Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

STABILITY OF NONLINEAR TIME SERIES: WHAT DOES NOISE HAVE TO DO WITH IT?

Daren B.H. Cline and Huay-min H. Pu
Department of Statistics
Texas A&M University

Abstract

We survey results on the stability of various nonlinear time series, both parametric and nonparametric. The emphasis will be on identifying the role that the "error term" has in determining stability. The error term can indeed affect stability, even when additive and for simple, common parametric models. The stability of the time series is not necessarily the same as that of its related (noiseless) dynamical system. In particular, this means that care must be taken to ensure that estimates are actually within the valid parameter space when analyzing a nonlinear time series.

Key Words: ergodicity, Markov chain, nonlinear time series

1 Introduction

Fitting time series with nonlinear models has become increasingly popular, especially since the emergence of nonparametric function estimation methods (Collomb and Härdle (1986), Härdle and Vieu (1992), Chen and Tsay (1993a,b), Tjøstheim and Auestad (1994a,b), Masry and Tjøstheim (1995)). No matter what model is fit, a critical part of the estimation procedure is determining whether the model is stable or whether the parameters are within the appropriate parameter space (Tjøstheim (1994)). Additionally, knowing the stability properties of a particular model makes it possible to develop simulation and resampling procedures to be used for inference.

For these procedures, as well as for the obvious questions of limit theorems and robustness, the nature of the noise (error) distribution is clearly a significant concern. What is not so clear, however, is how this distribution can affect — if it does at all — the stability question itself. Habit with linear models has made it seem as though the error distribution is essentially irrelevant for stability or, as in the case of bilinear and ARCH models, the magnitude of the error variance can appear to be all that is relevant. On the contrary, even when additive, noise often plays a large and critical role in determining the stability of many nonlinear time series models. In particular, the assumption that the process is stochastic is fundamental.

In a sense, the issue is between taking a stochastic view of the reasons underlying a process and taking a deterministic view. In the stochastic view, noise is not just a nuisance of observation and inference, a proxy for the uncertainty of scientific investigation. Rather it is integral to the behavior of the process. Of course this is no surprise to those who study stochastic processes: the effect of noise on the values of the time series persists due to dependence. This persistent effect, in turn, affects the stability of the process itself.

Determinism as a tenet of science is ingrained into us at our earliest experiences with the scientific method. Linear models, it turns out, naturally lend themselves to this point of view. One's objective as a scientist is to identify the hidden principle, to strip away the flesh (as it were) of noise and distracting factors and with Occam's razor lay bare the skeleton of a true mechanism for movement in the process. This is determinism. If instead we choose to view nature as stochastic (and we will not argue whether this is a wise choice) then we also recognize the biases of traditional determinism, and this includes our view of the stability of nonlinear time series.

In this paper, then, our objective is to identify the role that noise plays in determining stability conditions for nonlinear time series. This will involve general approaches to the problem of stability, illustrated with specific examples. We will also directly compare stability conditions for time series with those of dynamical systems which are deterministic and can sometimes be thought of as noiseless "skeletons" of the process (Tong (1990)). For some examples the stability conditions coincide and for others they do not.

As have most authors studying stability, starting with Priestley (1980), we take the approach of embedding the time series (say, $\{\xi_t\}$) in a suitable Markov chain referred to as the state space model and defining stability in terms of the ergodicity, null recurrence or transience of that chain. State space models may vary with the application and for our purposes they need not be observable. For example, an autoregressive type of process can be embedded in

$$\{X_t\} = \{(\xi_t, \dots, \xi_{t-p+1})\} \tag{1}$$

for some order p, whereas a process that also has a moving average component of order q could use the state space model

$$\{(X_t, U_t)\} = \{(\xi_t, \dots, \xi_{t-p+1}, e_t, \dots, e_{t-q+1})\}. \tag{2}$$

The time series is most suitably stable — and the properties of estimators behave best — when the Markov chain is geometrically ergodic (Nummelin (1984), Athreya and Pantula (1986), Chan (1989,1993a,b), Meyn and Tweedie (1992)), meaning the chain converges to its stationary distribution at a uniform geometric rate. Throughout the paper we set aside the questions of irreducibility and aperiodicity even though the errors can have a role in determining these properties as well. So unless we say otherwise, the stability conditions discussed below are to be taken in the context of a subspace on which the process is irreducible and/or with time lags according to the periodicity. Likewise, we assume continuity of the transitions in the sense of a T-chain (cf. Tuominen and Tweedie (1979), Meyn and Tweedie (1993), Cline and Pu (1998)).

In addition, there is a distinction between geometric ergodicity of a Markov chain and geometrically stable drift of a Markov chain. A chain with geometrically stable drift tends to decrease geometrically in magnitude when it becomes too large and it will satisfy a drift condition such as those in Theorems 1 and 2 below, but such drift is not necessary for geometric ergodicity. For most of the paper we will focus on geometrically stable drift as it most directly compares with the notion of geometric stability of a dynamical system skeleton (see (5)). In section 7, however, we return to this distinction in a discussion of the role that the noise distribution tails can play.

The error sequence will be denoted $\{e_t\}$ and is assumed iid. It may or may not contribute additively to the time series.

Sections 2 and 3 review stability of linear, bilinear and ARCH models, as well as stability of models that can be tied directly to their skeletons. Sections 4 and 5 go on to describe the standard uses of Foster-Lyapounov test functions with examples that behave like their skeletons and an example that does not. Section 6 presents a new approach to using such test functions to analyze models which either do not have skeletons or are characteristically different from their skeletons. Finally, in section 7 we discuss improvements possible when errors have sufficently light distribution tails.

2. Linear, Bilinear and ARCH Models.

Example 1. Viewing stability deterministically works especially well for a linear model. When in reduced form, the usual linear ARMA(p, q) model

$$\xi_t = a_1\xi_{t-1} + \cdots + a_p\xi_{t-p} + e_t + b_1e_{t-1} + \cdots + b_qe_{t-q} \tag{3}$$

has (2) as its state space model and

$$x_t^* = a_1x_{t-1}^* + \cdots + a_px_{t-p}^* \tag{4}$$

as its skeleton. The skeleton, in other words, is the time series stripped of its noise terms. Here, it is a linear dynamical system. A dynamical system $\{x_t^*\}$ is defined to be geometrically stable when a bounded solution exists for each initial condition and there exist $K < \infty$ and $\rho < 1$ such that

$$|x_t^*| \leq K(1 + \rho^t \|x_0\|) \quad \text{for all } t \geq 1 \text{ and } x_0 = (x_0^*, \ldots, x_{1-p}^*). \tag{5}$$

The system (4) is geometrically stable precisely when the solution is attracted to 0 regardless of the initial condition. An equivalent algebraic condition is that the eigenvalues of the so-called companion matrix,

$$A = \begin{pmatrix} a_1\, a_2 \cdots a_p \\ 1 \;\; 0 \cdots 0 \\ \vdots \;\; \ddots \; \ddots \;\; \vdots \\ 0 \cdots 1 \;\; 0 \end{pmatrix}.$$

have maximum modulus less than 1. Obviously, geometric stability for (3) (more precisely, for (2)) is thus identical to geometric stability for (4).

