Discussion Paper Series

RIEB

Kobe University

DP2014-02

Seeking Ergodicity in Dynamic Economies

Takashi KAMIHIGASHI
John STACHURSKI

January 30, 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]IPAG Business School, Paris, France

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

[c]Research School of Economics, Australian National University

January 30, 2014

ABSTRACT. In both estimation and calibration studies, the notion of ergodicity plays a fundamental role, permitting time series averages to be regarded as approximations to population means. As it turns out, many economic models routinely used for quantitative modeling do not satisfy the classical ergodicity conditions. In this paper we develop a new set of ergodicity conditions orientated towards economic dynamics. We also provide sufficient conditions suitable for a variety of applications. It's notable that the classical ergodicity results can be recovered as a special case of our main theorem.

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 Kevin Reffett and Tom Sargent, and from 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 very heart of quantitative economics, supporting a vast multitude of calculations, computations and theoretical results. Consistency of estimators is an obvious example (see, e.g., Hansen (1982)). Calibration is another. Many calibration studies would be essentially meaningless without ergodicity. Ergodicity is likewise fundamental to almost all 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, a process $\{X_t\}$ is typically defined to be ergodic if all bounded invariant functions are constant (see section 3 of this paper, or section 2.2 of Ljungqvist and Sargent (2012)). For a Markov process $\{X_t\}$ with stationary distribution π, ergodicity implies that, for any initial condition X_0,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} h(X_t) = \int h(x)\,\pi(dx) \quad \text{whenever} \quad \int |h|\,d\pi < \infty. \tag{1}$$

Here "$=$" is in the sense of almost sure equality, and h is required to be measurable. Details and intuition are given below. For a standard reference see, e.g., Meyn and Tweedie (2009), theorem 17.1.7.[2]

[2]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 Markovian version of ergodicity considered here, the initial condition is irrelevant. This is 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 Markov version of ergodicity. Similar ideas can be applied to other versions.

This is a powerful result, implying probability one convergence over an extremely wide class of functions h. In fact the result is in some senses too strong, because it fails to hold for a significant number of common economic models. Naturally these include models that are in no sense ergodic, such as Durlauf (1993). More worrying is the fact that it often fails for workhorse models where ergodicity is routinely assumed. 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 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), the distorted version in Greenwood and Huffman (1995), and the heterogeneous agent extension analyzed in Krusell and Smith (1998).[3]

This same point has been made in other ways in the literature on economic dynamics. For example, for any aperiodic model, ergodicity is equivalent to total variation convergence of the distribution of X_t to π from any initial condition, where π is the stationary distribution in (1).[4] Total variation convergence is itself very strict, and, as argued by Nancy Stokey and Robert Lucas, is often "more than we expect or care about" (Stokey and Lucas, 1989, p. 352). This is precisely why economists have developed alternative approaches to the study of stability for Markovian economic models, rather than simply adopting the classical total variation based approach. See, for example, Razin and Yahav (1979) and chapter 12 of Stokey and Lucas (1989), as well as subsequent work by Hopenhayn and Prescott (1992) and Bhattacharya and Majumdar (2001).[5]

[3]Strengthening the assumptions made by these authors can recover ergodicity in some cases (cf., e.g., Nishimura and Stachurski (2005)). However, as discussed in detail below, this is in fact not necessary. The key problem here is not that these models are in any sense pathological, or even badly behaved. Rather the problem is that the classical ergodicity conditions are unnecessarily strict.

[4]For a proof see, e.g., Meyn and Tweedie (2009), theorems 13.0.1 and 17.1.7.

[5]More recent contributions along the same lines can be found in Bhattacharya and Majumdar (2007), Szeidl (2013), and Kamihigashi and Stachurski (2013). For an overview of the classical approach, see Meyn and Tweedie (2009), chapter 13.

A more detailed discussion on the failure of classical ergodicity in standard economic environments is given in section 3.2.

In addressing this problem, it is important to understand that the models in question are not pathological, or even badly behaved in any sense that matters for quantitative economics. Instead, the problem is caused by the fact that ergodicity as defined above requires the convergence in (1) to hold *for every measurable function h that is integrable with respect to π.* This is an extremely taxing condition. The family of measurable functions is vast, and contains elements so irregular that they are closer to curiosities than objects likely to form the basis of moment conditions or other empirical comparisons. Paraphrasing Stokey and Lucas (1989), convergence for these functions is more than we expect or care about.

