Functional-Coefficient Autoregressive Models

RONG CHEN and RUEY S. TSAY*

In this article we propose a new class of models for nonlinear time series analysis, investigate properties of the proposed model, and suggest a modeling procedure for building such a model. The proposed modeling procedure makes use of ideas from both parametric and nonparametric statistics. A consistency result is given to support the procedure. For illustration we apply the proposed model and procedure to several data sets and show that the resulting models substantially improve postsample multi-step ahead forecasts over other models.

KEY WORDS: Arranged local regression; Consistency; Forecasting; Geometrical ergodicity; Nonlinear time series; Threshold autoregressive model.

Nonlinear time series analysis has gained much attention recently, due primarily to the fact that linear models such as the autoregressive moving average (ARMA) models of Box and Jenkins (1976) fail (a) to capture many nonlinear features of the processes commonly encountered in practice, (b) to provide a parsimonious model to describe nontrivial data, and (c) to make accurate multi-step ahead predictions. Because linearity is at best a convenient artifact, and because the world is full of nonlinear phenomena such as limit cycles and jump resonance, we need to study nonlinearity to explore the nature.

An immediate problem of departing from linearity is to formulate a class of well-parameterized nonlinear models that are simple yet sufficient in handling most nonlinear phenomena observed in practice. This is a tough problem to solve, however, because there is no unified theory applicable to all nonlinear models. The main difficulty is that unlike linear models where the functions involved can be treated fairly systematically, nonlinear models have too many possibilities. There are not only so many different types of nonlinear functions, but also so many different structures within a given class of functions. The common research in nonlinear time series analysis has focused on several classes of models, such as the threshold autoregressive (TAR) model of Tong (1983, 1990) and the exponential autoregressive (EXPAR) model of Haggan and Ozaki (1981).

In this article we are concerned with empirical modeling of nonlinear time series. In particular we focus on exploring the nonlinear feature of a time series in the process of model building. This is achieved by generalizing directly the linear autoregressive (AR) models and exploiting local characteristics of a given time series. The generalized model is referred to as the functional coefficient autoregressive (FAR) models. Most nonlinear AR models considered in the literature are special cases of the FAR model. It turns out that the FAR models are flexible enough to accommodate most nonlinear features considered in the literature and are simple enough to be treated relatively easily.

This article is organized as follows. Section 1 introduces the proposed model and some of its probabilistic properties. Section 2 proposes a procedure for building such a model. A consistency result of the proposed procedure is also given. Section 3 applies the modeling procedure to several examples to illustrate practical uses of the procedure and to demonstrate the advantages of using FAR models for multi-step ahead forecasts.

1. THE MODEL AND ITS PROBABILISTIC PROPERTIES

1.1 Definition

A time series x_t is said to follow a FAR model if it satisfies

$$x_t = f_1(\mathbf{X}_{t-1}^*)x_{t-1} + \cdots + f_p(\mathbf{X}_{t-1}^*)x_{t-p} + \varepsilon_t, \qquad (1)$$

where p is a positive integer, $\{\varepsilon_t\}$ is a sequence of iid random variables with mean 0 and variance σ^2 such that ε_t is independent of x_{t-i} for $i > 0$, $f_i(\mathbf{X}_{t-1}^*)$'s are measurable functions from $\Re^k$ to $\Re$, and

$$\mathbf{X}_{t-1}^* = (x_{t-i_1}, x_{t-i_2}, \ldots, x_{t-i_k})' \quad \text{with} \quad i_j > 0 \quad \text{for} \quad j = 1, \ldots, k. \qquad (2)$$

Here $\mathbf{X}_{t-1}^*$ is a threshold vector with $i_1, \ldots, i_k$ as the threshold lags (or delay parameters), and x_{t-i_j}'s are referred to as the threshold variables. Without loss of generality we may assume $\max(i_1, \ldots, i_k) \le p$ in studying properties of the FAR model; otherwise, we can simply extend the representation in (1) by adding some zero coefficient functions.

Model (1) is a special case of the state-dependent model of Priestley (1980), and hence it enjoys all the nice properties of a state-dependent model. But without any white-noise term ε_{t-i} in the threshold vector in (2) and with the simple AR structure in (1), empirical identification of such a process is much easier than the general state-dependent model.

This definition of a FAR model is very general. In application we expect that only simple models with a small number of threshold variables, say one or two, and a low AR order are used. From (1) it is easily seen that many existing nonlinear AR models are special cases of the FAR model. For example, with $f_i(\mathbf{X}_{t-1}^*) = a_i + b_i\exp(-c_i x_{t-d}^2)$, (1) reduces to an EXPAR model; with $f_i(\mathbf{X}_{t-1}^*) = \phi_i^{(1)} I(x_{t-d} \le c) + \phi_i^{(2)} I(x_{t-d} > c)$, where $I(\cdot)$ is the usual indicator function,

* Rong Chen is Assistant Professor, Department of Statistics, Texas A&M University, College Station 77843. Ruey S. Tsay is Professor of Statistics, Graduate School of Business, University of Chicago, IL 60637. This research was supported in part by National Science Foundation Grants DMS89-02177 and DMS91-03250 and by the Graduate School of Business, University of Chicago. The authors thank George Sugilara for providing the New York chickenpox data and an associate editor and two referees for providing helpful comments.

Journal of the American Statistical Association
March 1993, Vol. 88, No. 421, Theory and Methods

(1) reduces to a TAR model. Thus the proposed FAR model is quite general and flexible.

1.2 Some Probabilistic Properties of the FAR Model

In this subsection we study some probabilistic properties of the proposed FAR model. We use results of Tweedie (1975) and of Tjøstheim (1990).

Lemma 1.1 (Tweedie 1975). Let $\{\mathbf{X}_t\}$ be a ϕ-irreducible Markov chain on a normed topological space with any norm $\| \ \|$. If the transition probability $\Pr(x, \cdot)$ is strongly continuous, then the conditions for geometrical ergodicity are met if there exist a compact set $\mathbf{K}$ and a constant ρ between 0 and 1 such that

$$E(\|\mathbf{X}_{t+1}\| \mid \mathbf{X}_t = \mathbf{X}) < \infty \qquad \text{if } \mathbf{X} \in \mathbf{K}$$
$$< \rho \|\mathbf{X}\| \qquad \text{otherwise.}$$

In Lemma 1.1 ϕ is a measure on the probability space of $\mathbf{X}_t$. The idea here is that when the process goes far away from the origin, the expectation of the norm of the next state vector will be shrunken toward the origin under the assumed condition.

Lemma 1.2 (Tjøstheim 1990). Let x_t be an aperiodic Markov chain and let h be a fixed positive integer. Then

$$\{x_{th}\} \text{ is } \begin{cases} \text{recurrent} \\ \text{geometrically ergodic} \\ \text{transient} \end{cases} \Rightarrow \{x_t\} \text{ is } \begin{cases} \text{recurrent} \\ \text{geometrically ergodic} \\ \text{transient,} \end{cases}$$

where $\{x_{th}\}$ is the h step subsequence of $\{x_t\}$.

By Lemmas 1.1 and 1.2, we obtain the following theorems.

Theorem 1.1. Assume that the functions $f_i(\cdot)$ in (1) are bounded such that $|f_i(\cdot)| \le c_i$ and the density function of ε_t is positive everywhere on the real line $\Re^1$. If all the roots of the characteristic function

$$\lambda^p - c_1\lambda^{p-1} - \cdots - c_p = 0 \tag{3}$$

are inside the unit circle, then the FAR process in (1) is geometrically ergodic.

Theorem 1.2. Assume that the functions $f_i(\cdot)$ in (1) can be written as $f_i(\mathbf{X}) = g_i(\mathbf{X}) + h_i(\mathbf{X})$ such that $g_i(\cdot)$ and $h_i(\cdot)$ are bounded functions with $|g_i(\cdot)| < c_i$ and $h_i(\mathbf{X})x_i$ is uniformly bounded. Then if ε_t has positive support everywhere on the real line $\Re^1$ and if all the roots of the characteristic function $\lambda^p - c_1\lambda^{p-1} - \cdots - c_p = 0$ are inside the unit circle, the FAR process in (1) is geometrically ergodic.

Proofs of these two theorems are in the Appendix. Here we briefly discuss their implications in nonlinear time series analysis. First, the theorems only provide sufficient conditions for geometrical ergodicity. This is most clear by considering the linear AR model for which $f_i(\cdot) = \phi_i$ and the condition of the theorems requires that all of the roots of $\lambda^p - |\phi_1|\lambda^{p-1} - \cdots - |\phi_p| = 0$ are inside the unit circle, which obviously is not a necessary condition. Second, for some special FAR models, weaker conditions—even the necessary and sufficient condition—of geometrical ergodicity are available. For example, Petruccelli and Woolford (1984) and Chen and Tsay (1991) derived the necessary and sufficient condition of geometrical ergodicity for TAR(1) models with two regimes. Third, the ergodic region in the parameter space of x_t implied by the two theorems is often bounded. On the other hand, many FAR models, such as the TAR(1) models mentioned earlier, have unbounded ergodic regions.

1.3 Some Special Cases and Their Ergodic Conditions

To illustrate that the conditions of Theorems 1.1 and 1.2 are not too restrictive and to show that they are convenient tools for checking the geometrical ergodicity of some complicated nonlinear time series models, we consider some special FAR models.