Example 2. The usual condition for stability of bilinear models also is essentially algebraic though it does depend on the error variance σ^2. The model is expressed as

$$\xi_t = a'X_{t-1} + X_{t-1}'BU_{t-1} + e_t + c'U_{t-1},$$

where e_t has zero mean, $X_{t-1} = (\xi_{t-1}, \ldots, \xi_{t-p})$, $U_{t-1} = (e_{t-1}, \ldots, e_{t-q})$, B is a matrix and a and c are vectors. If B is subdiagonal, there are appropriately defined matrices A_1, B_1 such that if $A_1 \otimes A_1 + \sigma^2 B_1 \otimes B_1$ has spectral radius less than 1 then the state space process (e.g., (2)) is geometrically ergodic (Pham (1985,1986)). See Bhaskara Rao et al (1983), Guégan (1987), Liu and Brockwell (1988), Liu (1992) and Pham (1993) for a treatment with more general B. Weaker conditions may actually suffice (Liu (1992)). These also depend on the distribution of e_t, again principally through the scale parameter.

Example 3. Combined autoregressive and autoregressive conditionally heteroscedastic (AR-ARCH) models such as

$$\xi_t = a_0 + a_1\xi_{t-1} + \cdots + a_p\xi_{t-p} + \left(b_0 + b_1\xi_{t-1}^2 + \cdots + b_p\xi_{t-p}^2\right)^{1/2} e_t$$

likewise have an algebraic condition depending on σ^2:

$$\left(\sum_{i=1}^{p} |a_i|\right)^2 + \sigma^2 \sum_{i=1}^{p} b_i < 1$$

(Tong (1981), Quinn (1982), Lu (1998b), cf. also Diebolt and Guégan (1993), Lu (1996,1998a), Borkovec (2000)). The relevant Markov chain is (1). Liu, Li and Li (1997) provided a similar algebraic condition for stability of a nonlinear AR-ARCH model with piecewise constant coefficient functions. We return to a generalization of this example in section 6.

3. Nonlinear AR(p) Models.

A common nonlinear model is the autoregressive model with functional coefficients (FCAR),

$$\xi_t = a_0(X_{t-1}) + a_1(X_{t-1})\xi_{t-1} + \cdots + a_p(X_{t-1})\xi_{t-p} + c(e_t; X_{t-1}), \tag{6}$$

where $c(e_t; x)$ has mean 0 for each $x \in \mathbb{R}^p$ and the state space model (1) defines the embedding Markov chain. The coefficient functions $a_0(x), \ldots, a_p(x)$ usually are assumed bounded. The error terms may be additive ($c(e_t; x) = e_t$) but more generally $\{c(e_t; x)\}$ has some regularity condition such as being uniformly integrable across the choices of x. Self-exciting threshold (SETAR) models are special cases where the coefficient functions are piecewise constant, whereas *threshold-like* models only require the functions to be asymptotically piecewise constant, as $||x|| \to \infty$. Examples of all of these have found successful application (cf. Tong (1990)). Nonparametric fitting of the FCAR model is now common (cf. Tjøstheim (1994), Chen and Härdle (1995), Härdle et al (1997)).

The skeleton of (6) is the dynamical system,

$$x_t^* = a(x_{t-1}^*, \ldots, x_{t-p}^*), \tag{7}$$

where

$$a(x) = a_0(x) + a_1(x)x_1 + \cdots + a_p(x)x_p, \quad x = (x_1, \ldots, x_p). \tag{8}$$

Explicit conditions for geometric stability of (7) are difficult to state, and not always known. Chan and Tong (1985) and Chan (1990) have shown, however, that if (7) is geometrically stable and $a(x)$ is Lipschitz continuous then (6) is likewise geometrically stable. That is, $\{X_t\}$ is a geometrically ergodic Markov chain satisfying a geometric drift condition. Cline and Pu (1999a, Thm. 3.1; 1999b, Thm. 2.5) have extended this result to the case $a(x)$ is asymptotically Lipschitz as $\min_{i=1,\ldots,p} |x_i| \to \infty$, in other words, for x far away from the axial hyperplanes. It has also been extended to nonlinear ARMA models (Cline and Pu (1999c)).

The condition is essentially deterministic as it depends only on the dynamical system and not on the noise. For the models where $a(x)$ is sufficiently smooth, the condition is sharp and thus stability of the time series coincides with stability of the skeleton. This includes any threshold model where $a(x)$ is piecewise linear and continuous. Unfortunately, the smoothness condition disallows the many threshold models for which the coefficient function is piecewise linear but not continuous.

Example 4. An example where this approach gives good results is the simple SETAR(1) model

$$\xi_t = (a_{01} + a_{11}\xi_{t-1})1_{\xi_{t-1}<0} + (a_{02} + a_{12}\xi_{t-1})1_{\xi_{t-1}>0} + c(\xi_{t-1})e_t$$

which has a piecewise linear and Lipschitz continuous autoregression function. Assuming $c(x)$ is bounded, this process has geometrically stable drift if and only if

$$\max(a_{11}, a_{12}, a_{11}a_{12}) < 1,$$

agreeing exactly with the geometric stability of its skeleton (Petruccelli and Woolford (1984), Chan et al (1985), Guo and Petruccelli (1991)).

This example and others naturally bring the following questions to mind: when are stability of the time series and stability of the skeleton equivalent? That is, when is stability determined independently of the errors? Is it usual for stability of the time series and stability of the skeleton to be equivalent or does the error distribution normally play a critical role?

4. The Foster-Lyapounov Drift Condition.

The connection between stability of a stochastic process and stability of a dynamical system is not superficial, even if it is not always as simple as one might like. To verify that a dynamical system is stable, one approach is to show that within some finite time the system (or some appropriate function of it) is sure to "drift" toward an attracting set. (See (5), for example.) A simple condition to check this is known as Lyapounov's drift condition (La Salle (1976)): for some nonnegative function V, $K < \infty$ and compact set C

$$V(x_t^*) \leq V(x_{t-1}^*) + K1_C(x_{t-1}^*) \quad \text{for all } t \geq 1.$$

Foster (1953) likewise showed that if the mean transition of a Markov chain on the nonnegative integers was uniformly negative for large states then the chain is certain to drift toward the origin whenever it gets too large, and therefore is ergodic.