In summary, the problem we face is one of mismatch between the requirements of economic modeling on one hand and the conditions of the classical ergodic theorem on the other. The objective of this paper is to resolve this mismatch by developing a notion of ergodicity that parallels the classical one, but with more flexibility vis-à-vis the class of functions for which (1) is required to hold. This opens the possibility of excluding irrelevant functions that make the condition overly restrictive, while at the same time retaining those functions typically used as the basis of empirical comparisons. We do this is by first focusing on monotone functions, and from there extending to linear combinations of monotone functions, and then to continuous functions under additional assumptions.

The conditions for the resulting ergodicity theorems are satisfied by a variety of important economic models under standard assumptions. These include common formulations of the standard neoclassical optimal growth model, infinite horizon incomplete market models, stochastic endogenous growth models, many overlapping generation models, models of industry dynamics, and so on. Further details are given in section 4.1.

In addition to the general results discussed above, this paper provides additional results related to (a) implications of the theory, and (b) sufficient conditions. As an example of the former, we show that, under the ergodicity conditions of our main theorem, the empirical distribution associated with any sample converges to the stationary distribution with probability one. Regarding the latter, we provide sufficient conditions that include a large class of well known models, and can be used to check the conditions of the theorem in new applications.

It is also notable from a theoretical perspective that our ergodicity results include the classical ergodicity results as a special case. Although this statement might seem surprising given the preceding discussion, it holds because we prove our results in the context of an abstract partial order. As verified below, specializing the partial order to equality recovers the classical results.

The remainder of our 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 4 provides sufficient conditions for the form of ergodicity considered in the paper. Section 5 concludes. All proofs are deferred to section 6.

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.[6]

[6] 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.

2.1. **Model Primitives.** To begin, let $(S, \mathscr{B})$ be a measurable space, and let $\precsim$ 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 $x \precsim x'$ 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.

To ensure that $\mathscr{B}$ and $\precsim$ are sufficiently compatible, we assume that $(S, \precsim)$ is a normally ordered Polish space and that $\mathscr{B}$ is its Borel sets.[7] We also assume that $\precsim$ is closed, in the sense that its graph

$$\mathbb{G} := \{(x, x') \in S \times S : x \precsim x'\} \tag{2}$$

is closed in the product space $S \times S$ with its product topology.

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 $x \in S$ to $B \in \mathscr{B}$ over one unit of time. As is common in the literature, for any suitably integrable $h \colon S \to \mathbb{R}$, we let Qh be the function defined by

$$(Qh)(x) := \int h(y) Q(x, dy) \qquad (x \in S). \tag{3}$$

Intuitively, if h gives some measurement of the state, then $Qh(x)$ is its expected value next period, given that the current state is x.

A stochastic kernel Q is called **increasing** if Qh is increasing whenever $h \colon S \to \mathbb{R}$ is measurable, bounded and increasing (see, e.g., Stokey and Lucas (1989) or

[7] In particular, S is a separable and completely metrizable topological space, with the property that, 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).

Hopenhayn and Prescott (1992)). This condition is typically satisfied in models where, holding all shocks fixed, increases in the current state shift up the future state. Many such models were cited in the introduction, and other examples are given below.

2.2. **Markov Processes.** In this section we recall some well known definitions and results related to Markov processes. Let Q be a stochastic kernel on S. A discrete time S-valued stochastic process $\{X_t\}$ on probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called Q-**Markov** if

$$\mathbb{P}[X_{t+1} \in B \mid X_0, \dots, X_t] = Q(X_t, B) \qquad \mathbb{P}\text{-almost surely} \tag{4}$$

for all $t \geq 0$ and $B \in \mathscr{B}$. For example, let $\{X_t\}$ be generated by the vector-valued stochastic difference equation

$$X_{t+1} = F(X_t, \eta_{t+1}), \qquad \{\eta_t\} \overset{\text{IID}}{\sim} \phi, \qquad X_0 \text{ given.} \tag{5}$$

Here $F \colon S \times Z \to S$ is Borel measurable, S and Z are Borel subsets of $\mathbb{R}^n$, and ϕ is a Borel probability measure on Z. If we define

$$Q_F(x, A) := \phi\{z \in Z : F(x, z) \in A\}, \tag{6}$$

then Q_F is a stochastic kernel on $(S, \mathscr{B})$ and $\{X_t\}$ is Q_F-Markov.

