KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

KYOTO UNIVERSITY

KYOTO, JAPAN

ASYMPTOTICS OF STOCHASTIC RECURSIVE ECONOMIES UNDER MONOTONICITY

TAKASHI KAMIHIGASHI AND JOHN STACHURSKI

ABSTRACT. This paper presents a new mixing condition for dynamic economies with a Markov structure. The mixing condition is stated in terms of order, and generalizes a number of well-known conditions used to establish stability of monotone dynamic models. By generalizing the key insights of the original conditions, we derive a set of results with applications to many theoretical and time series models.

1. INTRODUCTION

In what was to become a seminal contribution to the theory of economic dynamics, Razin and Yahav (1979, theorem 1) proposed a technique for assesssing the stability of stochastic steady states (i.e., stationary distributions) for Markovian models with the property that transitions are monotone increasing[1] in the state variable. Their mixing condition was subsequently extended to n-dimensional process by Stokey and Lucas (1989), and to general partially ordered compact metric spaces by Hopenhayn and Prescott (1992). It has been widely used to assess the stability of theoretical and applied economic models.

In a related line of research, Bhattacharya and Lee (1988) and Bhattacharya and Majumdar (2001) developed a "splitting condition" for Markov models that is defined in terms of an ordering on the state space. Splitting is closely related to the mixing condition discussed above, and, using different methods, these authors showed that splitting can be used to prove stability.

Date: January 20, 2009.

[1]Here and below increasing is synonymous with nondecreasing.

Both of these classes of results provided stability conditions that are simple to state, and apply to a number of different economic settings. In this paper we seek to identify the essence of these monotone stability results and extend them to a broader class of applications. To do so, we introduce a new concept, referred to below as order mixing. A Markov process is called *order mixing* if, given any two independent realizations $(X_t)_{t \geq 0}$ and $(X'_t)_{t \geq 0}$, we have[2]

(1) $\mathbb{P}\{\exists t \geq 0 \text{ s.t. } X_t \leq X'_t\} = \mathbb{P}\{\exists t \geq 0 \text{ s.t. } X'_t \leq X_t\} = 1$

We show that, taken together, monotonicity and order mixing guarantee global stability whenever a stationary distribution exists.

The economic intuition for condition (1) is straightforward. For example, if the state variable represent household wealth in a given model of savings and investment, then (1) states that, regardless of initial wealth, any pair of households receiving mutually independent, idiosyncratic shocks will attain both orderings (household A is richer than household B and vice versa) at some point in time with probability one. This rules out multiple disjoint absorbing sets.

We provide a number of sufficient conditions via which order mixing can be verified in applications. Using these results, we demonstrate that the models considered in Razin and Yahav (1979, theorem 1), Bhattacharya and Lee (1988, theorem 2.1), Stokey and Lucas (1989, theorem 12.12), Bhattacharya and Majumdar (2001, theorem 2.1) and Hopenhayn and Prescott (1992, theorem 2) all satisfy our definition of order mixing. However, the set of order mixing models is larger those treated in these papers. Thus, we extend the domain of application for monotone stability methods to a broader class of economic and econometric models.[3]

It is worth noting that our method also provides a clear link to the classical theory of stability for Markov processes initiated by Doeblin (1938). In his analysis, stability is proved for systems such that, given

---

[2]The distributions of the initial conditions X_0 and X'_0 are permitted to be distinct, but both processes are updated according to the same transition law.

[3]We do not claim to encompass all of the theory in the papers cited above, each of which contains unique and important results. Our contribution is to generalize the order-theoretic mixing ideas that appear in all of these papers.

any two independent copies $(X_t)_{t \geq 0}$ and $(X'_t)_{t \geq 0}$ of the same process, we have

$$(2) \qquad \qquad \mathbb{P}\{\exists t \geq 0 \text{ s.t. } X_t = X'_t\} = 1$$

Intuitively such systems have a form of long run history independence, and Doeblin showed this property is sufficient for a strong form of stability. Our order mixing condition (1) is clearly weaker than (2). Nevertheless, we prove that (1) is sufficient for stability when paired with monotonicity of the law of motion.

Our proof of this result is based on coupling theory. Coupling is a means of comparing two distributions ψ and ψ' by constructing random variables X and X' such that $\mathscr{D}X = \psi$ and $\mathscr{D}X' = \psi'$. (Here $\mathscr{D}Y = \phi$ means that Y has distribution ϕ.) The essential idea is that many pairs X and X' have this property, and a careful choice can simplify the analysis of ψ and ψ' one wishes to perform.[4] For stability of Markov processes, the two distributions in question are the time t distribution ψ_t of X_t and the stationary distribution ψ^*. Here coupling is used to show that $\psi_t \to \psi^*$ as $t \to \infty$.

One of the advantages of the order mixing concept introduced in the paper is that through coupling we are able to provide a purely probabilistic proof of the main result, without any topological or order-theoretic arguments. By separating our theory from topological considerations, we are able to isolate more clearly the implications on monotone mixing.

The paper is structured as follows. Section 2 begins with an overview of the main theorem. Section 3 provides background on Markov processes, while section 4 gives a formal discussion of order mixing. Section 5 states the main stability result. Section 6 gives applications of our results. Remaining proofs are given in the appendix.

2. OUTLINE OF THE METHOD

We begin with an overview of stability via monotonicity and order mixing, setting aside measure-theoretic details until the next section.

---

[4]A general discussion of coupling techniques is given in Lindvall (1992).

To do so, we consider a model that evolves according to the stochastic recursive sequence $X_{t+1} = F(X_t, W_{t+1})$, where the shocks $(W_t)_{t \geq 1}$ are independently drawn from distribution ϕ on shock space Z, and the initial condition X_0 is given. A large class of economic models can be formulated in this manner for suitable choices of F and the state space.

We assume X_t takes values in a partially ordered space such as $\mathbb{R}^n$, and that the process is *monotone increasing*, in the sense that, for all $z \in Z$, we have $F(x, z) \leq F(x', z)$ whenever $x \leq x'$. We suppose further that a stationary distribution ψ^* exists. Our objective is to show that under order mixing, the distribution of X_t converges to ψ^*, independent of X_0.

To this end, let $X_{t+1}^* = F(X_t^*, W_{t+1}^*)$ be a second version of the process where $(W_t^*)_{t \geq 1}$ is another IID-ϕ sequence, and $\mathscr{D}X_0^* = \psi^*$. By the definition of stationarity, we have $\mathscr{D}X_t^* = \psi^*$ for all t. Consider also a third process $X_{t+1}^L = F(X_t^L, W_{t+1}^L)$ where $X_0^L = X_0$ and $W_t^L = W_t$ until the first time τ such that $X_t \leq X_t^*$. From then on we set $W_t^L = W_t^*$. Since this process starts off at X_0 and initially shares the same draws as the first process $(X_t)_{t \geq 0}$, it tracks $(X_t)_{t \geq 0}$ up until date τ. At this point in time we have

$$X_\tau^L = X_\tau \leq X_\tau^*$$

From τ on, $(X_t^L)_{t \geq 0}$ receives the same shocks as $(X_t^*)_{t \geq 0}$. Since these processes receive the same shocks, it follows from our monotonicity assumption that the time τ ordering $X_\tau^L \leq X_\tau^*$ is preserved at all future dates (i.e, $X_t^L \leq X_t^*$ for all $t \geq \tau$). See figure 1.

Observe that order mixing implies $X_t \leq X_t^*$ for some t with probability one. Equivalently, τ is finite with probability one. For large t, then, we have $X_t^L \leq X_t^*$ with high probabilily (because $t \geq \tau$ with high probability, and $X_t^L \leq X_t^*$ whenever $t \geq \tau$).

Another pertinent feature of (X_t^L) is that $\mathscr{D}X_t^L = \mathscr{D}X_t$ for *all* $t \geq 0$. The reason is that both processes start off at the same initial condition X_0, and both are subsequently updated by IID shocks from the same distribution ϕ. In the case of (X_t^L), the source of shocks changes from (W_t) to (W_t^*) at τ, but since the distributions of the shocks are the same, this change does not alter the distribution of X_t^L.
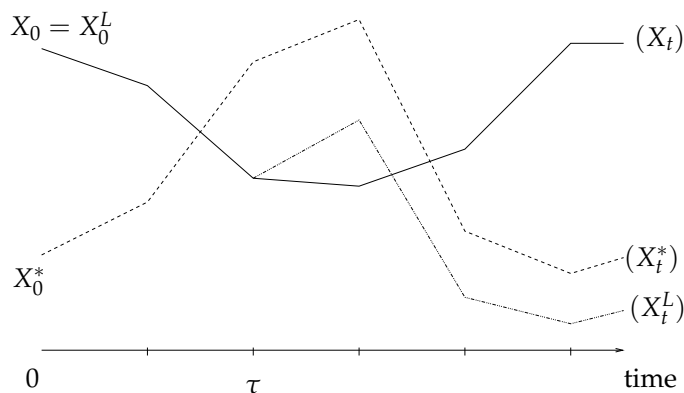
FIGURE 1. $X_t^L \leq X_t^*$ for all $t \geq \tau$

Regarding stability, the assertion we wish to prove is that $\int h d\psi_t \to \int h d\psi^*$ for every increasing bounded real-valued function h, where ψ_t is the distribution of X_t.[5] In our set up, this is equivalent to the assertion that $\lim_t \mathbb{E}h(X_t) = \lim_t \mathbb{E}h(X_t^*)$ for any such h. Let us fix h and consider the claim that $\lim_t \mathbb{E}h(X_t) \leq \lim_t \mathbb{E}h(X_t^*)$.

Our first observation is that since $\mathscr{D}X_t = \mathscr{D}X_t^L$, we need only show that $\lim_t \mathbb{E}h(X_t^L) \leq \lim_t \mathbb{E}h(X_t^*)$. Now if t is large, then by order mixing we have $X_t^L \leq X_t^*$ with high probability. Since h is increasing, this implies $h(X_t^L) \leq h(X_t^*)$ with high probability. Using monotonicity of $\mathbb{E}$ and taking limits, we obtain $\lim_t \mathbb{E}h(X_t^L) \leq \lim_t \mathbb{E}h(X_t^*)$, as was to be show.[6]

So far we have sketched the proof that $\lim_t \int h d\psi_t \leq \int h d\psi^*$ for every increasing bounded real-valued function h. The reverse inequality can now be obtained in a similar way, this time replace (X_t^L) with a process (X_t^U) which is eventually *larger* than X_t^*. These arguments are made precise throughout the remainder of the paper.

---

[5]For any common state space, this convergence is stricter than the standard notion of convergence in distribution. See below for details.

[6]For more precise arguments, see theorem 5.1 and its proof.