The method was generalized by Tweedie (1975,1976,1983a), Popov (1977), Nummelin and Tuominen (1982), Meyn and Tweedie (1992) and others (cf. Nummelin (1984), Meyn and Tweedie (1993)) to what is now called the Foster-Lyapounov drift condition for ergodicity of an irreducible, aperiodic Markov chain $\{X_t\}$: for some function V taking values in $[1, \infty)$, $K < \infty$ and "small" set C,

$$E\Big(V(X_t) - V(X_{t-1}) \;\Big|\; X_{t-1} = x\Big) \leq K1_C(x) - 1. \qquad (9)$$

(A set C is *small* if there exists $m \geq 1$ and measure ν such that $P(X_m \in B \mid X_0 = x) \geq \nu(B)$ for all $x \in C$ and all B (cf. Nummelin (1984), Meyn and Tweedie (1993)). Typically, compact sets on an appropriately defined topological space are small.) The function V is then called a Foster-Lyapounov test function.

Furthermore, a Markov chain (or more precisely, the sequence of distributions generated by the transition kernel from an initial distribution) is in fact a dynamical system on the space of probability distributions and thus (9) can be interpreted as an ordinary Lyapounov drift condition for that system. This is the connection with dynamical systems, therefore, to be exploited. Indeed, Meyn and Tweedie (1992,1993) have explored the depth of this concept, and especially for the stronger drift condition for geometric ergodicity: for some function V taking values in $[1, \infty)$, $K < \infty$, $\rho < 1$ and small set C,

$$E\left(V(X_t) \,\Big|\, X_{t-1} = x\right) \leq \rho V(x) + K 1_C(x). \tag{10}$$

This drift condition ensures, among other things, V-uniform ergodicity and a geometric rate of convergence of the marginal distributions to the stationary distribution (Nummelin (1984), Chan (1989), Meyn and Tweedie (1992)), a strong law of large numbers for $\frac{1}{n}\sum_{t=1}^{n} h(X_t)$ if $|h(x)| \leq V(x)$ (Meyn and Tweedie (1992)), and a central limit theorem for $\frac{1}{n}\sum_{t=1}^{n} h(X_t)$ if $(h(x))^2 \leq V(x)$ (Meyn and Tweedie (1992), Chan (1993a,b)).

If the test function satisfies $||x||^r \leq V(x) \leq M + K||x||^r$ for some finite K and M then (10) is a condition for geometrically stable drift of the chain. (See Theorem 1 in the next section.)

Example 5. An example of the application of the drift condition (10) is to the FCAR model (6). Chan and Tong (1985, 1986) have shown that if

$$\sum_{i=1}^{p} \sup_x |a_i(x)| < 1 \tag{11}$$

then (10) may be verified for a test function of the form $V(x) = 1 + \sum_{i=1}^{p} c_i |x_i|$. Chen and Tsay (1993a) prove geometric ergodicity with the same condition using a slightly different approach. Condition (11) depends on the FCAR representation but An and Huang (1996) have shown that the somewhat weaker

$$\limsup_{||x|| \to \infty} \frac{|a(x)|}{\max_{i \leq p} |x_i|} < 1 \tag{12}$$

also suffices, using $V(x) = 1 + \max_{i \leq p} |x_i|$ and the m-step approach discussed in the next section. Unfortunately, not even (12) includes all geometrically stable linear models. For example, the AR(2) model $\xi_t = a_1 \xi_{t-1} + a_2 \xi_{t-2} + e_t$

is stable if and only if $\max(|a_2|, |a_1| + a_2) < 1$ but (12) holds only if $|a_1| + |a_2| < 1$.

5. Approaches to Using Drift Tests.

The benefit realized in applying a drift condition clearly relies on the nature of the test function. The choice of a test function has been described as something of an art and the functions are often constructed specifically for the model at hand. Test functions which are based on a simple kind of norm (e.g., $V(x) = 1 + ||x||^r$ or $V(x) = 1 + \max_{i \le p} |x_i|^r$) usually do not provide sharp conditions for stability.

Tjøstheim (1990) suggested the *m-step* approach for overcoming this difficulty. The idea is to apply a drift condition such as (9) or (10) to the m-step process $\{X_{mt}\}$ rather than to $\{X_t\}$ and to employ the useful result that, under irreducibility and aperiodicity, stability of $\{X_{mt}\}$ is equivalent to stability of $\{X_t\}$. This approach is theoretically optimal if m is chosen large enough but it obviously requires higher order transitions and computation of the resulting expectations.

Another approach we call the *directional* method has proved useful for certain SETAR models (Petruccelli and Woolford (1984), Chen and Tsay (1991), Lim (1992), Cline and Pu (1999b)). Here we use $V(x) = 1 + \lambda(x)||x||^r$, where $r > 0$ and λ is bounded and bounded away from 0. Typically λ is primarily a function of the direction of x rather than its magnitude. Optimally choosing λ is equivalent to optimally choosing Tjøstheim's m (see below). Furthermore, the geometric drift condition can be expressed in terms of the drift of the logarithm of $V(X_t)$ (the *log-drift* condition).

For the following theorems let Λ be the class of all nonnegative measurable functions on $\mathbb{R}^p$ that are bounded and bounded away from 0.

Theorem 1 *Assume $\{X_t\}$ is an aperiodic, ϕ-irreducible T-chain in $\mathbb{R}^p$ such that $E\left(||X_1||^r/(1 + ||x||^r) \,\middle|\, X_0 = x\right)$ is bounded for some $r > 0$. The following are equivalent conditions, each sufficient for $\{X_t\}$ to be geometrically ergodic.*

(i) $\limsup_{||x|| \to \infty} E\left(\frac{\lambda(X_1)||X_1||^r}{\lambda(x)||x||^r} \,\middle|\, X_0 = x\right) < 1$ *for some* $\lambda \in \Lambda$, $r > 0$.

(ii) $\limsup_{||x|| \to \infty} E\left(\frac{1+||X_n||^r}{1+||X_m||^r} \,\middle|\, X_0 = x\right) < 1$ *for some* $r > 0$, $n > m \ge 0$.

(iii) $\limsup_{||x|| \to \infty} E\left(\log\left(\delta + \frac{\lambda(X_1)||X_1||}{\lambda(x)||x||}\right) \,\middle|\, X_0 = x\right) < 0$ *for some* $\delta > 0$, $\lambda \in \Lambda$.

(iv) $\limsup_{||x|| \to \infty} E\left(\log\left(\delta + \frac{1+||X_n||}{1+||X_m||}\right) \,\middle|\, X_0 = x\right) < 0$ *for some* $\delta > 0$, $n > m \ge 0$.