As is well-known (see, e.g., Stokey and Lucas (1989), p. 222), if $\{X_t\}$ is Q-Markov with $X_0 \sim \mu$, then the joint distribution induced by $\{X_t\}$ on the sequence space $(S^\infty, \mathscr{B}^\infty)$ is uniquely defined by the expression

$$\mathbb{P}^Q_\mu(B_0 \times \cdots \times B_n \times S \times S \times \cdots) = \int_{B_0} \mu(dx_0) \int_{B_1} Q(x_0, dx_1) \cdots \int_{B_{n-1}} Q(x_{n-2}, dx_{n-1}) \int_{B_n} Q(x_{n-1}, dx_n) \tag{7}$$

for any finite collection $\{B_i\}_{i=0}^n$ with $B_i \in \mathscr{B}$.[8] If $\mu = \delta_x$ then we simply write $\mathbb{P}^Q_x$.

[8] As is conventional, the integrals in (7) are computed from right to left, with the integrand written to the right of the integrating measure.

Without any loss of generality, we will take $(S^\infty, \mathscr{B}^\infty, \mathbb{P}_x^Q)$ as our probability space unless otherwise stated, and $\{X_t\}$ is just the identity map. This is a convenient way of creating a standard Q-Markov process, with joint distribution $\mathbb{P}_x^Q$.[9]

3. Ergodicity

In this section we state the classical ergodicity results for discrete time Markov processes and present an extention based on monotonicity. We then explore the implications of our ergodicity results in terms of their implications for convergence of sample averages generated by dynamic models.

3.1. **Classical Ergodicity.** To state the classical ergodicity results, recall the following definitions: Given stochastic kernel Q on S, a distribution $\pi \in \mathscr{P}$ is called **stationary** for Q if

$$\int Q(x, B)\pi(dx) = \pi(B)$$

for all $B \in \mathscr{B}$. 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, the dynamics for which are given by Q. A bounded measurable function $h\colon S \to \mathbb{R}$ is called **invariant** if

$$\int h(y)Q(x, dy) = h(x) \tag{8}$$

for all $x \in S$; that is, if h is a fixed point of (3). We define Q to be **ergodic** 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 presentation and terminology largely follows Ljungqvist and Sargent (2012)—see definition 2.2.3 of that reference.) The standard ergodicity result for general state discrete time Markov processes can now be stated:

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

[9] If $(E, \mathscr{E}, \mathbb{P})$ is any probabilty space and X is defined by $X(\omega) = \omega$, then X is an E-valued random element with distribution $\mathbb{P}$.

(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" simply means that $h \colon S \to \mathbb{R}$ is measurable, and $\int |h| \, d\mu < \infty$. 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.

Regarding the intuition behind theorem 3.1, let us call $A \in \mathscr{B}$ invariant if its indicator function $\mathbb{1}_A$ is invariant in the sense of (8). This can be restated as $Q(x, A) = \mathbb{1}_A(x)$ for all $x \in S$, or, equivalently, $Q(x, A) = 1$ for all $x \in A$ and $Q(x, A^c) = 1$ for all $x \in A^c$. In other words, both A and its complement are absorbing. If such a set exists, long run average outcomes will depend on whether the initial condition x lies in A or A^c. To recover the general convergence in (ii) of theorem 3.1, such a scenario must be ruled out. In particular, we must rule out nontrivial invariant sets (i.e., invariant sets that are neither the empty set not the whole space). The definition of ergodicity does just this, since it forces the indicator of any invariant set to be constant. The only sets with constant indicators are $\emptyset$ and S. In other words, the only invariant sets are trivial.

