ISSN 0819-2642
ISBN 0 7340 2583 1

**THE UNIVERSITY OF MELBOURNE**

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 927

FEBRUARY 2005

# Computable Bounds for Extreme Event Probabilities in Stochastic Economic Models

by

John Stachurski

Department of Economics
The University of Melbourne
Melbourne Victoria 3010
Australia.

# COMPUTABLE BOUNDS FOR EXTREME EVENT PROBABILITIES IN STOCHASTIC ECONOMIC MODELS

JOHN STACHURSKI

Abstract. The paper introduces a multiplicative drift condition for evaluating stochastic economic models. The drift condition is shown to permit computation of quantitative bounds for extreme event probabilities in terms of the model primitives. By way of illustration, the technique is applied to a simple threshold autoregression model of exchange rates.

## 1. Introduction

We consider a stochastic economic model which generates a sequence of state variables $(X_t)_{t=0}^{\infty}$ taking values in $S \subset \mathbb{R}^n$. Let the (marginal) distribution of $X_t$ be denoted by $\psi_t$, which is a probability distribution on $S$. We derive a uniform bound on the tails of each $\psi_t$ given suitable conditions on the primitives. In addition, when $\psi_t$ converges to some limiting distribution $\psi^*$ as $t \to \infty$, we derive similar bounds on the tails of $\psi^*$.

These tail bounds can be regarded as bounds on probabilities of extreme events. Extreme events are thought of as those which occur only infrequently, but potentially have large impact. A classic example is large movements in share prices. For example, the stock market crash on 19th October 1987 saw the Dow Jones index drop by 23% in one day, wiping out nearly US$1 trillion in market capitalization. The financial crisis that engulfed many Asian economies in the middle of

The author thanks Vance Martin and Lawrence Uren for helpful comments.

1997 likewise led to huge percentage changes in the exchange rates of several Asian currencies, with far-reaching economic consequences.

The potential impact of such events in financial markets has led to considerable research on extreme event probabilities. The standard methodology is Extreme Value Theory, which bounds the tails of the running maximum $M_t := \max\{X_0, \ldots, X_t\}$ generated by an independent and identically distributed sequence $(X_t)_{t=0}^{\infty}$. Unfortunately, the IID assumptions precludes application of this theory to dynamic economic models, which typically involve at least some degree of correlation. In this paper we permit all finite orders of correlation in $(X_t)_{t=0}^{\infty}$, and study not the running maximum but rather the state variables themselves, as well as any ergodic limit they might converge to.

The methodology is based on a new multiplicative drift condition (MDC) for Markov chains. Our MDC complements the more standard additive drift conditions, used extensively in the existing literature to establish stability and stationarity of stochastic processes.[1] It is the source of the computable tail bounds derived in the paper.

Previously, Borovkov (1998, Theorem 3.1) also studied bounds on the tails of the stationary distributions of Markov chains. His bounds are not directly comparable with those given here. The main difference is in the conditions on the primitives used to derive the bounds. The conditions used here exploit the MDC discussed above. This drift technique is intended to fit the kind of equilibrium structure typically available in economic models. For example, in our exchange rate application, the drift is due to arbitrage, which pushes the expected value of the rate towards its purchasing power parity equilibrium.

Our results have many applications for the modeling of financial variables. For example, heavy tails have been observed in many kinds

[1] See, for example, Meyn and Tweedie (1993), or Borovkov (1998).

of market returns data.[2] The property of having heavy tails is often linked informally with "chaotic" or highly nonlinear behavior in the model which describes motion of the system. One of the contributions of this paper is to show that a large class of highly nonlinear and discontinuous models in fact generate marginal and stationary distributions with exponentially decreasing tails. These models therefore cannot represent time series which empirically are observed to feature heavy tails.

Another potential application of this research is when the state variable is itself a distribution. For example, it often happens that in macroeconomic dynamics one wishes to study a situation where each entity in a given economic model has a vector of endogenously evolving attributes, such as income, wealth, asset holdings, human capital, wage rate, and so on. The state of the economy is given by the distribution of these attributes across the population. In this case, the size of the distribution tails provides a measure of dispersion.