Proof The equivalences follow from Cline and Pu (1999a, Lem. 4.1, Lem. 4.2) and the sufficiency for geometric ergodicity from condition (10) with test function $V(x) = 1 + \lambda(x)||x||^r$. □

If the noise terms in a FCAR(p) model do not dominate when the time series is very large then they may be at least partly ignored, as the following theorem suggests.

Theorem 2 *Assume $\{X_t\}$ is an aperiodic, ϕ-irreducible T-chain defined by (1) and (6) such that $a_0, \ldots, a_p$ are bounded, $\sup_{||x|| \le M} E\left(|c(e_1;x)|^r\right) < \infty$ for some $r > 0$ and all $M < \infty$, and $\lim_{||x|| \to \infty} E\left(|c(e_1;x)|^r/||x||^r\right) = 0$. Let a be given by (8) and $\theta(x) = (a(x), x_1, \ldots, x_{p-1})/(1 + ||x||)$ for $x = (x_1, \ldots, x_p)$. The following are equivalent conditions, each sufficient for $\{X_t\}$ to be geometrically ergodic.*

(i) $\limsup_{||x|| \to \infty} E\left(\frac{\lambda(X_1)}{\lambda(x)} \,\middle|\, X_0 = x\right) ||\theta(x)||^r < 1$ *for some* $r > 0$, $\lambda \in \Lambda$.

(ii) $\limsup_{||x|| \to \infty} E\left(\prod_{j=m}^{n} ||\theta(X_j)||^r \,\middle|\, X_0 = x\right) < 1$ *for some* $r > 0$, $n \ge m \ge 0$:

(iii) $\limsup_{||x|| \to \infty} E\left(\log\left(\delta + \frac{\lambda(X_1)}{\lambda(x)}||\theta(x)||\right) \,\middle|\, X_0 = x\right) < 0$ *for some* $\delta > 0$, $\lambda \in \Lambda$.

(iv) $\limsup_{||x|| \to \infty} E\left(\sum_{j=m}^{n} \log\left(\delta + ||\theta(X_j)||\right) \,\middle|\, X_0 = x\right) < 0$ *for some* $\delta > 0$, $n \ge m \ge 0$.

Proof This essentially is Theorem 3.2 in Cline and Pu (1999a). □

Similar theorems express equivalent conditions for transience of a Markov chain (e.g., Cline and Pu (2001, Thm. 2.2)).

Example 6. We refer to conditions (iii,iv) in Theorem 1 and (iii,iv) in Theorem 2 as "log-drift" conditions. A model which illustrates the benefits of using a log-drift condition is the Periodic FCAR(1) model,

$$\xi_t = a_0(\xi_{t-1}) + a_1(\xi_{t-1})\xi_{t-1} + e_t,$$

where the coefficient function $a_1(x)$ is periodic with period τ. This is an unusual model but there is a suggestion of periodicity in the coefficient functions fitted by Chen and Tsay (1993a) for the sunspot number data, and it was at the urging of Prof. Chen that we studied the stability of this particular model. At any rate, it is a useful example.

With $n = m = 1$, the condition in Theorem 2(iv) can be reexpressed as

$$\limsup_{||x|| \to \infty} E\left(\log(\delta + |a_1(\xi_1)|) \,\middle|\, \xi_0 = x\right) < 0 \quad \text{for some } \delta > 0. \tag{13}$$

(This can also be determined from the 2-step condition with $V(x) = 1 + |x|$.) The function $E(\log(\delta + |a_1(\xi_1)|) \mid \xi_0 = x)$ is close to being periodic in $|a_1(x)x|$. Thus the left-hand side of (13) is only a limsup, not a limit, which has the unfortunate consequence that condition (13) is not sharp.

The solution is to choose $n = m = 2$ instead. Assume $a_1(x)$ is continuously differentiable with a derivative that is 0 on a set of measure 0. Then the function $E(\log(\delta + |a_1(\xi_2)|) \mid \xi_0 = x)$ does in fact have a limit and therefore does lead to a sharp condition, namely if

$$\int_0^{\tau} \log(|a_1(u)|)du < 0 \tag{14}$$

then $\{\xi_t\}$ is geometrically ergodic and if $\int_0^{\tau} \log(|a_1(u)|)du > 0$ then $\{\xi_t\}$ is transient (Cline and Pu (1999a, Thm. 3.4; 2001, Thm. 3.2)).

Note that the skeleton process, $x_t = a_0(x_{t-1}) + a_1(x_{t-1})x_{t-1}$ is geometrically stable if and only if $\sup |a_1(x)| < 1$. Therefore, this is an example where stability of the time series does not coincide with stability of its skeleton. For a specific example, suppose $a_1(x) = c + d\cos(x)$ with $|c| + |d| > 1$, $|c| \leq |d| < 2$. Then the time series is geometrically ergodic but the skeleton is not stable.

Although condition (14) does not explicitly refer to the noise distribution, the noise does play a major role in determining the condition by causing the drift to be averaged.

Example 7. The directional method, on the other hand, seems to work well with many threshold models of order 1, including ones that employ a delay. Consider, for example, the TAR(1) time series with delay d given by

$$\xi_t = a_0(X_{t-1}) + a_1(X_{t-1})\xi_{t-1} + e_t, \qquad X_{t-1} = (\xi_{t-1}, \ldots, \xi_{t-d}),$$

$a_0(x)$ bounded and $a_1(x)$ depends only on $(\mathrm{sgn}(x_1), \ldots, \mathrm{sgn}(x_d))$, $x = (x_1, \ldots, x_d)$. There are thus 2^d regions $R_1, \ldots, R_{2^d}$ each corresponding to a coefficient: $a_{1j} = a_1(x)$, $x \in R_j$. Notice that the thresholds — boundaries of the regions — are the axial hyperplanes.

As long as the time series remains large (which is all that is of concern for stability), ξ_t avoids the thresholds and hence X_t cycles among some subset of the regions. There may be several such subsets possible, depending on the signs of the a_{1j}'s, and it is the drift of the "worst case" cycle that is critical. Thus the geometric drift condition for stability is exactly (Cline and Pu (1999b, Cor. 2.4))

$$\max_{cycle(R_{j_1}, \ldots, R_{j_k})} \prod_{i=1}^{k} a_{1j_i} < 1,$$

where the maximum is taken over the possible cycles described above. This corresponds precisely to the geometric stability of the skeleton. The test function for establishing this condition takes the directional form, $V(x) = 1 + \lambda(x)||x||^r$, and the optimal choice for $\lambda(x)$ is constant on each of the 2^d regions.

6. The Piggyback Method.

To this point we have presented examples studied with what one might call the traditional methods of drift analysis. Not all models yield to this analysis, however, including some surprisingly simple models. In this section we present a new approach, as yet somewhat informal, that employs a much more sophisticated m-step test function. We call it the *piggyback* method because it relies on finding a stable Markov chain similar to a process embedded in the one of interest and building a new test function on top of one that works for the known stable chain.

