Statistica Sinica **9**(1999), 1103-1118

GEOMETRIC ERGODICITY OF NONLINEAR TIME SERIES

Daren B. H. Cline and Huay-min H. Pu

Texas A & M University

Abstract: We identify conditions for geometric ergodicity of general, and possibly nonparametric, nonlinear autoregressive time series. We also indicate how a condition for ergodicity, with minimal side assumptions, may in fact imply geometric ergodicity. Our examples include models for which exponential stability of the associated (noiseless) dynamical system is not sufficient or not necessary, or both.

Key words and phrases: Ergodicity, Markov chain, nonlinear time series.

1. Introduction

We assume $\{X_t\}$ is a nonlinear time series with state dependent errors, defined on $\mathbb{R}^p$ by

$$X_t = \alpha(X_{t-1}) + \gamma(e_t; X_{t-1}), \qquad t \geq 1, \tag{1.1}$$

where $\{e_t\}$ is an i.i.d. sequence of random variables, independent of the initial state X_0. The process may be a $\mathbb{R}^p$-valued nonlinear autoregression of order 1, or it may be the state space vector $X_t = (\xi_t, \ldots, \xi_{t-p+1})$ for a nonlinear autoregression $\{\xi_t\}$ of order p. Either way, (1.1) ensures that $\{X_t\}$ is a time homogeneous Markov chain. We intend α and γ to be somewhat general and not necessarily parametric.

Knowing when such a time series is geometrically ergodic is very useful for analyzing the series. This is so first because it clarifies the parameter space for estimation purposes when the model is parametric and second, because it validates useful limit theorems such as the asymptotic normality of various estimators (Cf. Meyn and Tweedie (1993), Thm. 17.0.1). Sharp conditions for geometric ergodicity are known for some parametric (and near-parametric) models. For example, the stability of the self-exciting threshold autoregression (SETAR) model of order 1 has been completely characterized (Petrucelli and Woolford (1984), Chan, Petrucelli, Tong and Woolford (1985), Guo and Petrucelli (1991)). More recently, Cline and Pu (1999) characterize the stability of order 1 threshold models with coefficients depending on the signs of the most recent d values of the time series. (See also Chen and Tsay (1991) and Lim (1992) for special cases.)

It may be difficult, however, to obtain more than sufficient conditions for general (including nonparametric) models. Yet such models are increasingly being used (Chen and Tsay (1993a,b), Tjøstheim and Auestad (1994a,b)). See also the review by Härdle, Lütkepohl and Chen (1997).

In pioneering work, Chan and Tong (1985, 1994) and Chan (1990) have shown that $\{X_t\}$ is geometrically ergodic when the (noiseless) dynamical system given by

$$x_t = \alpha(x_{t-1}) \tag{1.2}$$

is exponentially stable, if $\alpha(x)$ is sufficiently smooth and $\gamma(e;x)$ is appropriately bounded. Under the stated conditions on α and γ, exponential stability of the dynamical system appears to be a sharp condition. Yet the conditions do not include many higher order or multidimensional threshold models. In addition, they do not include many nonparametric models. Furthermore, as we will show, exponential stability of the associated dynamical system is neither adequate for, nor required by, geometric stability of the time series. Interestingly, this is so even for some parametric threshold models.

Our objective, therefore, is to determine sufficient conditions for geometric ergodicity which are useful for generally defined models. In particular, we provide conditions for examples where the results of Chan and Tong either do not apply or are limited. Like other authors, we will apply the so-called Foster-Lyapounov drift conditions as developed by Foster (1953), Tweedie (1975, 1976), Nummelin (1984), Meyn and Tweedie (1992) and others, and collected together in Meyn and Tweedie (1993). We will extend their use, however, and suggest new applications. This allows us to improve the results of Chan and Tong and also to analyze new examples for which understanding the stability of (1.2) does not suffice.

In Section 2 we identify the assumptions we will be working under. Section 3 explores conditions for geometric ergodicity and provides examples. The final section contains proofs and useful lemmas.

2. Assumptions

When conditioning on the initial state we will denote probabilities and expectations as $P_x(\cdot) = \mathbf{P}(\cdot|X_0 = x)$ and $E_x(\cdot) = \mathbf{E}(\cdot|X_0 = x)$. Also we define

$$\theta(x) = \frac{\alpha(x)}{1 + ||x||},$$

where $||\cdot||$ denotes any norm on $\mathbb{R}^p$. In addition we will require $\{X_t\}$ to satisfy the following. (For the definition of a T-chain, which generalizes the strong Feller property of Markov chains, see Meyn and Tweedie (1993), p.127.)

Assumption 2.1. Assume $\{X_t\}$ is an aperiodic, ψ-irreducible T-chain satisfying (1.1). Let $r > 0$. Assume also that

(i) α is unbounded on $\mathbb{R}^p$ and θ is bounded on $\mathbb{R}^p$,
(ii) $\sup_{||x|| \le M} \mathbf{E}(||\gamma(e_1;x)||^r) < \infty$ for all $M < \infty$,
(iii) $\lim_{||x|| \to \infty} \mathbf{E}\left(\frac{||\gamma(e_1;x)||^r}{||x||^r}\right) = 0.$

There is no loss in generality in assuming α is unbounded; otherwise, it is easy to show the chain is geometrically ergodic. Models with additive errors satisfy the assumption if the errors have a finite rth moment and a density which is locally bounded away from 0. Even if the error term $\gamma(e_t; x)$ is state dependent, (1.1) defines an aperiodic, ψ-irreducible T-chain under fairly mild conditions (Cline and Pu (1998b)). Parametric examples include the SETAR models and the β-ARCH model of Guégan and Diebolt (1994).

3. Geometric Ergodicity

In the context of Markov chains, ergodic means aperiodic, ψ-irreducible and positive Harris recurrent. Geometric ergodicity refers to the rate of convergence to the invariant distribution. Processes which are ergodic but not geometrically ergodic often require a more subtle treatment than we provide here.