Section 2 formulates the problem. Section 3 sets out the multiplicative drift condition and derives some of its immediate consequences. Section 4 gives a number of applications which illustrate the method. The proofs are in Section 5.

## 2. Formulation of the Problem

Consider a process evolving in state space $S$, a Borel subset of $\mathbb{R}^n$. The law of motion is given by

$$X_{t+1} = h(X_t, \xi_{t+1}), \quad X_0 \equiv x_0 \in S \text{ given}, \quad (\xi_t)_{t=0}^{\infty} \text{ IID.} \tag{1}$$

The variables $X_t$ all take values in $S$, the shocks $\xi_t$ take values in $Z$, a Borel subset of $\mathbb{R}^k$, and $h$ is a measurable function mapping $S \times Z \to S$.

[2]A classic early reference is Mandelbrot (1963). For a more recent overview see Rachev (2001).

The shocks are generated on probability space $(\Omega, \mathscr{F}, \mathbf{P})$, and $\mathbf{E}$ is the expectations operator corresponding to $\mathbf{P}$.

In time series modeling and macroeconomic dynamics it is common to deal with seemingly more complex models than (1). For example, $X_{t+1}$ might depend on $X_t, \ldots, X_{t-j}$ for some $j$, and the shocks might themselves be correlated of some finite order. However, such models can always be rewritten in the form of (1) by suitably expanding the number of state variables. As a result, in all of what follows we concentrate only on models with this simple first order representation (1).

As a matter of notation, for topological space $T$, we let $\mathscr{B}(T)$ denote the Borel sets, and $\mathscr{P}(T)$ denote the probability measures on $(T, \mathscr{B}(T))$. The common distribution of $\xi_t$ is denoted by $\varphi \in \mathscr{P}(Z)$, while that of $X_t$ is denoted by $\psi_t \in \mathscr{P}(S)$. Also, $\mathbb{1}_B$ is the indicator function of $B$. Thus, for example, $\mathbf{E}\, \mathbb{1}_B \circ X_t = \psi_t(B)$ holds for every $B \in \mathscr{B}(S)$.

A common measure of convergence for elements of $\mathscr{P}(T)$ is via the *total variation* distance. For elements $\mu$ and $\nu$ in $\mathscr{P}(T)$ we define

$$\|\mu - \nu\|_{TV} = \sup_{B \in \mathscr{B}(T)} |\mu(B) - \nu(B)|.$$

For $(\mu_n)_{n=0}^{\infty} \subset \mathscr{P}(T)$ and $\mu \in \mathscr{P}(T)$ we say that $\mu_n$ converges to $\mu$ in total variation if $\|\mu - \nu\|_{TV} \to 0$ as $n \to \infty$. If $(X_n)_{n=0}^{\infty}$ and $X$ are $T$-valued random variables, we say that $X_n \to X$ in total variation if the distribution of $X_n$ converges in total variation to that of $X$ (as elements of $\mathscr{P}(T)$).[3]

We also define stationary distributions and ergodicity. A probability $\psi^* \in \mathscr{P}(S)$ is called *stationary* for (1) iff

$$\int \left[ \int \mathbb{1}\{z : h(x,z) \in B\} \varphi(dz) \right] \psi^*(dx) = \psi^*(B), \quad \forall B \in \mathscr{B}(S).$$

[3] It is well-known and easy to check that convergence in total variation is stronger than convergence in distribution in the usual sense. See, for example, Stokey, Lucas and Prescott (1989, Chapters 10–11).

If the current (i.e., time $t$) distribution is $\psi^*$, then the left hand side gives the probability that $X_{t+1} \in B$. Thus, if $\psi^*$ satisfies this equation, then this probability is $\psi^*(B)$, which is the same as it is today. Since this holds for all $B$, we have $\psi_t = \psi_{t+1} = \psi^*$.

