i) $\pi_n \xrightarrow{w} \pi$ with probability one.

(ii) Given any continuous bounded function $h$, we have $\frac{1}{n} \sum_{t=1}^n h(X_t) \to \int h \, d\pi$ with probability one.

## 4. Connections to the Literature

As discussed in the introduction, there is a large literature on asymptotics of recursive economic models based around order theoretic conditions. To date this literature has focused mainly on distributional properties, such as existence, uniqueness and stability of the stationary distribution. What we now show is that the assumptions that these authors work under all imply monotone ergodicity. Hence our conclusions on sample path convergence can be added to the distributional implications previously derived.

To begin the discussion, consider the "splitting condition" found, for example, in Bhattacharya and Majumdar (2001). Their environment consists of a sequence of IID random maps $\{\gamma_t\}$ from $S$ to itself, where $S$ is a subset of $\mathbb{R}^m$. The maps generate $\{X_t\}$ via $X_{t+1} = \gamma_{t+1}(X_t)$, or, more explicitly,

$$X_t = \circ_{i=1}^t \gamma_i(X_0) := \gamma_t \circ \cdots \circ \gamma_1(X_0).$$

The corresponding stochastic kernel is $Q(x, B) = \mathbb{P}\{\gamma_1(x) \in B\}$. The splitting condition runs as follows:

**Condition 4.1.** There exists a $c \in S$ and $k \in \mathbb{N}$ such that

$$\mathbb{P}\{\circ_{i=1}^k \gamma_i(y) \leq c, \; \forall y \in S\} > 0 \;\text{ and }\; \mathbb{P}\{\circ_{i=1}^k \gamma_i(y) \geq c, \; \forall y \in S\} > 0.$$

Closely related to splitting are the conditions of Razin and Yahav (1979), Stokey and Lucas (1989) and Hopenhayn and Prescott (1992). In particular, Hopenhayn and Prescott (1992) adopt the following restrictions:

**Condition 4.2.** $S$ is a compact metric space with closed partial order $\preceq$. $S$ has a least element $a$ and greatest element $b$. $Q$ is an increasing kernel on $S$ satisfying the following restriction:

$$\exists\, \bar{x} \in S \text{ and } k \in \mathbb{N} \text{ such that } \mathbb{P}_a^Q\{X_k \geq \bar{x}\} > 0 \text{ and } \mathbb{P}_b^Q\{X_k \leq \bar{x}\} > 0. \tag{5}$$

Szeidl (2013) studies a variety of economic models including the buffer stock savings model of Carroll (1997) in the following setting:

**Condition 4.3.** $S$ is an order interval in $\mathbb{R}^m$ with its usual pointwise order and $Q$ is increasing, uniformly asymptotically tight[11] and weakly mixing in the sense that there exists a $c \in S$ such that, $\forall\, x \in S$, we can find $j, k \in \mathbb{N}$ with $\mathbb{P}_x^Q\{X_j > c\} > 0$ and $\mathbb{P}_x^Q\{X_k < c\} > 0$.

In a separate study, Kamihigashi and Stachurski (2014) provide distributional results in the order theoretic framework using the following conditions:

**Condition 4.4.** $Q$ is increasing, order reversing and bounded in probability.

In condition 4.4, *order reversing* means that, for any given $x$ and $x'$ in $S$ with $x' \preceq x$ and any independent processes $\{X_t\}$ and $\{X_t'\}$ generated by $Q$ and starting at $x$ and $x'$ respectively, there exists a $t \in \mathbb{N}$ with $\mathbb{P}\{X_t \preceq X_t'\} > 0$. As usual, $Q$ is called *bounded in probability* if, given any $x \in S$ and any $\varepsilon > 0$, there exists a compact $K \subset S$ with $\mathbb{P}_x^Q\{X_t \in K\} \geq 1 - \varepsilon$ for all $t$

The same authors also consider distributional properties under so-called order mixing:

**Condition 4.5.** $Q$ is *order mixing* in the sense that, given any pair of independent Markov processes $\{X_t\}$ and $\{X_t'\}$ generated by $Q$, the event $\{X_t \preceq X_t'\}$ occurs with probability one.

The main result of this section is that all these sets of conditions are stricter than monotone ergodicity:

[11] As usual, $Q$ is called *uniformly asymptotically tight* if, for all $\delta > 0$, there exists a compact $C \subset S$ such that $\liminf \mathbb{P}_x^Q\{X_n \in C\} > 1 - \delta$ for all $x \in S$.

**Proposition 4.1.** *Any one of conditions 4.1–4.5 implies monotone ergodicity.*

The benefit of proposition 4.1 is that the sets of conditions given in conditions 4.1–4.5 have previously been applied to many different economic models, establishing useful distributional properties. Proposition 4.1 tells us that the conclusions of theorem 3.1, corollary 3.1 and theorem 3.2 are also valid for those models.

## 5. Sufficient Conditions

As discussed above, there are existing conditions in the literature that imply monotone ergodicity, and these suffice for many economic problems. However, for classes of economic models that possess certain monotonicity and continuity conditions, it is possible to develop another approach that is particularly straightforward and intuitive.

Consider a generic model of the form

$$X_{t+1} = F(X_t, \varepsilon_{t+1}), \qquad \{\varepsilon_t\} \overset{\text{IID}}{\sim} \phi, \qquad X_0 \text{ given}, \tag{6}$$

where $F\colon S \times E \to S$ is continuous, $S$ and $E$ are Borel subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, and $\phi$ is a Borel probability measure on $E$. In this section $S$ is always endowed with its usual pointwise order $\leq$. The stochastic kernel corresponding to (6) is

$$Q_F(x, A) := \phi\{\varepsilon \in E : F(x, \varepsilon) \in A\}, \tag{7}$$

**Assumption 5.1.** $S$ is a subset of $\mathbb{R}^n$ satisfying assumption 3.1. The shock distribution $\phi$ is supported on all of the shock space $E$.[12] The function $x \mapsto F(x, \varepsilon)$ is increasing for each $\varepsilon \in E$, and $Q_F$ is bounded in probability.

Each finite path of shock realizations $\{\varepsilon_t\}_{t=1}^k \subset E$ and initial condition $x \in S$ determines a path $\{x_t\}_{t=0}^k$ for the state variable up until time $k$ via $x_0 = x$ and $x_{t+1} = F(x_t, \varepsilon_{t+1})$. Let $F^k(x, \varepsilon_1, \ldots, \varepsilon_k)$ denote the value of $x_k$ determined in this way. Given vectors $x$ and $y$ in $S$, we write $x < y$ if $x_i < y_i$ for all $i$.

**Proposition 5.1.** *If assumption 5.1 is satisfied, then $Q_F$ is increasing and at least one stationary distribution exists. If, in addition, one of the following three conditions holds*

[12] That is, $\phi(E) = 1$, and $\phi(G) > 0$ whenever $G \subset E$ is nonempty and open. This entails no loss of generality, since the shock space can always be re-defined appropriately.

(i) *for any* $x, c \in S$, *there exists* $\{\varepsilon_1, \ldots, \varepsilon_k\} \subset E$ *such that* $F^k(x, \varepsilon_1, \ldots, \varepsilon_k) < c$

(ii) *for any* $x, c \in S$, *there exists* $\{\varepsilon_1, \ldots, \varepsilon_k\} \subset E$ *such that* $F^k(x, \varepsilon_1, \ldots, \varepsilon_k) > c$

(iii) *for any* $x, x' \in S$, *there exists* $\{\varepsilon_1, \ldots, \varepsilon_k\} \subset E$ *and* $\{\varepsilon'_1, \ldots, \varepsilon'_k\} \subset E$ *such that* $F^k(x, \varepsilon_1, \ldots, \varepsilon_k) < F^k(x', \varepsilon'_1, \ldots, \varepsilon'_k)$

*then* $Q_F$ *has exactly one stationary distribution and is monotone ergodic.*

Conditions (i)–(iii) are mixing conditions, and are related to the notions of upward reaching, downward reaching and order reversing processes introduced in Kamihigashi and Stachurski (2014). Unlike the latter, conditions (i)–(iii) exploit continuity to provide statements that are easier to check in applications.