Example 1. Consider the linear TAR process

$$x_t = \phi_1^{(i)}x_{t-1} + \cdots + \phi_p^{(i)}x_{t-p} + \varepsilon_t^{(i)}$$
$$\text{if } x_{t-d} \in \Omega_i, \qquad i = 1, \ldots, k,$$

where Ω_i's are nonoverlapping intervals on the real line such that $\cup_{i=1}^k \Omega_i = \Re$. Letting $f_j(\mathbf{X}_{t-1}^*) = \phi_j^{(i)}$ if $x_{t-d} \in \Omega_i$, we see that this TAR process is a special case of FAR models. By Theorem 1.1, the process is geometrically ergodic if all the roots of $\lambda^p - c_1\lambda^{p-1} - \cdots - c_p = 0$ are inside the unit circle, where $c_j = \max\{|\phi_j^{(1)}|, \ldots, |\phi_j^{(k)}|\}$. This is the result of Chan and Tong (1985) derived via the Lyapunov function in the nonlinear dynamical literature.

Example 2. Suppose that x_t follows the EXPAR model

$$x_t = [a_1 + b_1\exp(-c_1x_{t-d}^2)]x_{t-1} + \cdots$$
$$+ [a_p + b_p\exp(-c_px_{t-d}^2)]x_{t-p} + \varepsilon_t,$$

where $c_i \ge 0$ for $i = 1, \ldots, p$. By Theorem 1.1, the process is geometrically ergodic if all the roots of $\lambda^p - d_1\lambda^{p-1} - \cdots - d_p = 0$ are inside the unit circle where $d_i = \max\{|a_i|, |a_i + b_i|\}$. In particular, if $p = 1$, then the condition becomes $|a_1| < 1$ and $|a_1 + b_1| < 1$. Further, a weaker condition $|a_1| < 1$ is obtained by Theorem 1.2 if $d = p = 1$. Next, for $p = 2$ Theorem 1.1 says that if $|a_1| < 1 - |a_2|$; $|a_2| < 1$; $|a_1 + b_1| < 1 - |a_2 + b_2|$; $|a_2 + b_2| < 1$; $|a_1 + b_1| < 1 - |a_2|$; $|a_1| < 1 - |a_2 + b_2|$, then x_t is geometrically ergodic.

Example 3. Finally, consider the process

$$x_t = [a_1 + b_1\sin(\phi_1x_{t-d})]x_{t-1} + \cdots$$
$$+ [a_p + b_p\sin(\phi_px_{t-d})]x_{t-p} + \varepsilon_t.$$

For $p = 1$, Theorem 1.1 shows that if $\max\{|a_1 + b_1|, |a_1 - b_1|\} < 1$, then x_t is geometrically ergodic. For $p = 2$, Theorem 1.1 shows that a sufficient condition of ergodicity of x_t is $c_1 + c_2 < 1$, where $c_i = \max\{|a_i + b_i|, |a_i - b_i|\}$.

2. AN IDENTIFICATION PROCEDURE FOR FAR MODELS

The main difficulty in using the proposed FAR model in (1) is specifying the functional coefficients $f_i(\cdot)$. We consider an arranged local regression (ALR) to explore the local behavior of the AR coefficients so that tentative specifications of the functional forms of $f_i(\cdot)$ can be made. The basic idea of ALR is to generalize the technique of arranged autoregression of Tsay (1989) by incorporating various concepts of nonparametric regression such as the method of nearest neighbors to reveal the shape of $f_i(\cdot)$. By "local" we mean using a certain window so that data points close together in the threshold space $\mathbf{X}^*_{t-1}$ can be used to obtain a local estimate of $f_i(\cdot)$. Such an estimate provides information on the shape of $f_i(\cdot)$ on which the functional form can be specified. A consistency result is given to support practical use of the proposed ALR procedure. We then suggest a modeling procedure for building FAR models.

2.1 Arranged Local Regression

For simplicity we consider only the case $\mathbf{X}^*_{t-1} = x_{t-d}$ (i.e., a single threshold variable) in this article. Extensions of the procedure to the case of vector $\mathbf{X}^*_{t-1}$ are possible provided that a systematic way is used to partition the $\mathbf{X}^*_{t-1}$ space and some high-dimensional graphical display is used.

Let $\{x_1, x_2, \ldots, x_N\}$ be N consecutive observations of the FAR model

$$x_t = f_1(x_{t-d})x_{t-1} + \cdots + f_p(x_{t-d})x_{t-p} + \varepsilon_t. \qquad (4)$$

The main objective of model specification is to infer the functional forms of $f_i(\cdot)$ from the data; that is, to estimate the function $f_i(\cdot)$ at various values of x_{t-d}. To this end we consider a simple procedure. Suppose that we are interested in estimating $f_i(x)$ and the data contain a sufficient number of observations, say $x_{t_1}, x_{t_2}, \ldots, x_{t_k}$, such that $x_{t_i-d} = x$ for $i = 1, \ldots, k$. In other words, there are k observations for which the threshold variable x_{t-d} assumes the value x. Then the linear regression $x_t = a_1x_{t-1} + \cdots + a_px_{t-p} + \varepsilon_t$ with $t = t_1, t_2, \ldots, t_k$ and $a_i = f_i(x)$ can be used, and the ordinary least squares (OLS) estimate of a_i is an estimate of $f_i(x)$. Denote such an estimate of $f_i(x)$ by $\hat{f}_i(x)$. By plotting $\hat{f}_i(x)$ versus x, we obtain an estimate of the functional form of $f_i(\cdot)$.

In applications such an ideal situation hardly ever happens. But the idea of using data points for which the threshold variable x_{t-d} assumes values close to x is still helpful, because under the assumption that the functions $f_i(\cdot)$ are sufficiently smooth, those data points continue to provide useful information on $f_i(x)$. Basically, for a well-behaved function $f_i(x_{t-d})$, if $x_{t-d} = x + \Delta_t$ with Δ_t a small real number, we have $f_i(x_{t-d}) = f_i(x) + O(\Delta_t)$ so that

$$\begin{aligned} x_t &= f_1(x_{t-d})x_{t-1} + \cdots + f_p(x_{t-d})x_{t-p} + \varepsilon_t \\ &= f_1(x)x_{t-1} + \cdots + f_p(x)x_{t-p} + \varepsilon_t + O(\Delta_t). \end{aligned}$$

For small Δ_t the last term $O(\Delta_t)$ is small and can be ignored at the stage of tentative model specification. Alternatively, we can incorporate $O(\Delta_t)$ into ε_t, resulting in a nonhomogeneous linear regression for which the OLS estimates are consistent for the coefficients but inefficient. Such consistent estimates are helpful, because the objective here is simply to infer the functional form of $f_i(\cdot)$.

The preceding discussion leads us to consider an ALR for tentative model specification of FAR models. For the data $x_1, \ldots, x_N$, the threshold variable x_{t-d} can assume the values $x_1, \ldots, x_{N-d}$. Let $x_{(i)}$ be the ith order statistic of $\{x_1, \ldots, x_{N-d}\}$. Denote the time index of $x_{(i)}$ by t_i. The proposed ALR is as follows:

1. Select an interval length c to form a window and a minimum sample size K to control the number of observations in the window.
2. Form the window $[x_{(1)}, x_{(1)} + c]$. Suppose that $x_{(1)}, \ldots, x_{(k)}$ are in this window. Initialize the estimation procedure by fitting the linear regression

$$x_{t+d} = a_1x_{t+d-1} + \cdots + a_px_{t+d-p} + e_{t+d}$$
$$\text{with} \quad t = t_1, \ldots, t_k. \qquad (5)$$

Treat the OLS estimate of a_i as an estimate of $f_i(x_{(1)} + c)$ provided that $k > K$.

3. Move the window along the x_{t-d} axis until either a new data point enters the window or a point drops out of the window. If the sample size in the new window is greater than or equal to K, then fit a linear regression similar to the one in (5) to obtain an estimate $\hat{f}_i(x_{(j)} + c)$ of $f_i(x_{(j)} + c)$, where $x_{(j)} + c$ denotes the right end point of the window used. On the other hand, if the sample size in the window is less than K, then continue to move the window forward.
4. Repeat Step 3 until $x_{(N-d)}$ is processed.
5. Make scatterplots of the estimates $\hat{f}_i(x)$ versus x and specify the functional forms of $f_i(\cdot)$ from the plots.

Some remarks on the ALR procedure are in order:

- Both the interval length c and the minimum sample size K are needed to obtain informative estimates of $f_i(\cdot)$. Using the minimum sample size K alone to define the moving window may lead to some numerical difficulties, because K cannot control the design matrix of the linear regression used. For example, a cluster of observations around a particular point of x_{t-d} can make the $\mathbf{X}'\mathbf{X}$ matrix of the linear regression used close to being singular. Similarly, using the interval length c alone can also lead to numerical difficulties. In an application the choices of c and K typically depend on the sample size N, the order p, and the functions $f_i(\cdot)$. In this article we choose K to be a fraction of the sample size N and c to be a fraction of the sample range of x_t; that is, $c = [x_{\max} - x_{\min}]/h$, where h is a positive integer and $x_{\max}$ and $x_{\min}$ are the maximum and minimum of the data.