In some of the earliest work on general time series, Chan and Tong (1985, 1994) and Chan (1990) demonstrate that the chain (1.1) is geometrically ergodic if the dynamical system (1.2) is "exponentially stable" and $\alpha(x)$ is sufficiently smooth (e.g., Lipschitz continuous). (See also An and Huang (1996).) Tong (1990) devotes a significant part of his book to exploiting this result. We start by strengthening the results of Chan and Tong.

Let $\alpha_t : \mathbb{R}^p \to \mathbb{R}^p$ be recursively defined by $\alpha_1(x) = \alpha(x)$ and $\alpha_{t+1}(x) = \alpha(\alpha_t(x))$. The process defined by $x_t = \alpha_t(x_0)$ satisfies (1.2). Tong (1990) calls it the *skeleton* of the time series (1.1).

Theorem 3.1. *Suppose Assumption* 2.1 *holds and, in addition, suppose*

$$\lim_{\substack{||\alpha(x)||\to\infty \\ ||x-y||/||x||\to 0}} ||\theta(x) - \theta(y)|| = 0. \tag{3.1}$$

If, for some $n \geq 1$

$$\limsup_{||x||\to\infty} \frac{||\alpha_n(x)||}{||x||} < 1, \tag{3.2}$$

then $\{X_t\}$ is geometrically ergodic.

Both conditions (3.1) and (3.2) depend on $\alpha(\cdot)$ only for $||\alpha(x)||$ arbitrarily large. Condition (3.2) is essentially exponential stability of the dynamical system $x_t = \alpha(x_{t-1})1_{||\alpha(x_{t-1})||>M}$, where M is finite but arbitrary. Theorem 3.1 also generalizes the work of Chan and Tong by weakening the smoothness of α and by allowing the rth moment of $\gamma(e_1; x)$ to be $o(||x||^r)$, as $||x|| \to \infty$, rather than to be uniformly bounded.

However, even as strengthened, Theorem 3.1 presents an incomplete picture. It fails to include models in which α is not "close" to Lipschitz continuous.

In particular, even fairly simple threshold models with delay (or in multiple dimensions) cannot be handled with the theorem. More importantly, perhaps, Theorem 3.1 suggests that exponential stability of the deterministic sequence is the essential condition for geometric ergodicity (assuming appropriate behavior of the error term, as in Assumption 2.1). In some parametric models, it is, as in the threshold AR(1) models with delay (Cline and Pu (1999)). More generally, as we shall see, it is neither sufficient nor necessary.

Our objective henceforth is two-fold. First, we provide general methodologies for determining geometric ergodicity of a time series in Theorem 3.2 below. This theorem is intended to be flexible enough so that it may be used for a wide variety of nonlinear models. Theorems 3.1 and 3.3 are consequences of Theorem 3.2. The argument for Theorem 3.5 is inspired by the proof of Theorem 3.2(ii). Second, we illustrate the theorems' use for two types of models and simultaneously demonstrate that stability of a time series does not have to agree with stability of its skeleton.

Theorem 3.2. *Suppose* $\{X_t\}$ *satisfies* (1.1) *and Assumption* 2.1 *for some* $r > 0$. *The following are equivalent conditions, each sufficient for* $\{X_t\}$ *to be geometrically ergodic.*

(i) *For some* $\lambda : \mathbb{R}^p \to (0, \infty)$ *bounded and bounded away from 0,*

$$\limsup_{||\alpha(x)||\to\infty} E_x\left(\frac{\lambda(X_1)}{\lambda(x)}||\theta(X_1)||^r\right) < 1. \tag{3.3}$$

(ii) *For some* $r' > 0$ *and some* $n \geq m \geq 1$,

$$\limsup_{||\alpha(x)||\to\infty} E_x\left(\prod_{j=m}^{n}||\theta(X_j)||^{r'}\right) < 1.$$

(iii) *For some* $\delta > 0$ *and some* $n \geq m \geq 1$,

$$\limsup_{||\alpha(x)||\to\infty} E_x\left(\sum_{j=m}^{n}\log(\delta + ||\theta(X_j)||)\right) < 0.$$

Choosing λ appropriately is of course the key to applying Theorem 3.2(i). Parts (ii) and (iii) of the theorem demonstrate that optimal choices for λ may be identified, though actual computation can be difficult. The condition in part (iii) may be interpreted as a condition for ergodicity strengthened sufficiently (along with Assumption 2.1) to imply geometric ergodicity, thus generalizing Spieksma and Tweedie (1994).

The proof of Theorem 3.2 uses the well known drift condition for geometric ergodicity (cf. Meyn and Tweedie (1993), Thm. 15.0.1). Use of drift conditions involve a carefully chosen "test" function. It may be shown that using the test function $||x||^r$ leads to the following condition for geometric ergodicity:

$$\limsup_{||x||\to\infty} ||\theta(x)|| < 1. \tag{3.4}$$

Tjøstheim (1990), An and Huang (1996) and Lu (1996), among others, prove special cases of this result. Under Assumption 2.1, however, a simple argument using Fatou's Lemma shows (3.4) is at least as strong as the conditions in Theorem 3.2 and, unless $\theta(x)$ converges as $||x|| \to \infty$, Theorem 3.2 is generally preferable. This is very clearly demonstrated in Example 3.1 below.

Handpicked variations of the function $||x||^r$ have also often been used in specific cases. The hope is that our results illustrate some generality and coherency to identifying appropriate test functions. In particular we have found that using $||\alpha(x)||^r$ in combination with $||x||^r$, first suggested by Pu (1995), can lead to superior (that is, weaker) conditions. The choice of test function also has statistical consequences because it regulates the order of sample moments for which the Law of Large Numbers and the Central Limit Theorem apply (Meyn and Tweedie (1993), Thm. 17.0.1).

Tjøstheim (1990) notes that geometric ergodicity of the n-step process $\{X_{nt}\}$ is both necessary and sufficient for geometric ergodicity of $\{X_t\}$ (based on results from Nummelin (1984)). Applying drift conditions to the n-step process results in special cases of the condition in Theorem 3.2(ii).

