# Establishing Stationarity of Time Series Models via Drift Criteria

Dawn B. Woodard David S. Matteson
Shane G. Henderson
School of Operations Research and Information Engineering
Cornell University

January 8, 2010

**Abstract**

Time series models are often constructed by combining nonstationary effects such as trends with stochastic processes that are known (or believed) to be stationary. However, there are numerous time series models for which the stationarity of the underlying process is conjectured but not yet proven. We give an approachable introduction to the use of drift criteria (also known as Lyapunov function techniques) for establishing strict stationarity and ergodicity of such models. These conditions immediately imply consistent estimation of the mean and lagged covariances, and more generally the expectation of any integrable function. We demonstrate by proving stationarity and ergodicity for several novel and useful examples, including Poisson log-link Generalized Autoregressive Moving Average models.

## 1 Introduction

Stationarity is a fundamental concept in time series modeling, capturing the idea that the future is expected to behave like the past; this assumption is inherent in any attempt to forecast the future. Many time series models are created by combining nonstationary effects such as trends, covariate effects, and seasonality with a stochastic process that is known or believed to be stationary. Alternatively, they can be defined by partial sums or other transformations of a stationary process. Consistency of statistical inferences for particular models is then established via the relationship to the stationary process; this includes consistency of parameter estimators and of standard error estimators (Brockwell and Davis 1991, Chap. 7-8).

Stationarity of a stochastic process also ensures the existence of a spectral representation. This means that one can perform frequency-domain analysis of time series data after removing trends and other nonstationary components (Brockwell and Davis 1991, Chap. 4; Box, Jenkins, and Reinsel 2008, Chap. 2).

However, (strict) stationarity can be nontrivial to establish, and many time series models currently in use are based on processes for which it has not been proven. Strict stationarity (henceforth, "stationarity") of a stochastic process $\{X_n : n \in \mathbb{Z}\}$ means that the distribution of the random vector $(X_n, X_{n+1} \ldots, X_{n+j})$ does not depend on $n$, for any $j \geq 0$ (Billingsley 1995, p.494). Sometimes weak stationarity (constant, finite first and second moments of $X_n$) is proven instead, or simulations are used to argue for stationarity.

One approach to establishing strict stationarity and ergodicity (defined as in Billingsley 1995, p.494) is based on application of Lyapunov function methods (also known as drift criteria) to a general state space Markov chain that is related to the time series model. Such a strong statement of stationarity is quite useful, since it immediately implies consistent estimation of the mean and lagged covariances of the process, and more generally the expectation of any integrable function (Billingsley 1995, p.495). Lyapunov function methods have been previously applied to prove stationarity and ergodicity for SETAR models by Chan and Tong (1985), for multivariate ARMA models by Bougerol and Picard (1992), for threshold AR-ARCH models by Cline and Pu (2004), and for integrated GARCH models by Liu (2009). Most of these papers use the Lyapunov exponent, which can give a sufficient and necessary condition for stationarity (Bougerol and Picard, 1992; Cline and Pu, 2004; Liu, 2009); however, this condition has a complex form that can be difficult to reduce to simple ranges on the parameter values.

The constructive approach used, e.g., in Chan and Tong (1985) and Fokianos et al. (2009), is much more straightforward to understand and apply. We give an approachable introduction to this method, and demonstrate it on a few novel and important examples, including Poisson log-link Generalized Autoregressive Moving Average (GARMA) models (Benjamin et al., 2003). These have been used for predicting numbers of births (Léon and Tsai, 1998), modeling poliomyelitis cases (Benjamin et al., 2003), and predicting valley fever incidence (Talamantes et al., 2007), among other applications. The main stationarity result that currently exists for GARMA models is weak stationarity in the case of an identity link function; unfortunately this excludes the most popular of these models (Benjamin et al., 2003). Zeger and Qaqish (1988) have also used a connection to branching processes to show stationarity and ergodicity for a special case of the purely autoregressive Poisson log-link model.

In Section 2 we describe the most common methods currently used for establishing stationarity of time series models. Then we describe Lyapunov function methods in Section 3. In Section 4 we show how to use these methods to prove stationarity of particular GARMA models, and in Section 5 we illustrate that drift conditions can also be used to prove stationarity of models that cannot be written in a linear form, even after transformation of the mean function.