To see how proposition 5.1 can be useful, compare condition (iii) to the notion of order mixing, which requires that separate time series driven by their own set of idiosyncratic shocks become ordered eventually with probability one (see condition 4.5). Condition (iii) simply states that such an occurrence is *possible*. This kind of condition is typically much easier to verify.

## 6. Applications

6.1. **Stochastic Optimal Growth.** Variations of the Brock-Mirman optimal growth model (Brock and Mirman (1972)) are routinely applied in many fields of macroeconomic modeling (see, e.g., Ljungqvist and Sargent (2012)). The one sector model takes the form

$$\max_{\{k_{t+1}, c_t\}_{t \geq 0}} \mathbb{E} \left[ \sum_{t=0}^{\infty} \beta^t u(c_t) \right]$$

$$\text{s.t. } c_t + k_{t+1} \leq (1 - \delta) k_t + f(k_t, z_t), \tag{8}$$

$$z_{t+1} = g(z_t, \varepsilon_{t+1}) \tag{9}$$

for all $t \geq 0$. Here all variables are nonnegative, $\{z_t\}$ is an exogenous productivity process, $\{\varepsilon_t\}$ is a sequence of IID innovations and $\beta \in (0, 1)$. To avoid trivial cases the initial conditions $k_0$ and $z_0$ are always assumed to be strictly positive in what follows, and $g(z, \varepsilon) \in \mathbb{R}_{++}$ whenever $(z, \varepsilon) \in \mathbb{R}^2_{++}$. The function $g$ is assumed to be continuous and increasing in its first argument.

Even in very simple settings the optimal state process fails to satisfy classical ergodicity. For example, let $u(c) = \ln(c)$, let $f(k, z) = Ak^{\alpha}z$ and let $\delta = 1$. In this case it is well known that capital evolves according to $k_{t+1} = \alpha\beta A k_t^{\alpha} z_t$. Choose common parameter values such as $\alpha = 0.3$, $\beta = 0.99$ and $A = 1$. Suppose that $\{z_t\}$ is IID and takes values in $\{0.75, 1.25\}$, say. To study the dynamics of $\{k_t\}$ on $\mathbb{R}_{++}$, let $L$ be all $x \in \mathbb{R}_{++}$ such that $\ln(x)$ is rational. It is easy to see that $k_{t+1} \in L$ if and only if $k_t \in L$. Letting $\mathbb{1}_L$ be the indicator function of $L$, we can express this as

$$\mathbb{P}\{k_{t+1} \in L \mid k_t = k\} = \mathbb{1}_L(k) \tag{10}$$

for all $k \in \mathbb{R}_{++}$. This says precisely that $\mathbb{1}_L$ is an invariant function for capital.[13] Since $\mathbb{1}_L$ is bounded and invariant but not constant, classical ergodicity fails.

This example illustrates the strictness of the classical ergodicity conditions. Similar difficulties carry over to applying classical irreducibility-based conditions to establish distributional properties such as existence, uniqueness and stability of stationary distributions. As noted in the introduction, these kinds of issues have motivated the development of alternative approaches to understanding the distributional properties of economic models based around order theoretic ideas (e.g., Razin and Yahav (1979); Stokey and Lucas (1989); Hopenhayn and Prescott (1992); Bhattacharya and Majumdar (2001); Kamihigashi and Stachurski (2014)).

For this particular example the shocks are bounded and the state space can be chosen to be compact. Any of these references can be used to show the distributional properties listed above. In addition, from the results in Bhattacharya and Lee (1988), we can also infer convergence in probability of $\frac{1}{n}\sum_{t=1}^{n} h(k_t)$ to its stationary expectation $\int h(k)\pi(dk)$ whenever $h = h_1 - h_2$ for some monotone $h_1$ and $h_2$. If we can establish the conditions of proposition 5.1 this convergence will extend to any continuous function $h$, and to probability one convergence of the empirical distribution, via theorem 3.2.

The conditions of proposition 5.1 certainly hold for this simple example. Evidently $k \mapsto \alpha\beta A k^{\alpha} z$ is continuous and monotone increasing for any possible realization of $z$. Boundedness in probability holds because the state space can be chosen to

[13] If $h = \mathbb{1}_B$ for some $B$, then (3) reduces to $Q(x, B) = \mathbb{1}_B(x)$ for all $x \in S$, which in this case is (10).

be compact. Condition (iii) of proposition 5.1 can be established as follows. Recall that $z_t$ is assumed to take either a high or a low value. If only the high value occurs then $k_t$ converges to some point $\bar{k} \in \mathbb{R}_{++}$ regardless of the initial condition. If only the low value of the shock occurs, $k_t$ converges to some lower point $\underline{k}$, regardless of initial conditions. Thus, given any pair of initial conditions, we can choose shock sequences such that the ordering in condition (iii) of proposition 5.1 holds.

6.2. **Correlated Shocks.** Next we consider a more general version of the optimal growth model, where no useful results on convergence of time series sample averages are available in the existing literature. We maintain the log-linear form $k_{t+1} = \alpha\beta A k_t^\alpha z_t$ while returning to the general correlated case (9) for the productivity shock. We study asymptotics of the Markov process $X_t := (k_t, z_t)$ on $S = \mathbb{R}^2_{++}$. To focus on a noncompact environment, we begin by assuming that, for each $z \in \mathbb{R}_{++}$ and $m \in \mathbb{N}$, there exists an $\varepsilon$ in the support of $\varepsilon_t$ such that $g(z, \varepsilon) \geq m$. To ensure that capital has a stationary solution, we assume that $\alpha \in (0, 1)$ and that $\sup_t \mathbb{E} \, | \ln z_t | < \infty$.

The existence of a unique stationary distribution and monotone ergodicity can be shown in a straightforward way using proposition 5.1. The function $F$ corresponding to (6) is

$$F((k_t, z_t), \varepsilon_{t+t}) = (\alpha\beta A k_t^\alpha z_t, \; g(z_t, \varepsilon_{t+1})).$$

Since $g$ is continuous and increasing, $F$ is continuous and increasing in $(k_t, z_t)$ for each fixed value of $\varepsilon_{t+1}$. Since $\mathbb{E} \, | \ln z_t |$ is assumed to be bounded in $t$, to show that $\{X_t\}$ is bounded in probability on $S$ it is enough to show that $\mathbb{E} \, | \ln k_t |$ is also bounded in $t$ (see, e.g., Meyn and Tweedie (2009), p. 559). This follows easily from taking logs in $k_{t+1} = \alpha\beta A k_t^\alpha z_t$ and using the assumption that $\mathbb{E} \, | \ln z_t |$ is bounded in $t$.

The only nontrivial remaining step needed to check the conditions of proposition 5.1 is that one of conditions (i)–(iii) hold. We claim that (ii) holds. To see this, fix any initial condition $x = (k_0, z_0)$ and any other point $c = (k_c, z_c)$ in $S$. Let $k_1 := \alpha\beta A k_0^\alpha z_0$. Consider the value of the state two periods after starting at $(k_0, z_0)$ and receiving shocks $\varepsilon_1, \varepsilon_2$. The values are

$$k_2(\varepsilon_1) := \alpha\beta A k_1^\alpha g(z_0, \varepsilon_1) \;\text{ and }\; z_2(\varepsilon_1, \varepsilon_2) := g(g(z_0, \varepsilon_1), \varepsilon_2).$$

In view of our assumptions on the shock, we can choose values $\varepsilon_1$ and $\varepsilon_2$ in the support of the shock distribution such that $k_2(\varepsilon_1) > k_c$ and $z_2(\varepsilon_1, \varepsilon_2) > z_c$. In the notation of proposition 5.1 we can write this as

$$F^2((k_0, z_0), \varepsilon_1, \varepsilon_2) = (k_2(\varepsilon_1),\ z_2(\varepsilon_1, \varepsilon_2)) > (k_c, z_c) = c.$$

Thus (ii) holds and all the conditions of proposition 5.1 are verified. Hence the state process has a unique stationary distribution, and the sample mean $\frac{1}{n}\sum_{t=1}^{n} h(k_t)$ converges to its expectation $\mathbb{E}\, h(k_t)$ under the stationary distribution with probability one whenever $h$ is monotone and the expectation is finite (theorem 3.1). The same is true if $h$ is a finite moment $k_t^p$ or any other linear combination of monotone functions (corollary 3.1), or a continuous bounded function (theorem 3.2).