3. DISCRETE-TIME MARKOV PROCESSES

We begin with some fundamental properties of discrete time Markov processes on an arbitrary measurable space. Let $(E, \mathscr{E})$ be such a space, and let $\mathscr{P}(E)$ be the probability measures on $(E, \mathscr{E})$. A *stochastic kernel* on E is a function $Q \colon E \times \mathscr{E} \to [0, 1]$ such that

(1) $B \mapsto Q(x, B) \in \mathscr{P}(E)$ for all $x \in E$, and
(2) $x \mapsto Q(x, B)$ is $\mathscr{E}$-measurable for all $B \in \mathscr{E}$.

A discrete-time, E-valued stochastic process $\mathbf{X} = (X_t)_{t \geq 0}$ on probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is said to be Markov-(Q, μ) if X_0 has distribution $\mu \in \mathscr{P}(E)$ and $Q(x, \cdot) =: Q(x, dy)$ is the conditional distribution of X_{t+1} given $X_t = x$. More precisely:

Definition 3.1. E-valued stochastic process $\mathbf{X} = (X_t)_{t \geq 0}$ is called *Markov-(Q, μ)* if

$$\mathscr{D} X_0 = \mu \quad \text{and} \quad \mathbb{P}[X_{t+1} \in B \mid \mathscr{F}_t] = Q(X_t, B) \text{ for all } B \in \mathscr{E}$$

where $\mathscr{F}_t$ is the σ-algebra generated by the history $X_0, \ldots, X_t$.

In economics and econometrics, most Markov processes are generated by a stochastic recursive sequence (SRS). For example, suppose that $\mathbf{X} = (X_t)_{t \geq 0}$ is generated by the model

$$(3) \qquad X_{t+1} = F(X_t, W_{t+1}) \quad \text{and} \quad \mathscr{D} X_0 = \psi$$

where X_t takes values in E and the shock sequence $(W_t)_{t \geq 1}$ is IID and takes values in a measurable space $(Z, \mathscr{Z})$ according to distribution ϕ.[7] This process is Markov-(P, ψ) on E for P defined by

$$(4) \qquad P(x, B) = \int \mathbb{1}_B[F(x, z)] \phi(dz) \qquad (x \in E, \ B \in \mathscr{E})$$

Indeed, if $h \colon E \to \mathbb{R}$ is any bounded measurable function, then

$$\mathbb{E}[h(X_{t+1}) \mid \mathscr{F}_t] = \mathbb{E}[h[F(X_t, W_{t+1})] \mid \mathscr{F}_t] = \int h[F(X_t, z)] \phi(dz)$$

Specializing to the case of $h = \mathbb{1}_B$ and using (4) gives

$$\mathbb{P}[X_{t+1} \in B \mid \mathscr{F}_t] = \int \mathbb{1}_B[F(X_t, z)] \phi(dz) = P(X_t, B)$$

---

[7]The initial condition X_0 is independent of the shocks $(W_t)_{t \geq 1}$, and both are defined on some underlying probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

This confirms that **X** is Markov-(P, ψ).

Returning to the general case, let $E^{\infty} = \times_{t \geq 0} E$ be the set of all E-valued sequences $(x_t)_{t \geq 0}$ and let $\mathscr{E}^{\infty} = \otimes_{t \geq 0} \mathscr{E}$ be the product σ-algebra. The following result is standard (cf., e.g., Pollard, 2002, p. 101). It describes the joint distribution of any Markov-(Q, μ) process on the sequence space $(E^{\infty}, \mathscr{E}^{\infty})$.

Theorem 3.1. *For each $\mu \in \mathscr{P}(E)$ and stochastic kernel Q on $(E, \mathscr{E})$ there exists a unique probability measure $\mathbf{P}_{\mu}^{Q}$ on $(E^{\infty}, \mathscr{E}^{\infty})$ such that for any finite collection $(B_i)_{i=0}^{n}$ with $B_i \in \mathscr{E}$ the measure $\mathbf{P}_{\mu}^{Q}$ satisfies*

(5) $\quad \mathbf{P}_{\mu}^{Q}(B_0 \times \cdots \times B_n \times E \times E \times \cdots) =$

$$\int_{B_0} \mu(dx_0) \int_{B_1} Q(x_0, x_1) \cdots \int_{B_{n-1}} Q(x_{n-2}, dx_{n-1}) \int_{B_n} Q(x_{n-1}, dx_n)$$

If an E-valued process $\mathbf{X} := (X_t)_{t \geq 0}$ on probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is Markov-(Q, μ), then the distribution of $\mathbf{X}$ on $(E^{\infty}, \mathscr{E}^{\infty})$ is given by $\mathbf{P}_{\mu}^{Q}$. At least one Markov-(Q, μ) process exists.[8]

In (5) the integrals are computed from right to left, with the integrand written to the right of the integrating measure. If the initial condition μ is the probability measure $\delta_x \in \mathscr{P}(E)$ concentrated on $x \in E$, then it is traditional to write $\mathbf{P}_x^{Q}$ rather than $\mathbf{P}_{\delta_x}^{Q}$. A generating class argument applied to (5) shows that

(6) $\qquad \mathbf{P}_{\mu}^{Q}(B) = \int \mathbf{P}_x^{Q}(B) \mu(dx) \qquad (B \in \mathscr{E}^{\infty}, \ \mu \in \mathscr{P}(E))$

Let Q^n be the n-th order kernel, defined by

$$Q^1 := Q \qquad Q^n(x, B) := \int Q^{n-1}(y, B) Q(x, dy)$$

Note that each Q^n is a stochastic kernel in its own right, and that $Q^n(x, B)$ is used to represent the probability of transitioning from x to B in n steps. We will make use of the fact that if an E-valued process $(X_t)_{t \geq 0}$ on $(\Omega, \mathscr{F}, \mathbb{P})$ is Markov-(Q, μ), then for any $n \in \mathbb{N}$ the process $(X_{t \times n})_{t \geq 0}$ is Markov-(Q^n, μ). Also, for a suitably integrable $h \colon E \to \mathbb{R}$ we adopt the standard notation

$$Q^n h(x) := \int h(y) Q^n(x, dy) \qquad (x \in E)$$

---

[8]The last statement is easily verified using coordinate projections.

A probability measure $\mu^* \in \mathscr{P}(E)$ is called *stationary* for Q if

$$\mu^*(B) = \int Q(x, B)\mu^*(dx) \quad \text{for all } B \in \mathscr{E}$$

It can be verified from (5) that if μ^* is stationary for Q and $\mathbf{X} = (X_t)_{t \geq 0}$ is Markov-(Q, μ^*) then $\mathbf{X}$ is a stationary stochastic process, and, in particular, $\mathscr{D} X_t = \mu^*$ for all $t \geq 0$.

Let $B \subset E$ and let $V_B \subset E^\infty$ be the set of E-valued sequences which visit B at least once:

$$V_B := \bigcup_{n=0}^{\infty} \{(x_t)_{t \geq 0} : x_n \in B\}$$

A set $B \in \mathscr{E}$ is called *Harris recurrent* if $\mathbf{P}_x^Q(V_B) = 1$ for all $x \in E$.[9] Note that if $\mathbf{X} = (X_t)_{t \geq 0}$ is Markov-(Q, x) on probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and B is Harris recurrent for Q, then since $\mathbf{P}_x^Q$ is the distribution of $\mathbf{X}$ we have

$$\mathbb{P}\{\mathbf{X} \text{ ever enters } B\} := \mathbb{P}\{\mathbf{X} \in V_B\} = \mathbf{P}_x^Q(V_B) = 1$$

Let $W_B \subset E^\infty$ be those sequences which visit B infinitely often:

$$(7) \qquad W_B := \bigcap_{m=0}^{\infty} \bigcup_{n \geq m} \{(x_t)_{t \geq 0} : x_n \in B\}$$

We will make use of the following elementary facts concerning Harris recurrence.

Lemma 3.1. *If B is Harris recurrent for Q, then $\mathbf{P}_x^Q(W_B) = 1$ for all $x \in E$. In other words, if $x \in E$ and $\mathbf{X}$ is Markov-(Q, x), then $\mathbf{X}$ visits B infinitely often with probability one.*

Lemma 3.2. *Let Q be a stochastic kernel on E, let $\ell \in \mathbb{N}$ and let $B \in \mathscr{E}$. If B is Harris recurrent for Q^ℓ then B is Harris recurrent for Q.*

For completeness, both lemmas are proved in the appendix.

Finally, let us consider product processes and product kernels, which represent pairs of independent Markov processes. As before, let Q

---

[9]Definitions vary. In Meyn and Tweedie (1993), B is called Harris recurrent if $\mathbf{P}_x^Q(V_B) = 1$ for all $x \in B$.

be any stochastic kernel on E, and let $Q \times Q$ be the stochastic kernel on $E \times E$ defined by

$$(8) \qquad (Q \times Q)((x, x'), A) = \int \int \mathbb{1}_A(y, y') Q(x, dy) Q(x', dy')$$

for $(x, x') \in E \times E$ and $A \in \mathscr{E} \otimes \mathscr{E}$. The kernel $Q \times Q$ corresponds to the bivariate chain $(X_t, X'_t)_{t \geq 0}$ where $(X_t)_{t \geq 0}$ and $(X'_t)_{t \geq 0}$ are independent with stochastic kernel Q. An inductive argument confirms that for any $n \in \mathbb{N}$ we have $Q^n \times Q^n = (Q \times Q)^n$.

4. ORDER MIXING

We are now ready to define order mixing and investigate its properties. In what follows, we consider processes that take values in a measurable space $(S, \mathscr{S})$ with partial order $\leq$. We let $\mathscr{P}(S)$ denote the set of probability measures on $(S, \mathscr{S})$, and let ibS denote the set of increasing bounded functions from S into $\mathbb{R}$.[10] A set $B \subset S$ is called increasing if its indicator function $\mathbb{1}_B$ is an increasing function; that is, if $x \leq x'$ and $x \in B$ implies $x' \in B$. Increasing sets are assumed to be elements of $\mathscr{S}$, from which it follows that ibS is contained in the $\mathscr{S}$-measurable functions.

4.1. **Definitions.** In the product space $(S \times S, \mathscr{S} \otimes \mathscr{S})$, the sets

$$(9) \quad L := \{(x, x') \in S \times S : x \leq x'\}, \ U := \{(x, x') \in S \times S : x' \leq x\}$$

are assumed to be measurable (i.e., elements of $\mathscr{S} \otimes \mathscr{S}$). Using these sets, we can now give a precise definition to the concept of order mixing.

Definition 4.1. Let P be a stochastic kernel on S. The kernel P is defined to be *order mixing* if the sets L and U in (9) are Harris recurrent for the joint kernel $P \times P$.

This is just a formalization of the idea discussed in the introduction. If P is order mixing, then any independent bivariate chain $(X_t, X'_t)_{t \geq 0}$ starting at arbitrary initial conditions $(x, x') \in S \times S$ (that is, any Markov-$(P \times P, (x, x'))$ process) visits L at least once—which is to

---

[10]Thus, $h \in ibS$ if $\sup_{x \in S} |h(x)| < \infty$ and $h(x) \leq h(x')$ whenever $x \leq x'$.

say that $X_t \leq X'_t$ occurs at least once—with probability one. The same is true for the set U.