# 2 Common Methods for Establishing Stationarity

The most common method for constructing a stationary and ergodic stochastic process is by transforming another known stationary, ergodic process. Denoting the known process by $\{\epsilon_n : n \in \mathbb{Z}\}$, for any measurable function $f$ the process $\{X_n : n \in \mathbb{Z}\}$ defined by

$$X_n = f(\ldots, \epsilon_{n-1}, \epsilon_n, \epsilon_{n+1}, \ldots) \tag{1}$$

is also stationary and ergodic (Billingsley 1995, Thm. 36.4). For instance, suppose $\epsilon_n$ is an iid sequence and that

$$X_n = \sum_{\ell=0}^{\infty} \alpha_\ell \epsilon_{n-\ell}.$$

Then $X_n$ is strictly stationary and ergodic if $\sum_{\ell=0}^{\infty} \alpha_\ell^2 < \infty$ (Taniguchi and Kakizawa 2000, Thm. 1.3.4). For linear time series models, such as the autoregressive (AR) model, the autocovariance function is a primary tool in specifying the sequence of constants $\{\alpha_\ell : \ell \geq 0\}$.

Some techniques for the analysis of linear time series models easily generalize for the analysis of nonlinear models; however, it is typically more difficult to verify stationarity of nonlinear models. Suppose

$$X_n = f(X_{n-1}, \epsilon_n)$$

where $\epsilon_n$ is iid and independent of $X_0$, and $f$ is continuous in its first argument. If $E|f(x,\epsilon_n) - f(y,\epsilon_n)| < |x-y|$ for $x,y \in \mathbb{R}, x \neq y$, and $Ef(x,\epsilon_n)^2 \leq \alpha x^2 + \beta$ for $x \in \mathbb{R}$, in which $0 \leq \alpha < 1$ and $0 \leq \beta$, then the Markov chain $X = \{X_n : n \in \mathbb{N}\}$ has a unique stationary distribution $\pi$ and if $X_0 \sim \pi$ then $X$ is strictly stationary and ergodic (Lasota and Mackey, 1989). This can be applied to a wide class of nonlinear models (Taniguchi and Kakizawa 2000, Sec. 3.2), but does not include many other models which are commonly used in practice, such as the generalized autoregressive conditionally heteroscedastic (GARCH) model (Bollerslev, 1986).

# 3 Drift Conditions for Establishing Stationarity

Another approach to establishing that certain time series models are stationary and ergodic is to use drift conditions to show that a closely related Markov chain $X = \{X_n : n \in \mathbb{N}\}$ has the same properties, and then to "lift" the result for the Markov chain back to the time series model. Often the mean process of the model is Markovian, in which case one can take $X$ to be this mean process. Showing that $X$ is $\varphi$-irreducible, aperiodic and positive (Harris) recurrent implies that $X$ has a unique stationary distribution $\pi$, and that if $X_0 \sim \pi$ then $X$ is a stationary and ergodic process defined on the nonnegative integers. If $X$ is the mean process, this can show stationarity and ergodicity of the time series model itself.

In this section we present the key definitions and results for the use of Lyapunov functions; see e.g. Meyn and Tweedie (1993) for more details. Let $S$ denote the state space of $X$, which will often be uncountable, and let $\mathcal{F}$ be an appropriate $\sigma$-field on $S$. Define $P^n(x,A) = P(X_n \in A | X_0 = x)$ to be the $n$-step transition probability starting from state $X_0 = x$. The appropriate notion of irreducibility when dealing with a general state space is that of $\varphi$-irreducibility, since general state space Markov chains may never visit the same point twice:

**Definition 1.** *We say that $X$ is* $\varphi$-irreducible *if there exists a nontrivial measure $\varphi$ on $\mathcal{F}$ such that, whenever $\varphi(A) > 0$, $P^n(x,A) > 0$ for some $n = n(x,A) \geq 1$, for all $x \in S$.*

The notion of aperiodicity in general state space chains is the same as that seen in countable state space chains, namely that one cannot decompose the state space into a finite partition of sets where the chain moves successively from one set to the next in sequence, with probability 1. For a more precise definition, see Meyn and Tweedie (1993), Sec. 5.4.

We need one more definition before we can present the basic result.

**Definition 2.** *A set $A \in \mathcal{F}$ is called a* small *set if there exists an $m \geq 1$, a probability measure $\nu$ on $\mathcal{F}$, and a $\lambda > 0$ such that for all $x \in A$ and all $C \in \mathcal{F}$, $P^m(x,C) \geq \lambda\nu(C)$.*

Small sets are a fundamental tool in the analysis of general state space Markov chains because, among other things, they allow one to apply regenerative arguments to the analysis of a chain's long-run behavior. Regenerative theory is indeed the fundamental tool behind the following result, which is a special case of Theorem 14.0.1 in Meyn and Tweedie (1993). Let $E_x f(X_1)$ denote the expectation of $f(X_1)$ for the chain $X$ with deterministic initial state $X_0 = x$.