The process (1) is called *ergodic* if there exists a unique stationary distribution $\psi^* \in \mathscr{P}(S)$ for (1), independent of $x_0$, and, in addition, $\psi_t$ converges to $\psi^*$ in total variation. It is *geometrically ergodic* if, moreover, $\|\psi_t - \psi^*\|_{TV} = O(\varrho^t)$ for some $\varrho < 1$.

## 3. A Multiplicative Drift Condition

Our main results are derived from the following multiplicative drift condition. The first lemma gives an immediate implication of the condition. The second result develops connections between the drift condition and geometric ergodicity.

**Condition 3.1.** There exists a measurable function $w$ mapping $S \to [1, \infty)$ and constants $\beta \in [0, \infty)$ and $\alpha \in (0, 1)$ such that

$$\int w[h(x,z)]\varphi(dz) \leq \beta[w(x)]^\alpha \quad \text{for all } x \in S.$$

Most of the interesting consequences of this condition are derived from the following lemma. Its proof and those of all other results are deferred to Section 5.

**Lemma 3.1.** *Let $(X_t)_{t=0}^\infty$ be the sequence defined inductively by* (1)*. If $h$ and $\varphi$ satisfy Condition 3.1, then $\sup_{t \in \mathbb{N}} \mathbf{E}\, w(X_t) \leq w(x_0)\beta^{\frac{1}{1-\alpha}}$. If, moreover, $w$ is bounded on compact sets, and if $X$ is an $S$-valued random variable such that $X_t \to X$ in total variation, then $\mathbf{E}\, w(X) \leq w(x_0)\beta^{\frac{1}{1-\alpha}}$ also holds.*

Extreme event bounds can be constructed from the results in Lemma 3.1 via Chebychev's inequality. For example, suppose that $\psi$ is the distribution of random variable $X$, and that $\mathbf{E} w(X) = \int w d\psi < \infty$, where, for the sake of concreteness, we take $w(x) = e^{\|x\|}$. Then Chebychev's inequality implies that

$$\mathbf{P}\{\|X\| > r\} = \mathbf{P}\{e^{\|X\|} > e^r\} \leq \mathbf{E}e^{\|X\|}e^{-r}, \tag{2}$$

so finiteness of $\mathbf{E}w(X) = \mathbf{E}e^{\|X\|}$ gives $\mathbf{P}\{\|X\| > r\} = O(e^{-r})$.

We now specialize (1) to the common case where the shock $\xi_t$ is additive. Precisely, the state space $S = \mathbb{R}^n$, $\xi_t$ also takes values in $S$, and $h(x, z) = g(x) + z$, where $g \colon S \to S$ is a measurable function. Thus,

$$X_{t+1} = g(X_t) + \xi_{t+1}, \quad X_0 \equiv x_0 \in S. \tag{3}$$

If $w$ has a "Lyapunov function" shape, then Condition 3.1 also has stability implications. To state the precise result, we need the notion of a norm-like function. Here, a measurable real-valued function $w$ is called *norm-like* if the sublevel sets $\{x \in S : w(x) \leq a\}$ are bounded, $\forall a \in \mathbb{R}$.[4]

**Theorem 3.1.** *Let $(X_t)_{t=0}^{\infty}$ be the sequence defined inductively by* (3)*. If Condition 3.1 holds for norm-like $w$, and, in addition, the distribution $\varphi$ of $\xi_t$ admits a density representation which is continuous and strictly positive on $S$, then $(X_t)_{t=0}^{\infty}$ is geometrically ergodic.*

## 4. Applications

We begin this section by consider the Markov chain $(X_t)_{t=0}^{\infty}$ generated by the additive shock model (3). A number of general results are given, followed by an application to exchange rate dynamics. The first result uses a growth condition on the function $g$ to establish an exponentially decreasing bound on the tail of $\psi_t$.

[4] In more general topological spaces, the sublevel sets of norm-like functions are required to have compact closure.