For threshold models of order 1 the function $\theta(x)$ is essentially piecewise constant. This includes SETAR models of order 1. Sometimes $\theta(x)$ is only asymptotic to a piecewise constant function as in amplitude dependent exponential autoregressive (EXPAR) models (see Jones (1978), Ozaki and Oda (1978)). In such cases an explicit, optimal choice for λ is often possible, giving rise to sharp conditions. See, for example, Cline and Pu (1999).

Optimal choices for $\lambda(x)$ can be very difficult for more general models, which is why conditions (ii) and (iii) in Theorem 3.2 can be useful. For example, when θ is essentially periodic and sufficiently smooth we are able to apply Theorem 3.2(iii) quite explicitly and obtain sharp conditions for geometric ergodicity. To understand this interesting result, suppose X_t is real valued and satisfies

$$X_t = \beta_0(X_{t-1}) + \beta_1(X_{t-1})X_{t-1} + e_t, \tag{3.5}$$

where β_1 is periodic. (The sunspot example analyzed by Chen and Tsay (1993a) hints at the possibility of such periodicity.) Note that $E_x(\log(|\beta_1(X_1)|)) = \mathbf{E}(\log(|\beta_1(\alpha(x) + e_1)|))$ is a periodic function of $\alpha(x)$. It follows that any condition based on this (such as that in Theorem 3.2(iii) with $m = n = 1$) will not be very sharp. The next simplest condition (Theorem 3.2(iii) with $m = n = 2$) is essentially based on an asymptotic bound for $E_x(\log(|\beta_1(X_2)|))$. Surprisingly, this converges as $|\alpha(x)| \to \infty$ and therefore provides a sharp condition. This is shown next.

Assumption 3.3. Let $\beta : \mathbb{R} \to \mathbb{R}$ be periodic with period τ. For each $\epsilon > 0$, there is an open set $A_\epsilon \subset [0, \tau]$ having Lebesgue measure at least $\tau - \epsilon$ and such that β is continuously differentiable on A_ϵ and its derivative is bounded away from 0 on A_ϵ.

Theorem 3.4. *Suppose X_t is real valued and satisfies* (3.5). *Assume β_0 and β_1 are bounded, $\beta = \beta_1$ satisfies Assumption* 3.3, *e_t has density f on $\mathbb{R}$ which is locally bounded away from* 0 *and locally Riemann integrable, and $\mathbf{E}(|e_t|^r) < \infty$ for some $r > 0$. If*

$$\frac{1}{\tau}\int_0^\tau \log(|\beta_1(u)|)du < 0 \tag{3.6}$$

then $\{X_t\}$ is geometrically ergodic.

Example 3.1. Suppose $\beta_1(x) = a + b\cos(x)$. Using Gradshteyn and Ryzhik (1980, eq. 4.226.1, p.528) one may show

$$\frac{1}{2\pi}\int_0^{2\pi} \log(|a + b\cos(u)|)du = \begin{cases} \log(\frac{|b|}{2}), & \text{if } |a| \le |b|, \\ \log(\frac{|a|+\sqrt{a^2-b^2}}{2}), & \text{if } |a| > |b|. \end{cases}$$

Thus by Theorem 3.4, a sufficient condition for geometric ergodicity is $|b| < 2$ if $|a| \le |b|$ and $|a| + \sqrt{a^2 - b^2} < 2$ if $|a| > |b|$. In a separate paper on transience of time series (Cline and Pu (1998a)) we show that this condition is sharp.

Note that with $|a| + |b| > 1$ this model does not satisfy the assumptions of Chan and Tong (1985). This may be seen as follows. It suffices to let $\alpha(x) = \beta_1(x)x$. Suppose $a = k_1/m$ and $b = k_2/m > 1 - a$ where k_1, k_2 and m are non-negative integers. Then for all integers j and n, $|\alpha_n(2jm^n\pi)| = (a+b)^n|2jm^n\pi|$. Since α_n is continuous, it follows that there exists an unbounded set A_n with positive Lebesgue measure such that

$$\inf_{x\in A_n} \frac{|\alpha_n(x)|}{|x|} > 1 \quad \text{for each } n \ge 1.$$

Therefore, the skeleton process is not exponentially stable and (3.2) fails to hold as well. With slight adjustments in the argument, this is also correct for arbitrary a and b.

Having shown that exponential stability of the skeleton is not necessary for geometric ergodicity of the time series, one might naturally wonder if it is sufficient. It is not, as we now proceed to show. To motivate the idea, consider a bivariate threshold model where the threshold depends only on the first coordinate. That is, suppose $e_t = (e_{t,1}, e_{t,2})$ and

$$X_t = \begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{cases} \begin{pmatrix} \phi_1 X_{t-1,1} + e_{t,1} \\ \Phi_1 X_{t-1,2} + e_{t,2} \end{pmatrix}, & \text{if } X_{t-1,1} < 0, \\ \begin{pmatrix} \phi_2 X_{t-1,1} + e_{t,1} \\ \Phi_2 X_{t-1,2} + e_{t,2} \end{pmatrix}, & \text{if } X_{t-1,1} > 0. \end{cases} \tag{3.7}$$

The sub-process $\{X_{t,1}\}$ is itself an ordinary threshold model of order 1 and it drives the nonlinearity of $\{X_{t,2}\}$. Conditions for the entire process to be stable are going to depend not only on whether $\{X_{t,1}\}$ is stable but also on what its invariant distribution is, since the second component gets multiplied by Φ_1 or Φ_2 according to how often the first component is negative or positive.