**Theorem 1.** *Suppose that $X = \{X_n : n \in \mathbb{N}\}$ is $\varphi$-irreducible on $S$. Let $A \subset S$ be small, and suppose that there exists a function $V : S \to [0,\infty)$, $b \in (0,\infty)$ and $\epsilon > 0$ such that for all $x \in S$,*

$$E_x V(X_1) \leq V(x) - \epsilon + b\mathbf{1}_{\{x \in A\}}. \tag{2}$$

*Then $X$ is positive Harris recurrent.*

The function $V$ is called a Lyapunov function or energy function. The condition (2) is known as a drift condition, in that for $x \notin A$, the expected energy $V$ drifts towards zero by at least $\epsilon$. The indicator function in (2) asserts that from a state $x \in A$, any energy increase is bounded (in expectation).

Positive Harris recurrent chains possess a unique stationary probability distribution $\pi$. If $X_0$ is distributed according to $\pi$ then the chain $X$ is a stationary process. If the chain is also aperiodic then $X$ is ergodic, in which case if the chain is initialized according to some other distribution, then the distribution of $X_n$ will converge to $\pi$ as $n \to \infty$.

Hence, the drift condition (2), together with aperiodicity, establishes ergodicity. A stronger form of ergodicity, called geometric ergodicity, arises if (2) is replaced by the condition

$$E_x V(X_1) \leq \beta V(x) + bI(x \in A) \tag{3}$$

for some $\beta \in (0,1)$ and some $V : S \to [1,\infty)$. Indeed, (3) implies (2). Either of these criteria are sufficient for our purposes.

# 4 Generalized Autoregressive Moving Average Models

## 4.1 AR(1)

The simplest case of an autoregressive model is given by the AR(1) process, defined as $Y_n = \rho Y_{n-1} + \epsilon_n$ for $n \geq 1$, where $\rho \in (-1,1)$, and $\{\epsilon_n : n \geq 1\}$ are i.i.d. normal random variables with mean 0 and variance $\sigma^2$. The stationarity and ergodicity of the AR(1) process are well-known, but we use this as a simple example of the application of Lyapunov functions.

The underlying Markov chain can be taken as $\{X_n = Y_n : n \in \mathbb{N}\}$, which is easily seen to be $\varphi$-irreducible (take $\varphi$ to be uniform measure on $[0,1]$). The density of $\epsilon_n$ is positive everywhere, so the chain is aperiodic. We take $V(x) = |x|$, and $A = [-M, M]$ for some yet-to-be specified constant $M > 0$. The set $A$ is small: take $m = 1$, and $\nu$ to be the probability measure associated with a uniform distribution on any bounded interval. Now, for any $x \notin A$,

$$\begin{aligned}
E_x V(X_1) &= E|\rho x + \epsilon_1| \\
&\leq |\rho||x| + E|\epsilon_1| \\
&= V(x) - (1 - |\rho|)|x| + \sqrt{2\sigma^2/\pi} \\
&\leq V(x) - (1 - |\rho|)M + \sqrt{2\sigma^2/\pi},
\end{aligned}$$

so if we choose $M > 0$ large enough then the drift criterion (2) holds for $x \notin A$. For $x \in A$, $E_x V(X_1) = E|\rho x + \epsilon_1| \leq |\rho| M + \sqrt{2\sigma^2/\pi}$. Hence (2) holds for all $x$, and the chain $\{X_n : n \in \mathbb{N}\}$ is positive Harris recurrent and has a unique stationary distribution $\pi$. In fact, it is easy to verify that $\pi$ is normal with mean zero and variance $\sigma^2/(1 - \rho^2)$.

Assuming that $X_0 \sim \pi$, the chain $X$ (and also the time series model $Y$) is stationary and ergodic. We can therefore consistently estimate $\rho$ using the empirical one-step correlation $\hat{\rho}$ of $Y$, and can consistently estimate $\sigma^2$ using $(1 - \hat{\rho}^2)$ times the empirical variance of $Y_n$. One can also prove the consistency of more sophisticated estimators by showing that they approach these simple estimators $\hat{\rho}$ and $\hat{\sigma}^2$ in the limit.

## 4.2 A Poisson GARMA Model

Generalized Autoregressive Moving Average (GARMA) Models are a generalization of Autoregressive Moving Average Models to exponential-family distributions, allowing direct treatment of positive and count data, among others. GARMA models were stated in their most general form by Benjamin et al. (2003), based on earlier work by Zeger and Qaqish (1988) and Li (1994).

