Discussion Paper Series
RIEB
Kobe University

DP2014-22

Seeking Ergodicity in Dynamic Economies

Takashi KAMIHIGASHI
John STACHURSKI

May 19, 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

May 9, 2014

ABSTRACT. In estimation and calibration studies the concept of ergodicity plays a fundamental role. At the same time, a significant number of economic models do not satisfy the classical ergodicity conditions. Motivated by existing work on economic dynamics, we develop a new set of results on ergodicity using an order-theoretic approach. Our conditions are necessary and sufficient, and, by varying the notion of order, can include the classical Markov ergodic theorem as a special case. We discuss implications, sufficient conditions and economic applications.

JEL Classifications: C62, C63
Keywords: Ergodicity, consistency, calibration

1. INTRODUCTION

One of the most fundamental ways to connect theory with data is to match sample averages with population means. In static cross-sectional models this can usually be justified by appealing to the law of large numbers for independent random variables (with obvious exceptions—see, for example, Brock and Durlauf (2001) or Nirei (2006)). In the case of dynamic models, convergence of sample averages may or may not hold. The most general approach to this problem is via the concept

[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`

of ergodicity, which represents the notion that, in the limit, time series and cross-sectional averages coincide. This typically requires some form of asymptotic path independence, which in turn depends on the primitives that define the economic system, the kinds of shocks that affect it, and how agents react to these shocks.

The concept of ergodicity forms a foundation stone at the heart of quantitative economics, supporting a huge variety of computations and theoretical results. Consistency of estimators is an obvious example (see, e.g., Hansen (1982)), and simulation of stationary equilibria is another (e.g., Santos and Peralta-Alva (2005)). Ergodicity is likewise fundamental to calibration and most forms of simulation-based time series estimation (e.g., 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 ergodicity (see, e.g., Geweke (2005)).

The majority of dynamic models used in quantitative economic modeling are recursive. In this setting, perhaps the best known ergodicity result is the classical Markov ergodic theorem. For a Markov process $\{X_t\}$ with stationary distribution π, the theorem gives necessary and sufficient conditions under which

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} h(X_t) = \int h(x)\, \pi(dx) \tag{1}$$

almost surely for any π-integrable function h and any initial condition X_0.[2]

This is a powerful result, implying probability one convergence over an extremely wide class of functions h. In fact the result is in some sense too strong, in that it fails to hold for some well known economic models. For example, the convergence result in (1) cannot be established under the stated assumptions for the capital and income processes in the canonical stochastic optimal growth model of Brock

[2]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 π, and that π 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). For these reasons we focus our attention on the version of ergodicity considered in (1), although similar ideas can be applied to other versions.

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. In fact the problem is relatively general, and easy to illustrate. To this end, consider the simple dynamic system given by

$$X_{t+1} = \alpha X_t + \xi_{t+1}. \tag{2}$$

Here the state space is $\mathbb{R}$ and $\{\xi_t\}$ is IID. Assume α is a rational number in $(0, 1)$, and that ξ_t is rational with probability one, as is the case for many discrete distributions. Now let $h = \mathbb{1}_{\mathbb{Q}}$, the indicator function of the rationals, so that $h(x) = 1$ if x is rational and zero otherwise. In this setting, it is clear that if $X_0 \in \mathbb{Q}$, then $X_t \in \mathbb{Q}$ for all t, and hence $\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} h(X_t) = 1$. On the other hand, if X_0 is irrational, then so is X_t for all t, and hence $\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} h(X_t) = 0$. Thus (1) fails. The model is in some sense path dependent, in that the limit of the sample average depends on the value of the initial draw.

While the nature of this counterexample is relatively specific, the discussion is instructive vis-à-vis the economic applications discussed above. Just like those economic models, the system in (2) is essentially well behaved. It is mean reverting, with a unique stationary distribution (see, e.g., Bhattacharya and Majumdar (2003), theorem 3.2). In fact the instability in the sample mean we observe in this example is mainly caused by the irregularity of the function $h = \mathbb{1}_{\mathbb{Q}}$ we chose to test convergence. This function is neither continuous nor monotone on any open interval. For many regular functions the convergence in (1) will in fact hold. This point is salient from the perspective of economic applications, since sample mean convergence for highly irregular functions is typically irrelevant for the purpose of economic modeling. What matters for quantitative work is convergence for more regular functions, such as centered and uncentered moments, generalized moment conditions from Euler equations and so forth.

These considerations motivate us to generalize the classical Markov ergodic theorem along the following lines: We embed the dynamic model in an abstract partially ordered space, and prove that a natural extension of the Markov ergodic theorem holds in this setting. By controlling the partial order on the space, it becomes possible to control the set of test functions h for which the convergence (1) is required to hold. Under one particular choice of partial order, the set of test functions expands to the extent that we recover the classical theorem as a special case. Under other partial orders the set of test functions is restricted to more well behaved functions, as might be considered in certain economic applications. This allows us to treat a variety of models that fail to satisfy the classical conditions.