We will first present a sketch of the piggyback method and then provide three examples. The sketch, however, is quite rough because in fact the method is applied somewhat differently for each of the examples. (See the papers referenced below.) Indeed, the concept is elegant but its application is messy and as yet we do not know how generally useful it may prove to be.

The time series $\{\xi_t\}$ is embedded in a Markov chain $\{X_t\}$. At the same time we consider another Markov chain $\{Y_t\}$ similar to a simpler process embedded in $\{X_t\}$. The chain $\{Y_t\}$ is assumed to be geometrically ergodic and, in particular, to satisfy the geometric drift condition with test function $V_1(y)$. Its stationary distribution we denote G. If there is a function $H(y)$ which somehow exemplifies (or bounds) the relative change in magnitude of X_1 when X_0 is large and $Y_0 = y$ then, intuitively, the stationary value of $H(Y_t)$ will measure the geometric drift of $\{X_t\}$. Thus, a log-drift condition for geometric stability would be

$$\exp\left(\int \log(\delta + H(y))G(dy)\right) < 1 \quad \text{for some } \delta > 0. \tag{15}$$

To obtain a test function that will yield such a condition, we first define

$$h(y) = E((\delta + H(Y_1))^r | Y_0 = y)V_1(y)$$

and let $y(x)$ identify the "embedding" of Y_t into X_t. An integer m is chosen suitably large, a "correction" function $c(x)$ constructed and the ultimate test function is (something like)

$$V(x) = c(x)\left(\prod_{j=1}^{m} E(h(Y_j)|Y_0 = y(x))\right)^{1/m}.$$

The key point for this paper is that the piggyback method and the resulting condition for ergodicity capture the implicit stochastic behavior of $\{Y_t\}$, not the behavior of a deterministic skeleton of $\{X_t\}$ or $\{\xi_t\}$. Even if such a skeleton can be identified, its stability properties will not coincide with those of $\{\xi_t\}$. Alternatively, one may think of $\{X_t\}$ as having a sort of stochastic skeleton which must be analyzed for stability.

Example 8. Our first example is a bivariate threshold model (Cline and Pu (1999a, Ex. 3.2; 2001, Ex. 3.2)). Indeed it is the simplest such model that is not just two independent univariate models joined together. Suppose

$$X_t = \begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} a_1(X_{t-1,1})X_{t-1,1} \\ a_2(X_{t-1,1})X_{t-1,2} \end{pmatrix} + \begin{pmatrix} e_{t,1} \\ e_{t,2} \end{pmatrix},$$

where $a_i(x_1) = a_{i1}1_{x_1<0} + a_{i2}1_{x_1>0}$, $i = 1, 2$. Note that the nonlinearity of the second component $X_{t,2}$ is driven by the univariate TAR(1) process $\{X_{t,1}\}$. The latter is our "embedded" process and is stable when

$$\max(a_{11}, a_{12}, a_{11}a_{12}) < 1. \tag{16}$$

Let G be its stationary distribution. The function $|a_2(x_1)|$ plays the role of $H(y)$ so that the resulting (sharp) stability condition is, in addition to (16),

$$|a_{21}|^{G(0)}|a_{22}|^{1-G(0)} < 1,$$

which represents the stationary value of the relative change in magnitude of $X_{t,2}$ when it is very large. This condition neither implies nor is implied by the stability condition for the corresponding skeleton process: (16) plus

$$\max(|a_{21}|1_{a_{11}>0}, |a_{22}|1_{a_{12}>0}, |a_{21}a_{22}|1_{a_{11}<0, a_{12}<0}) < 1.$$

Example 9. The second example (cf. Cline and Pu (1999c)) is the threshold ARMA(1,q) model (TARMA) with a delay d:

$$\xi_t = a_0(X_{t-1}) + a_1(X_{t-1})\xi_{t-1} + e_t + b_1(X_{t-1})e_{t-1} + \cdots + b_q(X_{t-1})e_{t-q},$$

where $X_{t-1} = (\xi_{t-1}, \ldots, \xi_{t-d})$, $a_0, b_1, \ldots, b_q$ are bounded and $a_1(x)$ is (asymptotically) piecewise constant. We further assume the thresholds are affine which implies the regions on which $a_1(x)$ is constant are cones in $\mathbb{R}^d$. This is the simplest interesting example of a TARMA process. See also Brockwell, Liu and Tweedie (1992) and Liu and Susko (1992). In the case $q > 0$, the time series must be embedded in the Markov chain $\{(X_t, U_t)\}$ where X_t is as above and $U_t = (e_t, \ldots, e_{t-q+1})$.

Threshold ARMA models have not seen a lot of study, perhaps in part because the moving average terms can affect the irreducibility and periodicity properties of the chain in complicated ways as yet not well understood. (See, for example, Cline and Pu (1999c)). This by itself is a major role played by the noise but we will pass by it here.

Let the regions $R_1, \ldots, R_m$ be the partition of $\mathbb{R}^p$ such that $a_1(x)$ is constant on each region, with $a_{11}, \ldots, a_{1m}$ being the corresponding constants. There are basically two types of situations that arise in these models when ξ_t is very large: cyclical and noncyclical. For the cyclical situation, $\{X_t\}$ essentially cycles close to certain rays having the form

$$(\prod_{i=1}^{p-1} a_{1j_i}, \prod_{i=1}^{p-2} a_{1j_i}, \ldots, 1)x_1, \tag{17}$$

if all are in the interior of the conical regions. Noise plays no role in determining the stability in this situation since X_t avoids the thresholds; all that matters is the product of coefficients realized by moving through the cycle. For a model which is purely cyclical the stability condition is based on the "worst case" cycle, it is deterministic and corresponds to that of the skeleton process, and it is very much like that of the TAR(1) process with delay discussed in section 5.

The model may also have, however, situations where one or more of the rays of type (17) actually lie on a threshold. In such a case, X_t can fall on either side of the threshold, and thus into one of two possible regions, at random but depending on both the present error e_t and the past errors $e_{t-1}, \ldots, e_{t-q}$. If J_t denotes the region that X_t is in then $\{(J_t, U_t)\}$ behaves something like a Markov chain where the first component is one of a finite number of states and the second component is stationary. (If $q = 0$ then $\{J_t\}$ itself is like a finite state Markov chain.) We relate $\{(J_t, U_t)\}$ to such a Markov chain denoted, say, $\{(\tilde{J}_t, \tilde{U}_t)\}$. This chain is not necessarily irreducible or aperiodic but clearly every invariant measure is finite. Indeed it may be decomposed into a finite number of uniformly ergodic subprocesses. The coefficients $|a_{1j}|$ play the role of $H(y)$ in this model.