The most commonly used GARMA model for count data takes the following form for $n \in \mathbb{N}$:

$$Y_n|\mathbf{D}_n \sim \text{Poisson}(\mu_n) \tag{4}$$

$$\ln \mu_n = X_n'\beta + \sum_{j=1}^{p} \rho_j[\ln \max\{Y_{n-j}, c\} - X_{n-j}'\beta] + \sum_{j=1}^{q} \theta_j[\ln \max\{Y_{n-j}, c\} - \ln \mu_{n-j}] \tag{5}$$

where $c \in (0, 1)$ is chosen to prevent taking the logarithm of 0, where $X_n$ are the covariates at time $n$, and where

$$\mathbf{D}_n = (X_{n-p}, \ldots, X_n, Y_{n-\max\{p,q\}}, \ldots, Y_{n-1}, \mu_{n-q}, \ldots, \mu_{n-1})$$

are the present covariates and the past information (Benjamin et al., 2003). The second and third terms of (5) are the autoregressive and moving-average terms, respectively.

Since the covariates are time-dependent, the model for $\{Y_n : n \in \mathbb{N}\}$ is in general nonstationary, and interest is in proving stationarity in the absence of covariates. For simplicity we prove the case $p = 1$ and $q = 1$ here; the extension to $p > 1$ and $q > 1$ is sketched at the end of this section. Let $\rho = \rho_1$ and $\theta = \theta_1$, and denote the intercept by $\gamma$, yielding the model:

$$\ln \mu_n = \gamma + \rho[\ln \max\{Y_{n-1}, c\} - \gamma] + \theta[\ln \max\{Y_{n-1}, c\} - \ln \mu_{n-1}]. \tag{6}$$

We would like to prove that $\{Y_n : n \in \mathbb{N}\}$ is stationarity and ergodic by first proving these properties for $\{\mu_n : n \in \mathbb{N}\}$ (given a suitable initial distribution for $\mu_0$). However, establishing $\varphi$-irreducibility and finding a small set is seriously complicated by the fact that $\mu_n$ has discrete-valued random innovations. Given a particular value of $\mu_0$, the set of possible values for $\mu_1$ is countable. In fact, the set of states that are reachable from a fixed starting state is also countable, and distinct initial values can have distinct sets of reachable locations. Chains with these kinds of properties have been studied (see e.g. Meyn and Tweedie 1993, Chap. 6), but the key ideas we wish to convey become obscured because of the need for technical care and model-specific arguments.

A natural approach in the face of these difficulties is to study a (very) closely related process, where the innovations have a density with respect to Lebesgue measure. Specifically, we will show stationarity and ergodicity of the process defined by

$$\ln \mu_n = \gamma + \rho[\ln \max\{Y_{n-1}, c\} - \gamma] + \theta[\ln \max\{Y_{n-1}, c\} - \ln \mu_{n-1}] + Z_n \tag{7}$$

where $Z_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$ for some small $\sigma > 0$.

The practical impact of using (7) versus (6) is negligible, shown as follows. For any particular data sequence $\{Y_n : 0 \leq n \leq N\}$ the likelihood function using (6) is a product over $n$ of the Poisson density, i.e.

$$L = \prod_{n=0}^{N} e^{-\mu_n} \mu_n^{Y_n} / Y_n!$$

which is a continuous, differentiable function of $\ln \mu_n$ (for this likelihood to be well-defined we must assume some values for $\mu_{-1}$ and $Y_{-1}$). The likelihood obtained using (7) is the integral over the process $\{Z_n : n \geq 1\}$ of the same expression. Therefore we can choose $\sigma^2$ to be small enough that the difference between the likelihood calculated using (6) and the (marginal) likelihood calculated using (7) is arbitrarily small. The use of perturbations like $\{Z_n\}$ is explored in more detail in Fokianos et al. (2009).

With these definitions, we have

**Theorem 2.** *The Poisson log-link GARMA model given by (4) and (7) is stationary and ergodic, provided that $\rho \in (-1,1)$ and $\theta \in (-1,1)$, and assuming an appropriate initial distribution for $\mu_0$.*

To see this, define $X_n = \ln \mu_n$. We will show that the Markov chain $X = \{X_n : n \in \mathbb{N}\}$ is $\varphi$-irreducible, aperiodic, and positive Harris recurrent. These properties will imply that $X$ has a unique stationary distribution $\pi$, and that if $X_0 \sim \pi$ then $X$ is stationary and ergodic. Then $\{Y_n : n \in \mathbb{N}\}$ is also stationary and ergodic, seen as follows. Let $\{U_n : n \in \mathbb{N}\}$ be an iid sequence of uniform random variables that is independent of $X$; the bivariate process $\{(X_n, U_n) : n \in \mathbb{N}\}$ is stationary and ergodic. Notice that $Y_n$ depends on $X$ only through $X_n$ and $X_{n+1}$, so that one can generate it from its conditional distribution given those two values. This can be done by a deterministic function, i.e., $Y_n = g(X_n, X_{n+1}, U_n)$. Since this is a special case of (1), Thm. 36.4 of Billingsley (1995) shows that $\{Y_n : n \in \mathbb{N}\}$ is stationary and ergodic.