There are other possible approaches to the problem described above. Our decision to add structure via a notion of order is motivated by earlier order-theoretic work on the asymptotics of Markov models, as found, for example, in Razin and Yahav (1979), Stokey and Lucas (1989), Hopenhayn and Prescott (1992), Bhattacharya and Majumdar (2001) and Bhattacharya and Majumdar (2007). The usefulness of this approach to modeling economic dynamics is described at length in Stokey and Lucas (1989) and Bhattacharya and Majumdar (2007). By drawing connections with this line of research, our results are seen to extend to a variety of well known models.

In addition to the general results discussed above, this paper provides further results related to implications of the theory. We show that, under some additional restrictions on the state space, the empirical distribution associated with any sample converges to the stationary distribution with probability one. We also discuss sufficient conditions, providing conditions that can be used to check the conditions of the theorem in new applications.

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

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^∞ are called time series. A typical probability distribution on S^∞ is denoted by $\mathbb{P}_x^Q$. 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.[3]

2.1. **Model Primitives.** To begin, let $(S, \mathscr{B})$ be a measurable space, and let $\preceq$ be a partial order on S (i.e., reflexive, transitive and antisymmetric). Let $\mathscr{P}$ be the probability measures on $(S, \mathscr{B})$. Let $S^\infty := S \times S \times \cdots$, and let $\mathscr{B}^\infty$ be the product σ-algebra. 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 of S is called increasing if its indicator function is increasing, and decreasing if its indicator is decreasing.

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 the current state X_t is drawn from π and then X_{t+1} is drawn from $Q(X_t, \cdot)$, the distribution of X_{t+1} will again be π. As in many other studies (e.g., Brock and Mirman (1972), Stokey and Lucas (1989), Duffie et al.

[3]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.

(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 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.** 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

$$\mathbb{P}^Q_\mu(B_0 \times \cdots \times B_n \times S \times S \times \cdots) = \int_{B_0} \mu(dx_0) \int_{B_1} Q(x_0, dx_1) \cdots \int_{B_{n-1}} Q(x_{n-2}, dx_{n-1}) \int_{B_n} Q(x_{n-1}, dx_n) \tag{3}$$

for any finite collection $\{B_i\}_{i=0}^n \subset \mathscr{B}$.[4] In essence, $\mathbb{P}^Q_\mu$ is the joint distribution of the Markov process $\{X_t\}$ defined by drawing X_0 from μ 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 μ.

3. Ergodicity

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

[4] As is conventional, the integrals in (3) are computed from right to left, with the integrand written to the right of the integrating measure.

3.1. **Classical Ergodicity.** One way to understand ergodicity is via the behavior of invariant functions. Recall that a bounded measurable function $h\colon S \to \mathbb{R}$ is called **invariant** for Q if

$$\int h(y) Q(x, dy) = h(x) \tag{4}$$

for all $x \in S$. In essence, invariant functions convert Markov processes into martingales, because if h is invariant then $\mathbb{E}\left[h(X_{t+1}) \mid X_t\right] = \int h(y) Q(X_t, dy) = h(X_t)$. The converse is also true—a function h is invariant if $\{h(X_t)\}$ is a martingale for any choice of X_0.

Any constant function is clearly invariant. This is a trivial way to create a martingale. A stochastic kernel Q on S is said to be **ergodic** if this trivial method is the only way to create a martingale—more formally, if the only bounded invariant functions are the constant functions. (Note that definitions of ergodicity vary slightly, ranging over several equivalent and near-equivalent conditions. Our terminology largely follows Ljungqvist and Sargent (2012).) The classical Markov ergodic theorem runs as follows:

Theorem 3.1. *For any stochastic kernel Q with stationary distribution π, the following are equivalent:*

(i) *Q is ergodic.*

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

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

Here "π-integrable" means that $h\colon S \to \mathbb{R}$ is measurable, and $\int |h| \, d\pi < \infty$. We maintain this definition throughout. The proof of theorem 3.1 can be found in proposition 17.1.4 and theorem 17.1.7 of Meyn and Tweedie (2009), although we prove a more general result below.

3.2. **Monotone Ergodicity.** We now provide a generalization of theorem 3.1. By way of analogy with the standard definition, we will call a stochastic kernel Q **monotone ergodic** if the only *increasing* bounded invariant functions are the constant functions.

Theorem 3.2. *For any increasing stochastic kernel Q with stationary distribution π, the following conditions are equivalent:*

(i) *Q is monotone ergodic.*

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

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

The proof of theorem 3.2 is given in section 7. To see that theorem 3.2 is in fact a generalization of theorem 3.1, it suffices to set the partial order $\preceq$ on S to be 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 3.2 reduce to those of theorem 3.1.