From Definition 4.1 it is clear that order mixing is purely a restriction on the stochastic kernel P. Also, when proving order mixing of P it is sufficient to prove that P^ℓ is order mixing for some $\ell \in \mathbb{N}$:

Lemma 4.1. *Let P be a stochastic kernel on S. If P^ℓ is order mixing for some $\ell \in \mathbb{N}$ then P is also order mixing.*

4.2. **Order Inducing Sets.** In order to identify order mixing in applications, we introduce the concept of order inducing sets. Loosely speaking, processes which return to order inducing sets infinitely often are order mixing. Order inducing sets are defined as follows:

Definition 4.2. Let P be a stochastic kernel on S. A set $C \in \mathscr{S}$ is called *order inducing* for P if there exists an $m \in \mathbb{N}$ and an $\epsilon > 0$ such that for any $(x, x') \in C \times C$ we have

$$(10) \qquad (P \times P)^m((x, x'), L) \geq \epsilon \quad \text{and} \quad (P \times P)^m((x, x'), U) \geq \epsilon$$

where L and U are the sets defined in (9).

In other words, C is order inducing if there is an $m \in \mathbb{N}$ and an $\epsilon > 0$ such that, for any $x, x' \in C$, independent processes $(X_t)_{t \geq 0}$ and $(X'_t)_{t \geq 0}$ run from these initial conditions attain $X_m \leq X'_m$ and $X'_m \leq X_m$ with at least ϵ probability. Clearly, any measurable subset of an order inducing set is also order inducing.

The definition of order inducing sets is stated in terms of joint kernels, but conditions stated in terms of the individual kernel P are easily derived. For example, a set $C \in \mathscr{S}$ is order inducing whenever there exists a $c \in S$, and $\epsilon > 0$ and an $m \in \mathbb{N}$ such that, $\forall x \in C$,

$$(11) \qquad P^m(x, \{y : y \leq c\}) \geq \epsilon \quad \text{and} \quad P^m(x, \{y : y \geq c\}) \geq \epsilon$$

To see that (11) implies that C is order inducing, consider the first inequality in (10). Since $\{(y, y') : y \leq c \leq y'\} \subset L$, we have

$$(P \times P)^m((x, x'), L) \geq (P \times P)^m((x, x'), \{(y, y') : y \leq c \leq y'\})$$
$$= (P \times P)^m((x, x'), \{y : y \leq c\} \times \{y' : c \leq y'\})$$
$$= P^m(x, \{y : y \leq c\}) P^m(x', \{y' : c \leq y'\}) \geq \epsilon^2$$

The argument for the second inequality in (10) is similar.

Example 4.1. Consider the Monotone Mixing Condition (MMC) of Hopenhayn and Prescott (1992, p. 1397), which generalizes the ideas of Razin and Yahav (1979) and Stokey and Lucas (1989), and pertains to state spaces S with a least element a and greatest element b.[11] Recall that P is called *monotone increasing* if $x \mapsto P(x, B)$ is increasing for every increasing set B. In this setting, the MMC is said to hold for kernel P whenever there exists a $c \in S$ and $m \in \mathbb{N}$ such that $P^m(a, [c, b]) > 0$ and $P^m(b, [a, c]) > 0$.

Under the MMC, the entire state space S is order inducing. To verify this, observe from monotonicity of P that $x \mapsto P^m(x, [c, b])$ is monotone increasing. The assertion that S is order inducing now follows from the sufficient condition presented in (11). Indeed, for any $x \in S$, we have

$$P^m(x, \{y : y \geq c\}) = P^m(x, [c, b]) \geq P^m(a, [c, b]) > 0$$

The proof of the other case is similar.

Example 4.2. Bhattacharya and Lee (1988) and Bhattacharya and Majumdar (2001) consider a "splitting condition" and its relationship to stability. Their environment consists of a sequence of IID random maps $(\alpha_t)_{t \geq 1}$ on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and a process $(X_t)_{t \geq 0}$ generated by

$$X_t = \alpha_t X_{t-1} = \alpha_t \circ \cdots \circ \alpha_1(x)$$

where $x \in S$ is the initial condition. Their splitting condition requires the existence of a $c \in S$ and $m \in \mathbb{N}$ such that

(a) $\mathbb{P}\{\alpha_m \circ \cdots \circ \alpha_1(x) \leq c, \ \forall x \in S\} > 0$; and
(b) $\mathbb{P}\{\alpha_m \circ \cdots \circ \alpha_1(x) \geq c, \ \forall x \in S\} > 0$

Under this condition, the entire state space S is again order inducing. To see this, observe that (11) is satisfied for the same choice of c and m, with ϵ as the infimum of the two probabilities in (a) and (b) above. Indeed, if P is the corresponding stochastic kernel, then for any $x \in S$ we have

$$\begin{aligned} P^m(x, \{y : y \leq c\}) &= \mathbb{P}\{\alpha_m \circ \cdots \circ \alpha_1(x) \leq c\} \\ &\geq \mathbb{P}\{\alpha_m \circ \cdots \circ \alpha_1(x) \leq c, \ \forall x \in S\} \geq \epsilon \end{aligned}$$

---

[11] That is, $a \leq x \leq b$ for all $x \in S$.

The argument for the second inequality in (11) is similar. Hence S is order inducing.

Thus, existing results correspond to the case where, in our notation, the entire state space S is order inducing. As we will see, this is sufficient but not necessary for order mixing. Indeed, the set of order mixing models is considerably larger than the set of models such that S is order inducing, as the next example demonstrates.

Example 4.3. Consider the elementary AR(1) process

$$(12) \qquad\qquad X_{t+1} = aX_t + W_{t+1}$$

where $S = \mathbb{R}$ and $(W_t)_{t \geq 1}$ is a univariate and standard normal. For this model, S is not order inducing whenever $a \neq 0$, and hence neither the MMC nor the splitting condition is satisfied. Let's check this for $a > 0$. Given two processes (X_t) and (X_t') following (12) with $X_0 = x$ and $X_0' = x'$, one can show that $X_m = U + a^m x$ and $X_m' = U' + a^m x'$ for some normally distributed U and U'. It follows that for any fixed $m \in \mathbb{N}$ we have

$$
\begin{aligned}
\mathbb{P}\{X_m \leq X_m'\} &= \mathbb{P}\{U + a^m x \leq U' + a^m x'\} \\
&= \mathbb{P}\{U - U' \leq a^m(x' - x)\} \to 0 \quad \text{as } (x, x') \to (\infty, -\infty)
\end{aligned}
$$

Hence S is not order inducing. The case $a < 0$ is similar.

On the other hand, sets of the form $[-K, K]$ *are* order inducing, as can be seen from (11): Let $c = 0$ and $m = 1$. For any $x \in [-K, K]$ we have

$$
\begin{aligned}
\mathbb{P}\{X_1 \geq c \mid X_0 = x\} &= \mathbb{P}\{ax + W_1 \geq 0\} \\
&= \mathbb{P}\{W_1 \geq -ax\} \geq \mathbb{P}\{W_1 \geq |a|K\} > 0
\end{aligned}
$$

The reverse case $\mathbb{P}\{X_1 \leq c \mid X_0 = x\}$ follows from a similar argument.

While for the AR(1) model S is not order inducing, the model is order mixing whenever $|a| < 1$. If one fixes an $K \in \mathbb{N}$, then $[-K, K]$ is order inducing. Moreover, the mean reverting assumption $|a| < 1$ implies that if we run two independent copies $(X_t)_{t \geq 0}$ and $(X_t')_{t \geq 0}$ of the process, then the pair X_t and X_t' will return simultaneously to $[-K, K]$ infinitely often with probability one. After each visit there

is an $\epsilon > 0$ probability of achieving the orderings $X \leq c \leq X'$ and $X' \leq c \leq X$. These facts can be shown to imply order mixing.

Rather than making this argument rigorous, we now turn to the general case, using similar intuition to link order inducing sets to the property of order mixing.

4.3. **Sufficient Conditions.**

4.3. **Sufficient Conditions.** Our most general sufficient condition for order mixing is given next. To state the theorem, recall that $W_{C \times C}$ is defined in (7) as all sequences in $S \times S$ which visit $C \times C$ infinitely often.

Theorem 4.1. *A kernel P is order mixing if, for each pair $(x, x') \in S \times S$, there exists an order inducing set $C \subset S$ with*

$$\mathbf{P}^{P \times P}_{(x,x')}(W_{C \times C}) = 1$$

The theorem states that if, for each initial condition of the bivariate process $(X_t, X'_t)_{t \geq 0}$, there exists an order inducing set C such that the bivariate process visits $C \times C$ infinitely often with probability one, then P is order mixing. Although the proof is nontrivial, the intuition behind the theorem is straightforward: If C is order inducing, then each time X_t and X'_t are both in C, the event $X_{t+m} \leq X'_{t+m}$ occurs with some positive probability ϵ. Independent events that occur with positive probability infinitely often must occur eventually with probability one. Hence $\mathbb{P}\{\exists t \geq 0 \text{ s.t. } X_t \leq X'_t\} = 1$. The argument for $\mathbb{P}\{\exists t \geq 0 \text{ s.t. } X'_t \leq X_t\} = 1$ is similar.

It was previously mentioned without proof that if the entire state space S is order inducing for a given kernel P, then P is order mixing. This is an immediate corollary of theorem 4.1:

Corollary 4.1. *If S is order inducing, then P is order mixing.*

While theorem 4.1 is suitably general, it is not particularly convenient, because it is a restriction on the product kernel $P \times P$. The next two theorems are more specialized, but concern P rather than $P \times P$. To state the first one, we introduce the idea of order norm-like functions.

Definition 4.3. A measurable function $v\colon S \to \mathbb{R}_+$ is called *order norm-like* if, for every $K \in \mathbb{R}_+$, the sublevel set $\{x \in S : v(x) \le K\}$ is order inducing.

Example 4.4. In the case of the AR(1) model in Example 4.3, the function $v(x) = |x|$ is order norm-like, because for every $K \ge 0$ the set

$$C := \{x \in S : v(x) \le K\} = \{x \in S : |x| \le K\} = [-K, K]$$

is order inducing (see Example 4.3).

We can now state the following key result:

Theorem 4.2. *Let P be a stochastic kernel on S. The kernel P is order mixing whenever there exists an order norm-like function $v\colon S \to \mathbb{R}_+$ and constants $\ell \in \mathbb{N}$, $\alpha \in [0,1)$ and $\beta \in \mathbb{R}_+$ such that*

$$(13) \qquad P^{\ell}v(x) := \int v(y) P^{\ell}(x, dy) \le \alpha v(x) + \beta \qquad (x \in S)$$

The drift condition (13) is a form of mean reversion, implying that the process returns to sets on which v is bounded. By the definition of v, these sets are order inducing. In fact, (13) can be used to show that the process returns to an order inducing set C infinitely often. The proof is in the appendix.