Next we show that the Markov chain $X = \{X_n : n \in \mathbb{N}\}$ is $\varphi$-irreducible, aperiodic, and positive Harris recurrent. The first two properties are automatic since the Markov transition kernel has a (normal mixture) density that is everywhere positive. Next, define $A = [-M, M]$ for some constant $M > 0$ to be chosen later; we will show that $A$ is small by taking $m = 1$ and $\nu$ to be the uniform distribution on $A$ in Definition 2. For $x \in A$,

$$P_x(Y_0 = 0) = e^{-e^x} \geq e^{-e^M}$$

so that with probability at least $e^{-e^M}$,

$$X_1 = (1-\rho)\gamma + (\rho + \theta)\ln c - \theta x + Z_1.$$

Noticing that $((1-\rho)\gamma + (\rho+\theta)\ln c - \theta x) \in [a_1(M), a_2(M)]$ for deterministic values $a_1(M)$ and $a_2(M)$, it is straightforward to show that $\exists \lambda > 0$ such that $P(x, \cdot) \geq \lambda\nu(\cdot)$ for all $x \in A$.

Next we use the small set $A$ to prove a drift condition. Taking the energy function $V(x) = |x|$, we have the following results. First we give the drift condition for $x \in A$:

**Proposition 3.** *There is some constant $K(M) < \infty$ such that $E_x V(X_1) \leq K(M)$ for all $x \in A$.*

Then we give the drift condition for $x \notin A$, handling the cases $x < -M$ and $x > M$ separately:

**Proposition 4.** *There is some constant $K_2 < \infty$ such that for $M$ large enough, $E_x V(X_1) \leq |\theta| V(x) + K_2$ for all $x < -M$.*

**Proposition 5.** *There is some constant $K_3 < \infty$ such that for $M$ large enough, $E_x V(X_1) \leq |\rho| V(x) + K_3$ for all $x > M$.*

Propositions 4 and 5 give the overall drift condition for $x \notin A$ as follows. Define $\delta = \max\{|\theta|, |\rho|\}$; for any $x \notin A$ we have

$$\begin{aligned} E_x V(X_1) &\leq \delta V(x) + \max\{K_2, K_3\} \\ &= V(x) - (1-\delta)V(x) + \max\{K_2, K_3\} \\ &\leq V(x) - (1-\delta)M + \max\{K_2, K_3\}. \end{aligned}$$

Taking $M$ large enough then establishes negative drift.

Proposition 4 shows that for very negative $X_{n-1}$, $|\theta|$ controls the rate of drift, while Proposition 5 shows that for large positive $X_{n-1}$, $|\rho|$ controls the rate of drift. The former result is due to the fact that for very negative values of $X_{n-1}$ the autoregressive term is a constant,

$(\rho \ln c - \rho\gamma)$, so the moving-average term dominates. The latter result is due to the fact that for large positive $X_{n-1}$, the distribution of $Y_{n-1}$ concentrates around $\mu_{n-1}$, so that the moving-average term $\theta(\ln \max\{Y_{n-1}, c\} - \ln \mu_{n-1})$ is negligible and the autoregressive term dominates.

Extension to the cases $p > 1$ and $q > 1$ can be achieved by showing stationarity and ergodicity of the multivariate Markov chain with state vector $(\mu_n, \mu_{n-1}, \ldots, \mu_{n-\max\{p,q\}+1})$. Again this is done by finding a small set and energy function such that a drift condition holds, subject to appropriate restrictions on the parameters $(\rho_1, \ldots, \rho_p)$ and $(\theta_1, \ldots, \theta_q)$.

### 4.2.1 Proof of Prop. 3 and Prop. 4

We can bound $E_x V(X_1)$ as follows, where $\mathbf{1}_B$ is the indicator of the event $B$:

$$
\begin{aligned}
E_x V(X_1) &= E_x|(1-\rho)\gamma + (\rho+\theta)\ln\max\{Y_0, c\} - \theta x + Z_1| \\
&\leq (1-\rho)|\gamma| + |\rho| E_x|\ln\max\{Y_0, c\}| + |\theta| E_x|\ln\max\{Y_0, c\} - x| + E_x|Z_1| \\
&= (1-\rho)|\gamma| + |\rho| P_x(Y_0 = 0)|\ln c| + |\rho| E_x[(\ln Y_0)\mathbf{1}_{Y_0 \geq 1}] + \\
&\quad |\theta| P_x(Y_0 = 0)|\ln c - x| + |\theta| E_x[|\ln Y_0 - x|\mathbf{1}_{Y_0 \geq 1}] + \sqrt{2\sigma^2/\pi} \qquad (8) \\
&\leq (1-\rho)|\gamma| + |\rho||\ln c| + (|\rho| + |\theta|) E_x[(\ln Y_0)\mathbf{1}_{Y_0 \geq 1}] + \\
&\quad |\theta| P_x(Y_0 = 0)|\ln c - x| + |\theta| P_x(Y_0 \geq 1)|x| + \sqrt{2\sigma^2/\pi} \\
&\leq (1-\rho)|\gamma| + |\rho||\ln c| + (|\rho| + |\theta|) E_x Y_0 + |\theta| P_x(Y_0 = 0)|\ln c| + |\theta||x| + \sqrt{2\sigma^2/\pi} \\
&\leq (1-\rho)|\gamma| + |\rho||\ln c| + (|\rho| + |\theta|) e^x + |\theta||\ln c| + |\theta||x| + \sqrt{2\sigma^2/\pi}. \qquad (9)
\end{aligned}
$$

For $x \in A$ this is bounded above by some constant $K(M) < \infty$, proving Prop. 3. For $x < -M$ we have $e^x < 1$, so that (9) proves Prop. 4.

### 4.2.2 Proof of Prop. 5

We will show that for $x > M$ where $M$ is large, the autoregressive part of the GARMA model dominates and the moving-average portion of the model is negligible. In the bound (8), the autoregressive part of the model is captured by $|\rho| E_x[(\ln Y_0)\mathbf{1}_{Y_0 \geq 1}]$, while the moving-average part of the model corresponds to the terms $|\theta| P_x(Y_0 = 0)|\ln c - x|$ and $|\theta| E_x[|\ln Y_0 - x|\mathbf{1}_{Y_0 \geq 1}]$. For large positive $x$ the former is small because $P_x(Y_0 = 0)$ is small, and we will show that the latter is small by bounding the tail probabilities of $(\ln Y_0 - x)$.

By Jensen's inequality, for any $x > M$

$$
E_x[(\ln Y_0)\mathbf{1}_{Y_0 \geq 1}] \;\leq\; \ln E_x[Y_0 \mathbf{1}_{Y_0 \geq 1}] \;=\; \ln E_x Y_0 \;=\; x \;=\; V(x). \qquad (10)
$$

Also, we can make $M$ large enough that for any $\epsilon > 0$,

$$
\begin{aligned}
\sup_{x > M} P_x(Y_0 = 0)|\ln c - x| &= \sup_{x > M} e^{-e^x}|\ln c - x| \\
&= \sup_{x > M} e^{-e^x}(|\ln c| + x) \\
&= e^{-e^M}(|\ln c| + M) < \epsilon. \qquad (11)
\end{aligned}
$$

We also have the following, using Markov's inequality and Chebyshev's inequality and taking

$M > 1$:

$$
\begin{aligned}
E_x[|\ln Y_0 - x|\mathbf{1}_{Y_0\geq 1}] &= \int_0^\infty P_x(\ln Y_0 - x > y, Y_0 \geq 1)dy + \int_0^\infty P_x(x - \ln Y_0 > y, Y_0 \geq 1)\, dy \\
&= \int_0^\infty P_x(Y_0 > e^{x+y})\, dy + \int_0^\infty P_x(1 \leq Y_0 < e^{x-y})\, dy \\
&\leq 2\epsilon + \int_\epsilon^\infty P_x(Y_0 > e^{x+y})\, dy + \int_\epsilon^x P_x(Y_0 < e^{x-y})\, dy \\
&\leq 2\epsilon + \int_\epsilon^\infty E_x Y_0/e^{x+y}\, dy + \int_\epsilon^x P_x(|Y_0 - e^x| > e^x - e^{x-y})\, dy \\
&\leq 2\epsilon + \int_\epsilon^\infty e^{-y}\, dy + \int_\epsilon^x e^x/(e^x - e^{x-y})^2\, dy \\
&= 2\epsilon + e^{-\epsilon} + \int_\epsilon^x e^{-x}(1-e^{-y})^{-2}\, dy \\
&\leq 2\epsilon + e^{-\epsilon} + \int_\epsilon^x e^{-x}(1-e^{-\epsilon})^{-2}\, dy \;\;\leq\;\; 2\epsilon + e^{-\epsilon} + xe^{-x}(1-e^{-\epsilon})^{-2} \\
&\leq 2\epsilon + e^{-\epsilon} + Me^{-M}(1-e^{-\epsilon})^{-2} \;\;\leq\;\; 2\epsilon + e^{-\epsilon} + (1-e^{-\epsilon})^{-2}.
\end{aligned}
$$

Combining this with (8), (10), and (11), we have that for $x > M$,

$$
E_x V(X_1) \;\leq\; (1-\rho)|\gamma| + |\rho||\ln c| + |\rho|V(x) + |\theta|\left(3\epsilon + e^{-\epsilon} + (1-e^{-\epsilon})^{-2}\right) + \sqrt{2\sigma^2/\pi}
$$

showing the desired result.