**Proposition 4.1.** *Let* $B_r := \{x \in S : \|x\| \leq r\}$, *and let* $(X_t)_{t=0}^{\infty}$ *be the sequence defined inductively by* (3). *If* $g$ *satisfies the constraint*

$$\|g(x)\| \leq c + \gamma\|x\|, \quad \forall\, x \in S \tag{4}$$

*for* $c \in [0, \infty)$ *and* $\gamma \in (0, 1)$, *then for all* $t \in \mathbb{N}$ *and all* $r > 0$ *we have*

$$\psi_t\,(S \setminus B_r) = \mathbf{P}\{\|X_t\| > r\} \leq \left[e^c \int e^{\|z\|}\varphi(dz)\right]^{\frac{1}{1-\gamma}} e^{\|x_0\|-r}. \tag{5}$$

The growth condition (4) permits $g$ to be discontinuous and highly nonlinear. It is equivalent to the statement that there exists a hypersphere $B \subset S = \mathbb{R}^n$ centered on the origin such that $\|g(x)\|$ is bounded for $x \in B$, and on the complement of $B$ the map $g$ is contracting, in the sense that $\exists\gamma \in (0, 1)$ such that $\|g(x)\| \leq \gamma\|x\|$ for all $x \in S \setminus B$. Similar restrictions have been used elsewhere in economics and finance. See, for example, Duffie and Singleton (1993).

The second application adds sufficient mixing to imply geometric ergodicity. It is then shown that $\psi^*$, the stationary distribution of the state variable and the long-run equilibrium of the system, also inherits a similar tail bound.

**Proposition 4.2.** *Let* $(X_t)_{t=0}^{\infty}$ *be the sequence defined inductively by* (3). *If, in addition to the hypotheses of Proposition 4.1, the distribution* $\varphi$ *of* $\xi_t$ *admits a density representation which is continuous, strictly positive on* $S$, *and satisfies* $\int e^{\|z\|}\varphi(z)dz < \infty$, *then* $(X_t)_{t=0}^{\infty}$ *is geometrically ergodic, and the stationary distribution* $\psi^*$ *satisfies*

$$\psi^*\,(S \setminus B_r) \leq \left[e^c \int e^{\|z\|}\varphi(dz)\right]^{\frac{1}{1-\gamma}} e^{-r}. \tag{6}$$

As above, we are using the notation $B_r := \{x \in S : \|x\| \leq r\}$. Note that, in contrast to the previous bound in (5), this bound does not depend on $x_0$.

As an example, consider the (self-exciting) threshold autoregression model, which has recently found many applications in macroeconomic modeling.[5] It has the form

$$X_{t+1} = \sum_{k=1}^{K} (A_k X_t + b_k) \mathbb{1}\{X_t \in B_k\} + \xi_{t+1}, \tag{7}$$

where $(B_k)_{k=1}^K \subset \mathscr{B}(S)$ is a partition of $S = \mathbb{R}^n$, each $A_k$ is an $n \times n$ matrix, and each $b_k$ is an $n \times 1$ vector. The structure of the model is such that when the state is in the region $B_k$, the state variable follows the regime $x \mapsto A_k x + b_k$. This structure allows for significant nonlinearities.

Without any loss of generality, suppose that the first $1, \ldots, J$ elements of the partition $(B_k)_{k=1}^K$ are unbounded, and the remaining $J+1, \ldots, K$ are bounded. Let $B$ be the union of the bounded elements $B_{J+1}, \ldots, B_K$. Evidently $g$ is bounded on bounded sets, so $a := \sup_{x \in B} \|g(x)\|$ is finite. Finally, set $b := \sup_{1 \leq k \leq J} \|b_k\|$, and $\varrho := \max_{1 \leq k \leq J} \varrho_k$, where $\varrho_k$ is the spectral radius of $A_k$.