  The choices of c and K are similar in spirit to the window width selection in nonlinear smoothing techniques. With a large window width c, the estimated functions are smoother, but the biases of the estimates tend to increase. A large K provides more stable estimates but reduces the number of points at which estimates are computed, resulting in less informative scatterplots. Optimal selection of these two parameters deserves further investigation.

  Because the ALR is a fast algorithm, however, we suggest that several values of c and K be used in an application. Our limited experience shows that by examining the scatterplots of various choices of c and K, one can refine the choices. Usually a reasonable selection would result in the scatterplots showing certain continuous functions. Because the objective of the ALR procedure is to provide tentative specifications of some parametric models, the trial-and-error method works reasonably well, as shown by the examples in Section 3.

- We treat the OLS estimate of a_i in the linear regression (5) as an estimate of $f_i(x_{(1)} + c)$. In fact the OLS estimate can be regarded as an estimate of $f_i(x)$ for $x \in [x_{(1)}, x_{(1)} + c]$. In practice this may affect the tentative specification of the threshold values in reading the scatterplots of $\hat{f}_i(x)$ versus x. We use the right end point of the window involved mainly for the ease in handling TAR types of models.
- The standard errors of the OLS estimates of the linear regression used also should be computed. These standard errors are useful in reading the scatterplots of $\hat{f}_i(x)$ versus x. See Section 3 for the use of the estimated standard errors.
- In the linear regression (5) we do not use a constant term mainly because when $d \le p$ the term $f_d(x_t)x_t$ is close to a constant within a given window. (Recall that the window is designed to use data points with thresholds in the vicinity of x_t.) Thus adding a constant term in this case may result in some numerical problem. Such a numerical problem will not occur if $d > p$.

2.2 A Recursive Method

From Step 3 of the proposed ALR procedure, estimates of ALR can be computed recursively. Furthermore, it suffices to consider the recursion of adding or deleting a single data point in a linear regression.

Let ϕ_i be the vector of OLS estimates at the current stage in an ALR, let $\mathbf{X}_i$ be the corresponding matrix of regressors, let $\mathbf{y}_i$ be the response vector, and let n_i be the number of data points in the current window. Also let $\mathbf{P}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}$ and ss_i be the corresponding sum of squares of residuals. Let $\mathbf{x}_{i+1}$ be the next regressor vector obtained by either adding or deleting a data point, and let y_{i+1} be the new response variable. In other words, if the data point we are adding to (or deleting from) the current window is labeled by $x_{(v)}$, then $y_{i+1} = x_{t_v+d}$ and $\mathbf{x}_{i+1}' = (x_{t_v+d-1}, \ldots, x_{t_v+d-p})$. Furthermore, set $\omega = 1$ for deleting and $\omega = 0$ for adding a data point. Then we have

$$\phi_{i+1} = \phi_i + \mathbf{P}_i\mathbf{x}_{i+1}[\mathbf{x}_{i+1}'\mathbf{P}_i\mathbf{x}_{i+1} + (-1)^\omega]^{-1}(y_{i+1} - \mathbf{x}_{i+1}'\phi_i)$$

and

$$\mathbf{P}_{i+1} = \mathbf{P}_i - \mathbf{P}_i\mathbf{x}_{i+1}[\mathbf{x}_{i+1}'\mathbf{P}_i\mathbf{x}_{i+1} + (-1)^\omega]^{-1}\mathbf{x}_{i+1}'\mathbf{P}_i.$$

To update the covariance matrix of the coefficient estimates, use

$$n_{i+1} = n_i + (-1)^\omega,$$

$$ss_{i+1} = ss_i + \frac{(y_{i+1} - \mathbf{x}_{i+1}'\phi_i)^2}{\mathbf{x}_{i+1}'\mathbf{P}_i\mathbf{x}_{i+1} + (-1)^\omega},$$

$$\hat{\sigma}_{i+1}^2 = \frac{ss_{i+1}}{n_{i+1} - p},$$

and

$$\widehat{\operatorname{cov}}(\phi_{i+1}) = \hat{\sigma}_{i+1}^2\mathbf{P}_{i+1}.$$

These equations can be easily derived by using some standard matrix algebra. For details see Young (1984) or Chen (1990).

2.3 Properties of Arranged Local Regression

We next give a consistency result to support practical use of the proposed ALR procedure for building FAR models. Suppose that $G = \{x_1, x_2, \ldots, x_N\}$ is a realization of the FAR process in (4) and that $f_k^N(x)$, $k = 1, \ldots, p$ are estimates of the functional coefficients $f_k(x)$ obtained by the ALR method with window size c and a sufficiently large interval $[-n, n]$, outside which $f_k^N(x)$ is defined to be 0, based on the data set G. Note that only values of the functions on the data points $\{x_1, \ldots, x_N\}$ are available.

Definition 2.1. Let $\{f_i^N\}$ be a sequence of functional estimates of f_i. The estimates f_i^N are mean squared consistent if

$$\lim_{N\to\infty} E\left\{\frac{1}{N}\sum_{j=1}^{N} |f_i^N(x_j) - f_i(x_j)|^2\right\} = 0,$$

where the expectation is taking over the joint distribution of $x_1, \ldots, x_N$.

Theorem 2.1. For the FAR process in (4), if (a) ε_t are iid random variables with $E(\varepsilon_t) = 0$, and $\operatorname{var}(\varepsilon_t) < \infty$; (b) the process is ergodic, x_t has a positive possibility density function over the domain of all $f_i(x)$, and the covariance matrix of $x_{t-1}, \ldots, x_{t-p}$ given $x_{t-d} \in I$ is full rank where I is an open interval in the domain of x_t; (c) in their domains, $f_i(x)$ are Lipschitz continuous; that is, $\exists M_0$ such that $|f_i(x + \delta) - f_i(x)| \le M_0\delta$; and (d) $E|f_i(x_t)|^2 < M < \infty$, $i = 1, \ldots, p$, then there exists a sequence of local arranged regression estimates of the functional coefficients that is mean squared consistent; that is, $\forall \varepsilon > 0$, $\exists c > 0$, $n > 0$, $N^* > 0$ such that $\forall N > N^*$,

$$E\left\{\frac{1}{N-p}\sum_{j=p+1}^{N} |f_i^N(x_j) - f_i(x_j)|^2\right\} < \varepsilon.$$

Our proof of this theorem is lengthy and hence is omitted; details can be found in Chen (1990).

2.4 A Modeling Procedure

We now propose a four-step procedure for building FAR models, based on the arranged local regression and some scatterplots:

1. Select the delay parameter d and the AR order p. This can be done by various methods. For example, one may apply various nonlinearity tests to explore the nonlinear feature of the data from which the threshold variable x_{t-d} can be identified (see, for example, Tsay 1989). For the AR order p, one may entertain several possible values and use model selection criteria, such as the Akaike information criterion (AIC; Akaike 1974), to select a final value p.
2. Select the functional forms of $f_i(\cdot)$ by using the proposed ALR procedure.
3. Postulate a well-parameterized model and perform a conditional least squares estimation. Asymptotic inferences of the conditional least squares estimates can be based on the results of Klimko and Nelson (1978), Tjøstheim (1986), and Lai (1990).
4. Check the estimated model by various diagnostic statistics for nonlinear time series analysis, such as those in

Tong (1990). Go back to Step 2 or 3 if necessary. If several nonlinear models are entertained, choose the one that appears most plausible.

3. APPLICATIONS

We illustrate the proposed modeling procedure of Section 2 by applying it to some simulated and real examples. The simulated example is used to demonstrate the ALR procedure proposed and the scatterplots suggested.

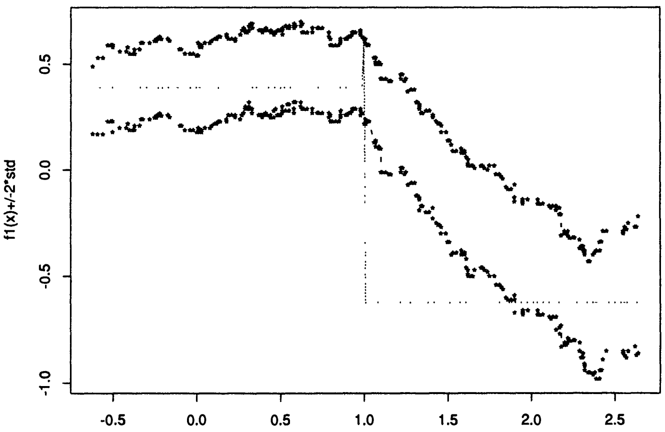
3.1 A Simulated Example

First, we consider a linear TAR model for which the coefficients $f_i(\cdot)$ are step functions. Two hundred observations were generated from the model