On the other hand, theorem 3.2 is more general than theorem 3.1. This is intuitively clear, since we can choose any partial order. To give a concrete example, recall the simple model (2) considered in the introduction. As shown in the introduction, this model is not ergodic. At the same time, it is monotone ergodic under the natural order $\leq$ whenever the shock ξ_t is non-degenerate. The proof will be easier once we have developed some sufficient conditions, so we delay it to section 5.2.

More generally, 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 familiar economic applications, where the relevant monotonicity conditions are satisfied, and convergence of sample means only matters for relatively regular functions.

3.3. **Uniqueness.** As suggested by the statement of theorem 3.2, monotone ergodicity is not sufficient to yield existence of a stationary distribution π. A more subtle question is uniqueness. Uniqueness requires some additional structure on the order $\preceq$ and the measurable sets $\mathscr{B}$. This is the purpose of the following assumption.

Assumption 3.1. The space S is a separable and completely metrizable topological space (also called a Polish space). The partial order $\preceq$ is closed, in the sense that its graph is closed in the product space $S \times S$ when endowed with its product topology.

In this topological setting we always take $\mathscr{B}$ to be the Borel sets. Assumption 3.1 is very standard and relatively weak (cf., e.g., Hopenhayn and Prescott (1992)), but still sufficient for uniqueness.

Proposition 3.1. *Let Q be an increasing stochastic kernel. If Q is monotone ergodic and assumption 3.1 is satisfied, then Q has at most one stationary distribution.*

3.4. **Convergence for Non-Monotone Functions.** One apparent concern with theorem 3.2 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 3.2 may be too small to be useful. For example, we might care about convergence of the second moment, which requires us to set $h(x) = x^2$. This function is not monotone.

Fortunately, it turns out that the convergence in theorem 3.2 extends to a larger set of functions, without additional assumptions. For example, let Q be a fixed stochastic kernel with stationary distribution π. Let $\mathscr{L}$ denote the linear span of the set of increasing π-integrable functions.[5]

Corollary 3.1. *If the conditions of theorem 3.2 hold, then*

$$\mathbb{P}^Q_\mu \left\{ \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} h(X_t) = \int h \, d\pi \right\} = 1, \qquad \forall \mu \in \mathscr{P}, \forall h \in \mathscr{L}. \tag{5}$$

Note that corollary 3.1 considers $\mathbb{P}^Q_\mu$ rather than $\mathbb{P}^Q_x$, so the convergence also applies to random X_0 with arbitrary distribution.

[5] 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 π-integrable f and g.

Example 3.1. Let $S = \mathbb{R}$ and let k be the number of finite moments possessed by the stationary distribution π. All k moment functions $h(x) = x^k$ lie in $\mathscr{L}$, as does any polynomial of order k or less.[6]

Example 3.2. If S is a closed interval in $\mathbb{R}$, then $\mathscr{L}$ contains all functions of bounded variation (see, e.g., Shiryaev (1996), p. 207).

3.5. **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** π_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 π, 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).[7]

To discuss continuous functions we need a notion of topology. We begin by adopting assumption 3.1, and let $\mathscr{B}$ be the Borel sets. We also need the following:

---

[6]If k is odd, then $h(x) := x^k$ is increasing. If k is even, then write $h(x) = x^k$ as $-h_1(x) + h_2(x)$, where $h_1(x) := -x^k \mathbb{1}\{x < 0\}$ and $h_k(x) := x^k \mathbb{1}\{x \geq 0\}$. Both h_1 and h_2 are increasing functions. Hence $h(x) \in \mathscr{L}$. Finally, if $p(x)$ is a polynomial of the form $p(x) = \sum_{i=1}^{k} a^i x^i$, then, since $\mathscr{L}$ is closed under linear combinations, $p \in \mathscr{L}$ is also true.

[7]The statement $\int h \, d\pi_n \to \int h \, d\pi$ for all continuous bounded h with probability one is a 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.

Assumption 3.2. The space $(S, \preceq)$ is normally ordered,[8] 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 3.2 adds a significant amount of structure relative to assumption 3.1. It is however satisfied for many common state spaces, such as when $S = \mathbb{R}^m$ with its usual pointwise order $\leq$, or more generally, when S is a cone in $\mathbb{R}^m$ with the usual pointwise order.

Theorem 3.3. *If assumptions 3.1–3.2 are satisfied, and, in addition, Q is increasing and monotone ergodic with stationary distribution π, 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

In sections 3.1–3.2 we described the connection between monotone ergodicity and classical ergodic theory. As discussed in the introduction, there are also connections between monotone ergodicity and the order theoretic work on economic dynamics found in references such as Razin and Yahav (1979), Bhattacharya and Lee (1988), Stokey and Lucas (1989), Hopenhayn and Prescott (1992), Bhattacharya and Majumdar (2001), Szeidl (2013), and Kamihigashi and Stachurski (2014). These papers introduce various order theoretic "mixing conditions". In this section we

[8] A topological space with partial order $\preceq$ 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 $x \in I$. See, e.g., Whitt (1980).

show that monotone ergodicity in the sense of theorem 3.2 is more general than all of these conditions.