**Proposition 4.3.** *If $\varrho < 1$, and if the distribution of $\xi_t$ is multivariate normal, then all of the conditions of Propositions 4.1 and 4.2 are satisfied. In particular, $(X_t)_{t=0}^\infty$ is geometrically ergodic, and the tail bounds* (5) *and* (6) *both hold when $c := a + b$ and $\gamma := \varrho$.*

To illustrate this result, consider Taylor's (2001) study of exchange rate dynamics and purchasing power parity (PPP). He uses a threshold autoregression of the form

$$X_{t+1} = \begin{cases} -\theta + \pi(X_t + \theta) + \xi_{t+1}, & \text{if } X_t < -\theta; \\ X_t + \xi_{t+1}, & \text{if } -\theta \leq X_t \leq \theta; \\ \theta + \pi(X_t - \theta) + \xi_{t+1}, & \text{if } X_t > \theta. \end{cases} \tag{8}$$

Here $X$ represents the proportional deviation of the real exchange rate from PPP. The idea of the model is that trade frictions result in a "band of inaction," given here by $[-\theta, \theta]$. In this band, transaction costs imply that no arbitrage is possible. Outside $[-\theta, \theta]$ there is drift

[5] See, for example, Hansen (2001), or Taylor (2001).

back towards the band, assuming that $\pi \in [0, 1)$. The shock $\xi_t$ is $N(0, \sigma^2)$.

Using the notation preceding Proposition 4.3, we can set $B = [-\theta, \theta]$, whence $a = \sup_{x \in B} |g(x)| = \theta$, and $b = \sup\{|(1-\pi)\theta|, |(-\pi+1)\theta|\} = (1-\pi)\theta$, so that $c = a + b = (2-\pi)\theta$. Also, $\varrho$ is the slope coefficient $\pi$. Applying these constants to Proposition 4.3 gives the equilibrium extreme value bound

$$\psi^* (S \setminus B_r) \leq \left[ e^{(2-\pi)\theta} \int e^{\|z\|} \varphi(dz) \right]^{\frac{1}{1-\pi}} e^{-r}, \tag{9}$$

where $\psi^*$ is the stationary distribution associated with the (geometrically ergodic) process (8).

## 5. Proofs

It is convenient to introduce some additional notation. Let $(\mathscr{F}_t)_{t=0}^\infty$ be any filtration to which $(\xi_t)_{t=0}^\infty$ is adapted. Also, if $w$ is a measurable real valued function on the state space $S$ which is either nonnegative or bounded, then we set $\mathbf{M}w(x) := \mathbf{E}w[h(x, \xi_t)] = \int_Z w[h(x, z)]\varphi(dz)$. The interpretation is that $\mathbf{M}w(x)$ is the expectation of $w(X_{t+1})$ when $X_t = x$. In fact we have $\mathbf{E}[w(X_{t+1}) \,|\, \mathscr{F}_t] = \mathbf{M}w(X_t)$. The intuition is clear and a formal proof is not difficult.[6]

*Proof of Lemma 3.1.* Pick any $t \in \mathbb{N}$. From the drift condition we get

$$\mathbf{M}\, w \circ X_t \leq \beta (w \circ X_t)^\alpha \text{ holds a.s. on } \Omega.$$

$$\therefore \quad \mathbf{E}\left[w \circ X_{t+1} \,|\, \mathscr{F}_t\right] \leq \beta (w \circ X_t)^\alpha.$$

$$\therefore \quad \mathbf{E}\, w \circ X_{t+1} \leq \beta \cdot \mathbf{E}\left[(w \circ X_t)^\alpha\right].$$

$$\therefore \quad \mathbf{E}\, w \circ X_{t+1} \leq \beta (\mathbf{E}\, w \circ X_t)^\alpha \quad (\because \text{ Jensen's inequality}).$$

Setting $y_t := \ln \mathbf{E}\, w \circ X_t$, it is easy to see that

$$y_t \leq y_0 + \frac{\ln \beta}{1 - \alpha}.$$

[6]See, for example, Taylor (1997, p. 225).

$$\therefore \quad \mathbf{E}\, w \circ X_t \leq w(x_0)\beta^{\frac{1}{1-\alpha}}. \tag{10}$$

Since $t$ is arbitrary the proof is done.