3.2. **Failure of Classical Ergodicity.** The conclusions of theorem 3.1 are very strong, since they imply almost sure convergence of sample means for every π-integrable function. As discussed in the introduction, this condition fails for a range of economic models that serve as popular instruments of quantitative economic modeling. To help illustrate this fact, we noted that for any aperiodic model, ergodicity in the sense of theorem 3.1 is equivalent to total variation convergence of the distribution of X_t to the stationary distribution π from any initial condition. To give one very simple example of how total variation convergence can fail, consider a stochastic model with dynamics $k_{t+1} = g(k_t, z_{t+1})$, where k_t is a vector of endogenous state variables, g is a policy function that depends on model primitives, and z_{t+1} is an exogenous state that follows a discrete state Markov chain. Following the contributions of Mehra and Prescott (1985), Huggett (1993), Krusell and Smith

(1998) and many other authors, such models are routinely employed to study the evolution of economic aggregates. Note that if k_0 is a fixed constant, then, given that the shocks are discrete, the distribution of the state vector (k_t, z_t) puts all its mass on finitely many points. On the other hand, the limiting distribution π can be absolutely continuous with respect to Lebesgue measure,[10] in which case the two distributions are mutually singular, and the total variation distance takes its maximal value for all t.[11]

In this example, failure of total variation convergence is partly due to the fact that the exogenous state follows a discrete Markov chain. However, even if this is not the case, problems can easily arise. For example, the ergodicity condition (1) is stricter than so called ψ-irreducibility, and establishing irreducibility of continuous state Markov models can be difficult or impossible, depending on the nature of the underlying primitives and the resulting policy functions (see, e.g., Zhang (2007)).

Returning to the logic stated above, ergodicity plus aperiodicity implies total variation convergence. The models we have just discussed are aperiodic, and total variation convergence fails. The implication is that classical ergodicity does not hold. It is possible to give many other examples of workhorse economic models where the same lines of argument are valid. These models simply fail to be ergodic under the standard theory.

3.3. **Monotone Ergodicity.** We now introduce an alternative notion of ergodicity well suited to economic dynamics. By way of analogy with the standard definition, we will call a stochastic kernel Q **monotone ergodic** if the only *increasing* bounded

[10]See, for example, Mitra et al. (2003).

[11]To gain some intuition, recall that the central limit theorem is stated using ordinary convergence in distribution rather than total variation convergence (TVC). This is because TVC often fails. For example, the binomial approximation to the normal distribution (the de Moivre-Laplace theorem) is not valid under TVC, because the limiting normal distribution is continuous, while the binomial distribution is discrete. Total variation distance between mutually singular distributions is always maximal, even when the distributions are arbitrarily close in some sense that respects the topology on the state space. Thus, the issue is not that the approximation is poor, but rather that the notion of total variation convergence is too strong. For further discussion see p. 292 of Dudley (2002).

invariant functions are the constant functions. With this definition in hand, we can generalize theorem 3.1 by providing an order theoretic version, the proof of which is given in section 6.

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,*

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

To see that theorem 3.2 is a generalization of theorem 3.1, it suffices to set the partial order $\precsim$ to be equality (i.e., $x \precsim y$ iff $x = y$). In this case, it's easily verified that every function from S to $\mathbb{R}$ is increasing, and, as a consequence, every stochastic kernel on S is increasing. In such a setting, the results of theorem 3.2 reduce to those of theorem 3.1.

For other partial orders, such as the standard pointwise partial order on $\mathbb{R}^n$, the family of increasing functions is a strict subset of the family of all functions. When such a partial order is chosen, convergence of sample means over increasing π-integrable functions is less restrictive than convergence over all π-integrable functions. Below we discuss how this difference is important in economic applications. We also show that the convergence in (ii) of theorem 3.2 can be extended to a much larger class of functions.

As a final comment before continuing, we note that some authors define ergodicity in terms of either sets or random variables that are invariant with respect to a shift operator defined on the underlying probability space. For completeness, in the proofs we provided order theoretic versions of these concepts too, and show their equivalence to monotone ergodicity.