Now let G be any stationary distribution for $\{(\tilde{J}_t, \tilde{U}_t)\}$ and define $\pi_j = \int_{\mathbb{R}^q} G(j, du)$. If (condition (15))

$$\prod_j |a_{1j}|^{\pi_j} < 1$$

regardless of the choice of G then $\{(X_t, U_t)\}$ is geometrically ergodic and, again, the condition is sharp. Because at least one ray lies on a threshold, the noncyclical models are special cases, but the stability condition for a noncyclical process can be quite different from that of nearby purely cyclical processes. See the parameter spaces for the TARMA(1,1) with delay 2 in Cline and Pu (1999c).

Example 10. The third example combines the nonlinearity of piecewise continuous coefficient functions with a piecewise conditional heteroscedascity, a model called the threshold AR-ARCH time series:

$$\xi_t = a(X_{t-1}) + b(X_{t-1})e_t + c(e_t; X_{t-1}),$$

where $a(x)$ and $b(x)$ are piecewise linear, $\{c(e_1; x)\}$ is uniformly integrable and $X_{t-1} = (\xi_{t-1}, \ldots, \xi_{t-p})$. We further suppose $a(x)$ and $b(x)$ are homogeneous, $b(x)$ is locally bounded away from zero except at $x = 0$, the thresholds are subspaces containing the origin and the regions of constant behavior are cones. Note that these assumptions need only hold asymptotically (in an appropiate sense) as x gets large. Once again, the Markov chain under study is $\{X_t\}$.

The basic idea on which we piggyback is that the process $\{X_t\}$ collapsed to the unit sphere behaves very much like a Markov chain. The compactness of the unit sphere serves to make this chain stable and then the stability condition for the original chain can be computed. More specifically, define

$$\xi_t^* = a(X_{t-1}) + b(X_{t-1})e_t, \quad X_t^* = (\xi_t^*, \ldots, \xi_{t-p+1}^*) \quad \text{and} \quad \theta_t^* = X_t^*/||X_t^*||.$$

Then, due to the homogeneity of $a(x)$ and $b(x)$, $\{\theta_t^*\}$ is a Markov chain on the unit sphere and is uniformly ergodic with stationary distribution G, say. By the piggyback method, therefore, X_t has geometrically stable drift if

$$E\left(\int \log(|a(\theta) + b(\theta)e_1|/|\theta_1|)G(d\theta)\right) < 0.$$

For a simple demonstration, suppose $p = 1$, $a(x) = (a_1 1_{x<0} + a_2 1_{x\geq 0})x$ and $b(x) = (b_1 1_{x<0} + b_2 1_{x\geq 0})x$, $b_i \neq 0$, $i = 1, 2$. Then $\{\theta_t^*\}$ is a Markov chain on $\{-1, 1\}$ with transition probabilities $p_{-1,1} = P(a_1 + b_1 e_1 \leq 0) = 1 - p_{-1,-1}$ and $p_{1,-1} = P(a_2 + b_2 e_1 \leq 0) = 1 - p_{1,1}$. If

$$E\left(\frac{p_{1,-1}}{p_{-1,1} + p_{1,-1}} \log(|a_1 + b_1 e_1|) + \frac{p_{-1,1}}{p_{-1,1} + p_{1,-1}} \log(|a_2 + b_2 e_1|)\right) < 0$$

then $\{\xi_t\}$ is geometrically ergodic. This example and generalizations of it will be considered fully in a forthcoming paper (Pu and Cline (2001)).

Example 11. A very simple model which has not been analyzed fully is the ordinary TAR(2) model with additive noise,

$$\xi_t = a_1(X_{t-1})\xi_{t-1} + a_2(X_{t-1})\xi_{t-2} + e_t,$$

where $X_{t-1} = (\xi_{t-1}, \xi_{t-2})$ and $a_1(x)$ and $a_2(x)$ are piecewise constant. The precise stability condition is not known even when there is but one affine threshold. The results of this section, however, suggest that the key will be to identify an appropriate stochastic skeleton process to study.

7. The Role of the Noise Distribution Tails.

Spieksma and Tweedie (1994) pointed out how, with appropriate assumptions on the error distribution tails, an ordinary drift condition (such as (9) with $V(x) = 1 + |x|$) can be boosted to ensure geometric ergodicity of the process. We generalize the result as follows.

Theorem 3 *Assume $\{X_t\}$ is an aperiodic, ϕ-irreducible T-chain in $\mathbb{R}^p$ and $V : \mathbb{R}^p \to [1, \infty)$ is locally bounded. Suppose there exists a random variable $W(x)$ for each x such that $V(X_1) \le W(x)$ whenever $X_0 = x$, $\{|W(x) - V(x)| + e^{r(W(x)-V(x))}\}$ is uniformly integrable for some $r > 0$ and*

$$\limsup_{||x|| \to \infty} E\left(W(x) - V(x)\right) < 0. \tag{18}$$

Then there exist $s > 0$ and $V_1(x) = e^{sV(x)}$ such that $\{X_t\}$ is V_1-uniformly ergodic (and hence geometrically ergodic).

Proof This follows directly from the drift condition for V-uniform ergodicity (cf. Meyn and Tweedie (1993, Thm. 16.0.1)) and uniform convergence (cf. Cline and Pu (1999a, Lem. 4.2)). (See also the proof to Theorem 4.) □

Essentially, this is the log-drift condition in another guise: if the test function in (10), for example, is replaced with $V_1(x) = e^{sV(x)}$ with some sufficiently small $s > 0$ then (18) is a log-drift version of the condition. As a bonus, if $V(x)$ is *norm-like*, satisfying $||x|| \le V(x) \le M + K||x||$, one gets strong laws and central limit theorems for all the sample moments (Meyn and Tweedie (1992), Chan (1993a,b)) and exponentially damping tails in the stationary distribution (Tweedie (1983a,b)).