Example 4.5. The AR(1) model satisfes the conditions of the theorem whenever $|a| < 1$, because $v(x) = |x|$ is order norm-like (see example 4.4) and

$$\int v(y) P(x, dy) = \mathbb{E}|ax + W|$$

$$\le |ax| + \mathbb{E}|W| = |a|v(x) + \mathbb{E}|W|$$

Lastly, we consider a sufficient condition for order mixing which is useful when S has a topology. Specifically, we assume that S is a Polish space, and take $\mathscr{S}$ to be the Borel sets. In this case a condition closely related to existence and stability of stationary distributions is tightness of the marginal distributions.[12] Following Meyn and Tweedie (1993, p. 145), we say that P is *bounded in probability* if $(P^t(x, \cdot))_{t \ge 1}$ is tight for every $x \in S$.

---

[12]Recall that a collection $\mathscr{M}$ of probabilities on S is called tight if, for all $\epsilon > 0$, there exists a compact set $K \subset S$ such that $\mu(K) \ge 1 - \epsilon$ for all $\mu \in \mathscr{M}$.

Theorem 4.3. *If P is bounded in probability and all compact subsets of S are order inducing, then P is order mixing.*

5. STABILITY VIA ORDER MIXING

Our main result concerning order mixing and stability is stated in terms of the SRS

$$(14) \qquad X_{t+1} = F(X_t, W_{t+1}), \qquad \mathscr{D}X_0 = \psi$$

Here X_t takes values in the partially ordered set S, and the shock sequence $(W_t)_{t \geq 1}$ is IID and takes values in a measurable space $(Z, \mathscr{Z})$ according to distribution ϕ. The function F is assumed to be $(\mathscr{S} \otimes \mathscr{Z}, \mathscr{S})$-measurable. The process $(W_t)_{t \geq 1}$ and the initial condition X_0 are independent and defined on probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The stochastic kernel P corresponding to (14) is as defined in (4).

Note that the formulation (14) is relatively general. Many models with additional lags and non-IID shocks can be expressed in the form (14) by readjusting the definition of the state variables. Further, if S is separable and completely metrizable (as is the case for every G_δ subset of $\mathbb{R}^k$), then every stochastic kernel P on S can be represented by an SRS in the form of (14).[13]

Regarding the process (14) we make the following assumptions:

Assumption 5.1. The process (14) is monotone increasing.

Assumption 5.2. The kernel P corresponding to (14) is order mixing.

We can now state the main result of the paper. In the statement of the theorem, ψ_t is the distribution of X_t, where $(X_t)_{t \geq 0}$ is Markov-(P, ψ).

Theorem 5.1. *If Assumptions 5.1 and 5.2 hold and ψ^* is stationary for P, then given any $\psi \in \mathscr{P}(S)$, we have*

$$(15) \qquad \int h d\psi_t \to \int h d\psi^* \ as \ t \to \infty, \quad \forall h \in ibS$$

---

[13]In other words, for each kernel P there exists a representation (14) such that the process $(X_t)_{t \geq 0}$ is Markov-(P, ψ). See Kifer (1986).

Here *ibS* is the set of increasing bounded functions from S to $\mathbb{R}$. For the benchmark case where S is a Borel subset of $\mathbb{R}^k$ and $\mathscr{S}$ is the Borel sets, the convergence in (15) is stronger than the standard definition of convergence in distribution (i.e., "weak" convergence).[14]

Corollary 5.1. *If Assumptions 5.1 and 5.2 hold and ibS separates the points of $\mathscr{P}(S)$, then P has at most one stationary distribution.*[15]

Proof of Corollary 5.1. The proof is trivial. If ψ^* and ψ^{**} are both stationary, then (15) implies that $\int h d\psi^* = \int h d\psi^{**}$ for every $h \in ibS$. Since *ibS* separates $\mathscr{P}(S)$, we have $\psi^* = \psi^{**}$. $\square$

5.1. **Existence.** The conditions of theorem 5.1 are too general to imply the existence of a stationary distribution—even when the state space is a well behaved space such as $\mathbb{R}$. For example, consider the random walk $X_{t+1} = X_t + W_{t+1}$ with $W_t \sim N(0,1)$. This process is order mixing (since the 2-dimensional random walk is recurrent) and monotone but has no stationary distribution.

However, drift conditions such as (13) are also used in the Markov process literature to establish existence. For example, (13) is known to imply existence of a stationary distribution when the sublevel sets of v are compact and F is continuous in its first argument (cf., e.g., Meyn and Tweedie, 1993, prop. 12.1.3). If the compact sets are also order inducing, then v is order norm-like, and theorem 5.1 implies uniqueness and global stability. The next theorem packages this result:

Theorem 5.2. *Let S be a Borel subset of $\mathbb{R}^k$, and suppose that there exists a measurable function $v\colon S \to \mathbb{R}_+$ and constants $\alpha \in [0,1)$ and $\beta \in \mathbb{R}_+$ such that the drift condition (13) holds. If*

(1) *all sublevel sets of v are compact,*
(2) *all compact subsets of S are order inducing, and*
(3) *the map $x \mapsto F(x,z)$ is increasing and continuous for all $z \in Z$,*

---

[14]Recall that $\psi_t \to \psi^*$ weakly if $\int h d\psi_t \to \int h d\psi^*$ for all bounded continuous $h\colon S \to \mathbb{R}$. See Torres (1990, Corollary 6.5) for details.

[15]The set *ibS* separates the points of $\mathscr{P}(S)$ if for any μ and ν in $\mathscr{P}(S)$ such that $\int h d\mu = \int h d\nu$ for all $h \in ibS$ we have $\mu = \nu$. This holds whenever the increasing subsets of S generate $\mathscr{S}$, as is the case for the Borel subset of $\mathbb{R}^k$.

then a unique, globally stable stationary distribution exists.

Example 5.1. A simple example is provided by the AR(1) model in Example 4.3 when $a \in [0, 1)$. We saw above that the drift condition (13) holds for $v(x) = |x|$. Sublevel sets of v take the form $[-K, K]$, and hence are compact. It was shown that sets of the form $[-K, K]$ are order inducing, and the argument is easily extended to include any compact set. The function $x \mapsto ax + z$ is increasing and continuous for every $z \in \mathbb{R}$.

Example 5.1 is for illustrative purposes only, as stability of the AR(1) model is trivial. However, many of the special properties of the AR(1) model can be relaxed without violating the conditions of theorem 5.2. We now turn to more sophisticated applications.

6. APPLICATIONS

6.1. **A Standard Result.** Stokey and Lucas (1989, thm. 12.12) prove a widely cited result, which can be translated into our notation as follows. Consider the SRS (14). Let S be the order interval $[a, b] \subset \mathbb{R}^k$. Suppose that $S \ni x \mapsto F(x, z) \in S$ is continuous and increasing for every $z \in Z$, and that the MMC holds (see above). Stokey and Lucas then show that the process has a unique stationary distribution which is globally stable in the sense of weak convergence.

This is a special case of theorem 5.2. To see this, let $v = 0$, so that all sublevel sets of v are equal to S. For this choice of v the drift condition (13) is trivially satisfied, and all sublevel set of v are compact. Moreover, under the MMC, the entire state space S is order inducing (see above). Since measurable subsets of order inducing sets are order inducing, all compact subsets of S are order inducing. Condition (3) of theorem 5.2 is true by assumption.

As suggested by the AR(1) example, theorem 5.2 is more general than the result of Stokey and Lucas. This will be further illustrated in the examples below.

6.2. **The Brock–Mirman Model.** Next let us consider benchmark single sector stochastic optimal growth model studied by Brock and

Mirman (1972). For the sake of concreteness we adopt the same assumptions as Zhang (2007), and show how his results can be obtained as a special case of theorem 5.2.

Let $(W_t)_{t \geq 1}$ be an IID sequence on $S := (0, \infty)$ with common distribution ϕ, and let $f \colon \mathbb{R}_+ \to \mathbb{R}_+$ be a standard neoclassical production function. Following Zhang (2007, Assumptions 1–5) we assume that

Assumption 6.1. The production function satisfies $f(0) = 0$, $f' > 0$, $f'' < 0$, $\lim_{x \to 0} f'(x) = \infty$ and $\lim_{x \to \infty} f'(x) = 0$.

Assumption 6.2. The distribution ϕ satisfies both $\int z\phi(dz) < \infty$ and $\int z^{-1}\phi(dz) < \infty$. The shocks are unbounded in the sense that for any $x \in S$ we have $\phi\{z \leq x\} > 0$ and $\phi\{z \geq x\} > 0$.[16]

Let u be a differentiable, strictly concave and strictly increasing utility function with $\lim_{x \to 0} u'(x) = \infty$, and consider the maximization problem

$$(16) \qquad \max_{(k_t)_{t \geq 0}} \mathbb{E}\left[\sum_{t \geq 0} \delta^t u(y_t - k_t)\right] \quad \text{s.t.} \quad y_{t+1} = f(k_t)W_{t+1}$$

where $k_t \in [0, y_t]$ for all t and $\delta \in (0, 1)$ is a discount factor. Under certain regularity conditions[17] on u there exists a unique optimal policy $\sigma \colon \mathbb{R}_+ \to \mathbb{R}_+$ such that $(k_t)_{t \geq 0}$ defined by $k_t = \sigma(y_t)$ solves (16). The policy σ is known to be continuous, interior and strictly increasing. It generates an optimal path for income defined by

$$(17) \qquad\qquad y_{t+1} = f(\sigma(y_t))W_{t+1} \quad \text{with } y_0 \text{ given}$$

To prove stability of (17) on $S = (0, \infty)$, let us begin by observing that sets of the form $[a, b] \subset S$ are all order inducing. To see this, pick any $C := [a, b] \subset S$ and fix $c \in C$. For any $y \in C$ we have

$$\begin{aligned}
\mathbb{P}\{y_{t+1} \geq c \mid y_t = y\} &= \phi\{z : f(\sigma(y))z \geq c\} \\
&= \phi\{z : z \geq c/f(\sigma(y))\} \geq \phi\{z : z \geq c/f(\sigma(a))\} > 0
\end{aligned}$$

---

[16]The case of bounded shocks can also be treated using order mixing. The proof is not much different from Hopenhayn and Prescott (1992, section 6B).

[17]We refer to conditions under which the dynamic programming problem is finite and well defined. See Kamihigashi (2007) for a comprehensive discussion.

where the second last inequality follows from $y \geq a$ and the last inequality follows from unboundedness of the shock. A similar argument shows that $\mathbb{P}\{y_{t+1} \leq c \,|\, y_t = y\}$ is also bounded away from zero over $y \in C$, thereby completing the proof that C is order inducing.