3.4. **Further Implications.** In this section our aim is to investigate the practical implications of monotone ergodicity. As a first step, we show that under monotone ergodicity, the almost sure convergence in the strong law of large numbers holds for a larger set of functions than just the increasing functions. To state the result, for fixed stochastic kernel Q with stationary distribution π we let $\mathscr{L}$ denote the linear span of the set of increasing π-integrable functions.[12] By taking finite unions of probability one sets and using linearity of sums and integrals, it is straightforward to show that, under monotone ergodicity, the strong law of large numbers extends to any such function:[13]

Corollary 3.1. *Let Q be an increasing stochastic kernel with stationary distribution π. Then Q is monotone ergodic if and only if*

$$\text{(9)} \qquad \text{for all } \mu \in \mathscr{P} \text{ and all } h \in \mathscr{L}, \quad \mathbb{P}^Q_\mu \left\{ \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n h(X_t) = \int h \, d\pi \right\} = 1.$$

Provided that the tails of π are not too heavy, the class $\mathscr{L}$ will contain many of the functions commonly used to compare models to data. Examples are given below. Moreover, since we now consider $\mathbb{P}^Q_\mu$ rather than $\mathbb{P}^Q_x$, the convergence applies to random X_0 with any distribution.

Example 3.1. Many calibration exercises involve comparison of moments. Consider, for example, a scalar model with $S = \mathbb{R}$, and let k be the number of finite moments possessed by the stationary distribution π. All k moment functions lie in $\mathscr{L}$, as does any polynomial up to order k.[14]

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).

---

[12] 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.

[13] As far as we know, the closest result to corollary 3.1 is theorem 3.1 of Bhattacharya and Lee (1988). This result gives sufficient conditions for the convergence in (9), and hence is a special case of our result. On the other hand, the authors also provide a version of the central limit theorem, which in turn implies that convergence occurs at a parametric rate. Clearly this is not possible in our setting, since our conditions are both necessary and sufficient.

[14] To give the idea, observe that $h(x) = x^2$ can be written as $-h_1(x) + h_2(x)$, where $h_1(x) := -x^2 \mathbb{1}\{x < 0\}$ and $h_2(x) := x^2 \mathbb{1}\{x \geq 0\}$. Both h_1 and h_2 are increasing functions. Thus, if π has finite second moment, then $x^2 \in \mathscr{L}$, and $n^{-1} \sum_{t=1}^n X_t^2 \to \int x^2 \pi(dx)$ with probability one for all initial conditions.

In some settings—particularly multivariate settings—continuity is more convenient to work with than monotonicity. It turns out that if S is compact and satisfies an additional mild restriction, 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 $\pi_n \xrightarrow{d} \pi$ with probability one, where $\xrightarrow{d}$ 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).[15]

With an additional restriction on S, we show that the desired convergence is obtained. The restriction strengthens our separability assumption for S, and is satisfied for many common state spaces (e.g., when $S = \mathbb{R}^m$, or more generally, when S is a cone in $\mathbb{R}^m$ with the usual partial order):

Assumption 3.1. 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 $a \precsim x \precsim a'$.

Theorem 3.3. *If assumption 3.1 is 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, $\pi_n \xrightarrow{d} \pi$ with probability one.

[15] 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.

4. Sufficient Conditions

So far we have focused on the implications of monotone ergodicity. Our next step is to show that monotone ergodicity is satisfied by many useful economic models.

4.1. **Order Mixing and Monotone Ergodicity.** We begin with a relatively general sufficent condition and then move to more specific ones. As a first step, we follow Kamihigashi and Stachurski (2012) in defining a stochastic kernel Q on S to be **order mixing** if, given any pair of independent Q-Markov processes $\{X_t\}$ and $\{X'_t\}$, the event $\{X_t \precsim X'_t\}$ occurs with probability one. 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 always be poorer than the second at some point in time, regardless of their initial wealth ranking.

We now state our first sufficient condition, for fixed stochastic kernel Q on S.

Theorem 4.1. *If Q is order mixing, then Q is monotone ergodic.*

This result is significant because many well-known models can be shown to be order mixing. For example, Razin and Yahav (1979), Stokey and Lucas (1989) and Hopenhayn and Prescott (1992) analyzed the dynamics of monotone economies satisfying a so-called "monotone mixing condition." These conditions imply order mixing, and hence, by theorem 4.1, every such economy is monotone ergodic. Similarly statements are true for the classes of models studied in Bhattacharya and Majumdar (2001) and Szeidl (2013).