To begin the discussion, consider the "splitting condition" approach 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,

$$X_t = \gamma_t \circ \cdots \circ \gamma_1(X_0).$$

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

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

$$\mathbb{P}\{\gamma_k \circ \cdots \circ \gamma_1(y) \leq c,\ \forall y \in S\} > 0 \ \text{ and } \ \mathbb{P}\{\gamma_k \circ \cdots \circ \gamma_1(y) \geq c,\ \forall y \in S\} > 0.$$

Here $\leq$ is the usual pointwise order on $\mathbb{R}^m$. In Bhattacharya and Majumdar (2001) and Bhattacharya and Majumdar (2007) it is shown that the splitting condition applies to many economic applications. The authors then establish many significant results, including a central limit theorem for models that satisfy the splitting condition when all the maps γ_t are monotone. This implies a weak law of large numbers with $1/\sqrt{n}$ consistency. We now add the following:

Proposition 4.1. *If assumption 4.1 is satisfied, then Q is monotone ergodic on $(S, \leq)$.*

As a result, when the maps γ_n are increasing, proposition 4.1 combined with theorem 3.2 strengthen the results of Bhattacharya and Majumdar (2001) by adding a strong law of large numbers (theorem 3.2 and corollary 3.1) and convergence for continuous functions (theorem 3.3).

Note that monotone ergodicity is significantly more general than assumption 4.1. For example, the AR(1) process in (2) satisfies monotone ergodicity on $(\mathbb{R}, \leq)$ but not assumption 4.1 whenever the shock ξ_t is unbounded. The same is true for various other autoregressive time series models, the optimal accumulation models studied in Nishimura and Stachurski (2005) and Zhang (2007), the wealth distribution model in Benhabib et al. (2011), the buffer stock model in Szeidl (2013), and the price coefficient process in Benhabib and Dave (2013). Section 5.2 of this paper gives a detailed example that illustrates these ideas.

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:

Assumption 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:

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

Hopenhayn and Prescott (1992) show that assumption 4.2 is sufficient for the existence of a unique, stable stationary distribution. We add the following result:

Proposition 4.2. *If assumption 4.2 is satisfied, then Q is monotone ergodic.*

Proposition 4.2 is significant because many well known models have been shown to satisfy assumption 4.2. These include most versions of the standard neoclassical optimal growth model with bounded shocks, as studied by Brock and Mirman (1972), Mirman and Zilcha (1975) and Hopenhayn and Prescott (1992), as well as by the infinite horizon incomplete market models typified by Huggett (1993), stochastic endogenous growth models such as that found in De Hek (1999), a wide variety of OLG models, such as those as found in Aghion and Bolton (1997), Piketty (1997), Owen and Weil (1998) and Morand and Reffett (2007), and industry models such as Cabrales and Hopenhayn (1997) and Cooley and Quadrini (2001).

Szeidl (2013) generalizes a number of the ideas in Hopenhayn and Prescott (1992). He takes S to be an order interval in $\mathbb{R}^m$ and $\preceq$ is the usual pointwise order $\leq$. He defines a stochastic kernel Q to be **weakly mixing** if there exists a point $c \in S$ such that given any $x \in S$ we can find integers j and k with $\mathbb{P}_x^Q\{X_j > c\} > 0$ and $\mathbb{P}_x^Q\{X_k < c\} > 0$. 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 $x \in S$. Szeidl (2013) shows existence, uniqueness and stability of the stationary distribution when Q is increasing, weakly mixing, uniformly asymptotically tight, and an additional regularity condition holds. He goes on to show how the conditions can be applied to a number of useful models, including the buffer stock saving model of Carroll (1997). Here we show that the following is also true:

Proposition 4.3. *Let $(S, \leq)$ be an order interval in $\mathbb{R}^m$ with the usual pointwise order. If Q is increasing, weakly mixing and uniformly asymptotically tight, then Q is monotone ergodic.*

As a consequence, the strong law of large numbers from theorem 3.2 is valid in this setting, as are the conclusions of corollary 3.1 and theorem 3.3.

Finally, recall that a stochastic kernel Q on S is defined to be **order mixing** if, 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 (see, e.g., Kamihigashi and Stachurski (2014)). For example, if X_t and X'_t represent the wealth of two households, whose inhabitants face labor income following idiosyncratic shock processes, then order mixing requires that, over an infinite horizon, the first household will have lower wealth than the second at some point in time, regardless of their initial ranking.

Proposition 4.4. *Let assumption 3.1 be satisfied, and let Q be a stochastic kernel on S. If Q is order mixing, then Q is monotone ergodic.*[9]

The conditions of theorem 3.1 and 3.2 of Kamihigashi and Stachurski (2014) both imply that Q is increasing, order mixing (see lemma 6.5 of that reference) and possesses a stationary distribution π. Hence all the conditions of theorem 3.2 of the present paper hold. It follows that the ergodicity results of the present paper extend to the applications treated in section 4 of Kamihigashi and Stachurski (2014).