In proving condition (ii) of proposition 5.1 we assumed for simplicity that the exogenous productivity process can be driven above any fixed number from any starting condition in one unit of time. This kind of scenario occurs in a variety of model specifications, such as those with lognormal innovations and a log linear shock process. However, condition (ii) also holds if the productivity process can be driven above any fixed number in *finite* time, with a suitable sequence of shocks. The argument is only slightly more elaborate.

Proposition 5.1 again holds if we assume instead that the exogenous process can be pushed arbitrarily close to zero from any initial condition in finite time, by verifying condition (i) of proposition 5.1 instead of condition (ii). In fact proposition 5.1 can be applied if we know only that we can select sequences for the shocks $\{\varepsilon_t\}$ that drive $\{z_t\}$ to either of two distinct possible values, regardless of initial conditions. This implies that we can also drive $\{k_t\}$ to either of two distinct values from any starting point by suitable choice of shocks. Hence condition (iii) of proposition 5.1 is valid.

One benefit of verifying the conditions of proposition 5.1 is that we know from theorem 3.2 that an empirical distribution computed from simulated time series converges weakly to the unique stationary distribution with probability one. In this setting the empirical distribution for capital given simulated sample $\{k_t\}_{t=1}^{n}$ can be written as $F_n(k) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{k_t \leq k\}$. Figure 1 shows two realizations of $F_n$, each computed from samples of size $n = 10^6$. The shock is a discretized AR(1)

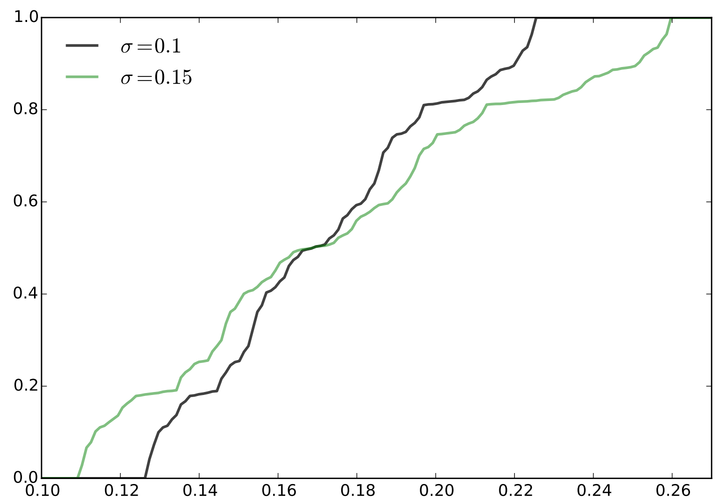


FIGURE 1. Stationary distribution of capital

process with 4 states. The two realizations correspond to simulations with different levels of volatility in the productivity process.[14]

6.3. **General Functional Forms.** Now let's return to the IID assumption on the shock process $\{z_t\}$ but drop the specifications on $u$ and $f$ that allowed us to derive closed form solutions. Suppose instead that $f$ is continuous and that $u$ and $k \mapsto f(k, z)$ are strictly increasing, strictly concave and continuously differentiable, and that $f(0, z) = 0$ for all $z$. Suppose for simplicity that $u$ is bounded and $\delta = 1$. Suppose further that $u'(0) = \infty$, that there exist positive $k$ and $r$ such that $\mathbb{E} f(k, z_t)$ and $\mathbb{E} f_1(k, z_t)^{-r}$ are both finite, and that

$$\lim_{k \to \infty} \mathbb{E} f_1(k, z_t) = 0 \quad \text{and} \quad \lim_{k \to 0} \mathbb{E} \ln[\beta f_1(k, z_t)] > 0.$$

Let $y_t = f(k_t, z_t)$. Under these conditions it is known that there exists a unique optimal investment policy $y \mapsto k(y)$ that is continuous, increasing and interior, and that the optimal income process $y_{t+1} = f(k(y_t), z_{t+1})$ is bounded in probability on $\mathbb{R}_{++}$ (Kamihigashi, 2007, theorem 2.1). To add monotone ergodicity we need only

[14]The shocks are discretized from a process which in logs has the form $z_{t+1} = \rho z_t + \sigma \varepsilon_{t+1}$. Discretization uses Tauchen's method. Parameters are $\alpha = 0.3$, $\beta = 0.96$, $A = 1$, $\rho = 0.8$ and either $\sigma = 0.1$ or $\sigma = 0.15$. For the code see `https://gist.github.com/jstac/d58d4f4273cf5a2deb03`.

provide conditions under which one of (i)–(iii) in proposition 5.1 hold. We can work with bounded shocks and condition (iii), but to focus on the unbounded case we assume that $z_t$ is supported on all of $\mathbb{R}_{++}$ and that $\lim_{z\to\infty} f(k,z) = \infty$ for any positive $k$. Condition (ii) then holds because if we fix initial condition $y \in \mathbb{R}_{++}$ and any other point $c \in \mathbb{R}_{++}$, there exists by assumption a $\bar{z}$ such that $f(k(y), \bar{z}) > c$.

6.4. **The Firm Size Distribution.** Rossi-Hansberg and Wright (2007) use a general equilibrium model to study firm size dynamics and their implications for the firm size distribution across different sectors and industries. Their model implies industry level firm size dynamics of the form

$$s_{t+1} = n^c + [1 - (1-\omega)(1-\beta(1-\alpha))]s_t - \beta(1-\alpha)\ln A_{t+1} \tag{11}$$

where $s_t$ is a measure of firm size, $n^c$ is a constant term, $\omega \in (0,1)$ is a parameter in the law of motion for human capital, $\alpha$ and $\beta$ are parameters from the production function taking values in $(0,1)$ and $\{A_t\}$ is a strictly positive IID sequence affecting accumulation of human capital (see Rossi-Hansberg and Wright (2007), p. 1645). Rossi-Hansberg and Wright (2007) study the asymptotics of $\{s_t\}$ by assuming that $A_t$ always takes values in a compact set (Rossi-Hansberg and Wright, 2007, proposition 5). We now do the same without compactification.

If $A_t$ is deterministic then dynamics are trivial to analyze, so suppose that the support of $A_t$ contains at least two distinct values. It follows that part (iii) of proposition 5.1 holds. To see why, note that (11) can be expressed as $s_{t+1} = F(s_t, \varepsilon_{t+1}) = as_t + b + \varepsilon_{t+1}$ where $a$ and $b$ are constants, $a \in (0,1)$, and $\{\varepsilon_t\}$ is IID and takes at least two distinct values $\underline{\varepsilon}$ and $\bar{\varepsilon}$. Without loss of generality suppose that $\underline{\varepsilon} < \bar{\varepsilon}$. Regardless of where we start the process, if we receive only shock $\underline{\varepsilon}$ we converge to $(b + \underline{\varepsilon})/(1-a)$. If we receive only shock $\bar{\varepsilon}$ we converge to the larger value $(b + \bar{\varepsilon})/(1-a)$. Thus we can choose shock sequences such that the ordering in (iii) of proposition 5.1 holds eventually, regardless of initial conditions.

The other conditions of proposition 5.1 are straightforward to verify in this context. Boundedness in probability holds because the coefficient $1 - (1-\omega)(1-\beta(1-\alpha))$ lies in $(0,1)$ and hence the process is mean reverting. Continuity and monotonicity are immediate. The conclusions of proposition 5.1 follow.