Based on these results, it can be shown that monotone ergodicity is satisfied by common formulations of the standard neoclassical optimal growth model, 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), and 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). Many models of industry and employment dynamics also fall into this category. Representative examples include Cabrales and Hopenhayn (1997), Cooley and Quadrini (2001) and Samaniego (2008). An exhaustive list is too large to give here.

4.2. **Verifying Order Mixing: New 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. To do so we need some additional definitions. As a start, given $a, b \in S$, we let $[a, b]$ denote the order interval $\{x \in S : a \precsim x \precsim b\}$. A subset K of S is called **order bounded** if there exists $a, b \in S$ with $K \subset [a, b]$.

Assumption 4.1. S is a Borel subset of $\mathbb{R}^n$ endowed with its usual partial order, and such that subsets of S are compact if and only if they are closed and order bounded.[16]

Consider now the generic model presented in (5). Without loss of generality, we suppose that the shock distribution ϕ is supported on all of the shock space Z.[17] We also assume the following:

Assumption 4.2. F is continuous, and $F(x, z)$ is increasing in x for each $z \in Z$. The corresponding kernel Q_F is bounded in probability.[18]

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

[16]For example, $\mathbb{R}^n$, $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ have this property, while $S = \mathbb{R} \setminus \{0\}$ does not.

[17]That is, $\phi(Z) = 1$, and $\phi(G) > 0$ whenever $G \subset Z$ is nonempty and open. Clearly Z can always be re-defined so that this assumption is valid.

[18]A stochastic kernel Q is called **bounded in probability** if, for all $x \in S$ it is the case that, given any $\epsilon > 0$, there exists a compact set $K \subset S$ with $\sup_t \mathbb{P}^Q_x\{X_t \notin K\} \leq \epsilon$. One way to ensure that this condition holds is to take the state space to be compact. A more general technique involves drift conditions, as found in Meyn and Tweedie (2009). The drift approach has been applied to a variety models in economics, finance and time series analysis.

$x_{t+1} = F(x_t, z_{t+1})$. Let $F^k(x, z_1, \ldots, z_k)$ denote the value of x_k determined in this way.[19] Given vectors x and y in S, we write $x < y$ if $x_i < y_i$ for all i.

Proposition 4.1. *If assumptions 4.1 and 4.2 are satsfied, then Q_F is increasing and at least one stationary distribution exists. If, in addition, one of the following three conditions holds*

(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 (2013). Unlike the latter, conditions (i)–(iii) exploit continuity to provide statements that are easier to check in applications.

To see how proposition 4.1 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 4.1). Condition (iii) simply states that such an occurrence is *possible*. This kind of condition is typically much easier to verify.

5. Conclusion

Many economic models routinely used for quantitative analysis fail to satisfy the classical ergodicity conditions. In this paper we develop an alternative set of ergodicity conditions orientated towards economic dynamics and explore their implications. Sufficient conditions for checking the conditions are provided. We also show that the classical ergodicity results are a special case of our theory.

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

6. Proofs

6.1. **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.

6.2. **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 $\theta(x_0, x_1, \ldots) = (x_1, x_2, \ldots)$. 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 $(x_0, x_1, \ldots, x_t, \ldots)$ 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 $\mathbb{P} = \mathbb{P}^Q_\mu$ yields the canonical Q-Markov process. 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 $(S, \precsim)$. In particular, we say that $\{x_t\} \precsim \{x'_t\}$ if $x_t \precsim x'_t$ 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 $h(x) := \mathbb{P}^Q_x(A)$ 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 σ-algebra). We will make use of the following lemma, which is proved in section 6.4.