Example 12. For example, consider the FCAR(p) model discussed in section 3. If the noise term $c(e_t; X_{t-1})$ is such that $\sup_x E(e^{r|c(e_1;x)|}) < \infty$ for some $r > 0$ then it frequently is possible to satisfy the requirements of Theorem 3 with a norm-like $V(x)$. To illustrate how this can work, consider the FCAR(1) process, $X_t = \xi_t = a_1(\xi_{t-1}) + c(e_t; \xi_{t-1})$ with

$$\begin{aligned} -L \le a_1(x) &\le a_{11}x + a_{01} \text{ if } x < -L, \\ L \ge a_1(x) &\ge a_{12}x + a_{02} \text{ if } x > L, \end{aligned} \tag{19}$$

where $a_{11}a_{12} = 1$, $a_{11} < 0$ and $L < \infty$. We assume here that $E(c(e_1; x)) = 0$ for all $x \in \mathbb{R}$ and $\sup_x E(e^{r|c(e_1;x)|}) < \infty$ for some $r > 0$. For the special case of equality on the right in (19) (the SETAR(1) model of Example 4), Chan et al (1985) showed $\{\xi_t\}$ is ergodic if and only if $\gamma \overset{\text{def}}{=} a_{11}a_{02} + a_{01} < 0$. We thus assume $\gamma < 0$. Let $\lambda_1 = \lambda_2^{-1} = \sqrt{-a_{11}}$ and choose $\delta_i \ge 1$, $i = 1, 2$ so that $-\lambda_1 a_{02} + \delta_1 - \delta_2 = \lambda_2 a_{01} + \delta_2 - \delta_1 = \lambda_2\gamma/2$. Define

$$V(x) = (\lambda_1|x| + \delta_1)1_{x<0} + (\lambda_2|x| + \delta_2)1_{x\ge 0}.$$

Then it is a simple computation to show that for some $\epsilon > 0$ and $K < \infty$, $V(X_1) - V(x) \le (\lambda_2 1_{x<0} - \lambda_1 1_{x\ge 0})c(e_1;x) + \lambda_2\gamma/2 + K|c(e_1;x)|1_{|c(e_1;x)|>\epsilon|x|}$ when $|X_0| = |x|$ is sufficiently large, which satisfies the conditions of Theorem 3 with the limit in (18) being $\lambda_2\gamma/2$. The time series is thus geometrically ergodic. On the other hand its skeleton, while stable, is not geometrically stable since $a_{11}a_{12} = 1$. In fact we would say both have only a linear drift.

Tanikawa (1999) studied this example and Cline and Pu (1999b) looked at similar first order threshold-like models, but with a possible delay. Using a similar approach but with stronger stability conditions, Diebolt and Guégan (1993) studied multivariate examples and An and Chen (1997) investigated FCAR(p) models with $p \ge 1$.

One of the drawbacks to a log-drift condition such as the one in Theorem 1(iii) is that it guarantees geometric ergodicity only with test functions of the form $V(x) = 1 + \lambda(x)||x||^r$ where r may be arbitrarily small and therefore it fails to imply needed limit theorems for sample moments. To be able to conclude V_1-uniform ergodicity with an exponential-like V_1, the condition must again be boosted and then the desired limit theorems will hold.

Theorem 4 *Assume $\{X_t\}$ is an aperiodic, ϕ-irreducible T-chain in $\mathbb{R}^p$ and $V : \mathbb{R}^p \to [1,\infty)$ is locally bounded and $V(x) \to \infty$ as $||x|| \to \infty$. Suppose there exists a random variable $W(x)$ for each x such that $V(X_1) \le W(x)$ whenever $X_0 = x$, $\{|\log(W(x)/V(x))| + e^{(W(x))^r - (V(x))^r}\}$ is uniformly integrable for some $r > 0$ and*

$$\limsup_{||x||\to\infty} E\left(\log(W(x)/V(x))\right) < 0. \tag{20}$$

Then there exist $s > 0$ and $V_1(x) = e^{(V(x))^s}$ such that $\{X_t\}$ is V_1-uniformly ergodic (and hence geometrically ergodic).

Proof For $v \ge w \ge 1$ and $0 < s < r$, we have $\frac{1}{s}\left(e^{w^s - v^s} - 1\right) \le \frac{1}{s}\left(\frac{w^s}{v^s} - 1\right)$ and $\log(w/v) \le \frac{1}{s}\left(\frac{w^s}{v^s} - 1\right) \le 0$. By the uniform integrability of $\{\log(W(x)/V(x))\}$, truncation and uniform convergence (as $s \downarrow 0$),

$$\begin{aligned}
&\limsup_{||x||\to\infty} E\left(\frac{1}{s}\left(e^{(W(x))^s - (V(x))^s} - 1\right) 1_{W(x)\le V(x)}\right) \\
&\le \limsup_{||x||\to\infty} E\left(\frac{1}{s}\left(\frac{(W(x))^s}{(V(x))^s} - 1\right) 1_{W(x)\le V(x)}\right) \\
&\le \limsup_{||x||\to\infty} E\left(\log(W(x)/V(x)) 1_{W(x)\le V(x)}\right) + \epsilon < -\epsilon,
\end{aligned} \tag{21}$$

for $\epsilon > 0$ and $s > 0$ small enough.

For $w \geq v \geq 1$ and $0 < s < r/2$, we have

$$0 \leq \log(w/v) \leq \frac{1}{s}\left(e^{w^s - v^s} - 1\right) \leq \frac{1}{r}\left(e^{w^r - v^r} - 1\right)$$

and if $w^r - v^r \leq K$, $v > M \geq 1$ then $\frac{1}{s}\left(e^{w^s - v^s} - 1\right) \leq \frac{Ke^{K/2}}{rM^{r/2}}$. By the uniform integrability of $\{e^{(W(x))^r - (V(x))^r}\}$, truncation and $V(x) \to \infty$ as $||x|| \to \infty$,

$$\begin{aligned} 0 &\leq \limsup_{||x|| \to \infty} E\left(\log(W(x)/V(x)) 1_{W(x) \geq V(x)}\right) \\ &\leq \limsup_{||x|| \to \infty} E\left(\frac{1}{s}\left(e^{(W(x))^s - (V(x))^s} - 1\right) 1_{W(x) \geq V(x)}\right) < \epsilon, \end{aligned} \tag{22}$$

for $\epsilon > 0$ and $s > 0$ small enough.

From (20)–(22), therefore, we conclude there exists $s > 0$ small enough that

$$\limsup_{||x|| \to \infty} E\left(\frac{1}{s}\left(e^{(V(X_1))^s - (V(x))^s} - 1\right) \,\Big|\, X_0 = x\right) < 0.$$

Also, $\sup_{||x|| \leq M} E\left(e^{(V(X_1))^s} \,\Big|\, X_0 = x\right) < \infty$ for all $M < \infty$, and hence geometric ergodicity is assured with test function V_1. □

Example 13. We again consider an FCAR(1) model, $\xi_t = a_1(\xi_{t-1}) + c(e_t; \xi_{t-1})$, satisfying (19) but now we assume $a_{11} < 0 < a_{11}a_{12} < 1$ and $|c(e_1; x)| \leq c_1|x|^\beta |e_1|$ where $c_1 > 0$, $0 < \beta < 1$ and $E(e^{\eta|e_1|}) < \infty$ for some $\eta > 0$. Let $\lambda_1 = \sqrt{-a_{11}}$, $\lambda_2 = \sqrt{-a_{12}}$ and $V(x) = 1 + (\lambda_1 1_{x<0} + \lambda_2 1_{x \geq 0})|x|$. Then for $|X_0| = |x|$ sufficiently large and some $\epsilon > 0$ and $K < \infty$, $V(X_1) \leq (1-\epsilon)V(x) + K|x|^\beta |e_1|$, which satisfies Theorem 4 with $r < 1 - \beta$. See also Diebolt and Guégan (1993) and Guégan and Diebolt (1994) for related results.