5. Sufficient Conditions and Applications

In this section we provide additional conditions for checking monotone ergodicity, and show how our results apply to certain economic applications. In the process we highlight models that cannot be treated with existing results from the literature.

[9]Here assumption 3.1 is only imposed to ensure that sets of the form $\{X_t \preceq X'_t\}$ are measurable. In the proofs for this section, we show that the conditions of proposition 4.4 imply monotone ergodicity, and that the conditions of propositions 4.1–4.3 imply the conditions of proposition 4.4.

5.1. **Sufficient Conditions.** As discussed above, there are existing conditions in the literature that imply order mixing, 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. Before starting we need the following definition. A stochastic kernel Q is called **bounded in probability** if, for all $x \in S$ and $\epsilon > 0$ there exists a compact set $K \subset S$ such that $\sup_t \mathbb{P}_x^Q\{X_t \notin K\} \leq \epsilon$. This is automatically true if S is compact.[10]

Consider now a generic model of the form

$$X_{t+1} = F(X_t, \xi_{t+1}), \qquad \{\xi_t\} \overset{\text{IID}}{\sim} \phi, \qquad X_0 \text{ given,} \tag{7}$$

where $F \colon S \times Z \to S$ is continuous, S and Z are Borel subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, and ϕ is a Borel probability measure on Z. In this section S is always endowed with its usual pointwise order $\leq$. The stochastic kernel corresponding to (7) is

$$Q_F(x, A) := \phi\{z \in Z : F(x,z) \in A\}, \tag{8}$$

Assumption 5.1. Subsets of S are compact if and only if they are closed and order bounded. The shock distribution ϕ is supported on all of the shock space Z.[11]

Assumption 5.2. $F(x,z)$ is increasing in x for each $z \in Z$, and Q_F is bounded in probability.

Observe that each finite path of shock realizations $\{z_t\}_{t=1}^k \subset Z$ and initial condition $X_0 = x \in S$ determines a path $\{x_t\}_{t=0}^k$ for the state variable up until time t via $x_{t+1} = F(x_t, z_{t+1})$. Let $F^k(x, z_1, \dots, z_k)$ denote the value of x_k determined in this way.[12] Given vectors x and y in S, we write $x < y$ if $x_i < y_i$ for all i.

Proposition 5.1. *If assumptions 5.1 and 5.2 are satisfied, then Q_F is increasing and at least one stationary distribution exists. If, in addition, one of the following three conditions holds*

[10] If not, boundedness in probability can often be established using drift conditions, as found in Meyn and Tweedie (2009).

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

[12] Formally, $F^1 := F$ and $F^{t+1}(x, z_1, \dots, z_{t+1}) := F(F^t(x, z_1, \dots, z_t), z_{t+1})$ for all $t \in \mathbb{N}$.

(i) *for any* $x, c \in S$, *there exists* $\{z_1, \ldots, z_k\} \subset Z$ *such that* $F^k(x, z_1, \ldots, z_k) < c$

(ii) *for any* $x, c \in S$, *there exists* $\{z_1, \ldots, z_k\} \subset Z$ *such that* $F^k(x, z_1, \ldots, z_k) > c$

(iii) *for any* $x, x' \in S$, *there exists* $\{z_1, \ldots, z_k\} \subset Z$ *and* $\{z'_1, \ldots, z'_k\} \subset Z$ *such that* $F^k(x, z_1, \ldots, z_k) < F^k(x', z'_1, \ldots, z'_k)$

then Q_F *is order mixing, and hence 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, the discussion at the top of section 4). Condition (iii) simply states that such an occurrence is *possible*. This kind of condition is typically much easier to verify.

5.2. **Applications.** A number of recent papers study "regime switching" type linear random coefficient systems of the form

$$
\begin{aligned}
w_{t+1} &= \alpha(s_{t+1}) w_t + \beta(s_{t+1}) \\
s_{t+1} &= g(s_t, \xi_{t+1})
\end{aligned}
$$

where $\{\xi_t\}$ is an IID shock sequence and α, β and g are given functions. Examples of this system can be found in the wealth distribution model of Benhabib et al. (2011), the price coefficient process in Benhabib and Dave (2013) and the inflation process in Farmer et al. (2009). In Benhabib et al. (2011), w_t is household wealth and s_t is discrete, with $\alpha(s) > 1$ for high values of s and $\alpha(s) < 1$ for low values of s. Hence wealth goes through periods of expansion and contraction. In fact, since it changes little of what follows, we assume that $s_t \in \{0, 1\}$, with $0 < \alpha(0) < 1 < \alpha(1)$. We suppose that wealth is nonnegative, that $\{s_t\}$ is irreducible, that $g(s, \xi)$ is increasing in s for each ξ, 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 π is the stationary distribution of s_t. See theorem 1 of Brandt (1986).