Let $v(x) := u'(x - \sigma(x)) + x$. By manipulating the Euler equation, Nishimura and Stachurski (2005) show that for this choice of v there exist constants $\ell \in \mathbb{N}$, $\alpha \in [0,1)$ and $\beta \in \mathbb{R}_+$ such that the drift condition (13) holds. Moreover, v is order-norm like, since (i) every sublevel set of v is contained in an interval $[a,b] \subset S$ for sufficiently inducing a and large b, (ii) subsets of order inducing sets are order inducing, and (iii) every interval $[a,b] \subset S$ is order inducing (as discussed above).

The remaining conditions of theorem 5.2 are easily verified, implying existence of a unique and globally stable stationary distribution. We have now obtained the results of Zhang (2007) as a consequence of theorem 5.2.

6.3. **Credit Constrained Growth.** Next let us consider a model of growth in a small open economy due to Matsuyama (2004). In Matsuyama's model production is deterministic and multiple steady states may exist. When production is stochastic the set of dynamics is richer. We show that when shocks are sufficiently large global stability holds.[18] The stability problem is challenging as a result of the nonlinearities inherent in the law of motion.

In the model, agents live for two periods, with a unit mass of young born at the start of each period. At time t, the old own a total stock of the capital good given by k_t. This is combined with the labor of the young to produce per capita output of the consumption good given by $f(k_t)\xi_t$, where ξ_t is a stochastic productivity term. The production function f satisfies $f(0) = 0$, $f' > 0$, $f'' < 0$, $f'(0) = \infty$ and $f'(\infty) = 0$. Factor markets are competitive, with return on capital given by

---

[18]Despite global stability, strong persistence of initial conditions remains, replicating the poverty trap dynamics observed by Matsuyama. However, "growth miracles" occur with nonzero probability.

$f'(k_t)\xi_t$ and wages by

$$w_t := w(k_t, \xi_t) := [f(k_t) - k_t f'(k_t)]\xi_t$$

The young now invest their wages, either in international credit markets at the risk free world interest rate R, or in a domestic project which converts one unit of the consumption good into one unit of the capital good next period. Due to indivisibilities agents can start at most one such project, so $0 \leq k_{t+1} \leq 1$.

From profit maximization (assume risk neutrality) we obtain the restriction

$$R \leq f'(k_{t+1})\mathbb{E}\left[\xi_{t+1}|\xi_t\right]$$

which says that expected return on the project must dominate the risk free rate. If the inequality is strict it means that all young agents invest in the project, whence $k_{t+1} = 1$. The excess cost of the project above wages, i.e., $1 - w_t$, is financed by borrowing abroad.

Due to imperfect credit markets, the liabilities $R(1 - w_t)$ of those agents borrowing abroad to finance the project cannot exceed a fraction λ of expected net worth at $t + 1$, which is $f'(k_{t+1})\mathbb{E}\left[\xi_{t+1}|\xi_t\right]$. Thus we have the additional restriction

$$R(1 - w_t) \leq \lambda f'(k_{t+1})\mathbb{E}\left[\xi_{t+1}|\xi_t\right]$$

which binds when $1 - w_t > \lambda$. We can combine these last two restrictions by setting

$$R \leq \Theta(w_t)f'(k_{t+1})\mathbb{E}\left[\xi_{t+1}|\xi_t\right]$$

where $\Theta(w)$ is equal to $\lambda/(1 - w)$ if $1 - w > \lambda$, and 1 otherwise. Again, if the inequality is strict then we have the corner solution $k_{t+1} = 1$. Letting g be the inverse of f' and incorporating the corner solution gives

$$k_{t+1} = \min\left\{g\left(\frac{R}{\Theta(w_t)\mathbb{E}\left[\xi_{t+1}|\xi_t\right]}\right), 1\right\}$$

Assume that the productivity process $(\xi_t)_{t \geq 0}$ follows the positively correlated AR(1) law

(18) $$\xi_{t+1} = \rho\xi_t + W_{t+1}$$

where $\rho \in [0, 1)$ and $(W_t)_{t \geq 1}$ is an IID sequence of nonnegative random variables with finite mean μ. From this law we can compute $\mathbb{E}[\xi_{t+1}|\xi_t]$ to give

$$(19) \qquad k_{t+1} = h(k_t, \xi_t) := \min \left\{ g \left(\frac{R}{\Theta(w(k_t, \xi_t))(\rho \xi_t + \mu)} \right), 1 \right\}$$

Equations (18) and (19) together with initial conditions k_0 and ξ_0 define a Markov process $(k_t, \xi_t)_{t \geq 0}$ on $S := [0, 1] \times \mathbb{R}_+$.

If the support of the innovation W is bounded then the law of motion for capital may have multiple steady states. If, on the other hand, the distribution has support equal to $\mathbb{R}_+$ (consider, for example, the lognormal density) then global stability can be established using theorem 5.2. This is significant because stability cannot be established using classical methods without additional assumptions.[19] Nor do the methods of Razin and Yahav (1979), Stokey, Lucas and Prescott (1989) and Hopenhayn and Prescott (1992) apply, since S is not compact. Finally, the process is not an average contraction as studied in, for example, Santos and Peralta-Alva (2005).

To prove stabilty using theorem 5.2, first observe that for $v \colon S \to \mathbb{R}_+$ defined by $v(k, \xi) = \xi$ we have

$$\mathbb{E}\left[v(k_1, \xi_1) \,|\, (k_0, \xi_0) = (k, \xi) \right]$$
$$= \mathbb{E}\left[\xi_1 \,|\, (k_0, \xi_0) = (k, \xi) \right] = \rho \xi + \mu = \rho v(k, \xi) + \mu$$

and since $\rho \in [0, 1)$ the drift condition (13) is satisfied. To check the remaining conditions of theorem 5.2, note that the joint law of motion

$$F((k, \xi), z) = (h(k, \xi), \rho \xi + z)$$

is increasing and continuous in (k, ξ) when z is held fixed. Thus the conditions of the theorem are satisfied if we can show that the sublevel sets

$$\{(k, \xi) \in S : v(k, \xi) \leq K\} = \{(k, \xi) \in S : \xi \leq K\} = [0, 1] \times [0, K]$$

are compact and order inducing in S. Since compactness is immediate we need only verify that sets of the form $[0, 1] \times [0, K]$ are order inducing. So in the definition of order inducing sets we take

---

[19]For example, irreducibility fails without additional assumptions on the shock process and even then is generally difficult to verify.

$C = [0,1] \times [0,K]$, $c = (1,K)$ and $m = 2$. First we show that the probability that $(k_2, \xi_2) \geq c = (1,K)$ is bounded below for any fixed initial condition $(k_0, \xi_0) = (k, \xi) \in C$. Since $\Theta(w) \geq \lambda$ for all w and g is decreasing we have

$$k_2 \geq g\left(\frac{R}{\lambda(\rho\xi_1 + \mu)}\right) \geq g\left(\frac{R}{\lambda\rho W_1}\right)$$

for any initial $(k, \xi) \in C$. In particular $k_2 \geq 1$ whenever

$$g\left(\frac{R}{\lambda\rho W_1}\right) \geq 1 \iff W_1 \geq \frac{R}{\lambda\rho f'(1)}$$

On the other hand, regarding ξ_2 we have

$$\xi_2 = \rho^2\xi + \rho W_1 + W_2$$

and hence $\xi_2 \geq K$ whenever $W_1 \geq K/\rho$. In summary,

$$(k_2, \xi_2) \geq c := (1,K) \text{ whenever } W_1 \geq \max\left\{\frac{R}{\lambda\rho f'(1)}, \frac{K}{\rho}\right\}$$

Since the support of W_1 is $\mathbb{R}_+$ the event on the right hand side has positive probability ϵ'. We have shown that

$$\mathbb{P}\{(k_2, \xi_2) \geq c \mid (k_0, \xi_0) = (k, \xi)\} \geq \epsilon' \qquad \forall\, (k, \xi) \in C$$

Similarly, one can establish that there is a positive ϵ'' such that

$$\mathbb{P}\{(k_2, \xi_2) \leq c \mid (k_0, \xi_0) = (k, \xi)\} \geq \epsilon'' \qquad \forall\, (k, \xi) \in C$$

and then set $\epsilon := \min\{\epsilon', \epsilon''\}$ in the definition of order inducing sets.

To check the last assertion, note that since $k_2 \leq 1$ always holds, we have $(k_2, \xi_2) \leq c = (1,K)$ whenever $\xi_2 \leq K$. Now observe that

$$\xi_2 = \rho^2\xi + \rho W_1 + W_2 \leq \rho^2 K + \rho W_1 + W_2$$

where the inequality follows from $(k, \xi) \in C$. Hence $\xi_2 \leq K$ when

$$\rho^2 K + \rho W_1 + W_2 \leq K \iff \rho W_1 + W_2 \leq K(1 - \rho^2)$$

Since the support of W is $\mathbb{R}_+$ this event occurs with positive probability ϵ''. That C is order inducing has now been established.

7. APPENDIX

This appendix collects all remaining proofs.

Proof of lemma 3.1. Meyn and Tweedie (1993, prop. 9.1.1) show that if $\mathbf{P}_x^Q(V_B) = 1$ for all $x \in B$, then $\mathbf{P}_x^Q(V_B) = \mathbf{P}_x^Q(W_B)$ for all $x \in E$. Thus if B is Harris recurrent, then $\mathbf{P}_x^Q(W_B) = 1$ for all $x \in E$. $\qquad\square$

Proof of lemma 3.2. Let Q, $\ell \in \mathbb{N}$ and $B \in \mathscr{E}$ be as in the statement of the lemma. Pick any $x \in E$. Let $\mathbf{X} = (X_t)_{t \geq 0}$ be an E-valued stochastic process on $(\Omega, \mathscr{F}, \mathbb{P})$ which is Markov-(Q, x). It follows that $\mathbf{Y} = (Y_t)_{t \geq 0}$ defined by $Y_t := X_{t \times \ell}$ is Markov-(Q^ℓ, x). We have

$$\mathbf{P}_x^Q(V_B) = \mathbb{P}\{\mathbf{X} \text{ ever enters } B\} \geq \mathbb{P}\{X_{t \times \ell} \in B \text{ for some } t \geq 0\}$$

$$= \mathbb{P}\{\mathbf{Y} \text{ ever enters } B\} = \mathbf{P}_x^{Q^\ell}(V_B) = 1$$

In other words, B is Harris recurrent for Q, as was to be shown. $\qquad\square$

Proof of lemma 4.1. Let P and $\ell \in \mathbb{N}$ be as in the statement of the lemma. We wish to show that the sets L and U defined in (9) are Harris recurrent for $P \times P$. By assumption, L is Harris recurrent for $P^\ell \times P^\ell$. But $P^\ell \times P^\ell = (P \times P)^\ell$, so L is Harris recurrent for $(P \times P)^\ell$. Lemma 3.2 then implies that L is Harris recurrent for $P \times P$. A similar argument shows that U is also Harris recurrent for $P \times P$. $\qquad\square$

In order to prove theorem 4.1, we begin with a lemma regarding visits of a sample path to a given set B. In the lemma, we take E to be any set, E^∞ the E-valued sequences, $B \subset E$ and $(x_t)_{t \geq 0}$ any element of E^∞. Define $N_t := \sum_{i=0}^t \mathbb{1}\{x_i \in B\}$ to be the number of visits to B up until date t, and let

$$J_1 := \inf\{t \geq 0 : x_t \in B\}, \quad J_{i+1} := \inf\{t \geq J_i + 1 : x_t \in B\}$$

so that J_i is the date of the i-th visit to B. We will be concerned below with visits to a set that are at least m periods apart, where m is a given integer. To this end, define

$$K_1 := J_1 \quad K_{i+1} := \inf\{t \geq K_i + m : x_t \in B\}$$

In what follows, the elements of $(K_i)_{i \geq 1}$ are called the *m-separated visits to B*. Regarding $(K_i)_{i \geq 1}$ and $(N_t)_{t \geq 0}$ we have the following relationship.