We consider somewhat more general models for our next result and return to the example at (3.7) afterward. For the following, write $x = (x_1, x_2)$ where $x_i \in \mathbb{R}^{p_i}$ and $p_1 + p_2 = p$. Also, let $||x|| = ||x_1||_1 + ||x_2||_2$ where $||\cdot||_i$ is a norm on $\mathbb{R}^{p_i}$, $i = 1, 2$. For each norm $||\cdot||_i$ we have a corresponding matrix norm, also denoted $||\cdot||_i$, and defined by

$$||A||_i = \sup_{x_i \neq 0} \frac{||Ax_i||_i}{||x_i||_i}.$$

We are interested in the Markov chain

$$X_t = \begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} A_1(X_{t-1,1})X_{t-1,1} + \gamma_1(e_t; X_{t-1,1}) \\ A_2(X_{t-1,1})X_{t-1,2} + \gamma_2(e_t; X_{t-1}) \end{pmatrix}, \tag{3.8}$$

where A_1 and A_2 are bounded matrix valued functions depending only on x_1. Note that $\{X_{t,1}\}$ is also a Markov chain. Assuming $\{X_{t,1}\}$ is geometrically ergodic, we ask what additional assumptions imply the stability of $\{X_t\}$.

Theorem 3.5. *Suppose $\{X_t\}$ satisfies* (3.8) *and Assumption* 2.1. *Assume also there exists locally bounded V_1 $\mathbb{R}^{p_1} \to [0, \infty)$, such that $V_1(x_1) \to \infty$ as $||x_1||_1 \to \infty$,*

$$\limsup_{||x_1||_1 \to \infty} E_x \left(\frac{V_1(X_{1,1})}{V_1(x_1)} \right) < 1, \tag{3.9}$$

and

$$\sup_{||x_1||_1 \leq M} E_x(V_1(X_{1,1})) < \infty \qquad \textit{for all } M < \infty, \tag{3.10}$$

thus verifying $\{X_{t,1}\}$ is geometrically ergodic with invariant distribution, say G_1. If

$$\int \log(||A_2(x_1)||_2) G_1(dx_1) < 0, \tag{3.11}$$

$\{X_t\}$ is geometrically ergodic.

Example 3.2. Let $p_1 = p_2 = 1$ and suppose $\{X_t\}$ is given by (3.7). Then (3.8) is satisfied with

$$A_1(x_1) = \phi_1 1_{x_1<0} + \phi_2 1_{x_1>0} \qquad \text{and} \qquad A_2(x_1) = \Phi_1 1_{x_1<0} + \Phi_2 1_{x_1>0}.$$

Note that this defines a threshold model on $\mathbb{R}^2$ with two matrices, $\mathrm{diag}(\phi_1, \Phi_1)$ and $\mathrm{diag}(\phi_2, \Phi_2)$, as "coefficients", depending only on the first component of the series. Thus $\{X_{t,1}\}$ is the univariate TAR(1) model which is well known to be geometrically ergodic precisely when

$$\max(\phi_1, \phi_2, \phi_1\phi_2) < 1. \tag{3.12}$$

Conditions (3.9) and (3.10) hold with $V_1(x_1) = |\lambda(x_1)x_1|^r$ and a two-valued function λ. By Theorem 3.5 a sufficient condition for geometric ergodicity is (3.12) and

$$|\Phi_1|^{G_1(0)}|\Phi_2|^{1-G_1(0)} < 1, \tag{3.13}$$

where G_1 is the invariant distribution for $\{X_{t,1}\}$. This condition is shown to be sharp in Cline and Pu (1998a).

Exponential stability of the deterministic sequence $\{\alpha_n(x)\}$ is, however, different. By considering how the second component of the skeleton process behaves for fixed ϕ_1 and ϕ_2, it is easily seen that $\{\alpha_n(x)\}$ is exponentially stable if and only if (3.12) holds and

$$\max(|\Phi_1| 1_{\phi_1>0}, |\Phi_2| 1_{\phi_2>0}, |\Phi_1\Phi_2| 1_{\phi_1\phi_2>0}) < 1. \tag{3.14}$$

To see that (3.13) and (3.14) differ, suppose $e_{t,1}$ has distribution F_1 such that $0 < F_1(u) < 1$ for all $u \in \mathbb{R}$. Then it must be that $0 < G_1(0) < 1$ and $G_1(0)$ depends only on ϕ_1, ϕ_2 and F_1. If, for example, $\phi_1 < 0 < \phi_2 < 1$ then (3.12) holds and $|\Phi_1| < |\Phi_2|^{1-1/G_1(0)} \leq 1$ implies (3.13) holds but not (3.14), while $|\Phi_1| \geq |\Phi_2|^{1-1/G_1(0)} > 1$ implies (3.14) holds but not (3.13). Thus, geometric ergodicity of $\{X_t\}$ and exponential stability of its skeleton are different and neither implies the other.

For a discussion on how one might actually evaluate $G_1(0)$, see Jones (1978) or Tong (1990, Ch. 4).

The examples we have presented are first order time series models. Sharp sufficient conditions for higher order models is still very much an open question. Known results include those in Chan and Tong (1986), Brockwell, Liu and Tweedie (1992), Chen and Tsay (1993a), An and Huang (1996) and Cline and Pu (1999).

4. Proofs

Proof of Theorem 3.1. This is a corollary to Theorem 3.2. The argument is completely analagous to that of Theorem 2.3 in Cline and Pu (1999), with $s(x) = \min(||\alpha(x)||, ||x||)$ and $h_M(x) = ||\theta(x)||1_{||s(x)||>M}$. It helps to note that $||\alpha_j(x)|| \to \infty$ implies $||x|| \to \infty$ and $||\alpha_i(x)|| \to \infty$ for each $i \leq j$.