6.5. **Wealth distributions.** Benhabib et al. (2011) study evolution of the wealth distribution in a general equilibrium model that produces a system of the form

$$\begin{aligned} w_{t+1} &= \alpha(z_{t+1})w_t + \beta(z_{t+1}) \\ z_{t+1} &= g(z_t, \varepsilon_{t+1}). \end{aligned}$$

Here $\{w_t\}$ is household wealth, $\{\varepsilon_t\}$ is an IID shock sequence, $\{z_t\}$ is an exogenous process and $\alpha$, $\beta$ and $g$ are given functions.[15] As in Benhabib et al. (2011), we take $z_t$ to be discrete, with $\alpha(z) > 1$ for high values of $z$ and $\alpha(z) < 1$ for low values. Hence wealth goes through periods of expansion and contraction. Since it changes little of what follows, we assume that $z_t \in \{0, 1\}$, with $0 < \alpha(0) < 1 < \alpha(1)$. We suppose that wealth is nonnegative, that $\mathbb{P}\{z_{t+1} = i \,|\, z_t = j\} > 0$ for all $i, j \in \{0, 1\}$, that $g(z, \varepsilon)$ is increasing in $z$ for each $\varepsilon$, and that $0 < \beta(0) \leq \beta(1)$. To prevent wealth from growing without limit, we assume that $\ln \alpha(0)\pi_0 + \ln \alpha(1)\pi_1 < 0$, where $\pi$ is the stationary distribution of $z_t$. See theorem 1 of Brandt (1986).

The endogenous state $w_t$ is is not in general irreducible and is naturally unbounded. Indeed, if $z_t$ remains in the high state for sufficiently long, then $w_t$ will exceed any given bound. As a result we take the state space for $w_t$ to be all of $[0, \infty)$, and the state space $S$ for the pair $X_t := (w_t, z_t)$ as $[0, \infty) \times \{0, 1\}$. Because of this unboundedness, the existing law of large number results based around condition 4.1 do not hold. Nor do the classical ergodic results hold here in general. (A counterexample analogous to the one developed in section 6.1 can also be applied here.)

On the other hand, the conditions of proposition 5.1 are easy to verify. Boundedness in probability is already known (Brandt, 1986, theorem 1). Continuity follows immediately from our assumptions, as does monotonicity. Condition (ii) of the proposition clearly holds too, since a sufficiently long sequence of high states for $z_t$ will drive $(w_t, z_t)$ above any given vector in $S$. Hence the system has a unique stationary distribution and is monotone ergodic.

6.6. **Consistency of Estimators.** A large number of standard results on consistency and asymptotic normality of estimators from time series econometrics rely

[15] Similar dynamics arise in models of prices and inflation. See, for example, Benhabib and Dave (2013) or Farmer et al. (2009).

on classical ergodicity (see, e.g., Hansen (1982)). Given that many economic models fail classical ergodicity, an important question is whether the same results can also be established after assuming only monotone ergodicity as defined in section 3.1. While it is beyond the scope of this paper to discuss estimation in detail, for the sake of illustration we now sketch how consistency of the OLS estimate can be derived in the simple scalar regression $Y_t = \beta X_t + \nu_t$ when $\{X_t\}$ is monotone ergodic.

For this purpose we will assume that $\{X_t\}$ is a Markov process with stationary distribution $\pi$ generated by an increasing monotone ergodic kernel $Q$, as was the case for the state processes in the applications in sections 6.1–6.5. Suppose in addition that $\{X_t\}$ is has finite nonzero second moment $s_X = \int x^2 \pi(dx)$. To study consistency recall that the difference between the true parameter and the OLS estimator $\hat{\beta}_n$ can be expressed as

$$\hat{\beta}_n - \beta = \left[ \frac{1}{n} \sum_{t=1}^{n} X_t^2 \right]^{-1} \frac{1}{n} \sum_{t=1}^{n} X_t \nu_t. \tag{12}$$

Under standard conditions on the error term, $\{X_t \nu_t\}$ is a martingale difference sequence, and $\frac{1}{n} \sum_{t=1}^{n} X_t \nu_t \xrightarrow{p} 0$, where $\xrightarrow{p}$ indicates convergence in probability. The remaining concern is the limit of $\hat{s}_X^{-1} := [\frac{1}{n} \sum_{t=1}^{n} X_t^2]^{-1}$. As shown in the discussion after corollary 3.1, the function $h(x) = x^2$ lies in $\mathscr{L}$ and hence we have $\frac{1}{n} \sum_{t=1}^{n} X_t^2 \to s_X$ with probability one, and therefore in probability. The continuous mapping theorem then gives $\hat{s}_X^{-1} \xrightarrow{p} s_X^{-1}$. We conclude that the right hand side of (12) converges to zero in probability, and hence that $\hat{\beta}_n \xrightarrow{p} \beta$, as was to be shown.

## 7. Conclusion

A significant number of economic models do not satisfy the classical ergodicity conditions. Motivated by a range of earlier studies of economic dynamics based on order-theoretic ideas, this paper develops a new condition called monotone ergodicity that is shown to be necessary and sufficient for probability one convergence of sample averages to population means over a certain class of functions. We show that monotone ergodicity is implied by a number of different conditions from the existing economics literature that were used to prove distributional properties.

Hence our results on convergence of sample averages provide more information on the dynamics of models that satisfy the conditions in these existing well known studies. Our results also provide information on sample path properties in settings where no previous sample path results were available.

A number of additional results related to implications of the theory are also provided. For example, we show that the empirical distribution associated with any sample converges to the stationary distribution with probability one. We also discuss sufficient conditions, providing a bridge from the abstract results in the paper to new applications. Several illustrations show how the results extend existing knowledge on the asymptotics of popular economic models.

## 8. Proofs

8.1. **Preliminaries.** For the proofs we adopt some additional notation. Let $bS$ denote the set of bounded measurable functions from $(S, \mathscr{B})$ to $\mathbb{R}$, let $ibS$ denote the set of increasing functions in $bS$, let $cbS$ denote the set of continuous functions in $bS$ and let $icbS := ibS \cap cbS$. We sometimes use inner product notation to represent integration, so that

$$\langle \mu, h \rangle := \int h(x) \mu(dx)$$

for all $h \colon S \to \mathbb{R}$ and measures $\mu$ on $(S, \mathscr{B})$ such that the integral is defined.

8.2. **Proofs from Section 3.** As mentioned in section 3, some authors define ergodicity in terms of shift-invariant events, and hence, for the sake of completeness, we prove a slightly more general form of theorem 3.1, encompassing monotone equivalents of these ideas. To begin, let the shift operator $\theta \colon S^\infty \to S^\infty$ be defined as usual by $\theta(x_0, x_1, \ldots) = (x_1, x_2, \ldots)$. Let $\theta^t$ denote the $t$-th composition of $\theta$ with itself, and let $\theta^0$ be the identity. Let $X$ be the first coordinate projection, sending $(x_0, x_1, \ldots, x_t, \ldots)$ into $x_0$. If $\mathbb{P}$ is any probability measure on the sequence space $(S^\infty, \mathscr{B}^\infty)$, then the $S$-valued stochastic process $\{X_t\}$ on $(S^\infty, \mathscr{B}^\infty, \mathbb{P})$ defined by $X_t := X \circ \theta^t$ has joint distribution $\mathbb{P}$. Specializing to $\mathbb{P} = \mathbb{P}_\mu^Q$ yields the canonical Markov process discussed in section 2.2. Here and below, $\{X_t\}$ is understood as being defined in this way and $(S^\infty, \mathscr{B}^\infty, \mathbb{P}_\mu^Q)$ is the probability space, unless otherwise stated.

A random variable is always a $\mathscr{B}^\infty$ measurable map from $S^\infty$ to $\mathbb{R}$. We endow $S^\infty$ with the pointwise order inherited from $(S, \preceq)$. In particular, we say that $\{x_t\} \preceq \{x'_t\}$ if $x_t \preceq x'_t$ in $S$ for all $t$. We will make use of the following lemma, which follows immediately from propositions 1 and 2 of Kamae et al. (1977).

**Lemma 8.1.** *If $Q$ is an increasing stochastic kernel on $S$, $E$ is an increasing set in $\mathscr{B}^\infty$ and $x, y \in S$, then $x \preceq y$ implies $\mathbb{P}^Q_x(E) \leq \mathbb{P}^Q_y(E)$.*

An event $A \in \mathscr{B}^\infty$ is called *shift-invariant* if $\theta^{-1}(A) = A$. It is called *trivial* if the function $h(x) := \mathbb{P}^Q_x(A)$ is constant on $S$ and takes values in $\{0, 1\}$. A family of sets in $\mathscr{B}^\infty$ is called trivial if every element of the family is trivial. A random variable $Y$ is called *shift-invariant* if it is measurable with respect to the family of shift-invariant sets (which form a $\sigma$-algebra). We will also make use of the following lemma, which is proved in section 8.5.