Lemma 6.1. *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 $\mathbb{P}^Q_x\{Y = \gamma\} = 1$ for all $x \in S$.*

Here is the generalization of theorem 3.2:

Theorem 6.1. *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*

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

Proof of theorem 3.2. (i) $\implies$ (ii). Let h be bounded, increasing and invariant. Define $Y := \limsup_t h(X_t)$. We then have $h(x) = \mathbb{E}^Q_x Y$ 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^∞. It now follows from our hypothesis and lemma 6.1 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 (8). 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) $\mathbb{P}^Q_x \left\{ \liminf_n \frac{1}{n} \sum_{t=1}^n h(X_t) = f(x) \right\} = 1$ 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,

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

A symmetric argument shows that $\mathbb{P}^Q_x \left\{ \limsup_n n^{-1} \sum_{t=1}^n h(X_t) \leq 0 \right\} = 1$ for all $x \in S$.[20] The claim in (iii) now follows.

(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 $x \in S$

[20] 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.

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 corollary 3.1. . Fix $x \in S$ and $h \in \mathscr{L}$. As per footnote 12, 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 (9) 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 (7).)

□

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, $\mu_n \to \mu$ 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 6.2. *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 6.3. *If the conditions of theorem 3.3 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 3.3. 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. (Recall that a sequence $\{\mu_n\} \subset \mathscr{P}$ 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.) 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 $I := \{y \in S : a \precsim y\}$ and $J := \{y \in S : y \precsim b\}$. Evidently

$$\pi_n([a,b]) = \pi_n(I \cap J) \geq \pi_n(I) + \pi_n(J) - 1. \tag{10}$$

Note that both I and J are increasing. By corollary 3.1, we can take F_a to be a subset of S^∞ with $\mathbb{P}^Q_x(F_a) = 1$ and $\pi_n(I) \to \pi(I)$ on F_a; and $F_b \subset S^\infty$ with $\mathbb{P}^Q_x(F_b) = 1$ and $\pi_n(J) \to \pi(J)$ on F_b. It follows from (10) 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 6.3. 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 6.2 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 $f \in icbS$.[21]

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

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

of lemma 6.6 of Kamihigashi and Stachurski (2013), 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. $\square$

6.3. **Proofs from section 4.**

Proof of theorem 4.1. Suppose instead that $h \in ibS$ is invariant, and, in addition, there exist distinct points $a, b \in \mathbb{R}$ with nonempty preimages under h. Without loss of generality, we set $a < b$. Using these two nonempty preimages, we can take $\{X_t\}$ and $\{X_t'\}$ to be independent Q-Markov processes defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and having X_0 and X_0' chosen such that $\mathbb{E}\, h(X_0) = b$ and $\mathbb{E}\, h(X_0') = a$.

As h is invariant, both $\{h(X_t)\}$ and $\{h(X_t')\}$ are martingales. For example, $\{h(X_t)\}$ is a martingale because

$$\mathbb{E}\,[h(X_{t+1}) \mid X_0, \ldots, X_t] = \mathbb{E}\,[h(X_{t+1}) \mid X_t] = Qh(X_t) = h(X_t),$$

where the first and second equalities are by the Markov property and the last is by the fact that h is invariant. Since h is bounded, these martingales are also bounded, and hence, by the martingale convergence theorem, converge $\mathbb{P}$-almost surely to respective limits Y_∞ and Y_∞'. Since finite intersections of probability one sets have probability one, the pair $(h(X_t), h(X_t'))$ converges to (Y_∞, Y_∞') in $\mathbb{R}^2$ with probability one.

Let $\{X_t \precsim X_t' \text{ i.o.}\}$ be the event that $X_t \precsim X_t'$ occurs infinitely often.[22] The assumption that Q is order mixing tells us that $X_t \precsim X_t'$ 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 $\mathbb{P}\{X_t \precsim X_t' \text{ i.o.}\} = 1$. Since h is increasing, this yields

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

By taking finite intersections again, we can now claim existence of a probability one set on which $h(X_t) \leq h(X_t')$ infinitely often and $(h(X_t), h(X_t'))$ converges to (Y_∞, Y_∞'). On such a set we must have $Y_\infty \leq Y_\infty'$. We have now shown that $Y_\infty \leq Y_\infty'$ holds $\mathbb{P}$-a.s., and hence $\mathbb{E}\, Y_\infty \leq \mathbb{E}\, Y_\infty'$.