Proof of Theorem 3.2. (i) The proof involves demonstrating the drift condition of Meyn and Tweedie (1993, Thm. 15.0.1), with test function $V(x) = \lambda(x)||\alpha(x)||^r + \epsilon||x||^r$. Given $X_0 = x$, we have

$$\begin{aligned} V(X_1) &= \lambda(X_1)||\theta(X_1)||^r(1 + ||\alpha(x) + \gamma(e_1; x)||)^r + \epsilon||\alpha(x) + \gamma(e_1; x)||^r \\ &\leq (\lambda(X_1)||\theta(X_1)||^r + \epsilon)(1 + ||\alpha(x)|| + ||\gamma(e_1; x)||)^r. \end{aligned}$$

It is easily seen that, under Assumption 2.1,

$$\limsup_{\substack{||x||\to\infty \\ ||\alpha(x)||\leq M}} E_x\left(\frac{V(X_1)}{V(x)}\right) = 0$$

for any $M < \infty$. Furthermore, Assumption 2.1 and (3.3) imply

$$\limsup_{\substack{||x||\to\infty \\ ||\alpha(x)||>M}} E_x\left(\frac{V(X_1)}{V(x)}\right) \leq \limsup_{\substack{||x||\to\infty \\ ||\alpha(x)||>M}} E_x\left(\frac{\lambda(X_1)||\theta(X_1)||^r + \epsilon}{\lambda(x)}\right) < 1,$$

if M is chosen large enough and ϵ is chosen small enough. Hence, for some $M_1 < \infty$,

$$\sup_{||x||>M_1} E_x\left(\frac{V(X_1)}{V(x)}\right) < 1.$$

Also, by Assumption 2.1, $\sup_{||x||\leq M_1} E_x(V(X_1)) < \infty$, and since compact sets are petite (cf. Meyn and Tweedie (1993), Thm. 6.2.5), geometric ergodicity holds by the drift condition of Meyn and Tweedie (1993, Thm. 15.0.1).

The equivalency of (ii) to (i) is a corollary of the following more general result, with $b(x, y) = ||\theta(x)||$ and $s(x) = ||\alpha(x)||$, and part (iii) is implicit in the proof.

Lemma 4.1. *Let $\{X_t\}$ be a Markov chain on $\mathbb{X}$ and assume $b : \mathbb{X}^2 \to \mathbb{R}_+$ such that $\sup_x E_x(b(X_1, x)) < \infty$ and $s : \mathbb{X} \to \mathbb{R}_+$ is unbounded.*

i) *If there exists $r > 0$, $m \geq 1$ and $n \geq m$ such that*

$$\limsup_{s(x)\to\infty} E_x\left(\prod_{j=m}^{n} b^r(X_j, X_{j-1})\right) < 1,$$

then there exists $r' > 0$ and $\lambda : \mathbb{X} \to \mathbb{R}_+$, bounded and bounded away from zero, such that

$$\limsup_{s(x)\to\infty} E_x\left(\frac{\lambda(X_1)}{\lambda(x)} b^{r'}(X_1, x)\right) < 1. \tag{4.1}$$

ii) *The converse is true if $s(x) \to \infty$ implies $\min(1/s(X_1), b(X_1, x)) \to 0$ in probability.*

Proof.

(i) We may assume $r \le 1$. By induction, it is easy to see that

$$\sup_x E_x\left(\prod_{j=m}^{n} (\delta + b(X_j, X_{j-1}))\right) < \infty$$

for any $\delta > 0$, and thus there exists $\delta > 0$ such that

$$\limsup_{s(x)\to\infty} E_x\left(\prod_{j=m}^{n} (\delta + b(X_j, X_{j-1}))^r\right) < 1.$$

It follows that

$$\limsup_{s(x)\to\infty} E_x\left(\sum_{j=m}^{n} \log(\delta + b(X_j, X_{j-1}))\right) < 0. \tag{4.2}$$

Let

$$\nu(x) = \sum_{j=1}^{n-1} \min(1 - \tfrac{j+1-m}{n+1-m}, 1) E_x(\log(\delta + b(X_j, X_{j-1}))).$$

Hence ν is bounded and, by (4.2),

$$\limsup_{s(x)\to\infty} E_x(\nu(X_1) + \log(\delta + b(X_1, x)) - \nu(x)) < 0.$$

By Lemma 4.2 below, there exists $r' > 0$ such that

$$\limsup_{s(x)\to\infty} E_x\left(\left(\frac{e^{\nu(X_1)}}{e^{\nu(x)}}(\delta + b(X_1, x))\right)^{r'}\right) < 1.$$

From this we see it suffices to take $\lambda(x) = e^{r'\nu(x)}$.

(ii) Let $r = r' < 1$ and let $c < 1$ be the limit in the lefthand side of (4.1). Define

$$B_n(x) = E_x\left(\frac{\lambda(X_n)}{\lambda(x)} \prod_{j=1}^{n} b^r(X_j, X_{j-1})\right)$$

and note that $B_n(x) = E_x\left(\frac{\lambda(X_1)}{\lambda(x)} b^r(X_1, x) B_{n-1}(X_1)\right)$. Choose $L < \infty$ so that $1/L < \lambda(x) < L$ and $E_x(b^r(X_1, x)) < L$ for all x. Note that $B_n(x) < L^{3n}$ for all x and n. Now suppose

$$\limsup_{s(x)\to\infty} B_j(x) \le c^j, \qquad j = 1, \dots, n-1,$$

for some $n \ge 2$. Given $\epsilon > 0$, fix $M < \infty$ so that $\sup_{s(x)>M} B_j(x) \le (c+\epsilon)^j$ for $j = 1, \dots, n-1$. Since by assumption,

$$\limsup_{s(x)\to\infty} P_x(s(X_1) \le M,\ b(X_1, x) > \epsilon) = 0,$$

we have

$$\begin{aligned}\limsup_{s(x)\to\infty} B_n(x) &= \limsup_{s(x)\to\infty} E_x\left(\frac{\lambda(X_1)}{\lambda(x)} b^r(X_1, x) B_{n-1}(X_1)(1_{s(X_1)>M} + 1_{b(X_1,x)\le\epsilon})\right)\\ &\le \epsilon^r L^{3n-1} + (c+\epsilon)^n.\end{aligned}$$

By induction we conclude that

$$\limsup_{s(x)\to\infty} E_x\left(\frac{\lambda(X_n)}{\lambda(x)} \prod_{j=1}^{n} b^r(X_j, X_{j-1})\right) \le c^n \qquad \text{for all } n \ge 1.$$

Since λ is bounded and bounded away from zero,

$$\limsup_{s(x)\to\infty} E_x\left(\prod_{j=1}^{n} b^r(X_j, X_{j-1})\right) \le c^n L^2 < 1$$

for all n large enough.