# 5 Nonlinear Models

Next we illustrate the utility of drift conditions in showing stationarity of models that cannot be expressed in a linear form, even after transformation of the mean function.

## 5.1 A Poisson Exponential Autoregressive Model

Consider a first order conditionally Poisson exponential autoregressive model with identity link function, defined as

$$
\begin{aligned}
Y_n | Y_{n-1}, \mu_{n-1}, Z_n &\sim \text{Poisson}(\mu_n), \quad Z_n \overset{\text{iid}}{\sim} \text{Uniform}(0,\epsilon), \\
\mu_n &= [\omega + \beta \exp(-\gamma \mu_{n-1}^2)]\mu_{n-1} + \alpha Y_{n-1} + Z_n,
\end{aligned}
$$

in which $\omega, \beta, \gamma, \alpha > 0$ (Fokianos et al., 2009). Exponential autoregressive models are attractive in modeling because of their threshold-like behavior: for large $\mu_{n-1}$, the functional coefficient for $\mu_{n-1}$ is approximately $\omega$, and for small $\mu_{n-1}$ it is approximately $\omega + \beta$; however, the transition between these regimes is smooth. This model may be considered a nonlinear generalization of what Ferland et al. (2006) and others have referred to as Integer-GARCH. The auxiliary innovations $Z_n$, for which $\epsilon > 0$ is chosen arbitrarily small, are introduced to simplify verification of the $\varphi$-irreducibility condition, and can be justified as in Section 4.2. As in Fokianos et al. (2009), for $\omega + \alpha < 1$ one can verify a drift condition on $\{\mu_n : n \in \mathbb{N}\}$. This shows that, with an appropriate initial distribution for $\mu_0$, the process $\{Y_n : n \in \mathbb{N}\}$ is stationarity and ergodic.

## 5.2 A Poisson Threshold Model

Finally we consider a first order conditionally Poisson threshold model with identity link function that we have found useful in our own applications (Matteson et al., 2010). We again use a perturbed version of the model to simplify verification of the $\varphi$-irreducibility condition, defining the model as

$$\begin{aligned} Y_n|Y_{n-1},\mu_{n-1},Z_n &\sim \text{Poisson}(\mu_n), \quad Z_n \overset{\text{iid}}{\sim} \text{Uniform}(0,\epsilon) \\ \mu_n &= \omega + \alpha Y_{n-1} + \beta\mu_{n-1} + (\gamma Y_{n-1} + \eta\mu_{n-1})\,\mathbf{1}_{\{Y_{n-1}\notin(L,U)\}} + Z_n \end{aligned}$$

where the threshold boundaries satisfy $0 < L < U < \infty$. To ensure positivity of $\mu_n$ we assume $\omega, \alpha, \beta > 0$, $(\alpha+\gamma) > 0$, and $(\beta+\eta) > 0$. Additionally we take $\eta \leq 0$ and $\gamma \geq 0$, so that when $Y_{n-1}$ is outside the range $(L,U)$ the mean process $\mu_n$ is more adaptive, i.e. puts more weight on $Y_{n-1}$ and less on $\mu_{n-1}$.

We will show that $\{Y_n : n \in \mathbb{N}\}$ is stationarity and ergodic under the restriction $(\alpha + \beta + \gamma + \eta) < 1$. This can be proven via extension of results in Fokianos et al. (2009) for the non-threshold linear model. However, a much simpler proof is as follows, where $X_n = \mu_n$. Take the state space of the Markov chain $X = \{X_n : n \in \mathbb{N}\}$ to be $S = [\frac{\omega}{1-\beta-\eta}, \infty)$. Define $A = [\frac{\omega}{1-\beta-\eta}, \frac{\omega}{1-\beta-\eta} + M]$ for any $M > 0$, and define $m$ to be the smallest positive integer such that $M(\beta+\eta)^{m-1} < \epsilon/2$. Then

$$\begin{aligned} &\inf_{x\in A} \Pr(Y_0 = Y_1 = \ldots = Y_{m-2} = 0|X_0 = x) > 0 \qquad \text{and} \\ &\Pr\left(Z_1 + Z_2 + \ldots + Z_{m-1} < \frac{\epsilon}{2} - M(\beta+\eta)^{m-1}\right) > 0. \end{aligned}$$