---

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

On the other hand, Y_∞ and Y'_∞ are the limits of martingales $\{h(X_t)\}$ and $\{h(X'_t)\}$, so $\mathbb{E}\, Y_\infty = \mathbb{E}\, h(X_t) = \mathbb{E}\, h(X_0)$ and $\mathbb{E}\, Y'_\infty = \mathbb{E}\, h(X'_t) = \mathbb{E}\, h(X'_0)$. Recalling our original construction of these sequences, we see that $\mathbb{E}\, Y_\infty = b$, while $\mathbb{E}\, Y'_\infty = a$. Therefore $\mathbb{E}\, Y_\infty > \mathbb{E}\, Y'_\infty$. Contradiction. □

Now we turn to the proof of proposition 4.1. As in Kamihigashi and Stachurski (2013), we will say that a stochastic kernel Q is **order reversing** if, given a pair $x' \precsim x$ and any independent Markov-Q processes $\{X_t\}$ and $\{X'_t\}$ starting at x and x' respectively, there exists a $t \in \mathbb{N}$ with $\mathbb{P}\{X_t \precsim X'_t\} > 0$. In other words, the initial ordering is reversed at some point in time with positive probability.

Proof of proposition 4.1. Let $\{\eta_t\}$ and $\{\eta'_t\}$ be IID draws from ϕ and independent of each other. Consider first condition (iii). We claim that Q_F is order reversing. To see this, fix $x' \precsim x$. Let $\{z_t\}_{t=1}^k$ and $\{z'_t\}_{t=1}^k$ be as in the statement of the proposition. Define the constant

$$\gamma := \mathbb{P}\{F^k(x, \eta_1, \ldots, \eta_k) < F^k(x', \eta'_1, \ldots, \eta'_k)\}.$$

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

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

This leads to the estimate

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

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 (2013) 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 (2013). The order reversing property then follows from Kamihigashi and Stachurski (2013), proposition 3.2, and the rest of the arguments are unchanged. □

6.4. **Proofs of Remaining Lemmata.** Finally, we complete the proof of all remaining lemmata stated in this section.

Proof of lemma 6.1. 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.[23] 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. □

Proof of lemma 6.2. 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 $h \in \mathscr{H} \subset bS$. 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 $\limsup_n \langle \nu_n, h\rangle \leq \langle \nu, h\rangle$. □

Proof of lemma 6.3. Let A be the countable subset of S in assumption 3.1. For $a \in A$, let $I_a := \mathbb{1}\{y \in S : a \precsim y\}$. 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

$$\sup\{\ell(x) : \ell \in \mathscr{K} \text{ and } \ell \leq f\} = f(x) \quad \text{for any } x \in S. \tag{11}$$

[23]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.

Indeed, if (11) 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 (11), fix $x \in S$ and $\epsilon > 0$. By continuity of f and assumption 3.1, we can find an $a \in A$ with $a \precsim x$ 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 (11) is valid.

To complete the proof of lemma 6.3, 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 $\mathbb{P}^Q_x\{n^{-1}\sum_{t=1}^n g(X_t) \to \int g \, d\pi\} = 1$, 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 6.3 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.

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.

——— (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.

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.

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.

Dudley, Richard M (2002) *Real analysis and probability*, Vol. 74: Cambridge University Press.

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.

Durlauf, Steven N (1993) "Nonergodic economic growth," *The Review of Economic Studies*, Vol. 60, pp. 349–366.

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, 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.

——— (2013) "Stochastic Stability in Monotone Economies," *Theoretical Economics*.

Krusell, Per and Anthony A Smith, Jr (1998) "Income and wealth heterogeneity in the macroeconomy," *Journal of Political Economy*, Vol. 106, pp. 867–896.

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

Mehra, Rajnish and Edward C Prescott (1985) "The equity premium: A puzzle," *Journal of Monetary Economics*, Vol. 15, pp. 145–161.

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.

Mitra, Tapan, Luigi Montrucchio, and Fabio Privileggi (2003) "The nature of the steady state in models of optimal growth under uncertainty," *Economic Theory*, Vol. 23, pp. 39–71.

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.

Samaniego, Roberto M (2008) "Can technical change exacerbate the effects of labor market sclerosis?" *Journal of Economic Dynamics and Control*, Vol. 32, pp. 497–528.

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.