**Lemma 8.2.** *Let $\mathscr{G} \subset \mathscr{B}^\infty$ be a $\sigma$-algebra, let $i\mathscr{G}$ be the increasing sets in $\mathscr{G}$, and let $Y$ be an increasing, $\mathscr{G}$-measurable random variable. If $i\mathscr{G}$ is trivial, then there exists a $\gamma \in \mathbb{R}$ such that $\mathbb{P}^Q_x\{Y = \gamma\} = 1$ for all $x \in S$.*

Here is the generalization of theorem 3.1. It does not use the Polish assumption on $(S, \preceq)$. In particular, $(S, \mathscr{B}, \preceq)$ can be any partially ordered measurable space.

**Theorem 8.1.** *For any increasing stochastic kernel $Q$ with stationary distribution $\pi$, the following conditions are equivalent:*

(i) *Every increasing shift-invariant set is trivial.*
(ii) *$Q$ is monotone ergodic.*
(iii) *For every $x \in S$ and increasing $\pi$-integrable function $h$, we have*

$$\mathbb{P}^Q_x \left\{ \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} h(X_t) = \int h \, d\pi \right\} = 1.$$

*Proof of theorem 3.1.* (i) $\implies$ (ii). Let $h$ be bounded, increasing and invariant. Define $Y := \limsup_t h(X_t)$. We then have $h(x) = \mathbb{E}^Q_x Y$ for all $x \in S$, as shown in theorem 17.1.3 of Meyn and Tweedie (2009). Notice that $Y$ is shift invariant, since, for each $a \in \mathbb{R}$, the set $A := \{Y \leq a\}$ satisfies $\theta^{-1}(A) = A$. Notice also that $Y$ is increasing on the sample space $S^\infty$. It now follows from our hypothesis and lemma 8.2 that there exists a $\gamma \in \mathbb{R}$ such that $\mathbb{P}^Q_x\{Y = \gamma\} = 1$ for all $x \in S$. Hence $h(x) = \mathbb{E}^Q_x(Y) = \gamma$ for all $x \in S$. Thus $h$ is constant, as was to be shown.

(ii) $\implies$ (iii). Let $h$ be any increasing function in $L_1(\pi)$. Without loss of generality, we assume that $\int h \, d\pi = 0$. Define

$$E_h := \left\{ \liminf_n \frac{1}{n} \sum_{t=1}^{n} h(X_t) \geq 0 \right\}$$

and $H(x) := \mathbb{P}^Q_x(E_h)$. It is clear that $E_h$ is shift-invariant, and hence, by theorem 17.1.3 of Meyn and Tweedie (2009), the function $H$ is invariant in the sense of (3). From the fact that $h$ is increasing, the set $E_h$ is increasing on $S^\infty$. Using the hypothesis that $Q$ is increasing and applying lemma 8.1, we see that $H$ is increasing. Evidently $H$ is bounded. It now follows from (ii) that $H$ is constant, with $H(x) \equiv \alpha$ for some $\alpha \in [0, 1]$.

Seeking a contradiction, suppose that $\alpha < 1$. In view of theorem 17.1.2 of Meyn and Tweedie (2009), there exists a measurable function $f \colon S \to \mathbb{R}$ and a set $F_h \in \mathscr{B}$ such that

(a) $\int f(x)\pi(dx) = 0$
(b) $\pi(F_h) = 1$
(c) $\mathbb{P}^Q_x \left\{ \liminf_n \frac{1}{n} \sum_{t=1}^n h(X_t) = f(x) \right\} = 1$ for all $x \in F_h$.

Fix $x \in F_h$. Since $\alpha < 1$, we have

$$\mathbb{P}^Q_x \left\{ \liminf_n \frac{1}{n} \sum_{t=1}^{n} h(X_t) < 0 \right\} = 1 - H(x) = 1 - \alpha > 0.$$

In conjunction with (c), this implies that

$$\left\{ \liminf_n \frac{1}{n} \sum_{t=1}^{n} h(X_t) < 0 \right\} \cap \left\{ \liminf_n \frac{1}{n} \sum_{t=1}^{n} h(X_t) = f(x) \right\} \neq \emptyset.$$

Hence $f(x) < 0$. Since $x \in F_h$ was arbitrary, we have $f < 0$ on $F_h$. From (b) we have $\pi(F_h) = 1$, so

$$\int f(x)\pi(dx) = \int_{F_h} f(x)\pi(dx) < 0.$$

This inequality is impossible by (a).

We have now contradicted $\alpha < 1$, which implies that $H$ is everywhere equal to 1. In other words,

$$\mathbb{P}_x^Q \left\{ \liminf_n \frac{1}{n} \sum_{t=1}^n h(X_t) \geq 0 \right\} = 1, \qquad \forall x \in S.$$

A symmetric argument shows that $\mathbb{P}_x^Q \left\{ \limsup_n n^{-1} \sum_{t=1}^n h(X_t) \leq 0 \right\} = 1$ for all $x \in S$.[16] The claim in (iii) now follows.

(iii) $\implies$ (i). Let $A$ be increasing and shift-invariant. Let $h(x) := \mathbb{P}_x^Q(A)$. Our aim is to show that $h$ is constant and equal to either zero or one. Fixing $x \in S$ and applying theorem 17.1.3 of Meyn and Tweedie (2009), we can write $\mathbb{1}_A = \lim_t h(X_t)$, where equality holds $\mathbb{P}_x^Q$-a.s. As a consequence,

$$\mathbb{1}_A = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^n h(X_t).$$

Since $A$ and $Q$ are both increasing, lemma 8.1 tells us that $h$ is increasing. Clearly it is $\pi$-integrable. Applying (iii), we see that $\mathbb{1}_A = \int h \, d\pi$ holds $\mathbb{P}_x^Q$-a.s. In particular, the indicator of $A$ is constant $\mathbb{P}_x^Q$-a.s., and the value of the constant does not depend on $x$. Being an indicator, the constant value is either zero or one. Hence either $h = 0$ or $h = 1$. $\square$

*Proof of proposition 3.1.* Suppose that $Q$ is increasing and monotone ergodic on $(S, \preceq)$, and that $\pi_1$ and $\pi_2$ are both stationary for $Q$. Since a sequence cannot converge almost surely to two different limits, theorem 3.1 implies that $\int h d\pi_1 = \int h d\pi_2$ for every bounded measurable increasing function $h$ from $S$ to $\mathbb{R}$. Moreover, the Polish assumption implies that if $\pi_1$ and $\pi_2$ are two probability measures on $\mathscr{B}$ satisfying this condition, then $\pi_1 = \pi_2$. See, for example, theorem 2 of Kamae et al. (1978). $\square$

*Proof of corollary 3.1.* . Fix $x \in S$ and $h \in \mathscr{L}$. As per footnote 7, we can write $h$ as $h = h_1 - h_2$, where $h_1$ and $h_2$ are increasing and $\pi$-integrable. By theorem 3.1, for $h_1$ and $h_2$ there exist events $F_1$ and $F_2$ with $\mathbb{P}_x^Q(F_i) = 1$ and $n^{-1} \sum_t^n h_i(X_t) \to \int h_i d\, \pi$ on $F_i$. Setting $F := F_1 \cap F_2$ and applying linearity, we obtain $n^{-1} \sum_t^n h(X_t) \to \int hd\, \pi$

[16] In this case, the analogous function $H$ is bounded and invariant, but decreasing rather than increasing. Under (ii), such a function is also constant, because $-H$ is bounded, invariant and increasing. The rest of the argument is essentially the same.

on $F$. Evidently $\mathbb{P}^Q_x(F) = 1$. Hence (4) holds with $\mu = \delta_x$ for any $x \in S$. This extends to general $\mu$ via the identity

$$\mathbb{P}^Q_\mu(B) = \int \mathbb{P}^Q_x(B)\mu(dx) \quad \text{for all} \quad B \in \mathscr{B}^\infty \text{ and } \mu \in \mathscr{P}.$$

(The last equality can be obtained via a generating class argument applied to (2).)