Therefore $\inf_{x\in A} P^{m-1}(x,B) > 0$, where $B = [\frac{\omega}{1-\beta-\eta}, \frac{\omega}{1-\beta-\eta} + \frac{\epsilon}{2}]$. Taking $\nu = \text{Unif}(\frac{\omega}{1-\beta-\eta} + \frac{\epsilon}{2}, \frac{\omega}{1-\beta-\eta} + \epsilon)$ in Definition 2 then establishes $A$ as a small set. A similar argument can be used to show $\varphi$-irreducibility and aperiodicity.

Taking the energy function $V(x) = x$,

$$\begin{aligned} E_x V(X_1) &= (\alpha+\beta)V(x) + \gamma E_x[Y_{n-1}\mathbf{1}_{\{Y_{n-1}\notin(L,U)\}}] + \eta x P_x[Y_{n-1} \notin (L,U)] + (\omega + \epsilon/2) \\ &\leq (\alpha+\beta+\gamma)V(x) + \eta x - \eta x P_x[Y_{n-1} \in (L,U)] + (\omega + \epsilon/2). \end{aligned}$$

As $x \to \infty$ we have $xP_x[Y_{n-1} \in (L,U)] \to 0$, so for sufficiently large $M$, $x > M$ implies that $-\eta x P_x[Y_{n-1} \in (L,U)] \leq \frac{\epsilon}{2}$. Thus

$$E_x V(X_1) \leq (\alpha+\beta+\gamma+\eta)V(x) + (\omega+\epsilon).$$

This is bounded for $x \in A$, and taking $M$ large enough this has negative drift for $x \notin A$. Therefore, assuming an appropriate initial distribution for $\mu_0$, the chain $X$ is stationary and ergodic. As shown in Section 4.2, this implies that the time series model $\{Y_n : n \in \mathbb{N}\}$ is also stationary and ergodic.

# Acknowledgments

This work was supported in part by National Science Foundation Grant Number CMMI-0926814.

# References

Benjamin, M. A., Rigby, R. A., and Stasinopoulos, D. M. (2003), "Generalized autoregressive moving average models," *Journal of the American Statistical Association*, 98, 214–223.

Billingsley, P. (1995), *Probability and Measure*, 3rd edn, New York: Wiley.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bougerol, P., and Picard, N. (1992), "Strict stationarity of generalized autoregressive processes," *Annals of Probability*, 20, 1714–1730.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008), *Time Series Analysis: Forecasting and Control*, 4th edn, New York: Wiley.

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd edn, New York: Springer-Verlag.

Chan, K. S., and Tong, H. (1985), "On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations," *Advances in Applied Probability*, 17, 666–678.

Cline, D. B. H., and Pu, H. H. (2004), "Stability and the Lyapunov exponent of threshold AR-ARCH models," *Annals of Applied Probability*, 14, 1920–1949.

Ferland, R., Latour, A., and Oraichi, D. (2006), "Integer-valued GARCH process," *Journal of Time Series Analysis*, 27, 923–942.

Fokianos, K., Rahbek, A., and Tjostheim, D. (2009), "Poisson autoregression," *Journal of the American Statistical Association*, 104, 1430–1439.

Lasota, A., and Mackey, M. C. (1989), "Stochastic perturbation of dynamical systems: the weak convergence of measures," *Journal of Mathematical Analysis and Applications*, 138, 232–248.

Léon, L. F., and Tsai, C. (1998), "Assessment of model adequacy for Markov regression time series models," *Biometrics*, 54, 1165–1175.

Li, W. K. (1994), "Time series models based on Generalized Linear Models: some further results," *Biometrics*, 50, 506–511.

Liu, J. (2009), "Stationarity of a family of GARCH processes," *Econometrics Journal*, 12, 436–446.

Matteson, D. S., Woodard, D. B., and Henderson, S. G. (2010), Time series models for predicting ambulance demand,. In preparation.

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

Talamantes, J., Behseta, S., and Zender, C. S. (2007), "Statistical modeling of valley fever data in Kern County, California," *International Journal of Biometeorology*, 51, 307–313.

Taniguchi, M., and Kakizawa, Y. (2000), *Asymptotic Theory of Statistical Inference for Time Series*, New York: Springer-Verlag.

Zeger, S. L., and Qaqish, B. (1988), "Markov regression models for time series: a quasi-likelihood approach," *Biometrics*, 44, 1019–1031.