Lemma 4.2. *Suppose $\{Y_x\}$ is a collection of nonnegative random variables indexed by x. Then $\sup_x \mathbf{E}(Y_x^r) < 1$ for some $r > 0$ if and only if*

$$\sup_x \mathbf{E}(Y_x^r) < \infty \textit{ for some } r > 0, \quad \textit{and} \quad \sup_x \mathbf{E}(\log(\delta + Y_x)) < 0 \textit{ for some } \delta > 0.$$

If $\log(Y_x) \ge T$ for each x, where $\mathbf{E}(|T|) < \infty$, it suffices to let $\delta = 0$.

Proof. This follows by standard arguments.

For what follows $\lfloor x \rfloor$ is the largest integer not greater than x and $\mathrm{frac}(x)$ is that portion of x which exceeds $\lfloor x \rfloor$.

Lemma 4.3. *Suppose β is bounded, periodic with period $\tau = 1$, and satisfies Assumption* 3.3. *Suppose Y has a density g on $\mathbb{R}$ which is locally Riemann integrable. Then for any $y \in \mathbb{R}$,*

$$\mathrm{frac}((n+Y)\beta(n+Y) + y) \Longrightarrow \textit{Uniform}(0,1) \qquad \textit{as } |n| \to \infty.$$

Proof. The idea behind this proof is that $x\beta(x)$ spreads its values so thinly for large x, that the fractional part of $(x+Y)\beta(x+Y)$ can be anything essentially with uniform probability.

Given $\epsilon > 0$ let A_ϵ be as in Assumption 3.3. We may assume without loss that A_ϵ is a finite union of disjoint open intervals, say $I_1, \ldots, I_M$. Now fix j and $k \in \mathbb{Z}$ and define

$$h(u) = g(k+u)1_{I_j}(u), \qquad u \in [0,1].$$

Define β_j on $[0,1]$ in such a way that β_j agrees with β on I_j, is linear off I_j and is continously differentiable. It is thus strictly monotone and β_j' is bounded away from 0. Let $b_* = \max_{[0,1]} \beta_j(u)/\beta_j'(u)$.

Now fix $0 < a < b < 1$. For $n > |k|$ and integer m, let

$$J_{n,m} = \{u : 0 \le u \le 1, \tfrac{m+a}{n+k+u} \le \beta_j(u) \le \tfrac{m+b}{n+k+u}\},$$

$$h_{n,m} = \min_{u \in J_{n,m}} h(u), \qquad b_{n,m} = \max_{u \in J_{n,m}} |\beta'(u)|.$$

Note that for large n and all m, $J_{n,m}$ is an interval with length at least $\frac{b-a}{(n+k+1+b_*)b_{n,m}}$. Let $W_n = \mathrm{frac}((n+Y)\beta(n+Y)) = \mathrm{frac}((n+Y)\beta(Y))$ and compute

$$\begin{aligned}
&\liminf_{n\to\infty} P(W_n \in (a,b), \mathrm{frac}(Y) \in I_j, \lfloor Y \rfloor = k) \\
&\qquad = \liminf_{n\to\infty} \sum_m \int_{I_j} 1_{\left(\frac{m+a}{n+k+u}, \frac{m+b}{n+k+u}\right)}(\beta(u)) g(k+u)du \\
&\qquad = \liminf_{n\to\infty} \sum_m \int_{J_{n,m}} h(u)du \\
&\qquad \ge \liminf_{n\to\infty} \sum_m h_{n,m} \frac{b-a}{(n+k+1+b_*)b_{n,m}} \\
&\qquad = (b-a)\int_0^1 h(u)du \\
&\qquad = (b-a)P(\mathrm{frac}(Y) \in I_j, \lfloor Y \rfloor = k).
\end{aligned}$$

The penultimate equality follows from the fact that h is Riemann integrable and β' is continuous on I_j and bounded away from 0. By Fatou's Lemma and then monotone convergence,

$$\liminf_{n\to\infty} P(W_n \in (a,b)) \ge b-a \qquad \text{for all } 0 < a < b < 1,$$

from which it follows that $W_n \Longrightarrow U \sim$Uniform(0,1) as $n \to \infty$.

Applying the same to $\beta(-x)$ and noting that $1-U \overset{\mathrm{d}}{=} U$, the above holds as $n \to -\infty$. Furthermore, since the limit is uniform, we also have $\mathrm{frac}((n+Y)\beta(n+Y)+y) \Longrightarrow$ Uniform(0,1), as $|n| \to \infty$, for any $y \in \mathbb{R}$.

Proof of Theorem 3.4. The proof applies Lemma 4.3 to show that $\mathrm{frac}(X_2/\tau)$ is effectively uniform in distribution when $\alpha(x)$ is large. Assumption 2.1 is met since the errors are additive, have finite rth moment and have density locally bounded away from 0. (See Chan (1993) or Cline and Pu (1998b).) Let $\alpha(x) = \beta_1(x)x$ and $\gamma(e_1; x) = \beta_0(x) + e_1$.

Fix $w \in [0, \tau)$ and apply Lemma 4.3 with $\beta(x) = \beta_1(\tau x)$ and $Y = (w+e_1)/\tau$ to get

$$\mathrm{frac}(\tfrac{n\tau+w+e_1}{\tau}\beta_1(n\tau + w + e_1) + y) \Longrightarrow U{\sim}\mathrm{Uniform(0,1)} \qquad \text{as } |n| \to \infty.$$

Since the limit is independent of w and y,

$$\mathrm{frac}(\tfrac{\alpha(x)+\beta_0(x)+e_1}{\tau}\beta_1(\alpha(x) + \beta_0(x) + e_1) + e_2) \Longrightarrow U \qquad \text{as } |\alpha(x)| \to \infty.$$

Given $X_0 = x$, we have $X_2 = (\alpha(x) + \beta_0(x) + e_1)\beta_1(\alpha(x) + \beta_0(x) + e_1) + e_2$. Hence by bounded convergence, for each $\delta > 0$,

$$\limsup_{|\alpha(x)|\to\infty} E_x(\log(\delta + |\beta_1(X_2)|)) = \frac{1}{\tau}\int_0^\tau \log(\delta + |\beta_1(u)|)du.$$

Then by (3.6) and monotone convergence,

$$\limsup_{|\alpha(x)|\to\infty} E_x(\log(\delta + |\beta_1(X_2)|)) < 0$$

for some $\delta > 0$ small enough. By Theorem 3.2(iii), the process is geometrically ergodic.