□

Now we turn to the proof of theorem 3.2. In the proof, we let $ic(S, [0,1])$ be the functions in $icbS$ taking values in $[0,1]$. As usual, $\mu_n \xrightarrow{w} \mu$ means that $\langle \mu_n, f\rangle \to \langle \mu, f\rangle$ for all $f \in cbS$. Also, we require the following definition: Letting $\mathscr{G}$ and $\mathscr{H}$ be sets of bounded measurable functions, we say that $\mathscr{H}$ is *monotonically approximated by* $\mathscr{G}$ if, for all $h \in \mathscr{H}$, there exist sequences $\{g^1_n\}$ and $\{g^2_n\}$ in $\mathscr{G}$ with $g^1_n \uparrow h$ and $g^2_n \downarrow h$ pointwise. The proofs of the next two lemmas are given at the end of this section.

**Lemma 8.3.** *If $\mathscr{H}$ is monotonically approximated by $\mathscr{G}$, then $\mathscr{G}$ is convergence determining for $\mathscr{H}$, in the sense that if $\{\nu_n\}$ and $\nu$ are elements of $\mathscr{P}$, and $\langle \nu_n, g\rangle \to \langle \nu, g\rangle$ for all $g \in \mathscr{G}$, then $\langle \nu_n, h\rangle \to \langle \nu, h\rangle$ for all $h \in \mathscr{H}$.*

**Lemma 8.4.** *If the conditions of theorem 3.2 hold, then there exists a countable class $\mathscr{G}$ such that $\mathbb{P}^Q_x\{n^{-1}\sum_{t=1}^n g(X_t) \to \int g\, d\pi\} = 1$ for every $g \in \mathscr{G}$, and, moreover, $ic(S, [0,1])$ is monotonically approximated by $\mathscr{G}$.*

*Proof of theorem 3.2.* Fix $x \in S$. Let $\pi_n$ be the empirical distribution. As a first step of the proof, we claim that $\{\pi_n\}$ is tight with probability one.[17] To see this, fix $\varepsilon > 0$, and let $K$ be a compact subset of $S$ with $\pi(K) \geq 1 - \varepsilon$. By assumption, compact subsets of $S$ are order bounded, and so we have $a, b \in S$ with $K \subset [a, b]$. Let $I := \{y \in S : a \preceq y\}$ and $J := \{y \in S : y \preceq b\}$. Evidently

$$\pi_n([a,b]) = \pi_n(I \cap J) \geq \pi_n(I) + \pi_n(J) - 1. \tag{13}$$

Note that both $I$ and $J$ are increasing. By corollary 3.1, we can take $F_a$ to be a subset of $S^\infty$ with $\mathbb{P}^Q_x(F_a) = 1$ and $\pi_n(I) \to \pi(I)$ on $F_a$; and $F_b \subset S^\infty$ with $\mathbb{P}^Q_x(F_b) = 1$

[17] Recall that $\{\mu_n\} \subset \mathscr{P}$ is called *tight* if, for all $\varepsilon > 0$, there exists a compact $K \subset S$ such that $\mu_n(K) \geq 1 - \varepsilon$ for all $n$.

and $\pi_n(J) \to \pi(J)$ on $F_b$. It follows from (13) that on $F := F_a \cap F_b$ we have

$$\liminf_{n\to\infty} \pi_n([a,b]) \geq \pi(I) + \pi(J) - 1 \geq 2\pi(K) - 1 \geq 1 - \varepsilon.$$

Since closed and bounded order intervals are compact by assumption, it follows that $\{\pi_n\}$ is tight on the probability one set $F$.

As the second step of the proof, we claim there exists a probability one set $F'$ such that, for any given $\omega \in F'$, we have $\langle \pi_n^\omega, f\rangle \to \langle \pi, f\rangle$ for all $f \in icbS$. To see that this is so, let $\mathscr{G}$ be as in lemma 8.4. Since $\mathscr{G}$ is countable and the law of large numbers holds for every element of $\mathscr{G}$, there exists a probability one set $F' \subset \Omega$ such that, for each $\omega \in F'$, we have $\langle \pi_n^\omega, g\rangle \to \langle \pi, g\rangle$ for all $g \in \mathscr{G}$. Fix $\omega \in F'$. Since $ic(S,[0,1])$ is monotonically approximated by $\mathscr{G}$, lemma 8.3 implies that $\langle \pi_n^\omega, f\rangle \to \langle \pi, f\rangle$ for all $f \in ic(S,[0,1])$. It immediately follows that $\langle \pi_n^\omega, f\rangle \to \langle \pi, f\rangle$ for all $f \in icbS$.[18]

Now let $F''$ be the probability one set $F \cap F'$. For any $\omega \in F''$, the sequence of distributions $\{\pi_n^\omega\}$ is tight, and satisfies $\langle \pi_n^\omega, f\rangle \to \langle \pi, f\rangle$ for all $f \in icbS$. In view of lemma 6.6 of Kamihigashi and Stachurski (2014), we then have $\langle \pi_n^\omega, f\rangle \to \langle \pi, f\rangle$ for all $f \in cbS$. This concludes the proof of theorem 3.2. □

8.3. **Proofs from Section 4.** In this section we give the proof of proposition 4.1. The strategy is to first show that condition 4.5 implies monotone ergodicity, and then show that conditions 4.1–4.4 all imply condition 4.5.

*Proof of proposition 4.1.* First we show that if condition 4.5 is satisfied then $Q$ is monotone ergodic. To see this let $h \in ibS$ be invariant, and let $x$ and $x'$ be any two points in $S$. We aim to show that $h(x) = h(x')$, and hence that $h$ is constant. To this end, let $\{X_t\}$ and $\{X_t'\}$ be independent $Q$-Markov processes defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$, with $X_0 = x$ and $X_0' = x'$. Since $h$ is bounded and invariant, both $\{h(X_t)\}$ and $\{h(X_t')\}$ are bounded martingales. By the martingale convergence theorem, there exist random variables $Y$ and $Y'$ such that $h(X_t) \to Y$ and $h(X_t') \to Y'$ $\mathbb{P}$-almost surely.

[18] If $f \in icbS$, then there exists a $g \in ic(S,[0,1])$ and constants $a, b$ such that $f = a + bg$.

Let $\{X_t \preceq X'_t \text{ i.o.}\}$ be the event that $X_t \preceq X'_t$ occurs infinitely often.[19] Since $Q$ is order mixing, $X_t \preceq X'_t$ at least once with probability one. As shown in proposition 9.1.1 of Meyn and Tweedie (2009), this in turn implies the seemingly stronger result $\mathbb{P}\{X_t \preceq X'_t \text{ i.o.}\} = 1$. Since $h$ is increasing, this implies that $\{h(X_t) \leq h(X'_t) \text{ i.o.}\}$ has probability one. It now follows that $Y \leq Y'$ holds $\mathbb{P}$-a.s., and hence $\mathbb{E}\, Y \leq \mathbb{E}\, Y'$.

By the dominated convergence theorem and the martingale property, we have $\mathbb{E}\, Y = \mathbb{E}\, h(X_t) = \mathbb{E}\, h(X_0) = h(x)$. Similarly, $\mathbb{E}\, Y' = h(x')$. We have now shown that $h(x) \leq h(x')$. A symmetric argument gives $h(x') \leq h(x)$, as can be seen by swapping the roles of $X_t$ and $X'_t$ in the proof above. We conclude that $h(x) = h(x')$, as was to be shown.

Next we claim that conditions 4.1–4.4 all imply condition 4.5. That is is true for condition 4.1 was established in section 4.1 of Kamihigashi and Stachurski (2012). That condition 4.2 implies condition 4.5 is immediate from remark 2.4 and lemma 6.5 of Kamihigashi and Stachurski (2014). That condition 4.4 implies condition 4.5 follows from lemma 6.5 of Kamihigashi and Stachurski (2014).

That condition 4.3 implies condition 4.5 is more subtle. We prove that condition 4.3 implies condition 4.4, which, as shown above, implies condition 4.5. In verifying condition 4.4, note that $Q$ is increasing by assumption, and that uniform asymptotic tightness clearly implies boundedness in probability. Hence it remains only to show that $Q$ is order reversing under condition 4.3.