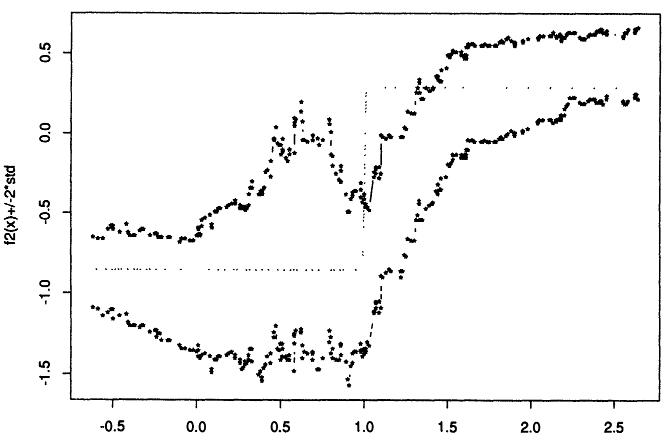
$$x_t = .4x_{t-1} - .8x_{t-2} + \varepsilon_t \qquad \text{if } x_{t-2} \le 1$$

$$= -.6x_{t-1} + .2x_{t-2} + \varepsilon_t \qquad \text{if } x_{t-2} > 1,$$

where ε_t are iid $N(0, 1)$. Figure 1 presents scatterplots of the estimated functional coefficients plus and minus their two standard errors. These results were obtained using $d = 2$, $p = 2$, and the proposed ALR procedure with $c = (x_{\max} - x_{\min})/5$ and $K = 40$. The dashed lines are the conditional least squares estimates of a TAR(2) model given the delay parameter $d = 2$ and the threshold value 1. These lines are used to denote the underlying true, but "unknown," coefficients. From the plots it is seen that the functional coefficients $f_1(x)$ and $f_2(x)$ are relatively stable when $x < 1$ and change smoothly to another constant level for $x > 1$, suggesting that a TAR model with a threshold value 1 might be appropriate for the data. Thus for this particular instance, the proposed modeling procedure effectively suggests the generating model of the process. Note that the smooth transition from one constant level to another showing in Figure 1 is due to the overlapping windows used in the ALR procedure. To check this overlapping effect, one can simply redo the ALR procedure starting with $x_{t-2} > 1$.



(a)



(b)

Figure 1. Scatterplots of Local Estimates of AR Coefficients for the Simulated Data. The upper and lower "" are least squares estimates plus and minus two standard errors; and the dotted line denotes the least squares estimates of fitting the generating model. The x axis is the threshold variable x_{t-2}.*

Table 1. Nonlinearity Tests For the Chickenpox Data

Delay	TAR-F	P value	d_1	d_2	CUSUM	New-F	P value	d_1	d_2
1	3.73	.00001	13	443	.47338	2.54	.00001	40	416
2	3.82	.00001	13	443	.02016	2.53	.00001	40	416
3	5.00	.00000	13	443	.26609	3.14	.00000	40	416
4	7.11	.00001	13	443	.59702	3.23	.00000	40	416
5	6.23	.00001	13	443	.00388	2.64	.00000	40	416
6	6.12	.00000	13	443	.25340	2.42	.00001	40	416
7	3.46	.00004	13	443	.00154	2.41	.00001	40	416
8	2.72	.00105	13	443	.00712	1.89	.00125	40	416
9	6.65	.00000	13	443	.02651	3.00	.00000	40	416
10	13.44	.00000	13	443	.00003	4.95	.00000	40	416
11	11.87	.00000	13	443	.00000	4.26	.00000	40	416
12	6.81	.00000	13	443	.00123	2.60	.00000	40	416

3.2 Chickenpox Data

Next we analyze the monthly records of cases of chickenpox in New York City for 1949–1972. (See Sugilara and May 1990 for further information on the data.) There are 533 observations in total; we took the natural log transformation to stabilize the variability.

First, we perform several nonlinearity tests, including the F test of Tsay (1986), the CUSUM test of Petruccelli and Davies (1986), the augmented F test of Luukkenon, Saikkonen, and Terasvirta (1988), and the threshold F test of Tsay (1989). All of the tests indicate that the process has strong nonlinearity. Table 1 gives some results of the tests where TAR-F is the threshold test, New-F denotes a generalized F test of Tsay (1991), CUSUM denotes the p value of the CUSUM test, and d_1 and d_2 are the degrees of freedom of the associated F tests.

From Table 1 we see that the process shows certain strong nonlinearity when the delay parameter d is 10, 11, or 12. Because the data are monthly observations, we use $d = 12$ in this article. Figure 2 shows the scatterplots of coefficient estimates of the model

$$x_t = f_1(x_{t-12})x_{t-1} + f_2(x_{t-12})x_{t-24} + \varepsilon_t \qquad (6)$$

via the proposed ALR procedure. The window size c used was one-tenth of the sample range of the data, and $K = 30$. Model (6) was used, because lags 1 and 24 are the most significant lags in our preliminary analysis of the process.

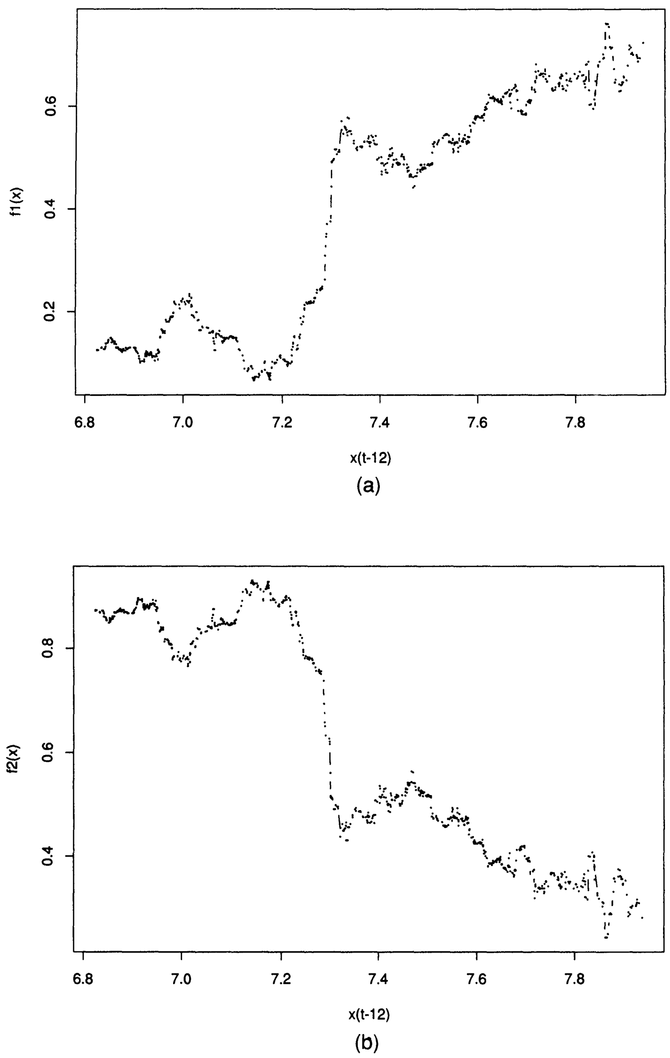


Figure 2. Scatterplots of Local Estimates of the Model in Equation (6). The x axis is the threshold variable x_{t-12}.

The plots in Figure 2 show that the coefficient estimates are relatively stable except for a big change after $x_{t-12} = 7.2$. Thus a TAR model with a threshold value 7.2 was tentatively entertained. In the process of model checking, we further detected a second possible change point around 6.8. This change point was not detected in Figure 2 primarily because it is at the very beginning of the scatterplots. Further, some AR coefficients are added to improve the fit. The modified model with conditional least squares estimates (standard deviations) is

$$\begin{aligned} x_t &= c_1 + \phi_1^{(1)} x_{t-1} + \phi_3^{(1)} x_{t-3} + \phi_9^{(1)} x_{t-9} + \phi_{24}^{(1)} x_{t-24} + \varepsilon_t^{(1)} \\ &\qquad\qquad \text{if } x_{t-12} < 6.85, \\ &= c_2 + \phi_1^{(2)} x_{t-1} + \phi_2^{(2)} x_{t-2} + \phi_9^{(2)} x_{t-9} + \phi_{24}^{(2)} x_{t-24} + \varepsilon_t^{(2)} \\ &\qquad\qquad \text{if } 6.85 \le x_{t-12} < 7.21, \\ &= c_3 + \phi_1^{(3)} x_{t-1} + \phi_2^{(3)} x_{t-2} + \phi_8^{(3)} x_{t-8} + \phi_9^{(3)} x_{t-9} \\ &\qquad + \phi_{24}^{(3)} x_{t-24} + \varepsilon_t^{(3)} \qquad \text{if } x_{t-12} \ge 7.21, \end{aligned}$$

where

$$\begin{aligned} \hat{c}_1 &= 2.63(.44), & \hat{\phi}_1^{(1)} &= .28(.04), \\ \hat{\phi}_3^{(1)} &= -.10(.02), & \hat{\phi}_9^{(1)} &= .07(.02), \\ \hat{\phi}_{24}^{(1)} &= .35(.06), & \hat{c}_2 &= .72(.68), \\ \hat{\phi}_1^{(2)} &= .87(.12), & \hat{\phi}_2^{(2)} &= -.51(.10), \\ \hat{\phi}_9^{(2)} &= .12(.04), & \hat{\phi}_{24}^{(2)} &= .42(.07), \\ \hat{c}_3 &= 23(.36), & \hat{\phi}_1^{(3)} &= .85(.06), \\ \hat{\phi}_2^{(3)} &= -.15(.05), & \hat{\phi}_8^{(3)} &= -.13(.04), \\ \hat{\phi}_9^{(3)} &= .26(.04), \text{ and } & \hat{\phi}_{24}^{(3)} &= .14(.04). \end{aligned}$$

For the three regimes, the residual variances are .0018, .0096, and .0099 and the number of observations are 135, 117, and 257. The overall residual variance is .00769, which is about 5% of the sample variance of x_t. Figure 3 presents time plots of the fitted values and residuals. From the plots, the model fits the process reasonably well. Both the autocorrelation function (ACF) and partial autocorrelation function (PACF)