When the errors are bounded, an otherwise unstable model can sometimes be stable. See Chan and Tong (1994) for an example.

References

An, H.Z. and Chen, S.G. (1997). A note on the ergodicity of non-linear autoregressive model, *Stat. Prob. Letters* 34, 365–372.

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

Athreya, K.B. and Pantula, S.G. (1986). Mixing properties of Harris chains and autoregressive processes, *J. Appl. Probab.* 23, 880–892.

Bhaskara Rao, M., Subba Rao, T. and Walker, A.M. (1983). On the existence of some bilinear time series models, *J. Time Series Anal.* 4, 95–110.

Borkovec, M. (2000). Extremal behavior of the autoregressive process with ARCH(1) errors, *Stoch. Proc. Appl.* 85, 189-207.

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. (1989). A note on the geometric ergodicity of a Markov chain, *Adv. Appl. Probab.* 21, 702–704.

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. (1993a). A review of some limit theorems of Markov chains and their applications, *Dimensions, Estimation and Models*, ed. by H. Tong, World Scientific (Singapore), 108–135.

Chan, K.-S. (1993b). On the central limit theorem for an ergodic Markov chain, *Stoch. Proc. Appl.* 47, 113–117.

Chan, K.-S., Petruccelli, 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 Series Anal.* 7, 179–190.

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

Chen, R. and Härdle, W. (1995). Nonparametric time series analysis, a selective review with examples, *Bulletin of the International Statistical Institute, 50th session of ISI*, August, 1995, Beijing, China.

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. Stat. Assoc.* 88, 298–308.

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

Cline, D.B.H. and Pu, H.H. (1998). Verifying irreducibility and continuity of a nonlinear time series, *Stat. & Prob. Letters* 40, 139–148.

Cline, D.B.H. and Pu, H.H. (1999a). Geometric ergodicity of nonlinear time series, *Stat. Sinica* 9, 1103–1118.

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

Cline, D.B.H. and Pu, H.H. (1999c). Stability of threshold-like ARMA time series, Statistics Dept., Texas A&M Univ.

Cline, D.B.H. and Pu, H.H. (2001). Geometric transience of nonlinear time series, *Stat. Sinica* 11, 273–287.

Collomb, G. and Härdle, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations, *Stoch. Proc. Appl.* 23, 77–89.

Diebolt, J. and Guégan, D. (1993). Tail behaviour of the stationary density of general non-linear autoregressive processes of order 1, *J. Appl. Probab.* 30, 315–329.

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

Guégan, D. (1987). Different representations for bilinear models, *J. Time Series Anal.* 8, 389–408.

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

Guo, M. and Petruccelli, 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.

Härdle, W. and Vieu, P. (1992). Kernel regression smoothing of time series, *J. Time Series Anal.* 13, 209–232.

La Salle, J.P. (1976). *The Stability of Dynamical Systems*, CMBS 25, Society for Industrial and Applied Mathematics (Philadelphia).

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

Liu, J. (1992). On stationarity and asymptotic inference of bilinear time series models, *Stat. Sinica* 2, 479–494.

Liu, J. and Brockwell, P.J. (1988). On the general bilinear time series models, *J. Appl. Probab.* 25, 553–564.

Liu, J., Li, W.K. and Li, C.W. (1997). On a threshold autoregression with conditional heteroscedastic variances, *J. Stat. Plan. Inference* 62, 279–300.

Liu, J. and Susko, E. (1992). On strict stationarity and ergodicity of a nonlinear ARMA model, *J. Appl. Probab.* 29, 363–373.

Lu, Z. (1996). A note on the geometric ergodicity of autoregressive conditional heteroscedasticity (ARCH) model, *Stat. Prob. Letters* 30, 305–311.

Lu, Z. (1998a). Geometric ergodicity of a general ARCH type model with applications to some typical models, in *Advances in Operations Research and Systems Engineering*, J. Gu, G. Fan and S. Wang, eds., Global-Link Informatics Ltd., 76–86.

Lu, Z. (1998b). On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term, *Stat. Sinica* 8, 1205–1217.

Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality, *Econometric Theory* 11, 258–289.

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.

Nummelin, E. and Tuominen, P. (1982). Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory, *Stoch. Proc. Appl.* 12, 187–202.

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

Pham D.T. (1985). Bilinear Markovian representation and bilinear models, *Stoch. Proc. Appl.* 20, 295–306.

Pham, D.T. (1986). The mixing property of bilinear and generalised random coefficient autoregressive models, *Stoch. Proc. Appl.* 23, 291–300.

Pham, D.T. (1993). Bilinear times series models, *Dimensions, Estimation and Models*, ed. by H. Tong, World Scientific Publishing (Singapore), 191–223.

Popov, N. (1977). Conditions for geometric ergodicity of countable Markov chains, *Soviet Math. Dokl.* 18, 676–679.

Priestley, M.B. (1980). State-dependent models: A general approach to non-linear time series analysis, *J. Time Series Anal.* 1, 47–71.

Pu, H.H. and Cline, D.B.H. (2001). Stability of threshold AR-ARCH models, tech. rpt., Dept. Stat., Texas A&M University (forthcoming).

Quinn, B.G. (1982). A note on the existence of strictly stationary solutions to bilinear equations, *J. Time Series Anal.* 3, 249–252.

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

Tanikawa, A. (1999). Geometric ergodicity of nonlinear first order autoregressive models, *Stoch. Models* 15, 227–245.

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

Tjøstheim, D. (1994). Non-linear time series: a selective review, *Scand. J. Stat.* 21, 97–130.

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. (1981). A note on a Markov bilinear stochastic process in discrete time, *J. Time Series Anal.* 2, 279–284.

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

Tuominen, P. and Tweedie, R.L., (1979). Markov chains with continuous components, *Proc. London Math. Soc.* (3) 38, 89–114.

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.

Tweedie, R.L. (1983a). Criteria for rates of convergence of Markov chains with application to queueing and storage theory, in *Probability, Statistics and Analysis*, London Math. Society Lecture Note Series, ed. by J.F.C. Kingman and G.E.H. Reuter, Cambridge Univ. Press (Cambridge).

Tweedie, R.L. (1983b). The existence of moments for stationary Markov chains, *Adv. Appl. Probab.* 20, 191–196.