The endogenous state w_t is naturally unbounded and its state space cannot be compactified. Indeed, if s_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 $X_t := (w_t, s_t)$ as $[0, \infty) \times \{0, 1\}$. Because of this unboundedness, the existing law of large number results based around assumption 4.1 do not hold. Nor do the classical ergodic results hold here in general, since the counterexample in equation (2) is a special case of the current model.

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 is obvious, as is monotonicity. Condition (ii) of the proposition clearly holds too, since a sufficiently long sequence of high states for s_t will drive (w_t, s_t) above any given vector in S. Hence the system has a unique stationary distribution and is monotone ergodic. (This also verifies a claim from section 3.2 that the model in (2) is monotone ergodic under the standard order $\leq$.)

6. Conclusion

A significant number of economic models do not satisfy the classical ergodicity conditions. Motivated by earlier work on economic dynamics using an order-theoretic approach, 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. By varying the notion of order, we show that our result can recover the classical Markov ergodic theorem as a special case. At the same time, we show that monotone ergodicity is implied by a number of different conditions from the existing economics literature. Hence our results also extend to a variety of well known models that fail to satisfy the classical conditions.

A number of additional results related to implications of the theory are also provided. For example, we show that, under some additional restrictions on the state space, 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. To illustrate, we analyze a regime switching model that has been used to study variables such as wealth, prices and inflation. We use the sufficient conditions to prove monotone ergodicity and hence a strong law of large numbers. To the best of our knowledge, this cannot be established using any results from the existing literature.

7. PROOFS

7.1. Preliminaries. For the proofs we adopt some additional notation. Let

- bS denote the set of bounded measurable functions from $(S, \mathscr{B})$ to $\mathbb{R}$
- ibS denote the set of increasing functions in bS.
- cbS denote the set of continuous functions in bS.
- $icbS := ibS \cap cbS$.

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 μ on $(S, \mathscr{B})$ such that the integral is defined.

7.2. Proofs from Section 3. As alluded to 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 3.2, encompassing monotone equivalents of these ideas.

To begin, let the shift operator $\theta \colon S^\infty \to S^\infty$ be defined as usual by $\theta(x_0, x_1, \dots) = (x_1, x_2, \dots)$. Let θ^t denote the t-th composition of θ with itself, and let θ^0 be the identity. Let X be the first coordinate projection, sending $(x_0, x_1, \dots, x_t, \dots)$ 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 $\mathbb{P} = \mathbb{P}^Q_\mu$ yields the canonical Markov process discussed in section 2.2. Here and below, $\{X_t\}$ is understood as being defined in this way and $(S^\infty, \mathscr{B}^\infty, \mathbb{P}^Q_\mu)$ is the probability space, unless otherwise stated. A random variable is always a $\mathscr{B}^\infty$ measurable map from S^∞ to $\mathbb{R}$. We endow S^∞ 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.

An event $A \in \mathscr{B}^\infty$ is called **shift-invariant** if $\theta^{-1}(A) = A$. It is called **trivial** if the function $h(x) := \mathbb{P}_x^Q(A)$ 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 σ-algebra). We will make use of the following lemma, which is proved in section 7.5.

Lemma 7.1. *Let $\mathscr{G} \subset \mathscr{B}^\infty$ be a σ-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 $\mathbb{P}_x^Q\{Y = \gamma\} = 1$ for all $x \in S$.*

Here is the generalization of theorem 3.2:

Theorem 7.1. *For any increasing stochastic kernel Q with stationary distribution π, 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 π-integrable function h, we have*

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

Proof of theorem 3.2. (i) $\implies$ (ii). Let h be bounded, increasing and invariant. Define $Y := \limsup_t h(X_t)$. We then have $h(x) = \mathbb{E}_x^Q Y$ 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^∞. It now follows from our hypothesis and lemma 7.1 that there exists a $\gamma \in \mathbb{R}$ such that $\mathbb{P}_x^Q\{Y = \gamma\} = 1$ for all $x \in S$. Hence $h(x) = \mathbb{E}_x^Q(Y) = \gamma$ 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 $H(x) := \mathbb{P}_x^Q(E_h)$. 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 (4). From the fact that h is increasing, the set E_h is increasing on S^∞. Using the hypothesis that Q is increasing and applying proposition 2 of Kamae et al. (1977), 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) $\mathbb{P}_x^Q \left\{\liminf_n \frac{1}{n} \sum_{t=1}^n h(X_t) = f(x)\right\} = 1$ for all $x \in F_h$.

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

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

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 $x \in S$.[13] The claim in (iii) now follows.

[13] 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.