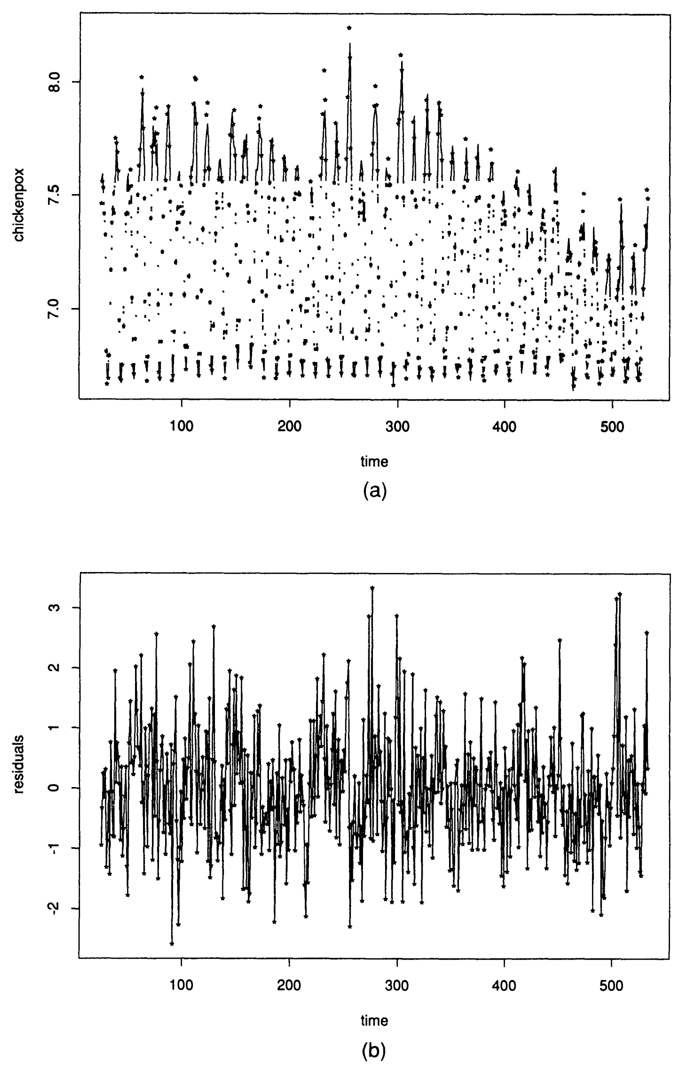


*Figure 3. Time Plots for the Chickenpox Example. (a) The fitted values in solid line and * denotes the data; (b) the residuals.*

of the residuals show no significant serial correlations in the residuals. We also applied the BDS test of Brock, Dechert, and Scheinkman (1987) to the standardized residuals. With $\varepsilon/\hat{\sigma} = 1.0$, the U statistics of BDS test are -1.11, -1.06, $-.60$, and $-.51$ for the dimension parameter $m = 2$, 3, 4, and 5. Under the iid hypothesis the U statistics are asymptotically standard normal. Thus for this particular instance, the BDS test cannot reject the iid assumption. We used this test statistic for nonlinear model checking mainly because it has been found useful in testing iid hypothesis (see Hsieh 1989).

For comparision, we also fit a linear ARMA model and a seasonal ARMA model to the data. The seasonal model is used to model the periodic behavior of the data. The best linear ARMA model is

$$x_t - \phi_1 x_{t-1} - \phi_{12} x_{t-12} - \phi_{13} x_{t-13} - \phi_{24} x_{t-24} = \varepsilon_t - \theta_7 \varepsilon_{t-7} - \theta_{12} \varepsilon_{t-12},$$

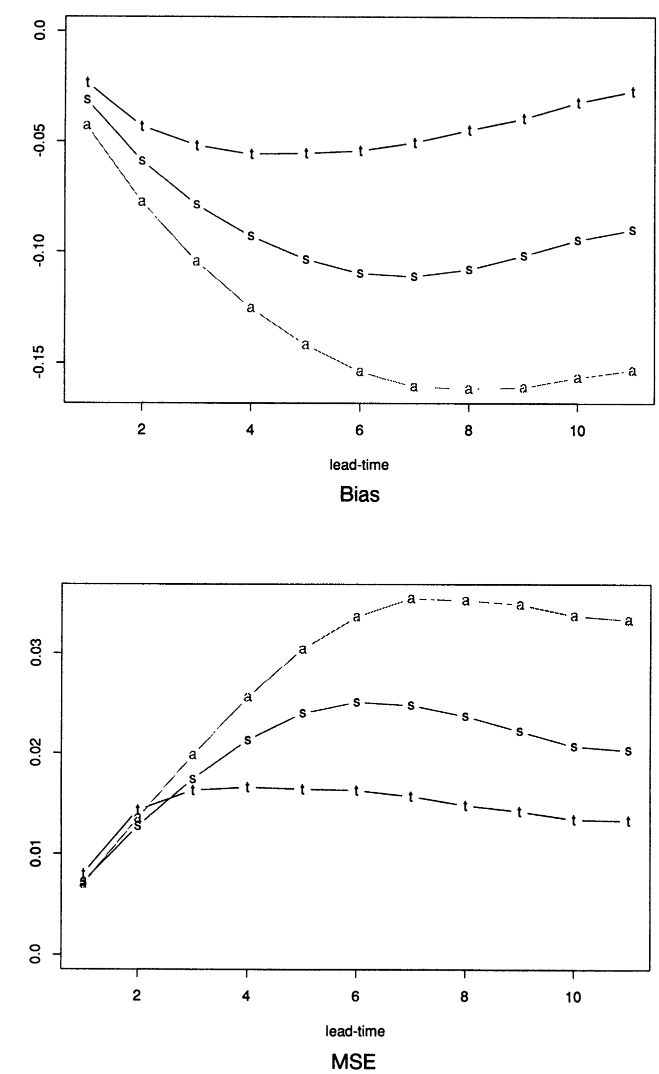


Figure 4. Biases and Mean Squared Errors of Postsample Multistep Forecasts for the Chickenpox Data. Here "t," "s," and "a" denote the threshold, seasonal ARMA, and regular ARMA models. The x axis is the lead-time in forecasts.

with exact maximum likelihood estimates (standard deviations) of the parameters

$$\hat{\phi}_1 = .79(.02), \quad \hat{\phi}_{12} = .90(.03), \quad \hat{\phi}_{13} = -.77(.03),$$
$$\hat{\phi}_{24} = .08(.03), \quad \hat{\theta}_7 = .07(.03), \quad \text{and} \quad \hat{\theta}_{12} = .79(.03),$$

and the estimated residual variance .00711, which is about 4.5% of the original variation. The best seasonal ARMA model is

$$(1 - \phi_1 B - \phi_7 B^7 - \phi_9 B^9 - \phi_{12} B^{12} - \phi_{13} B^{13})(1 - B^{12}) x_t = (1 - \theta_1 B - \theta_{12} B^{12} - \theta_{24} B^{24}) \varepsilon_t,$$

where B is the backshift operator such that $Bx_t = x_{t-1}$. The exact maximum likelihood estimates (standard deviations) of the parameters are

$$\hat{\phi}_1 = .71(.04), \quad \hat{\phi}_7 = -.09(.03),$$
$$\hat{\phi}_9 = .08(.03), \quad \hat{\phi}_{12} = -.55(.10),$$
$$\hat{\phi}_{13} = .31(.09), \quad \hat{\theta}_1 = -.14(.05),$$
$$\hat{\theta}_{12} = .36(.11), \quad \text{and} \quad \hat{\theta}_{24} = .38(.10),$$

with an estimated residual variance of .00735, which is almost the same as that of the linear ARMA model. For this particular series the linear models have slightly smaller residual variances than does the TAR model obtained by the proposed modeling procedure of Section 2. On the other hand, forecasting comparison suggests that the TAR model outperforms the two linear models in postsample multistep forecasts. More specifically, for each of the three models entertained, the first 450 observations were used to reestimate the parameters of the models. The estimated models are then used to make one- to 11-step ahead postsample forecasts for the last 70 observations. The average bias and mean squared error (MSE) of the postsample forecasts of the three models entertained are shown in Figure 4. From the plot it is clear that the TAR model substantially outperforms the two linear models in multi-step ahead forecasts, even though the linear models work fine for one- and two-step ahead forecasts.

3.3 Sunspot Numbers

The square root transformed series $y_t = 2(\sqrt{1 + x_t} - 1)$ of annual sunspot numbers for the period 1700–1979 has been analyzed in detail by Tong (1990) and other researchers. In this example we reanalyze the data via the proposed procedure and compare the results obtained with those of Tong (1990).