Now let $X_t \to X$ in total variation, and let $\psi^* \in \mathscr{P}(S)$ be the distribution of $X$. By the above argument we have $\mathbf{E}\, w \circ X_t \leq J$ for all $t$, where $J$ is the constant on the right hand side of (10). Convergence in total variation implies that for every bounded measurable $h\colon S \to \mathbb{R}$ we have $\int h d\psi_t \to \int h d\psi^*$. So let $s_n$ be the indicator function of the ball of radius $n$, and let $h_n := s_n \cdot w$, which is bounded by hypothesis. Then

$$\begin{aligned}\int w\, d\psi^* &= \lim_n \int h_n\, d\psi^* \qquad (\because \text{ Monotone Convergence Theorem}) \\ &= \lim_n \lim_t \int h_n\, d\psi_t \quad (\because\ h_n \text{ is bounded and measurable}) \\ &\leq \lim_n \lim_t \int w\, d\psi_t \leq J.\end{aligned}$$

□

*Proof of Theorem 3.1.* To establish geometric ergodicity we use the conditions of Theorem 15.0.1 in Meyn and Tweedie (1993). Precisely, the Markov chain $(X_t)_{t=0}^{\infty}$ generated on $S$ by (3) and starting at initial state $X_0 \equiv x \in S$ is geometrically ergodic whenever it is irreducible, aperiodic, and there exists a $r > 1$ and a petite set $C \subset S$ such that

$$\sup_{x \in C} \mathbf{E}\, r^{\tau_C^x} < \infty, \quad \text{where } \tau_C^x := \min\{t \geq 1 \,:\, X_t \in C\}. \tag{11}$$

The random variable $\tau_C^x$ is called the return time to $C$. The superscript $x$ indicates its dependence on the initial condition $x$. Clearly $\tau_C^x$ is a stopping time with respect to $(\mathscr{F}_t)_{t=0}^{\infty}$.

For definitions of irreducibility and aperiodicity see Meyn and Tweedie (1993, §§4.2.1 and §§5.4.3 respectively). We omit formal statement of these definitions and their verification, but a sufficient condition for a Markov chain to be irreducible and aperiodic is that any set $B \in \mathscr{B}(S)$

of positive Lebesgue measure can be reached in one step from any $x \in S$ with positive probability, which is to say that

$$\int \mathbb{1}\{g(x) + z \in B\} \varphi(z) dz = \int_{B-g(x)} \varphi(z) dz > 0.$$

This is immediate from the assumption that $\varphi > 0$ almost everywhere.

We also omit the definition of petite sets (see Meyn and Tweedie, 1993, §§5.5.2), but for a set $C \in \mathscr{B}(S)$ to be petite it is sufficient that there exists a measurable function $f \colon S \to [0, \infty)$ with $\int_S f > 0$ and

$$x \in C \quad \text{implies} \quad \varphi(y - g(x)) \geq f(y), \ \forall y \in S. \tag{12}$$

Let $C$ be any bounded set, and let $\delta := \inf_{x,y \in C \times C} \varphi(y - g(x))$. If $C$ has positive measure, and if $\delta > 0$, then we can take $f := \delta \mathbb{1}_C$, because if $x \in C$ then by the definition of $\delta$ we have $\varphi(y - g(x)) \geq f(y) = \delta \mathbb{1}_C(y)$.[7] But $\delta > 0$ must always hold for bounded $C$, because if $C$ is bounded then it must be contained in some ball of size $L$, so that when $(x, y) \in C \times C$ we have

$$\|y - g(x)\| \leq \|y\| + \|g(x)\| \leq \|y\| + c + \gamma \|x\| \leq c + (1 + \gamma) L =: M.$$

Thus $\delta = \inf_{x,y \in C \times C} \varphi(y - g(x)) \geq \inf_{\|z\| \leq M} \varphi(z)$, which is strictly positive because $\varphi$ is strictly positive and continuous. We conclude that all bounded sets of positive measure are petite.

Thus, it remains only to verify condition (11) for some $r > 0$ and some bounded set $C$ with positive measure. Evidently it is sufficient to prove

$$\exists \lambda < 1 \text{ and } N < \infty \text{ s.t. } \mathbf{P}\{\tau_C^x \geq t\} \leq N \lambda^t, \quad \forall x \in C, \tag{13}$$

because then

$$\sup_{x \in C} \mathbf{E}\, r^{\tau_C^x} \leq \sup_{x \in C} \sum_t r^t \mathbf{P}\{\tau_C^x \geq t\} \leq N \sum_t r^t \lambda^t,$$

which is finite whenever $r \in (1, 1/\lambda)$.