Lemma 7.1. *Let $n \in \mathbb{N}$. If $N_t > n \times m$, then $K_n + m \leq t$.*

In other words, if, as of date t, the number of visits to B exceeds $n \times m$, then the n-th m-separated visit to B occurs prior to $t - m$.

Proof of lemma 7.1. Fix $n \in \mathbb{N}$. As $K_{i+1} \geq K_i + m$, we have $K_n \geq K_1 + (n-1)m$, and hence $K_n + m \leq K_1 + n \times m = J_1 + n \times m$. As a result, the proof will be complete if we can show that

(20) $$J_1 + n \times m \leq t$$

In general, $J_{i+1} \geq J_i + 1$ and hence $J_{i+k} \geq J_i + k$.

$$\therefore \quad J_i \leq J_{i+k} - k$$

$$\therefore \quad J_1 \leq J_{n \times m + 1} - n \times m$$

Since $N_t > n \times m$, we have $J_{n \times m + 1} \leq t$. From this and the previous inequality we obtain $J_1 \leq t - n \times m$. This is precisely (20), which completes the proof of lemma 7.1. $\square$

We will also make use of the following lemma.

Lemma 7.2. *Suppose that $\mathbf{P}_x^Q(W_B) = 1$ for some $x \in E$ and $B \in \mathcal{E}$, and that $\mathbf{X} = (X_t)_{t \geq 0}$ is Markov-(Q, x). If $N_t := \sum_{i=0}^{t} \mathbb{1}\{X_i \in B\}$ counts the number of visits to B prior to t, then for any $j \in \mathbb{N}$ we have $\mathbb{P}\{N_t \leq j\} \to 0$ as $t \to \infty$.*

Proof. Let N_t and $\mathbf{X}$ be as defined in the lemma. Evidently

$$\cup_j \cap_t \{N_t \leq j\} = \{\mathbf{X} \in W_B\}^c$$

Since $\mathbb{P}\{\mathbf{X} \in W_B\}^c = \mathbf{P}_x^Q(W_B^c) = 0$, for any $j \in \mathbb{N}$ we have

$$\lim_{t \to \infty} \mathbb{P}\{N_t \leq j\} = \mathbb{P} \cap_t \{N_t \leq j\} = 0$$

$\square$

We are now ready to prove theorem 4.1. In the proof we will make use of the strong Markov property,[20] which states that if σ is a stopping time and $B \in \mathcal{E}$ then any Markov-(Q, μ) process $(X_t)_{t \geq 0}$ satisfies

(21) $$\mathbb{P}[X_{\sigma+n} \in B \mid \mathscr{F}_\sigma] = Q^n(X_\sigma, B) \qquad (n \geq 0)$$

[20]See Meyn and Tweedie (1993, prop. 3.4.6).

Proof of theorem 4.1. Fix $(x, x') \in S \times S$. Let V_L (resp., V_U) be the set of sequences in $(S \times S)^\infty$ which visit L (resp., U) at least once, where L and U are defined in (9). To establish order mixing, we must show that

$$\mathbf{P}^{P \times P}_{(x,x')}(V_L) = \mathbf{P}^{P \times P}_{(x,x')}(V_U) = 1$$

Let $(X_t, X'_t)_{t \geq 0}$ be any Markov-$(P \times P, (x, x'))$ process. If

$$\tau := \inf\{t \geq 0 : X_t \leq X'_t\} \quad \text{and} \quad \sigma := \inf\{t \geq 0 : X'_t \leq X_t\}$$

then, since $(X_t, X'_t)_{t \geq 0}$ has distribution $\mathbf{P}^{P \times P}_{(x,x')}$, it suffices to prove that

$$\mathbb{P}\{\tau < \infty\} = \mathbb{P}\{\sigma < \infty\} = 1$$

We will restrict attention to the case of τ, as the proof for σ is similar. To show that $\mathbb{P}\{\tau < \infty\} = 1$, let C be the set in the statement of the theorem. Since C is order inducing, there exists an $m \in \mathbb{N}$ and an $\epsilon > 0$ such that

$$(22) \qquad (P \times P)^m((x, x'), L) \geq \epsilon, \qquad \forall (x, x') \in C \times C$$

Let N_t be the number times $(X_t, X'_t)_{t \geq 0}$ visits $C \times C$ up until date t:

$$N_t := \sum_{i=0}^{t} \mathbb{1}\{(X_i, X'_i) \in C \times C\}$$

Consider the decomposition

$$(23) \quad \mathbb{P}\{\tau > t\} = \mathbb{P}\{\tau > t, N_t \leq n \times m\} + \mathbb{P}\{\tau > t, N_t > n \times m\}$$

Let us consider first the second term on the right hand side of (23). We claim that for n, t with $n \times m \leq t$ we have

$$(24) \qquad\qquad \mathbb{P}\{\tau > t, N_t > n \times m\} \leq (1 - \epsilon)^n$$

To see this, let $(K_i)_{i \geq 1}$ be the *m*-separated visits of $(X_t, X'_t)_{t \geq 0}$ to $C \times C$. That is,

$$K_1 := \inf\{t \geq 0 : (X_t, X'_t) \in C \times C\}$$

$$K_{i+1} := \inf\{t \geq K_i + m : (X_t, X'_t) \in C \times C\}$$

In view of lemma 7.1, $\{N_t > n \times m\} \subset \{K_n + m \leq t\}$. As a result,

$$(25) \qquad \mathbb{P}\{\tau > t, N_t > n \times m\} \leq \mathbb{P}\{\tau > t, K_n + m \leq t\}$$

Consider the set $\{\tau > t, K_n + m \leq t\}$. If a path is in this set, then as $\tau > t$, for any index j with $j \leq t$ we have $X_j \not\leq X'_j$. In addition,

$K_i + m \leq K_n + m \leq t$ for any $i \leq n$, so we have $X_{K_i+m} \not\leq X'_{K_i+m}$ for every $i \leq n$.

(26) $\therefore$ $\mathbb{P}\{\tau > t, K_n + m \leq t\} \leq \mathbb{P}\left[\bigcap_{i=1}^{n} \{X_{K_i+m} \not\leq X'_{K_i+m}\}\right]$

We now show that the term on the right hand side is bounded by $(1 - \epsilon)^n$, which verifies the claim in (24). This can be demonstrated as follows.

$$\mathbb{P}\left[\bigcap_{i=1}^{n} \{X_{K_i+m} \not\leq X'_{K_i+m}\}\right] = \mathbb{P}\left[\mathbb{P}\left[\bigcap_{i=1}^{n} \{X_{K_i+m} \not\leq X'_{K_i+m}\} \mid \mathscr{F}_{K_n}\right]\right]$$

$$= \mathbb{P}\left[\bigcap_{i=1}^{n-1} \{X_{K_i+m} \not\leq X'_{K_i+m}\} \, \mathbb{P}[X_{K_n+m} \not\leq X'_{K_n+m} \mid \mathscr{F}_{K_n}]\right]$$

Using the strong Markov property (21) we have

$$\mathbb{P}[X_{K_n+m} \leq X'_{K_n+m} \mid \mathscr{F}_{K_n}] = (P \times P)^m((X_{K_n}, X'_{K_n}), L) \geq \epsilon$$

where the last inequality uses (22) and the fact that both X_{K_n} and X'_{K_n} are by definition contained in the order inducing set C.

$$\therefore \quad \mathbb{P}[X_{K_n+m} \not\leq X'_{K_n+m} \mid \mathscr{F}_{K_n}] \leq (1 - \epsilon)$$

$$\therefore \quad \mathbb{P}\left[\bigcap_{i=1}^{n} \{X_{K_i+m} \not\leq X'_{K_i+m}\}\right] \leq (1 - \epsilon)\mathbb{P}\left[\bigcap_{i=1}^{n-1} \{X_{K_i+m} \not\leq X'_{K_i+m}\}\right]$$

Continuing to iterate backwards in this way yields

$$\mathbb{P}\left[\bigcap_{i=1}^{n} \{X_{K_i+m} \not\leq X'_{K_i+m}\}\right] \leq (1 - \epsilon)^n$$

Combining this inequality with (25) and (26) verifies the claim in (24). Returning to (23), then, we have

$$\mathbb{P}\{\tau > t\} = \mathbb{P}\{\tau > t, N_t \leq n \times m\} + (1 - \epsilon)^n$$
$$\leq \mathbb{P}\{N_t \leq n \times m\} + (1 - \epsilon)^n$$

for any t and any n such that $n \times m \leq t$. Fix $n \in \mathbb{N}$. Since $C \times C$ is Harris recurrent for the bivariate process, lemma 7.2 implies that $\mathbb{P}\{N_t \leq n \times m\} \to 0$ as $t \to \infty$. Hence

$$\limsup_{t \to \infty} \mathbb{P}\{\tau > t\} \leq (1 - \epsilon)^n, \quad \forall n \in \mathbb{N}$$

Therefore $\lim_t \mathbb{P}\{\tau > t\} = 0$, and $\mathbb{P}\{\tau < \infty\} = 1$. $\square$

Next we turn to the proof of theorem 4.2. In the proof we can and do assume that $\ell = 1$. The reasoning is as follows. Suppose that the theorem is true for the case $\ell = 1$. Now pick any other $\ell \in \mathbb{N}$ and suppose the conditions of the theorem hold. Since $Q := P^\ell$ is a stochastic kernel in its own right, and since the theorem holds for $\ell = 1$, the kernel Q is order mixing. But if $Q = P^\ell$ is order mixing then P itself is order mixing by lemma 4.1. In all of what follows we set $\ell = 1$ in (13) without further comment.

To prove theorem 4.2 we will make use of the next two lemmas.