Proof of Theorem 3.5. Throughout the proof, $E_{x_1}(\cdot)$ will refer to expectations for the process $\{X_{t,1}\}$ conditioned on $X_{0,1} = x_1$. Likewise, $E_{X_{1,1}}(\cdot)$ will refer to expectations conditioned on $X_{1,1}$. The proof consists of finding a somewhat complicated test function to verify the drift condition for geometric ergodicity.

There is no loss in assuming $V_1(x_1) \ge 1$ for all X_1; if not, replace V_1 with $1+V_1$. That $\{X_{t,1}\}$ is geometrically ergodic is immediate by the drift conditions (3.9)–(3.10) and the fact that compact sets are petite (Meyn and Tweedie (1993), Thm. 15.0.1 and Thm. 6.2.5). Furthermore,

$$K_1 = \int V_1(x_1)G_1(dx_1) < \infty.$$

By (3.11) and the invariance of G_1, there exists $\delta > 0$ and $r > 0$ such that the conditions in Assumption 2.1(ii,iii) hold and

$$\rho_1 = \int E_{x_1}\left((\delta + ||A_2(X_{1,1})||_2))^r\right) G_1(dx_1) = \int (\delta + ||A_2(x_1)||_2)^r\, G_1(dx_1) < 1.$$

Choose $\epsilon \in (0,1)$ so that $\rho_1^{1-\epsilon} K_1^\epsilon < 1$ and let $s = (1-\epsilon)r$. By Hölder's inequality,

$$\rho_2 = \int E_{x_1}\left((\delta + ||A_2(X_{1,1})||_2))^s\right) V_1^\epsilon(x_1) G_1(dx_1) \le \rho_1^{1-\epsilon} K_1^\epsilon < 1.$$

Note that $E_{x_1}\left((\delta + ||A_2(X_{1,1})||_2))^s\right)$ is bounded. Define

$$h(x_1) = E_{x_1}\left((\delta + ||A_2(X_{1,1})||_2))^s\right) V_1^\epsilon(x_1).$$

Since $V_1^\epsilon(x_1)$ also satisfies the drift condition for geometric ergodicity, there are $K_2 < \infty$ and $q < 1$ such that

$$E_{x_1}(h(X_{n,1})) \le \rho_2 + K_2 q^n V_1^\epsilon(x_1) \qquad \text{for all } n \ge 1,\ x_1 \in \mathbb{R}^{p_1}, \tag{4.3}$$

by Meyn and Tweedie (1993, Thm. 15.0.1).

Let $b_1 = \sup_{x_1} ||A_1(x_1)||_1$ and $b_2 = \sup_{x_1} ||A_2(x_1)||_2$. Choose $M_1 < \infty$ and $n \ge 1$ so that $\rho_2 + K_2 q^n < 1$ and $(\rho_2 M_1^{-\epsilon} + K_2 q^n)(1 + b_1/\delta)^s < 1$. By (4.3), we have

$$\sup_{V_1(x_1) \le M_1} \frac{E_{x_1}(h(X_{n,1}))}{V_1^\epsilon(x_1)} < 1 \quad \text{and} \quad \sup_{V_1(x_1) > M_1} \frac{E_{x_1}(h(X_{n,1}))}{V_1^\epsilon(x_1)} < \left(1 + \frac{b_1}{\delta}\right)^{-s}. \tag{4.4}$$

By some careful algebra,

$$\begin{aligned} R_1(x, X_1) &\stackrel{\text{def}}{=} \frac{\delta + ||A_2(X_{1,1}) X_{1,2}||_2 + \delta ||X_1||}{(\delta + ||A_2(x_1) x_2||_2 + \delta ||x||)(\delta + ||A_2(X_{1,1})||_2)} \\ &\le 1 + \frac{1 + b_1 ||x_1||_1 + (1 + b_2/\delta) ||\gamma(e_1; x)||}{\delta + \delta ||x||}. \end{aligned}$$

From this, Assumption 2.1, and the fact that $V_1(x_1) \to \infty$ as $||x_1||_1 \to \infty$,

$$\limsup_{\substack{||x|| \to \infty \\ V_1(x_1) \le M_1}} E_x(R_1^s(x, X_1)) \le 1 \quad \text{and} \quad \limsup_{\substack{||x|| \to \infty \\ V_1(x_1) > M_1}} E_x(R_1^s(x, X_1)) \le (1 + \frac{b_1}{\delta})^s. \tag{4.5}$$

Hence, by (4.4) and (4.5),

$$\limsup_{||x|| \to \infty} \frac{E_{x_1}(h(X_{n,1}))}{V_1^\epsilon(x_1)} E_x(R_1^s(x, X_1)) < 1. \tag{4.6}$$

Now let

$$R_2(x, X_1) = \frac{(\delta + ||A_2(X_{1,1})||_2)^s}{E_{x_1}((\delta + ||A_2(X_{1,1})||_2)^s)}$$

and

$$V(x) = \left((\delta + ||A_2(x_1) x_2||_2 + \delta ||x||)^s \prod_{j=0}^{n-1} E_{x_1}(h(X_{j,1})) \right)^{\frac{1}{n+2}}.$$

Then, by Hölder's inequality, the fact that $\{X_{t,1}\}$ is Markov, and (4.6),

$$\begin{aligned}
&\limsup_{||x||\to\infty} E_x\left(\frac{V(X_1)}{V(x)}\right)\\
&=\limsup_{||x||\to\infty} E_x\left(\left(R_1^s(x,X_1)R_2(x,X_1)\frac{E_{X_{1,1}}(h(X_{n,1}))}{V_1^\epsilon(x_1)}\prod_{j=1}^{n-1}\frac{E_{X_{1,1}}(h(X_{j,1}))}{E_{x_1}(h(X_{j,1}))}\right)^{\frac{1}{n+2}}\right)\\
&\le \limsup_{||x||\to\infty} \left(E_x(R_1^s(x,X_1))\right)^{\frac{1}{n+2}}\left(\frac{E_{x_1}(h(X_{n,1}))}{V_1^\epsilon(x_1)}\right)^{\frac{1}{n+2}} < 1.
\end{aligned}$$

In a similar manner it is easy to see, by (3.10), (4.3), and the fact that V_1 is locally bounded, that $\sup_{||x||\le M} E_x(V(X_1)) < \infty$ for all $M < \infty$. Thus, by Meyn and Tweedie (1993, Thm. 15.0.1), the process is geometrically ergodic.