[7] Consider the two cases $y \in C$ and $y \notin C$.

To establish (13), let $\alpha$ and $\beta$ be as in Condition 3.1, and let $d$ be any number such that $d > \max\{\beta^{\frac{1}{1-\alpha}}, 1\}$, and that $C := \{x \in S : w(x) \leq d\}$ has positive measure. Note that $C$ is bounded, in view of the fact that $w$ is norm-like. Let $\lambda := \beta d^{\alpha-1}$. Note that $\lambda < 1$. Note also that

$$\text{if } x \notin C, \text{ then } \beta w(x)^\alpha \leq \lambda w(x), \tag{14}$$

because $x \notin C$ implies $w(x) > d$, and so $\beta w(x)^{\alpha-1} \leq \beta d^{\alpha-1} = \lambda$.

Note finally that if for such $\lambda$ and $C$ we define $Y_t := w \circ X_t \cdot \mathbb{1}\{\tau_C^x \geq t+1\}$, then

$$\mathbf{E}[Y_{t+1} \,|\, \mathscr{F}_t] \leq \lambda Y_t. \tag{15}$$

This is because

$$\begin{aligned}
\mathbf{E}[Y_{t+1} \,|\, \mathscr{F}_t] &= \mathbf{E}[w \circ X_{t+1} \cdot \mathbb{1}\{\tau_C^x \geq t+2\} \,|\, \mathscr{F}_t] \\
&\leq \mathbf{E}[w \circ X_{t+1} \cdot \mathbb{1}\{\tau_C^x \geq t+1\} \,|\, \mathscr{F}_t] \\
&= \mathbf{E}[w \circ X_{t+1} \,|\, \mathscr{F}_t] \cdot \mathbb{1}\{\tau_C^x \geq t+1\} \quad (\because\ \tau_C^x \text{ is a stopping time}) \\
&= [\mathbf{M}\, w \circ X_t] \cdot \mathbb{1}\{\tau_C^x \geq t+1\} \\
&\leq \beta (w \circ X_t)^\alpha \cdot \mathbb{1}\{\tau_C^x \geq t+1\},
\end{aligned}$$

and since $\tau_C^x \geq t+1$ implies that $X_t \notin C$, (14) now gives

$$\mathbf{E}[Y_{t+1} \,|\, \mathscr{F}_t] \leq \lambda \cdot w \circ X_t \cdot \mathbb{1}\{\tau_C^x \geq t+1\}, \tag{16}$$

which is (15).

We are now ready to complete the proof. Pick any $x \in C$. Since $\tau_C^x \geq t+1$ implies that $X_t \notin C$, which in turn gives $w \circ X_t > d > 1$, we have

$$\mathbf{P}\{\tau_C^x \geq t+1\} \leq \mathbf{E}\left[w \circ X_t \cdot \mathbb{1}\{\tau_C^x \geq t+1\}\right] = \mathbf{E}\, Y_t. \tag{17}$$

Moreover, taking expectations of both hand sides of (15) gives $\mathbf{E}\, Y_{t+1} \leq \lambda \mathbf{E}\, Y_t$, which in turn gives $\mathbf{E}\, Y_t \leq \lambda^t \mathbf{E}\, Y_0$. Since $\tau_C^x \geq 1$ is true by definition, this becomes $\mathbf{E}\, Y_t \leq \lambda^t w(x) \leq \lambda^t d$, where the second inequality follows from the fact that $x \in C$. From (17), then

$$\mathbf{P}\{\tau_C^x \geq t+1\} \leq \lambda^t d = N \lambda^{t+1},$$

where $N := d/\lambda$. This proves (13), and hence the theorem. □

*Proof of Proposition 4.1.* If $\int e^{\|z\|}\varphi(dz) = \infty$ then the bound is trivial. Suppose instead that it is finite. We claim that Condition 3.1 is satisfied for $w(x) := e^{\|x\|}$, $\beta := e^c \int e^{\|z\|}\varphi(dz)$ and $\alpha := \gamma$. To show this we must prove that