Lemma 7.3. *Let $(E, \mathscr{E})$ be any measurable space, let $B \in \mathscr{E}$, and let Q be a stochastic kernel on E. If there exists a measurable function $w \colon E \to [1, \infty)$ and a $\lambda \in [0, 1)$ such that $Qw(x) \leq \lambda w(x)$ whenever $x \notin B$, then B is Harris recurrent for Q.*

Proof. Pick any $x \in E$, and let $(X_t)_{t \geq 0}$ be Markov-(Q, x). Define $\eta := \min\{t \geq 0 : X_t \in B\}$. We will show that $\mathbb{P}\{\eta > t\} \to 0$ as $t \to \infty$. To this end, let $M_t := w(X_t)\mathbb{1}\{\eta > t - 1\}$ with $M_0 := w(x)$. This process is a supermartingale. Indeed,

$$
\begin{aligned}
\mathbb{E}[w(X_{t+1})\mathbb{1}\{\eta > t\} \,|\, \mathscr{F}_t] &= \mathbb{E}[w(X_{t+1}) \,|\, \mathscr{F}_t]\mathbb{1}\{\eta > t\} \\
&= Qw(X_t)\mathbb{1}\{\eta > t\} \\
&\leq \lambda w(X_t)\mathbb{1}\{\eta > t\} \\
&\leq \lambda w(X_t)\mathbb{1}\{\eta > t - 1\}
\end{aligned}
$$

Taking expectations we obtain $\mathbb{E}M_{t+1} \leq \lambda \mathbb{E}M_t$, and, iterating backwards to M_0, $\mathbb{E}M_t \leq \lambda^t w(x)$ for all t. As a result we have

$$
\mathbb{P}\{\eta > t\} \leq \mathbb{E}\mathbb{1}\{\eta > t\}w(X_{t+1}) = \mathbb{E}M_{t+1} \leq \lambda^{t+1}w(x) \to 0
$$

It follows that $\mathbb{P}\{\eta = \infty\} = 0$, and B is Harris recurrent. $\square$

Lemma 7.4. *If P satisfies the conditions of theorem 4.2, then there exists an order inducing set C, a measurable function $w \colon S \times S \to [1, \infty)$ and constant $\lambda \in [0, 1)$ such that*

$$
(27) \qquad (P \times P)w(x, x') \leq \lambda w(x, x') \text{ whenever } (x, x') \notin C \times C
$$

Proof. Let P be any stochastic kernel on S, and let v, α and β be as in theorem 4.2. We can without any loss of generality assume that $v \geq 1$, as simple algebra shows that if the conditions in theorem 4.2

hold for given v, α and β, then they also hold for v', α' and β', where $v' := v + 1$, $\alpha' := \alpha$ and $\beta' := \beta + 1$.

So suppose that $v \geq 1$. Choose any d such that $\alpha + \beta/d < 1$. Let $C := \{x \in S : v(x) \leq 2d\}$, which is order inducing by the defintion of v. Define the function w and constant λ by

$$w(x, x') := \frac{v(x) + v(x')}{2}, \quad \lambda := \alpha + \frac{\beta}{d}$$

Evidently $w \geq 1$ and $\lambda < 1$. Thus it remains only to show that (27) holds on $S \times S$. In doing so we will make use of the inequality

$$(P \times P)w(x, x') \leq \alpha w(x, x') + \beta, \quad (x, x') \in S \times S$$

which can be verified with some straightforward calculations.

Now pick any $(x, x') \notin C \times C$. Then $v(x) + v(x') > 2d$, and hence $w(x, x') > d$.

$$\therefore \quad \frac{(P \times P)w(x, x')}{w(x, x')} \leq \alpha + \frac{\beta}{w(x, x')} < \alpha + \frac{\beta}{d} =: \lambda$$

This proves (27), and hence lemma 7.4. □

Proof of theorem 4.2. In view of lemmas 7.3 and 7.4, there exists an order inducing set C such that $C \times C$ is Harris recurrent for $P \times P$. Order mixing now follows from lemma 3.1 and theorem 4.1. □

Proof of theorem 4.3. Our first claim is that if P is bounded in probability on S, then $P \times P$ is bounded in probability on $S \times S$.[21] To see this, pick any $(x, x') \in S \times S$. Fix $\epsilon > 0$. Since P is bounded in probability, we can choose compact sets K and K' such that $P^n(x, K) \geq (1 - \epsilon)^{1/2}$ and $P^n(x', K') \geq (1 - \epsilon)^{1/2}$ for all n. It follows that

$$(P \times P)^n((x, x'), K \times K') = (P^n \times P^n)((x, x'), K \times K')$$
$$= P^n(x, K)P^n(x', K') \geq 1 - \epsilon, \quad \forall n \in \mathbb{N}$$

Since $K \times K$ is compact in $S \times S$, the sequence of measures in tight, and $P \times P$ is bounded in probability as claimed.

By Meyn and Tweedie (1993, prop. 12.1.1), for any given pair $(x, x') \in S \times S$, there exists a compact set $M \subset S \times S$ such that $\mathbf{P}^{P \times P}_{(x,x')}(W_M) =$

---

[21]The topology on $S \times S$ is the product topology.

1. Since M is compact we can choose a compact $C \subset S$ such that $M \subset C \times C$. The set C is order inducing by assumption, and $\mathbf{P}^{P \times P}_{(x,x')}(W_{C \times C}) = 1$. theorem 4.1 now implies that P is order mixing. $\qquad\square$

Proof of theorem 5.1. Now let us turn to the proof of theorem 5.1, an outline of which was presented in section 2. Let ψ^* and ψ be as in the statement of the theorem. Let $(W_t)_{t \geq 1}$ and $(W_t^*)_{t \geq 1}$ to be independent Z-valued processes defined on probability space $(\Omega, \mathscr{F}, \mathbb{P})$, all having distribution ϕ. Let $\mathbf{X} = (X_t)_{t \geq 0}$ and $\mathbf{X}^* = (X_t^*)_{t \geq 0}$ be defined by

$$X_{t+1} = F(X_t, W_{t+1}), \quad \mathscr{D}X_0 = \psi$$
$$X_{t+1}^* = F(X_t^*, W_{t+1}^*), \quad \mathscr{D}X_0 = \psi^*$$

where X_0 and X_0^* are mutually independent. The first process is Markov-(P, ψ) on S, while the second is Markov-(P, ψ^*). By construction, the process $\mathbf{X}^* := (X_t^*)_{t \geq 0}$ is stationary, satisfying $\mathscr{D}X_t^* = \psi^*$ for all $t \geq 0$. We aim to establish (15).

To begin, we introduce the pair of stopping times

(28) $\quad \tau := \inf\{t \geq 0 : X_t \leq X_t^*\} \quad \text{and} \quad \sigma := \inf\{t \geq 0 : X_t^* \leq X_t\}$

with the usual convention that $\inf \varnothing = \infty$. By the order mixing assumption we have

$$\mathbb{P}\{\tau < \infty\} = \mathbb{P}\{\sigma < \infty\} = 1$$

To verify this claim, note that the distribution of the bivariate chain $(X_t, X_t^*)_{t \geq 0}$ in the sequence space $((S \times S)^\infty, (\mathscr{S} \otimes \mathscr{S})^\infty)$ is precisely $\mathbf{P}^{P \times P}_{\psi \times \psi^*}$, where the product kernel $P \times P$ is defined in (8), and the joint distribution $\mathbf{P}^{P \times P}_{\psi \times \psi^*}$ is defined in (5).[22] This is true because $(X_t, X_t^*)_{t \geq 0}$ is Markov-$(P \times P, \psi \times \psi^*)$, as can be shown by checking the requirements of Definition 3.1. (The measure-theoretic details are routine and hence omitted.)

Now let V_L the set of sequences $(x_t, x_t')_{t \geq 0}$ in $(S \times S)^\infty$ such that $x_t \leq x_t'$ for some t. Evidently

$$\{\omega \in \Omega : \tau(\omega) < \infty\} = \{\omega \in \Omega : (X_t(\omega), X_t^*(\omega))_{t \geq 0} \in V_L\}$$
$$\therefore \quad \mathbb{P}\{\tau < \infty\} = \mathbf{P}^{P \times P}_{\psi \times \psi^*}(V_L)$$

---

[22]Here $\psi \times \psi^*$ is the product measure of ψ and ψ^* on $S \times S$.

Since P is order mixing we have $\mathbf{P}^{P \times P}_{(x,x')}(V_L) = 1$ for all pairs $(x, x') \in S \times S$. Applying (6) now gives

$$\mathbb{P}\{\tau < \infty\} = \mathbf{P}^{P \times P}_{\psi \times \psi^*}(V_L) = 1$$

The proof that $\mathbb{P}\{\sigma < \infty\} = 1$ is similar and hence omitted.

To continue with the proof of stability, consider the two auxillary processes $\mathbf{X}^L = (X^L_t)_{t \geq 0}$ and $\mathbf{X}^U = (X^U_t)_{t \geq 0}$ defined by

$$(29) \qquad X^L_{t+1} = \begin{cases} F(X^L_t, W_{t+1}) & \text{if } t < \tau \\ F(X^L_t, W^*_{t+1}) & \text{if } t \geq \tau \end{cases} \quad \text{and} \quad X^L_0 = X_0$$

$$(30) \qquad X^U_{t+1} = \begin{cases} F(X^U_t, W_{t+1}) & \text{if } t < \sigma \\ F(X^U_t, W^*_{t+1}) & \text{if } t \geq \sigma \end{cases} \quad \text{and} \quad X^U_0 = X_0$$

They can be understood as follows: $\mathbf{X}^L$ is identical to the original process $\mathbf{X}$ defined in (3) until $t = \tau$; that is, until the first time that $\mathbf{X}$ falls below $\mathbf{X}^*$. At τ its source of shocks switches from $(W_t)_{t \geq 1}$, the shocks driving $\mathbf{X}$, to $(W^*_t)_{t \geq 1}$, the shocks driving $\mathbf{X}^*$. The process $\mathbf{X}^U$ is similar, but switches shocks when $\mathbf{X}$ first *exceeds* $\mathbf{X}^*$.

Two properties of $\mathbf{X}^L$ and $\mathbf{X}^U$ are crucial for the proof:

(i) $\mathscr{D}X^L_t = \mathscr{D}X^U_t = X_t$ for all $t \geq 0$; and
(ii) if $t \geq \tau$ then $X^L_t \leq X^*_t$, while if $t \geq \sigma$ then $X^U_t \geq X^*_t$.

The first claim follows from the next lemma.

Lemma 7.5. *The processes $\mathbf{X}^L$ and $\mathbf{X}^U$ defined in (29) and (30) respectively are both Markov-(P, ψ) on S.*

The intuition was discussed in section 2. The formal proof is somewhat routine and given below.