References

An, H. Z. and Huang, F. C. (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statist. Sinica* **6**, 943-956.

Brockwell, P. J., Liu, J. and Tweedie, R. L. (1992). On the existence of stationary threshold autoregressive moving-average processes. *J. Time Series Anal.* **13**, 95-107.

Chan, K. S. (1990). Deterministic stability, stochastic stability, and ergodicity. Appendix 1 in *Non-linear Time Series Analysis: A Dynamical System Approach* (By H. Tong). Oxford University Press, London.

Chan, K. S. (1993). A review of some limit theorems of Markov chains and their applications. *Dimensions, Estimation and Models* (Edited by H. Tong), 108-135. World Scientific, Singapore.

Chan, K. S., Petrucelli, J. D., Tong, H. and Woolford, S. W. (1985). A multiple threshold AR(1) model. *J. Appl. Probab.* **22**, 267-279.

Chan, K. S. and Tong, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Probab.* **17**, 666-678.

Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *J. Time Ser. Anal.* **7**, 179-190.

Chan, K. S. and Tong, H. (1994). A note on noisy chaos. *J. Roy. Statist. Soc.* **56**, 301-311.

Chen, R. and Tsay, R. S. (1991). On the ergodicity of TAR(1) process. *Ann. Appl. Probab.* **1**, 613-634.

Chen, R. and Tsay, R. S. (1993a). Functional-coefficient autogressive models. *J. Amer. Statist. Assoc.* **88**, 298-308.

Chen, R. and Tsay, R. S. (1993b). Nonlinear additive ARX models. *J. Amer. Statist. Assoc.* **88**, 955-967.

Cline, D. B. H. and Pu, H. H. (1998a). Geometric transience of nonlinear time series. Technical Report 289, Dept. of Statistics, Texas A&M University, College Station, TX.

Cline, D. B. H. and Pu, H. H. (1998b). Verifying irreducibility and continuity of a nonlinear time series. *Statist. Probab. Lett.* **40**, 139-148.

Cline, D. B. H. and Pu, H. H. (1999). Stability of nonlinear AR(1) time series with delay. *Stoch. Proc. Appl.* **82**, 307-333.

Foster, F. G. (1953). On the stochastic matrices associated with certain queueing processes. *Ann. Math. Stat.* **24**, 355-360.

Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Table of Integrals, Series and Products*. Academic Press, Orlando, FL.

Guégan, D. and Diebolt, J. (1994). Probabilistic properties of the β-ARCH model. *Statist. Sinica* **4**, 71-87.

Guo, M. and Petrucelli, J. D. (1991). On the null-recurrence and transience of a first order SETAR model. *J. Appl. Probab.* **28**, 584-592.

Härdle, W., Lütkepohl, H. and Chen, R. (1997). A review of nonparametric time series analysis. *Intern'l. Stat. Review* **65**, 49-72.

Jones, D. A. (1978). Nonlinear autoregressive processes. *Proc. Roy. Soc.* **A360**, 71-95.

Lim, K. S. (1992). On the stability of a threshold AR(1) without intercepts. *J. Time Ser. Anal.* **13**, 119-132.

Lu, Z. (1996). Geometric ergodicity of a vector non-linear AR model mixed with an ARCH term. Preprint, Inst. Stat., Univ. Catholique de Louvain, Leuven, Belgium.

Meyn, S. P. and Tweedie, R. L. (1992). Stability of Markovian processes I: criteria for discrete-time chains. *Adv. Appl. Probab.* **24**, 542-574.

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

Nummelin, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press, Cambridge.

Ozaki, T. and Oda, H. (1978). Non-linear time series model identification by Akaike's information criterion. *Proc. IFAC Workshop on Information and Systems, Compeign, France, October 1977.*

Petrucelli, J. D. and Woolford, S. W. (1984). A threshold AR(1) model. *J. Appl. Probab.* **21**, 270-286.

Pu, H. H. (1995). Recurrence and transience for autoregressive nonlinear time series. Dissertation, Texas A&M University, College Station, TX.

Spieksma, F. M. and Tweedie, R. L. (1994). Strengthening ergodicity to geometric ergodicity for Markov chains. *Stochastic Models* **10**, 45-74.

Tjøstheim, D. (1990). Non-linear time series and Markov chains. *Adv. Appl. Probab.* **22**, 587-611.

Tjøstheim, D. and Auestad, B. H. (1994a). Nonparametric identification of nonlinear time series: projections. *J. Amer. Stat. Assoc.* **89**, 1398-1409.

Tjøstheim, D. and Auestad, B. H. (1994b). Nonparametric identification of nonlinear time series: selecting significant lags. *J. Amer. Stat. Assoc.* **89**, 1410-1419.

Tong, H. (1990). *Non-linear Time Series Analysis: A Dynamical System Approach*. Oxford University Press, London.

Tweedie, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stoch. Proc. Appl.* **3**, 385-403.

Tweedie, R. L. (1976). Criteria for classifying general Markov chains. *Adv. Appl. Probab.* **24**, 542-574.

Department of Statistics, Texas A&M University, College Station TX 77843-3143, U.S.A.

E-mail: dcline@stat.tamu.edu

Department of Statistics, Texas A&M University, College Station TX 77843-3143, U.S.A.

E-mail: pu@stat.tamu.edu

(Received October 1997; accepted December 1998)