First, as for the chickenpox example, we carried out various nonlinearity tests to confirm that the process is indeed nonlinear. The tests showed that lags 3 and 8 are potential threshold variables. In the following analysis we used x_{t-3} as the threshold variable. Further, the data indicate that except for x_{t-3}, lags x_{t-1}, x_{t-2}, and x_{t-8} are highly correlated with x_t. Based on these findings, we entertain a FAR(8) model

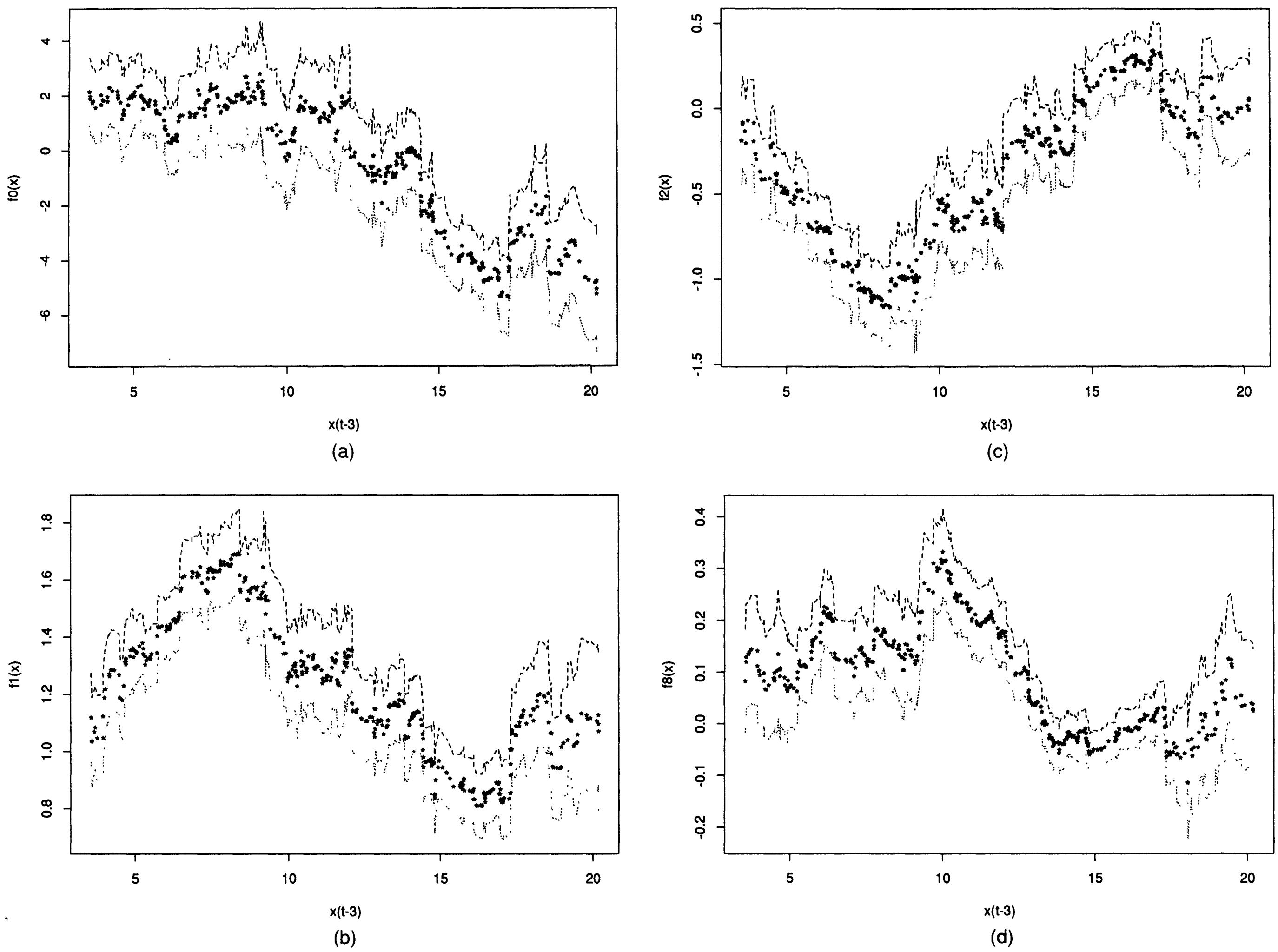


Figure 5. Scatterplots of Local Estimates of AR Coefficients for the Transformed Sunspot Numbers in Equation (7). The dotted line denotes the least squares estimates, and the solid lines are estimates plus and minus two standard errors. The x axis is the threshold variable x_{t-3}.

with the threshold variable x_{t-3} and nonzero coefficients at lags 1, 2, and 8; that is,

$$x_t = f_0(x_{t-3}) + f_1(x_{t-3})x_{t-1} + f_2(x_{t-3})x_{t-2} + f_8(x_{t-3})x_{t-8} + \varepsilon_t. \quad (7)$$

Figure 5 shows the scatterplots of the estimated functional coefficients in (7) along with limits of two standard errors obtained via the proposed ALR procedure. The window length used is one-ninth of the sample range, and the minimum data points in each window is $K = 25$. We see from the scatterplots that the estimated functions behave differently for x_{t-3} less than or greater than a number around 10.5. In the first threshold regime, the function $f_0(\cdot)$ appears to be a constant, $f_1(\cdot)$ and $f_2(\cdot)$ behave as absolute functions, (i.e., symmetric functions with a sharp turning point), and $f_8(\cdot)$ is approximately a constant. In the upper threshold regime, it seems that there is another change point around 17.5. But a further study indicates that this change point is not pronounced, because only a few data points are beyond 17.5. For simplicity we ignore this possible change point. Thus in this upper regime, $f_0(x)$ behaves like a linear function in x and other coefficient functions are approximately constant. Consequently, we tentatively specify the following coefficient functions for the transformed sunspot numbers:

$$\begin{aligned} f_0(x) &= a_1 \quad \text{if } x < r \\ &= a_2 + a_3 x \quad \text{if } x \geq r, \\ f_1(x) &= b_1 + b_2|x - v| \quad \text{if } x < r \\ &= b_3 \quad \text{if } x \geq r, \\ f_2(x) &= c_1 + c_2|x - v| \quad \text{if } x < r \\ &= c_3 \quad \text{if } x \geq r, \end{aligned}$$

and

$$\begin{aligned} f_8(x) &= d_1 \quad \text{if } x < r \\ &= d_2 \quad \text{if } x \geq r, \end{aligned}$$

where r is around 10.5.

In sum, the specified FAR model for the transformed sunspot numbers is

$$\begin{aligned} x_t &= a_1 + (b_1 + b_2|x_{t-3} - v|)x_{t-1} \\ &\quad + (c_1 + c_2|x_{t-3} - v|)x_{t-2} + d_1 x_{t-8} + \varepsilon_t^{(1)}, \quad \text{if } x_{t-3} < r \\ &= a_2 + a_3 x_{t-3} + b_3 x_{t-1} + c_3 x_{t-2} + e_1 x_{t-6} \\ &\quad + d_2 x_{t-8} + \varepsilon_t^{(2)} \quad \text{otherwise,} \end{aligned}$$

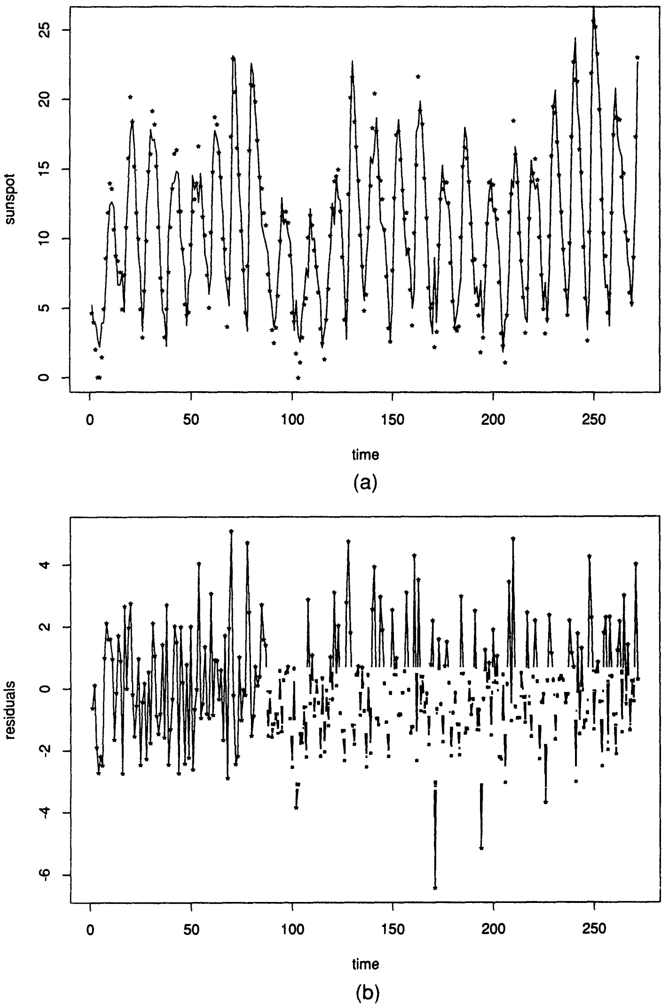


*Figure 6. Time Plots for Transformed Sunspot Numbers: (a) Observed and fitted values with * denotes the data; (b) the residuals of fitted model.*

where the x_{t-6} term in the upper threshold region is used to further improve the fit. The conditional least squares estimates (standard deviations) are

$$\hat{r} = 10.3, \qquad \hat{v} = 6.6,$$

$$\hat{a}_1 = 1.23(.73), \quad \hat{a}_2 = .92(.75), \quad \hat{a}_3 = -.24(.06),$$

$$\hat{b}_1 = 1.75(.13), \quad \hat{b}_2 = -.17(.05), \quad \hat{b}_3 = .87(.07),$$

$$\hat{c}_1 = -1.28(.20), \quad \hat{c}_2 = .27(.05), \quad \hat{c}_3 = .17(.10),$$

$$\hat{d}_1 = .20(.04), \quad \hat{d}_2 = .04(.02), \quad \text{and} \quad \hat{e}_1 = -.06(.03),$$