For the rest of this proof let $Q^n(x, B) := \mathbb{P}^Q_x\{X_n \in B\}$ for all $x \in S$, $n \in \mathbb{N}$ and $B \in \mathscr{B}$. Fix $x$ and $x'$ in $S$ with $x \leq x'$, let $\{X_t\}$ and $\{X'_t\}$ be independent Markov processes generated by $Q$ and starting and $x$ and $x'$ respectively. Let $c$ be as in the definition of weak mixing. As a first step, we claim that

$$\exists\, j \in \mathbb{N} \text{ s.t. } Q^{nj}(x, (c, \infty)) > 0, \quad \forall\, n \in \mathbb{N}. \tag{14}$$

[19] That is, $\{X_t \preceq X'_t \text{ i.o}\} := \bigcap_{m=0}^{\infty} \bigcup_{t \geq m} \{X_t \preceq X'_t\}$.

To see that this is so, define $a := \min\{x, c\}$. By weak mixing there is a $j \in \mathbb{N}$ with $Q^j(a, (c, \infty)) > 0$. Now note that, by the Chapman-Kolmogorov equations,

$$\begin{aligned} Q^{2j}(a, (c, \infty)) &= \int Q^j(a, dy) Q^j(y, (c, \infty)) \\ &\geq \int \mathbb{1}\{y > c\} Q^j(a, dy) Q^j(y, (c, \infty)). \end{aligned}$$

Since $y > c$ implies that $y > a$, $Q$ is increasing and $Q^j(a, (c, \infty)) > 0$, it follows that $Q^j(y, (c, \infty))$ is strictly positive on $\{y > c\}$. Moreover, $Q^j(a, dy)$ puts positive measure on $\{y > c\}$. Hence the integral is strictly positive, and $Q^{2j}(a, (c, \infty)) > 0$ is established. An induction argument generalizes this result to all $n$, and (14) is established. A symmetric argument shows that $\exists \, k \in \mathbb{N}$ with $Q^{nk}(x', (-\infty, c)) > 0$ for all $n \in \mathbb{N}$. Combining this result and (14), we see that for $t = jk$ we have

$$Q^t(x', (-\infty, c)) \cdot Q^t(x, (c, \infty)) > 0. \tag{15}$$

Finally, since $\{X_t\}$ and $\{X'_t\}$ are independent, we obtain

$$\mathbb{P}\{X'_t \leq X_t\} \geq \mathbb{P}\{X'_t < c < X_t\} = \mathbb{P}\{X'_t < c\}\mathbb{P}\{c < X_t\}.$$

Combined with (15) this shows that $Q$ is order reversing as claimed. $\square$

### 8.4. **Proofs from Section 5.**

*Proof of proposition 5.1.* Let $\{\varepsilon_t\}$ and $\{\varepsilon'_t\}$ be IID draws from $\phi$ and independent of each other. Consider first condition (iii). We claim that $Q_F$ is order reversing (recall the proof of proposition 4.1. To see this, fix $x' \preceq x$. Let $\{\varepsilon_t\}_{t=1}^k$ and $\{\varepsilon'_t\}_{t=1}^k$ be as in the statement of the proposition. Define the constant

$$\gamma := \mathbb{P}\{F^k(x, \varepsilon_1, \ldots, \varepsilon_k) < F^k(x', \varepsilon'_1, \ldots, \varepsilon'_k)\}.$$

We aim to show that $\gamma > 0$. By hypothesis, $F^k(x, \varepsilon_1, \ldots, \varepsilon_k) < F^k(x', \varepsilon'_1, \ldots, \varepsilon'_k)$. By continuity of $F$, there exist open neighborhoods $N_t$ of $\varepsilon_t$ and $N'_t$ of $\varepsilon'_t$ such that

$$\tilde{\varepsilon}_t \in N_t \text{ and } \tilde{\varepsilon}'_t \in N'_t \text{ for } t \in \{1, \ldots, k\} \implies F^k(x, \tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_k) < F^k(x', \tilde{\varepsilon}'_1, \ldots, \tilde{\varepsilon}'_k).$$

This leads to the estimate

$$\gamma \geq \mathbb{P} \cap_{t=1}^n \{\varepsilon_t \in N_t \text{ and } \varepsilon'_t \in N'_t\} = \prod_{t=1}^n \phi(N_t)\phi(N'_t).$$

Since $E$ is the support of $\phi$, this last term is positive, and $\gamma > 0$.

The inequality $\gamma > 0$ tells us directly that $Q_F$ is order reversing. Since $Q_F$ is also increasing and bounded in probability, lemma 6.5 of Kamihigashi and Stachurski (2014) implies that $Q_F$ is order mixing. Existence of a stationary distribution follows from theorem 3.2 of the same reference.

The proof of the proposition under conditions (i)–(ii) is similar. For example, an argument similar to the one just given shows that condition (i) implies that $Q_F$ is downward reaching in the sense of Kamihigashi and Stachurski (2014). The order reversing property then follows from Kamihigashi and Stachurski (2014), proposition 3.2, and the rest of the arguments are unchanged. $\square$

8.5. **Remaining Proofs.** Finally, we complete the proofs of lemmas 8.2–8.4.

*Proof of lemma 8.2.* Assume the conditions of the lemma. In particular, let $i\mathscr{G}$ be trivial, and let $Y$ be increasing and $\mathscr{G}$-measurable. Fixing $c \in \mathbb{R}$, let $F_x(c) := \mathbb{P}^Q_x\{Y \leq c\}$. Given the assumptions on $Y$, the set $\{Y \leq c\}$ is decreasing and in $\mathscr{G}$. Sinc $i\mathscr{G}$ is trivial, the decreasing sets in $\mathscr{G}$ must also be trivial.[20] Hence the distribution function $F_x(c)$ is either zero or one. Letting $\gamma := \inf\{c \in \mathbb{R} : F_x(c) = 1\}$ and applying right-continuity, we have $F_x(\gamma) = 1$ and $F_x(c) = 0$ for any $c < \gamma$. Hence $\mathbb{P}^Q_x\{Y = \gamma\} = 1$. By the definition of triviality, $\gamma$ does not depend on $x$. $\square$

*Proof of lemma 8.3.* Let $\{\nu_n\}$ and $\nu$ be probability measures on $S$, and suppose that $\langle \nu_n, g\rangle \to \langle \nu, g\rangle$ for all $g \in \mathscr{G} \subset bS$. We claim that $\langle \nu_n, h\rangle \to \langle \nu, h\rangle$ for all $h \in \mathscr{H} \subset bS$. To see this, pick any $h \in \mathscr{H}$, and choose sequences $\{g^1_n\}$ and $\{g^2_n\}$ in $\mathscr{G}$ with $g^1_n \uparrow h$ and $g^2_n \downarrow h$. Clearly

$$\liminf_n \langle \nu_n, h\rangle \geq \liminf_n \langle \nu_n, g^1_k\rangle = \langle \nu, g^1_k\rangle \quad \text{for all } k.$$

$$\therefore \quad \liminf_n \langle \nu_n, h\rangle \geq \sup_k \langle \nu, g^1_k\rangle = \lim_k \langle \nu, g^1_k\rangle = \langle \nu, h\rangle.$$

A symmetric argument applied to $\{g^2_n\}$ yields $\limsup_n \langle \nu_n, h\rangle \leq \langle \nu, h\rangle$. $\square$

*Proof of lemma 8.4.* Let $A$ be the countable subset of $S$ in assumption 3.1. For $a \in A$, let $I_a := \mathbb{1}\{y \in S : a \preceq y\}$. Let $\mathscr{K}$ be the set of functions $\ell = rI_a$ for some

[20] Just observe that if $D \in \mathscr{G}$ is decreasing, then $D^c$ is increasing, and hence $h(x) = \mathbb{P}^Q_x(D^c) = 1 - \mathbb{P}^Q_x(D)$ is constant in $\{0, 1\}$. The claim follows.