(iii) $\implies$ (i). Let A be increasing and shift-invariant. Let $h(x) := \mathbb{P}^Q_x(A)$. 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 $\mathbb{P}^Q_x$-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, proposition 2 of Kamae et al. (1977) tells us that h is increasing. Clearly it is π-integrable. Applying (iii), we see that $\mathbb{1}_A = \int h \, d\pi$ holds $\mathbb{P}^Q_x$-a.s. In particular, the indicator of A is constant $\mathbb{P}^Q_x$-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 $h = 1$. □

Proof of proposition 3.1. Suppose that Q is increasing and monotone ergodic on $(S, \preceq)$, and that π_1 and π_2 are both stationary for Q. Since a sequence cannot converge almost surely to two different limits, theorem 3.2 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, assumption 3.1 implies that if π_1 and π_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). □

Proof of corollary 3.1. . Fix $x \in S$ and $h \in \mathscr{L}$. As per footnote 5, we can write h as $h = h_1 - h_2$, where h_1 and h_2 are increasing and π-integrable. By theorem 3.2, for h_1 and h_2 there exist events F_1 and F_2 with $\mathbb{P}^Q_x(F_i) = 1$ 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 $n^{-1} \sum_t^n h(X_t) \to \int hd\,\pi$ on F. Evidently $\mathbb{P}^Q_x(F) = 1$. Hence (5) holds with $\mu = \delta_x$ for any $x \in S$. This extends to general μ 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 (3).)

□

Now we turn to the proof of theorem 3.3. 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 $\{g_n^1\}$ and $\{g_n^2\}$ in $\mathscr{G}$ with $g_n^1 \uparrow h$ and $g_n^2 \downarrow h$ pointwise. The proofs of the next two lemmas are given at the end of this section.

Lemma 7.2. *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 ν 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 7.3. *If the conditions of theorem 3.3 hold, then there exists a countable class $\mathscr{G}$ such that $\mathbb{P}_x^Q\{n^{-1}\sum_{t=1}^n g(X_t) \to \int g \, d\pi\} = 1$ for every $g \in \mathscr{G}$, and, moreover, $ic(S, [0,1])$ is monotonically approximated by $\mathscr{G}$.*

Proof of theorem 3.3. Fix $x \in S$. Let π_n be the empirical distribution. As a first step of the proof, we claim that $\{\pi_n\}$ is tight with probability one.[14] To see this, fix $\epsilon > 0$, and let K be a compact subset of S with $\pi(K) \geq 1 - \epsilon$. 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

$$\pi_n([a,b]) = \pi_n(I \cap J) \geq \pi_n(I) + \pi_n(J) - 1. \tag{9}$$

Note that both I and J are increasing. By corollary 3.1, we can take F_a to be a subset of S^∞ with $\mathbb{P}_x^Q(F_a) = 1$ and $\pi_n(I) \to \pi(I)$ on F_a; and $F_b \subset S^\infty$ with $\mathbb{P}_x^Q(F_b) = 1$ and $\pi_n(J) \to \pi(J)$ on F_b. It follows from (9) that on $F := F_a \cap F_b$ we have

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

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 7.3. 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 7.2 implies that $\langle \pi_n^\omega, f\rangle \to$

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

$\langle \pi, f \rangle$ 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 $f \in icbS$.[15]

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 3.3. □

7.3. **Proofs from Section 4.** It is convenient to start with the proof of proposition 4.4, and then return to the proofs of propositions 4.1–4.3.

Proof of proposition 4.4. 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.

Let $\{X_t \preceq X'_t \text{ i.o.}\}$ be the event that $X_t \preceq X'_t$ occurs infinitely often.[16] 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

$$\mathbb{P}\{h(X_t) \leq h(X'_t) \text{ i.o.}\} = 1.$$

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. □

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

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

Proof of proposition 4.1. In view of proposition 4.4, it is enough to show that Q is order mixing under assumption 4.1. This was established in section 4.1 of Kamihigashi and Stachurski (2012). $\square$

Proof of proposition 4.2. The result is immediate from remark 2.4 and lemma 6.5 of Kamihigashi and Stachurski (2014), plus proposition 4.4 of the present paper. $\square$

Proof of proposition 4.3. As in Kamihigashi and Stachurski (2014), we will say that a stochastic kernel Q on partially ordered space $(S, \preceq)$ is **order reversing** if, given a pair $x' \preceq x$ and any independent Markov processes $\{X_t\}$ and $\{X'_t\}$ generated by Q and starting at x and x' respectively, there exists a $t \in \mathbb{N}$ with $\mathbb{P}\{X_t \preceq X'_t\} > 0$. In view of proposition 4.4 and lemma 6.5 of Kamihigashi and Stachurski (2014), to verify proposition 4.3, it is enough to show that Q is order reversing and bounded in probability under the conditions of proposition 4.3.

To establish boundedness in probability (see section 5.1 for the definition), we use the assumption that Q is uniformly asymptotically tight. Fix $x \in S$ and $\delta > 0$. We claim the existence of a compact $K \subset S$ such that $\mathbb{P}^Q_x\{X_n \in K\} \geq 1 - \delta$ for all n. To see this, let C be as in the definition of uniform asymptotic tightness. By uniform asymptotic tightness, there exists an N such that $\mathbb{P}^Q_x\{X_n \in C\} > 1 - \delta$ for all $n \geq N$. In addition, for $n < N$, choose K_n to be a compact subset of S such that $\mathbb{P}^Q_x\{X_n \in K_n\} > 1 - \delta$. Clearly $K := (\cup_{n=1}^{N-1} K_n) \cup C$ is compact and satisfies $\mathbb{P}^Q_x\{X_n \in K\} \geq 1 - \delta$ for all n.