$$\int \exp(\|g(x) + z\|)\varphi(dz) \leq \beta e^{\alpha\|x\|}.$$

By the growth condition on $g$ we have

$$\|g(x) + z\| \leq \|g(x)\| + \|z\| \leq c + \gamma\|x\| + \|z\|.$$

$$\therefore \quad \int \exp(\|g(x) + z\|)\varphi(dz) \leq e^c \int e^{\|z\|}\varphi(dz) e^{\gamma\|x\|} = \beta e^{\alpha\|x\|}.$$

As a result, we can apply Lemma 3.1, which gives

$$\sup_t \mathbf{E}\, e^{\|X_t\|} \leq \left[e^c \int e^{\|z\|}\varphi(dz)\right]^{\frac{1}{1-\gamma}} e^{\|x_0\|}. \tag{18}$$

The bound (5) now follows from the Chebychev bound (2). □

*Proof of Proposition 4.2.* In the proof of Proposition 4.1 we already established that Condition 3.1 holds for $w(x) = e^{\|x\|}$, $\beta := e^c \int e^{\|z\|}\varphi(dz)$ and $\alpha := \gamma$. Clearly $w$ is norm-like. As a result, all of the conditions of Theorem 3.1 are satisfied, and the process is geometrically ergodic. Regarding (6), Lemma 3.1 and the Chebychev bound (2) give

$$\psi^* (S \setminus B_r) \leq \left[e^c \int e^{\|z\|}\varphi(dz)\right]^{\frac{1}{1-\gamma}} e^{\|x_0\| - r}. \tag{19}$$

Since this bound holds for all $x_0$ we can minimize over $x_0 \in S$. Doing so gives (6). □

*Proof of Proposition 4.3.* We need to verify the conditions of Propositions 4.1 and 4.2. The only one which is not clear is that (4) holds for

$c = a + b$ and $\gamma = \varrho$, where here $g(x) = \sum_{k=1}^{K}(A_k x + b_k)\mathbb{1}_{B_k}(x)$. For $x \notin B$ we have

$$\|g(x)\| = \left\| \sum_{k=1}^{J}(A_k x + b_k)\mathbb{1}_{B_k}(x) \right\|$$
$$\leq \sup_{1 \leq k \leq J} \|A_k x + b_k\| \leq \sup_{1 \leq k \leq J} \|A_k x\| + \sup_{1 \leq k \leq J} \|b_k\| \leq \gamma\|x\| + b.$$

As a result, whether $x \in B$ or $x \in S \setminus B$ we have

$$\|g(x)\| \leq a + \gamma\|x\| + b = c + \gamma\|x\|.$$

□

## References


[1] Borovkov, A.A. (1998): *Ergodicity and Stability of Stochastic Processes,* Wiley, New York.

[2] Duffie, D. and K. J. Singleton (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica,* 61 (4) 929–952.

[3] Hansen, B. (2001): "Threshold Autoregression with Unit Root," *Econometrica,* 69 (6), 1555–96.

[4] Mandelbrot, B.B. (1963): "The Variation of Certain Speculative Prices," *Journal of Business,* 36, 394–419.

[5] Meyn, S. P. and Tweedie, R. L. (1993): *Markov Chains and Stochastic Stability,* Springer-Verlag: London.

[6] Rachev, S.T. (ed., 2001): *Handbook of Heavy Tailed Distributions in Finance,* Elsevier, North-Holland.

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

[8] Taylor, J.C. (1997): *An Introduction to Measure and Probability,* Springer-Verlag, New York.

[9] Taylor, A.M. (2001): "Potential Pitfalls for the Purchasing Power Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price," *Econometrica,* 69 (2), 473–498.



Department of Economics, University of Melbourne, VIC 3010, Australia

*E-mail address*: `j.stachurski@econ.unimelb.edu.au`