Claim (ii) follows from monotonicity. Consider for example the statement $t \geq \tau$ implies $X^L_t \leq X^*_t$. By the definition of τ we have $X_\tau \leq X^*_\tau$. Since $X^L_\tau = X_\tau$, this implies that $X^L_\tau \leq X^*_\tau$ also holds. The source of shocks for $\mathbf{X}^L$ now switches from $(W_t)_{t \geq 1}$ to $(W^*_t)_{t \geq 1}$, so X^L_τ and X^*_τ are updated with the same shocks. Monotonicity implies that processes updated with the same shocks maintain their initial order. In particular,

$$X^L_\tau \leq X^*_\tau \quad \text{implies} \quad X^L_{\tau+1} = F(X^L_\tau, W^*_{\tau+1}) \leq F(X^*_\tau, W^*_{\tau+1}) = X^*_{\tau+1}$$

Iterating forward now gives $X_t^L \leq X_t^*$ for all $t \geq \tau$.

To establish global stability of ψ^* we will show that

(31) $\qquad \forall h \in ibS, \quad \limsup_{t \to \infty} \int h d\psi_t \leq \int h d\psi^* \leq \liminf_{t \to \infty} \int h d\psi_t$

To do so, fix any $h \in ibS$, and consider the left hand inequality in (31). In view of (ii) above, on the set $\{\tau \leq t\}$ we have $X_t^L \leq X_t^*$, and hence $h(X_t^L) \leq h(X_t^*)$. By order mixing the probability of the set $\{\tau \leq t\}$ converges to one as $t \to \infty$, and as a result we obtain the inequality in the next lemma. (The proof is below.)

Lemma 7.6. *We have* $\limsup_{t \to \infty} \mathbb{E}h(X_t^L) \leq \limsup_{t \to \infty} \mathbb{E}h(X_t^*)$.

It follows that

$$\limsup_{t \to \infty} \int h d\psi_t \leq \limsup_{t \to \infty} \mathbb{E}h(X_t^*) \qquad (\because \ \mathscr{D}X_t^L = \mathscr{D}X_t = \psi_t)$$

$$\therefore \quad \limsup_{t \to \infty} \int h d\psi_t \leq \int h d\psi^* \qquad (\because \ \mathscr{D}X_t^* = \psi^*, \ \forall\, t \geq 0)$$

Thus we have established the inequality on the left hand side of (31). The proof of the right hand inequality is similar.[23] $\qquad \square$

Proof of lemma 7.5. We prove only the case of $\mathbf{X}^L$, as that of $\mathbf{X}^U$ is similar. Pick any bounded measurable function $h \colon S \to \mathbb{R}$. We have

$$h(X_{t+1}^L) = h(X_{t+1}^L)\mathbb{1}\{t < \tau\} + h(X_{t+1}^L)\mathbb{1}\{t \geq \tau\}$$

$$\therefore \quad \mathbb{E}[h(X_{t+1}^L) \mid \mathscr{F}_t] =$$
$$\mathbb{E}[h(X_{t+1}^L)\mathbb{1}\{t < \tau\} \mid \mathscr{F}_t] + \mathbb{E}[h(X_{t+1}^L)\mathbb{1}\{t \geq \tau\} \mid \mathscr{F}_t]$$

From the definition of τ we have

$$h(X_{t+1}^L)\mathbb{1}\{t < \tau\} = h[F(X_t^L, W_{t+1})]\mathbb{1}\{t < \tau\}$$

$$h(X_{t+1}^L)\mathbb{1}\{t \geq \tau\} = h[F(X_t^L, W_{t+1}^*)]\mathbb{1}\{t \geq \tau\}$$

Taking expectations gives

$$\mathbb{E}[h(X_{t+1}^L)\mathbb{1}\{t < \tau\} \mid \mathscr{F}_t] = \mathbb{E}[h[F(X_t^L, W_{t+1})]\mathbb{1}\{t < \tau\} \mid \mathscr{F}_t]$$

$$= \mathbb{E}[h[F(X_t^L, W_{t+1})] \mid \mathscr{F}_t]\mathbb{1}\{t < \tau\} = \int h[F(X_t^L, z)]\phi(dz)\mathbb{1}\{t < \tau\}$$

---

[23]Use $\mathbf{X}^U$ in place of $\mathbf{X}^L$ and σ in place of τ.

A similar calculation shows that

$$\mathbb{E}[h(X_{t+1}^L)\mathbb{1}\{t \geq \tau\} \mid \mathscr{F}_t] = \int h[F(X_t^L, z)]\phi(dz)\mathbb{1}\{t \geq \tau\}$$

Adding the last two expressions gives

$$\mathbb{E}[h(X_{t+1}^L) \mid \mathscr{F}_t] = \int h[F(X_t^L, z)]\phi(dz)$$

Specializing to the case $h = \mathbb{1}_B$ gives

$$\mathbb{P}[X_{t+1}^L \in B \mid \mathscr{F}_t] = \int \mathbb{1}_B[F(X_t^L, z)]\phi(dz) = P(X_t^L, B)$$

This proves our claim that $(X_t^L)_{t \geq 0}$ is Markov-(P, ψ) on S. $\qquad\square$

Proof of lemma 7.6. We claim that

$$\limsup_{t \to \infty} \mathbb{E}h(X_t^L) \leq \limsup_{t \to \infty} \mathbb{E}h(X_t^*) = \int h d\psi^* \quad (h \in ibS)$$

Note that since h is increasing we have $h(X_t^L) \leq h(X_t^*)$ on the set $\{\tau \leq t\}$. Thus $h(X_t^L)\mathbb{1}\{\tau \leq t\} \leq h(X_t^*)\mathbb{1}\{\tau \leq t\}$, and hence

$$\mathbb{E}h(X_t^L)\mathbb{1}\{\tau \leq t\} \leq \mathbb{E}h(X_t^*)\mathbb{1}\{\tau \leq t\}$$

Since $h(X_t^L) = h(X_t^L)\mathbb{1}\{\tau \leq t\} + h(X_t^L)\mathbb{1}\{\tau > t\}$ we have

$$\limsup_{t \to \infty} \mathbb{E}h(X_t^L)$$
$$\leq \limsup_{t \to \infty} \mathbb{E}h(X_t^L)\mathbb{1}\{\tau \leq t\} + \limsup_{t \to \infty} \mathbb{E}h(X_t^L)\mathbb{1}\{\tau > t\}$$

By assumption h is bounded by some constant M, and

$$\limsup_{t \to \infty} \mathbb{E}h(X_t^L)\mathbb{1}\{\tau > t\} \leq M \limsup_{t \to \infty} \mathbb{P}\{\tau > t\} = 0$$

$$\therefore \quad \limsup_{t \to \infty} \mathbb{E}h(X_t^L) \leq \limsup_{t \to \infty} \mathbb{E}h(X_t^L)\mathbb{1}\{\tau \leq t\}$$

$$\therefore \quad \limsup_{t \to \infty} \mathbb{E}h(X_t^L) \leq \limsup_{t \to \infty} \mathbb{E}h(X_t^*)\mathbb{1}\{\tau \leq t\}$$

The inequality in (7) will thus be verified if we can show that

$$(32) \qquad\qquad \limsup_{t \to \infty} \mathbb{E}h(X_t^*)\mathbb{1}\{\tau \leq t\} = \int h d\psi^*$$

This equality holds because

$$\mathbb{E}h(X_t^*)\mathbb{1}\{\tau \le t\} = \mathbb{E}h(X_t^*) - \mathbb{E}h(X_t^*)\mathbb{1}\{\tau > t\}$$
$$= \int h d\psi^* - \mathbb{E}h(X_t^*)\mathbb{1}\{\tau > t\}$$

And boundedness of h gives $\mathbb{E}h(X_t^*)\mathbb{1}\{\tau > t\} \to 0$ as $t \to \infty$. $\qquad\square$

REFERENCES

[1] Bhattacharya, R. N. and O. Lee (1988): "Asymptotics of a Class of Markov Processes which are Not in General Irreducible," *The Annals of Probability,* 16 (3), 1333–1347.

[2] Bhattacharya, R. N. and M. Majumdar (2001): "On a Class of Stable Random Dynamical Systems: Theory and Applications," *Journal of Economic Theory,* 96, 208–229.

[3] Brock, W. A. and L. J. Mirman (1972): "Optimal Economic Growth and Uncertainty: the Discounted Case," *Journal of Economic Theory,* 4, 479–513.

[4] Doeblin, W. (1938): "Exposé de la theorie des chaîns simples constantes de Markov à un nombre fini d'états," *Rev. Math. Union Interbalkanique,* 2, 77–105.

[5] Hopenhayn, H.A. and E. C. Prescott (1992): "Stochastic Monotonicity and Stationary Distributions for Dynamic Economies," *Econometrica,* 60, 1387–1406.

[6] Kamihigashi, T. (2007): "Stochastic Optimal Growth with Bounded or Unbounded Utiltity and Bounded or Unbounded Shocks," *Journal of Mathematical Economics,* 43, 477–500.

[7] Kifer, Y. (1986): *Ergodic Theory of Random Transformations,* Birkhäuser, Boston.

[8] Lindvall, T. (1992): *Lectures on the Coupling Method,* Dover Publications, New York.

[9] Matsuyama, K. (2004): "Financial Market Globalization, Symmetry-Breaking, and Endogenous Inequality of Nations," *Econometrica,* 72, 853-884.

[10] Meyn, S. and R. L. Tweedie (1993): *Markov Chains and Stochastic Stability,* Springer-Verlag, New York.

[11] Nishimura, K. and J. Stachurski (2005): "Stability of Stochastic Optimal Growth Models: A New Approach," *Journal of Economic Theory,* 122 (1), 100–118.

[12] Pollard, D. (2002): *A User's Guide to Measure Theoretic Probability,* Cambridge University Press, Cambridge.

[13] Razin, A. and J. A. Yahav (1979): "On Stochastic Models of Economic Growth," *International Economic Review,* 20, 599–604.

[14] Rosenthal, J. S. 2002. "Quantitative Convergence Rates of Markov Chains: A Simple Account. *Electronic Communications in Probability* 7, 123–28.

[15] Santos, M. S. and A. Peralta-Alva (2005): "Accuracy of Simulations or Stochastic Dynamic Models," *Econometrica,* 73, 1939–1976.

[16] Stokey, N.L., Lucas, R.E., and Prescott, E.C. (1989): *Recursive Methods in Economic Dynamics,* Harvard University Press, Massachusetts.

[17] Torres, R. (1990): "Stochastic Dominance," manuscript, Northwestern University.

[18] Zhang, Y. (2007): "Stochastic Optimal Growth with a Non-Compact State Space," *Journal of Mathematical Economics,* 43, 115–129.

RESEARCH INSTITUTE FOR ECONOMICS AND BUSINESS ADMINISTRATION, KOBE UNIVERSITY, 2-1, ROKKODAI CHO, NADA-KU, KOBE, 657-8501, JAPAN

E-mail address: tkamihig@rieb.kobe-u.ac.jp

INSTITUTE OF ECONOMIC RESEARCH, KYOTO UNIVERISTY, YOSHIDA HONMACHI, SAKYO-KU, KYOTO 606–8501, JAPAN

E-mail address: john@keir.kyoto-u.ac.jp