Table 2. Multistep Postsample Forecasts of Annual Sunspot Numbers for 1980–1987

		TAR		*FAR*	
Year	*Observation*	*Prediction*	*Error*	*Prediction*	*Error*
1980	154.7	160.1	−5.4	168.5	−13.8
1981	140.5	141.8	−1.3	140.6	−.01
1982	115.9	96.4	19.5	105.9	10.0
1983	66.6	61.8	4.8	70.9	−3.3
1984	45.9	31.1	14.8	42.1	3.8
1985	17.9	18.1	−.2	22.5	−4.6
1986	13.4	18.9	−5.5	12.1	1.3
1987	29.2	29.9	−.7	7.5	21.7

where the estimates of the structure parameters r and v were obtained by minimizing the sum of squares of residuals. The residual standard errors for the two regimes are 2.11 and 1.50. The model has 12 coefficient parameters plus two structure parameters. The overall residual variance is 3.27, which is about 12% less than Tong's TAR model (1990, pp. 420–429) that has 16 parameters and one structure parameter. In terms of AIC, this FAR model improves substantially over Tong's model. The residuals and fitted values, shown in Figure 6, appear to be reasonable. The residual ACF and PACF fail to suggest any significant serial correlations in the residuals. Also, with $\varepsilon/\hat{\sigma} = 1$ the U statistics of the BDS test for the standardized residuals are $-.46$, $-.10$, $.32$, and $-.01$ for the dimension parameter $m = 2, 3, 4, 5$. Again, the test fails to reject the iid assumption.

For further comparision we used the model to predict the sunspot numbers for 1980–1987. This was done by taking anti-square root transformation from the postsample forecasts of the transformed series. The results, along with the forecasts given in Tong (1990), are shown in Table 2. From the table, the FAR model outperforms Tong's TAR model in 5 of 8 years; however, the FAR model predicted poorly for 1987. Because 1987 is a turning point for the sunspot number and the delay parameter of the FAR model is 3, compared to 8 for the TAR model, the inaccuracy of the FAR model in predicting the number for 1987 is understandable.

We also compared the out-of-sample multistep forecasts of the FAR model with a TAR model and a linear AR(12) model. The TAR model used has two regimes each with 12 lagged variables. This modified TAR model produces better out-of-sample forecasts than Tong's model. We used the transformed data for 1700–1955 to estimate the three models

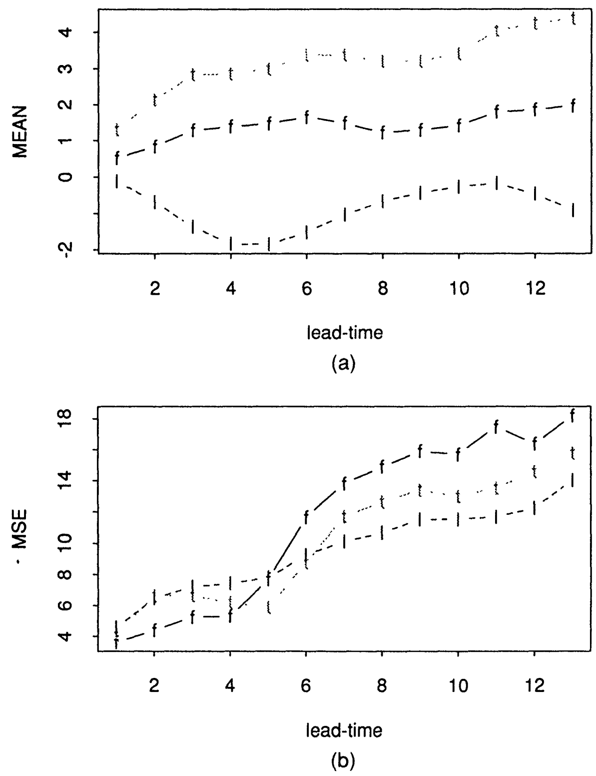


Figure 7. Biases and Mean Squared Errors of Postsample Multistep Forecasts for the Sunspot Numbers. Here "t," "l," and "f" denote threshold, linear, and FAR models. The x axis is the lead time in forecasts.

and then used the estimated models to make naive out-of-sample forecasts for one- to 13-step ahead for 1956–1974. The average biases and the MSEs of the forecasts are shown in Figure 7. From the plots it is seen that for MSE the FAR model is the best in short term forecast and the linear AR model is the best in long term forecast. In terms of bias, the FAR model seems much better than the other two models.

4. CONCLUSION

In this article we generalized directly the linear AR model to a class of FAR models. Some properties of the proposed model and a modeling procedure were discussed. Applications of the proposed model to real examples suggested that the FAR model can substantially improve postsample multistep forecasts over other time series models.

APPENDIX: PROOFS OF THEOREMS 1.1 AND 1.2

Let $\mathbf{X}_t = (x_t, \ldots, x_{t-p+1})'$. Because $\mathbf{X}^*_{t-1}$ in (2) is a subvector $\mathbf{X}_{t-1}$, we can write $f(\mathbf{X}^*_{t-1})$ as $f(\mathbf{X}_{t-1})$. Letting $\mathbf{Z}_t = (\varepsilon_t, 0, \ldots, 0)'$ and

$$\mathbf{A}(\mathbf{X}) = \begin{pmatrix} f_1(\mathbf{X}) & f_2(\mathbf{X}) & \cdots & f_{p-1}(\mathbf{X}) & f_p(\mathbf{X}) \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \tag{A.1}$$

we can rewrite the FAR model in (1) as

$$\mathbf{X}_t = \mathbf{A}(\mathbf{X}_{t-1})\mathbf{X}_{t-1} + \mathbf{Z}_t. \tag{A.2}$$

For Model (A.2), Chan and Tong (1985) proved that under some mild condition, $\{\mathbf{X}_t\}$ is a ϕ-irreducible and aperiodic Markov chain. Next, in the $\Re^p$ space with the Euclidean norm $\|\mathbf{X}\| = \sqrt{x_1^2 + \cdots + x_p^2}$, we have the following lemma.

Lemma A.1. If $|f_i(\mathbf{X})| \le c_i$, then $\forall \mathbf{Y} = (y_1, \ldots, y_p)' \in \Re^p$, we have

$$\|\mathbf{A}(\mathbf{X})\mathbf{Y}\| \le \|\mathbf{C}|\mathbf{Y}|\|,$$

where $|\mathbf{Y}| = (|y_1|, \ldots, |y_p|)'$ and

$$\mathbf{C} = \begin{pmatrix} c_1 & c_2 & \cdots & c_{p-1} & c_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{A.3}$$

Proof. Let $\mathbf{A}(\mathbf{X})\mathbf{Y} = (d_1, \ldots, d_p)'$ and $\mathbf{C}|\mathbf{Y}| = (g_1, \ldots, g_p)'$. Then we have $|d_i| = g_i = |y_{i-1}|$ for $i = 2, \ldots, p$, and for $i = 1$,

$$\begin{aligned} |d_1| &= |f_1(\mathbf{X})y_1 + \cdots + f_p(\mathbf{X})y_p| \\ &\le |f_1(\mathbf{X})y_1| + \cdots + |f_p(\mathbf{X})y_p| \\ &\le c_1|y_1| + \cdots + c_p|y_p| = g_1. \end{aligned}$$

Hence $\|\mathbf{A}(\mathbf{X})\mathbf{Y}\| \le \|\mathbf{C}|\mathbf{Y}|\|$.

Proof of Theorem 1.1. First, for the matrix $\mathbf{C}$ defined in (A.3), the determinant is $|\lambda\mathbf{I} - \mathbf{C}|_p = \lambda^p - c_1\lambda^{p-1} - \cdots - c_p$ (see, for example, Anderson 1971, p. 180). Hence all the roots of the characteristic function (3) are eigenvalues of the matrix $\mathbf{C}$. Let $\lambda_{\max}$ be the maximum eigenvalue in modulus of $\mathbf{C}$. Because $|\lambda_{\max}| < 1$, and $\|\mathbf{C}^n\|^{1/n} \to |\lambda_{\max}|$, we can find $1 > \delta > 0$ and an $h > 0$, such that $\|\mathbf{C}^h\| < 1 - \delta$. Consequently, by Lemma A.1 we have

$$\begin{aligned} &E(\|\mathbf{X}_{t+h}\| \mid \mathbf{X}_t = \mathbf{X}) \\ &= E\left(\left\| \prod_{i=0}^{h-1} \mathbf{A}(\mathbf{X}_{t+i})\mathbf{X}_t + \sum_{i=1}^{h} \left[\prod_{j=i}^{h-1} \mathbf{A}(\mathbf{X}_{t+j}) \right] \mathbf{Z}_{t+i} \right\| \mid \mathbf{X}_t = \mathbf{X} \right) \\ &\le \|\mathbf{C}^h|\mathbf{X}|\| + \mathbf{E}\left\| \sum_{i=1}^{h} \mathbf{C}^{h-i}|\mathbf{Z}_{t+i}| \right\| \\ &\le \|\mathbf{C}^h\|\|\mathbf{X}\| + \mathbf{E}\left\| \sum_{i=1}^{h} \mathbf{C}^{h-i}|\mathbf{Z}_{t+i}| \right\| \\ &\le (1-\delta)\|\mathbf{X}\| + E\left\| \sum_{i=1}^{h} \mathbf{C}^{h-i}|\mathbf{Z}_{t+i}| \right\|. \end{aligned}$$