Next we turn to order reversing. 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

$$\exists\, j \in \mathbb{N} \text{ s.t. } Q^{nj}(x, (c, \infty)) > 0, \;\; \forall\, n \in \mathbb{N}. \tag{10}$$

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 (10) is established.

A symmetric argument now shows that

$$\exists\, k \in \mathbb{N} \text{ s.t. } Q^{nk}(x', (-\infty, c)) > 0, \quad \forall\, n \in \mathbb{N}. \tag{11}$$

Combining (10) and (11), we see that for $t = jk$ we have

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

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

$$\begin{aligned} \mathbb{P}\{X'_t \leq X_t\} &\geq \mathbb{P}\{X'_t < c < X_t\} \\ &= \mathbb{P}\{X'_t < c\} \mathbb{P}\{c < X_t\} \\ &= Q^t(x', (-\infty, c)) Q^t(x, (c, \infty)) > 0. \end{aligned}$$

Hence Q is order reversing as claimed. □

7.4. **Proofs from Section 5.**

Proof of proposition 5.1. Let $\{\xi_t\}$ and $\{\xi'_t\}$ be IID draws from ϕ and independent of each other. Consider first condition (iii). We claim that Q_F is order reversing (recall the proof of proposition 4.3. To see this, fix $x' \preceq x$. Let $\{z_t\}_{t=1}^k$ and $\{z'_t\}_{t=1}^k$ be as in the statement of the proposition. Define the constant

$$\gamma := \mathbb{P}\{F^k(x, \xi_1, \ldots, \xi_k) < F^k(x', \xi'_1, \ldots, \xi'_k)\}.$$

We aim to show that $\gamma > 0$. By hypothesis, $F^k(x, z_1, \dots, z_k) < F^k(x', z'_1, \dots, z'_k)$. By continuity of F, there exist open neighborhoods N_t of z_t and N'_t of z'_t such that

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

This leads to the estimate

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

Since Z is the support of ϕ, 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. $\square$

7.5. **Remaining Proofs.** Finally, we complete the proofs of lemmas 7.1–7.3.

Proof of lemma 7.1. 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 $F_x(c) := \mathbb{P}_x^Q\{Y \leq c\}$. 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.[17] 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 $\mathbb{P}_x^Q\{Y = \gamma\} = 1$. By the definition of triviality, γ does not depend on x. $\square$

Proof of lemma 7.2. Let $\{\nu_n\}$ and ν 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 $h \in \mathscr{H} \subset$

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

bS. To see this, pick any $h \in \mathscr{H}$, and choose sequences $\{g_n^1\}$ and $\{g_n^2\}$ in $\mathscr{G}$ with $g_n^1 \uparrow h$ and $g_n^2 \downarrow h$. Clearly

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

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

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

Proof of lemma 7.3. Let A be the countable subset of S in assumption 3.2. 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 $r \in \mathbb{Q} \cap [0,1]$ 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 3.2, 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

$$\sup\{\ell(x) : \ell \in \mathscr{K} \text{ and } \ell \leq f\} = f(x) \quad \text{for any } x \in S. \tag{12}$$

Indeed, if (12) 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 (12), fix $x \in S$ and $\epsilon > 0$. By continuity of f and assumption 3.2, we can find an $a \in A$ with $a \preceq x$ and $f(x) - \epsilon < f(a)$. Let $r \in \mathbb{Q}$ be such that $f(x) - \epsilon < r < f(a)$ 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, $f(x) - \epsilon < r = \ell(a) \leq \ell(x)$. Since ϵ was arbitrary we conclude that (12) is valid.

To complete the proof of lemma 7.3, 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 7.3 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.

——— (2003) "Random dynamical systems: a review," *Economic Theory*, Vol. 23, pp. 13–38.

——— (2007) *Random dynamical systems: theory and applications*: Cambridge University Press.

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, classics in applied mathematics, vol. 7," *Society for Industrial and Applied Mathematics (SIAM), 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*, Vol. 537: Wiley. com.

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 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*.

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.

Mirman, Leonard J and Itzhak Zilcha (1975) "On optimal growth under uncertainty," *Journal of Economic Theory*, Vol. 11, pp. 329–339.

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.

Nishimura, Kazuo and John Stachurski (2005) "Stability of stochastic optimal growth models: a new approach," *Journal of Economic Theory*, Vol. 122, pp. 100–118.

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.

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

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

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."

Whitt, Ward (1980) "Uniform conditional stochastic order," *Journal of Applied Probability*, pp. 112–123.

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