$r \in \mathbb{Q} \cap [0,1]$ and $a \in A$. Let $\mathscr{G}_1$ be all functions $g = \max_{\ell \in F} \ell$ where $F \subset \mathscr{K}$ is finite. Clearly $\mathscr{G}_1$ is countable, and, by theorem 3.1, every $g \in \mathscr{G}_1$ satisfies $\mathbb{P}_x^Q\{n^{-1}\sum_{t=1}^n g(X_t) \to \int g \, d\pi\} = 1$. We claim that for each $f \in ic(S, [0,1])$ there exists a sequence $\{g_n\}$ in $\mathscr{G}_1$ converging up to $f$. To verify this claim it suffices to show that

$$\sup\{\ell(x) : \ell \in \mathscr{K} \text{ and } \ell \leq f\} = f(x) \quad \text{for any } x \in S. \tag{16}$$

Indeed, if (16) is valid, then take $\{\ell_k\}$ to be an enumeration of all $\ell \in \mathscr{K}$ with $\ell \leq f$ and choose $g_n = \max_{1 \leq k \leq n} \ell_k$.

To establish (16), fix $x \in S$ and $\varepsilon > 0$. By continuity of $f$ and assumption 3.1, we can find an $a \in A$ with $a \preceq x$ and $f(x) - \varepsilon < f(a)$. Let $r \in \mathbb{Q}$ be such that $f(x) - \varepsilon < r < f(a)$ and let $\ell(x) := rI_a$. Since $\ell \leq f(a)I_a$ and $f$ is increasing we have $\ell \leq f$. On the other hand, $f(x) - \varepsilon < r = \ell(a) \leq \ell(x)$. Since $\varepsilon$ was arbitrary we conclude that (16) is valid.

To complete the proof of lemma 8.4, we show existence of a class of functions $\mathscr{G}_2$ such that $\mathscr{G}_2$ is countable, every $g \in \mathscr{G}_2$ satisfies $\mathbb{P}_x^Q\{n^{-1}\sum_{t=1}^n g(X_t) \to \int g \, d\pi\} = 1$, and, for each $f \in ic(S, [0,1])$, there exists a sequence $\{g_n\}$ in $\mathscr{G}_2$ converging down to $f$. The claim in lemma 8.4 is then satisfied with $\mathscr{G} := \mathscr{G}_1 \cup \mathscr{G}_2$. We omit the details, since the construction of $\mathscr{G}_2$ is entirely symmetric to the construction of $\mathscr{G}_1$. $\square$

## References


Aghion, Philippe and Patrick Bolton (1997) "A theory of trickle-down growth and development," *The Review of Economic Studies*, Vol. 64, pp. 151–172.

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu (2011) "The distribution of wealth and fiscal policy in economies with finitely lived agents," *Econometrica*, Vol. 79, pp. 123–157.

Benhabib, Jess and Chetan Dave (2013) "Learning, large deviations and rare events," *Review of Economic Dynamics*.

Bhattacharya, Rabi and Mukul Majumdar (2001) "On a class of stable random dynamical systems: theory and applications," *Journal of Economic Theory*, Vol. 96, pp. 208–229.

Bhattacharya, Rabi N and Oesook Lee (1988) "Asymptotics of a class of Markov processes which are not in general irreducible," *The Annals of Probability*, pp. 1333–1347.

Brandt, Andreas (1986) "The stochastic equation Yn+ 1= AnYn+ Bn with stationary coefficients," *Advances in Applied Probability*, pp. 211–220.

Breiman, Leo (1992) *Probability*: Society for Industrial and Applied Mathematics, Pennsylvania.

Brock, William A and Steven N Durlauf (2001) "Discrete choice with social interactions," *The Review of Economic Studies*, Vol. 68, pp. 235–260.

Brock, William A and Leonard Jay Mirman (1972) "Optimal economic growth and uncertainty: The discounted case," *Journal of Economic Theory*, Vol. 4, pp. 479–513.

Cabrales, Antonio and Hugo A Hopenhayn (1997) "Labor-market flexibility and aggregate employment volatility," in *Carnegie-Rochester Conference Series on Public Policy*, Vol. 46, pp. 189–228, Elsevier.

Carroll, Christopher D (1997) "Buffer-stock saving and the life cycle/permanent income hypothesis," *The Quarterly Journal of Economics*, Vol. 112, pp. 1–55.

Cooley, Thomas F and Vincenzo Quadrini (2001) "Financial markets and firm dynamics," *American Economic Review*, pp. 1286–1310.

De Hek, Paul A (1999) "On endogenous growth under uncertainty," *International Economic Review*, Vol. 40, pp. 727–744.

Duffie, Darrell, John Geanakoplos, Andreu Mas-Colell, and Andrew McLennan (1994) "Stationary markov equilibria," *Econometrica*, pp. 745–781.

Duffie, Darrell and Kenneth J Singleton (1993) "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, pp. 929–952.

Farmer, Roger EA, Daniel F Waggoner, and Tao Zha (2009) "Indeterminacy in a forward-looking regime switching model," *International Journal of Economic Theory*, Vol. 5, pp. 69–84.

Geweke, John (2005) *Contemporary Bayesian econometrics and statistics*: Wiley.

Greenwood, Jeremy and Gregory W Huffman (1995) "On the existence of nonoptimal equilibria in dynamic stochastic economies," *Journal of Economic Theory*, Vol. 65, pp. 611–623.

Hansen, Lars Peter (1982) "Large sample properties of generalized method of moments estimators," *Econometrica*, pp. 1029–1054.

Hansen, Lars Peter and Thomas J Sargent (2010) "Wanting Robustness in Macroeconomics," *Handbook of Monetary Economics*, Vol. 3, pp. 1097–1157.

Hopenhayn, Hugo A and Edward C Prescott (1992) "Stochastic monotonicity and stationary distributions for dynamic economies," *Econometrica*, pp. 1387–1406.

Huggett, Mark (1993) "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of Economic Dynamics and Control*, Vol. 17, pp. 953–969.

Kamae, Teturo, Ulrich Krengel et al. (1978) "Stochastic partial ordering," *The Annals of Probability*, Vol. 6, pp. 1044–1049.

Kamae, Teturo, Ulrich Krengel, and George L O'Brien (1977) "Stochastic inequalities on partially ordered spaces," *The Annals of Probability*, pp. 899–912.

Kamihigashi, Takashi (2007) "Stochastic optimal growth with bounded or unbounded utility and with bounded or unbounded shocks," *Journal of Mathematical Economics*, Vol. 43, pp. 477–500.

Kamihigashi, Takashi and John Stachurski (2012) "An order-theoretic mixing condition for monotone Markov chains," *Statistics & Probability Letters*, Vol. 82, pp. 262–267.

——— (2014) "Stochastic stability in monotone economies," *Theoretical Economics*, Vol. 9, pp. 383–407.

Ljungqvist, Lars and Thomas J Sargent (2012) *Recursive macroeconomic theory*, 3rd edition.

Meyn, Sean P and Richard L Tweedie (2009) *Markov chains and stochastic stability*: Cambridge University Press, 2nd edition.

Morand, Olivier F and Kevin L Reffett (2007) "Stationary Markovian equilibrium in overlapping generation models with stochastic nonclassical production and Markov shocks," *Journal of Mathematical Economics*, Vol. 43, pp. 501–522.

Nirei, Makoto (2006) "Threshold behavior and aggregate fluctuation," *Journal of Economic Theory*, Vol. 127, pp. 309–322.

Owen, Ann L and David N Weil (1998) "Intergenerational earnings mobility, inequality and growth," *Journal of Monetary Economics*, Vol. 41, pp. 71–104.

Piketty, Thomas (1997) "The dynamics of the wealth distribution and the interest rate with credit rationing," *The Review of Economic Studies*, Vol. 64, pp. 173–189.

Razin, Assaf and Joseph A Yahav (1979) "On stochastic models of economic growth," *International Economic Review*, Vol. 20, pp. 599–604.

Rossi-Hansberg, Esteban and Mark LJ Wright (2007) "Establishment size dynamics in the aggregate economy," *The American Economic Review*, Vol. 97, pp. 1639–1666.

Santos, Manuel S and Adrian Peralta-Alva (2005) "Accuracy of simulations for stochastic dynamic models," *Econometrica*, Vol. 73, pp. 1939–1976.

Stokey, Nancy and Robert E Lucas (1989) *Recursive Methods in Economic Dynamics (with EC Prescott)*: Harvard University Press.

Szeidl, Adam (2013) "Invariant distribution in buffer-stock saving and stochastic growth models."

Zhang, Yuzhe (2007) "Stochastic optimal growth with a non-compact state space," *Journal of Mathematical Economics*, Vol. 43, pp. 115–129.