Because $E\|\mathbf{Z}_t\|$ is bounded, we know that $\mathbf{E}\|\sum_{i=1}^{h} \mathbf{C}^{h-i}|\mathbf{Z}_{t+i}|\|$ is bounded and that the bound does not depend on $\mathbf{X}$. Thus we can find a sufficiently large $M > 0$ such that when $\|\mathbf{X}\| > M$,

$$(1-\delta)\|\mathbf{X}\| + E\left\| \sum_{i=1}^{h} \mathbf{C}_{h-i}|\mathbf{Z}_{t+i}| \right\| < (1-\delta_1)\|\mathbf{X}\|,$$

where $0 < \delta_1 < 1$. Hence the compact set $K = \{(x_1, \ldots, x_p) \mid \sqrt{x_1^2 + \cdots + x_p^2} \le M\}$ satisfies that when $\mathbf{X} \notin K$, $E(\|\mathbf{X}_{t+h}\| \mid \mathbf{X}_t = \mathbf{X}) < (1-\delta_1)\|\mathbf{X}\|$. By Lemmas 1.1 and 1.2, $\{\mathbf{X}_t\}$ is geometrically ergodic.

Proof of Theorem 1.2. Let

$$\mathbf{G}(\mathbf{X}) = \begin{pmatrix} g_1(\mathbf{X}) & \cdots & g_{p-1}(\mathbf{X}) & g_p(\mathbf{X}) \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix},$$

$$\mathbf{H}(\mathbf{X}) = \begin{pmatrix} h_1(\mathbf{X}) & \cdots & h_{p-1}(\mathbf{X}) & h_p(\mathbf{X}) \\ 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix},$$

and C be the matrix defined in (A.3). Then

$$\begin{aligned} &E(\|\mathbf{X}_{t+h}\| \mid \mathbf{X}_t = \mathbf{X}) \\ &= E\Bigg(\Bigg\| \prod_{i=0}^{h-1} \mathbf{G}(\mathbf{X}_{t+i})\mathbf{X}_t + \mathbf{H}(\mathbf{X}_{t+h-1})\mathbf{X}_{t+h-1} \\ &\qquad + \sum_{i=0}^{h-2} \left[\prod_{j=i+1}^{h-1} \mathbf{G}(\mathbf{X}_{t+j}) \right] \mathbf{H}(\mathbf{X}_{t+i})\mathbf{X}_{t+i} \\ &\qquad + \sum_{i=1}^{h} \left[\prod_{j=i}^{h-1} \mathbf{G}(\mathbf{X}_{t+j}) \right] \mathbf{Z}_{t+i} \Bigg\| \mid \mathbf{X}_t = \mathbf{X} \Bigg) \\ &\le \|\mathbf{C}^h|X|\| + E\Bigg(\Bigg\| |\mathbf{H}(\mathbf{X}_{t+h-1})\mathbf{X}_{t+h-1}| \\ &\qquad + \sum_{i=0}^{h-2} \mathbf{C}^{h-i-1}|\mathbf{H}(\mathbf{X}_{t+i})\mathbf{X}_{t+i}| \Bigg\| \mid \mathbf{X}_t = \mathbf{X} \Bigg) \\ &\qquad + E\left(\sum_{i=1}^{h} \|\mathbf{C}^{h-i}|\mathbf{Z}_{t+i}|\| \right) \\ &\le \|\mathbf{C}^h\|\|\mathbf{X}\| + \sum_{i=0}^{h-1} \|\mathbf{C}^{h-i-1}\| E(\|\mathbf{H}(\mathbf{X}_{t+i})\mathbf{X}_{t+i}\| \mid \mathbf{X}_t = \mathbf{X}) \\ &\qquad + E\left(\sum_{i=1}^{h} \|\mathbf{C}^{h-i}\mathbf{Z}_{t+i}\| \right) \\ &\le (1-\delta)\|\mathbf{X}\| + \sum_{i=0}^{h-1} \|\mathbf{C}^{h-i-1}\| E(\|\mathbf{H}(\mathbf{X}_{t+i})\mathbf{X}_{t+i}\| \mid \mathbf{X}_t = \mathbf{X}) \\ &\qquad + E\left(\sum_{i=1}^{h} \|\mathbf{C}^{h-i}\mathbf{Z}_{t+i}\| \right). \end{aligned}$$

Because $h_i(\mathbf{X})x_t$ is bounded, we have that $\|\mathbf{H}(\mathbf{X})\mathbf{X}\|$ is bounded for any $\mathbf{X}$. Hence

$$\sum_{i=0}^{h-1} \|\mathbf{C}^{h-i-1}\| E(\|\mathbf{H}(\mathbf{X}_{t+i})\mathbf{X}_{t+i}\| \,|\, \mathbf{X}_t = \mathbf{X})$$

is bounded, and the bound does not depend on $\mathbf{X}$. Similarly, we can prove that $E(\sum_{i=1}^{h} \|\mathbf{C}^{h-i}\mathbf{Z}_{t+i}\|)$ is bounded. The theorem then follows the same lines as Theorem 1.1.

[*Received February 1991. Revised June 1992.*]

REFERENCES

Akaike, H. (1974), "A New Look at Statistical Model Identification," *IEEE Transactions on Automatic Control,* AC-19, 716–722.

Anderson, T. W. (1971), *The Statistical Analysis of Time Series,* New York: John Wiley.

Brock, W., Dechert, D., and Scheinkman, J. (1987), "A Test for Independence Based on the Correlation Dimension," unpublished manuscript, University of Chicago, Dept. of Economics.

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control,* San Francisco: Holden-Day.

Chan, K. S., and Tong, H. (1985), "On the Use of the Deterministic Lyapunov Function for the Ergodicity of Stochastic Difference Equations," *Advanced Applied Probability,* 17, 667–678.

Chen, R. (1990), "Two Classes of Nonlinear Time Series Models," unpublished Ph.D. thesis, Carnegie Mellon University, Dept. of Statistics.

Chen, R., and Tsay, R. S. (1991), "On the Ergodicity of TAR(1) Processes," *Annals of Applied Probability,* 1, 613–634.

Haggan, V., and Ozaki, T. (1981), "Modeling Nonlinear Vibrations Using an Amplitude-Dependent Autoregressive Time Series Model," *Biometrika,* 68, 189–196.

Hsieh, D. A. (1989), "Testing for Nonlinear Dependence in Daily Foreign Exchange Rates," *Journal of Business,* 62, 339–368.

Klimko, L. A., and Nelson, P. I. (1978), "On Conditional Least Squares Estimation For Stochastic Processes," *The Annals of Statistics,* 6, 629–643.

Lai, T. L. (1990), "Asymptotic Properties of Nonlinear Least Squares Estimates in Stochastic Regression Models," unpublished technical report, Stanford University, Dept. of Statistics.

Luukkonen, R., Saikkonen, P., and Terasvirta, T. (1988), "Testing Linearity against Smooth Transition Autoregressive Models," *Biometrika,* 75, 491–499.

Petruccelli, J., and Davis, N. (1986), "A Portmanteau Test for Self-Exciting Threshold Autoregressive-Type Nonlinearity in Time Series," *Biometrika,* 73, 687–694.

Petruccelli, J., and Woolford, S. W. (1984), "A Threshold AR(1) Model," *Journal of Applied Probability,* 21, 270–286.

Priestley, M. B. (1980), "State-Dependent Models: A General Approach to Nonlinear Time Series Analysis," *Journal of Time Series Analysis,* 1, 47–71.

Sugihara, G., and May, R. M. (1990), "Nonlinear Forecasting as a Way of Distinguishing Chaos From Measurement Error in Time Series," *Nature,* 344, 734–741.

Tjøstheim, D. (1986), "Estimation in Nonlinear Time Series Models," *Stochastic Processes and Their Applications,* 21, 251–273.

——— (1990), "Nonlinear Time Series and Markov Chains," *Advanced Applied Probability,* 22, 587–611.

Tong, H. (1983), *Threshold Models in Nonlinear Time Series Analysis* (Lecture Notes in Statistics 21), New York: Springer-Verlag.

——— (1990), *Nonlinear Time Series Analysis: A Dynamical System Approach,* London: Oxford University Press.

Tsay, R. S. (1986), "Nonlinearity Tests for Time Series," *Biometrika,* 73, 461–466.

——— (1989), "Testing and Modeling Threshold Autoregressive Processes," *Journal of the American Statistical Association,* 84, 231–240.

——— (1991), "Detecting and Modeling Nonlinearity in Univariate Time Series Analysis," *Statistica Sinica,* 1, 431–451.

Tweedie, R. L. (1975), "Sufficient Conditions for Ergodicity and Recurrence of Markov Chain on a General State-Space," *Stochastic Processes and Their Applications,* 3, 385–403.

Young, P. (1984), *Recursive Estimation and Time Series Analysis,* New York: Springer-Verlag.