FUNCTIONAL COEFFICIENT AUTOREGRESSIVE MODELS: ESTIMATION AND TESTS OF HYPOTHESES

BY RONG CHEN AND LON-MU LIU[1]

University of Illinois at Chicago

First Version received May 1998

Abstract. In this paper, we study nonparametric estimation and hypothesis testing procedures for the functional coefficient AR (FAR) models of the form $X_t = f_1(X_{t-d})X_{t-1} + \cdots + f_p(X_{t-d})X_{t-p} + \varepsilon_t$, first proposed by Chen and Tsay (1993). As a direct generalization of the linear AR model, the FAR model is a rich class of models that includes many useful parametric nonlinear time series models such as the threshold AR models of Tong (1983) and exponential AR models of Haggan and Ozaki (1981). We propose a local linear estimation procedure for estimating the coefficient functions and study its asymptotic properties. In addition, we propose two testing procedures. The first one tests whether all the coefficient functions are constant, i.e. whether the process is linear. The second one tests if all the coefficient functions are continuous, i.e. if any threshold type of nonlinearity presents in the process. The results of some simulation studies as well as a real example are presented.

Keywords. Continuity test; linearity test; local linear estimation; nonparametric estimation; one-sided kernel; threshold model.

1. INTRODUCTION

Nonlinear time series analysis has been one of the major areas of research in time series for more than two decades now. Many nonlinear parametric models such as the threshold AR (TAR) model of Tong (1983, 1990), the exponential AR (EXPAR) model of Haggan and Ozaki (1981) and the smooth transition AR (STAR) model of Granger and Teräsvirta (1993) and Teräsvirta (1994) have been proposed and successfully applied in many important real-life problems. Tong (1990) and Priestley (1988) provided many foundations of parametric nonlinear time series analysis. A more recent review of the subject can be found in Tjøsteim (1994).

It is noted that although in many applications background knowledge can often shed lights on finding an appropriate model, other applications lack such knowledge and often require trial-and-error type of model selection procedures. To overcome the subjectivity in model selection, Chen and Tsay (1993) proposed a class of models referred to as functional coefficient autoregressive (FAR) models which assumes the form of

$$X_t = f_1(X_{t-d})X_{t-1} + \cdots + f_p(X_{t-d})X_{t-p} + \varepsilon_t \tag{1}$$

© 2001 Blackwell Publishers Ltd., 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

where ε_t is white noise with finite variance σ^2 and is independent of X_s for all $s < t$. It is a direct extension of the linear AR model, but allows the coefficients varying according to a threshold variable X_{t-d}. They suggested using nonparametric procedures to determine the functions in the model, hence allowing 'data to speak for themselves' regarding the model to be used. It is noted that many of the successful parametric nonlinear models belong to the FAR family. For example, if the functions $f_i(x)$ in (1) are step functions $f_i(x) = a_i + b_i I(x \geqslant c)$, we have the TAR model. When $f_i(x) = a_i + b_i e^{-\gamma x^2}$, the model becomes an EXPAR model. STAR and many other models also belong to this class. Hence, nonparametric determination of the functional forms in model (1) may provide objective guildlines on choosing an appropriate parametric model. It also allows researchers to develop new models that are useful in their applications by specifying a parametric form for the coefficient functions based on the nonparametric estimates. In addition, nonparametric estimators can also be the final solution to the problem on hand.

Nonparametric procedures have been used extensively in time series analysis. Györfi *et al.* (1989), Tjøsteim (1994), Härdle *et al.* (1997) and Hart (1996) have given selective reviews on this topic. These procedures extend many nonparametric procedures developed in regression context into time series analysis.

In this paper, Sections 2–4 concentrate on three aspects of the FAR models. In Section 2, we propose a local linear estimator for estimating the coefficient functions nonparametrically. It is similar to the moving window procedure proposed by Chen and Tsay (1993), though we use Kernel weight functions. We systematically study the asymptotic properties of the estimator. Note that this procedure is slightly different from local polynomial curve estimation procedures of Cleveland and Devlin (1988), Fan and Gijbels (1996) and Tsybakov (1986). Here, we are interested in estimating the coefficient functions. Hastie and Tibshirani (1993) proposed similar estimation procedures in regression context for 'varying coefficient models', which is similar to the FAR model.

In Section 3, we develop a procedure to test if the coefficient functions are constant functions. It is basically a linearity test since, when all the coefficient functions are constant, the FAR model becomes a linear AR model. There are many linearity tests available in the literature. For example, Keenan (1985), Tsay (1986) and Luukkonen *et al.* (1988) proposed different forms of Lagrange multiplier type of tests. Chan and Tong (1986) and Tsay (1989) considered testing threshold type of nonlinearity. Nonparametrically, Hjellvik and Tjøstheim (1995, 1996) and Hjellvik *et al.* (1997) developed linearity tests by comparing nonparametric and linear estimates of $E[X_t|X_{t-k}]$. Here, we handle this problem within the FAR model framework.

In Section 4, we develop another testing procedure, to detect if there are any discontinuous points in the coefficient functions. This is of interest due to the fact that all threshold type of models have jump points in the coefficient functions. Since the class of threshold models is one of the most important and

© Blackwell Publishers Ltd 2001

widely used classes of nonlinear time series models, it is certainly important to be able to detect if there is any threshold type of nonlinearity when one uses FAR models as a tool for model selection. The test is also of interest when the nonparametric estimate is treated as a final solution of the problem. Most of the nonparametric estimators are designed to estimate continuous functions. They are not consistent at discontinuous points. In finite samples, they tend to have large bias in the neighbourhood of the discontinuous points. Hence, it is important to detect the existence of jump points, so as to select suitable nonparametric estimators. The proposed testing procedure is based on nonparametric estimation of the coefficient functions with one-sided kernels and the fact that, at a discontinuous point, estimates with left-side kernels and right-side kernels are significantly different while, at continuous points, they are not. The proposed procedure is similar to that of Müller (1992) in a regression setting.

All the simulation studies for the above proposed procedures are presented within their respective sections. The analysis of a real-life example, the chickenpox series, is presented in Section 5.

2. NONPARAMETRIC ESTIMATION OF THE FAR MODEL

We begin this section by mentioning that, in practice, the FAR model

$$X_t = f_0(X_{t-d}) + f_1(X_{t-d})X_{t-l_1} + \cdots + f_p(X_{t-d})X_{t-l_p} + \varepsilon_t$$

is often used. It is slightly more general than the model (1) by allowing skipped AR lags. In the cases that the threshold lag d is also one of the AR lags, it creates ambiguity since the model includes both $f_0(X_{t-d})$ and $f_i(X_{t-d})X_{t-d}$ if $l_i = d$. Hence, one of those terms should be removed. From our experience, estimation of the model with the f_0 term is more stable, particularly when 0 is in the range of threshold variable X_{t-d}.

However, for simplicity and clearer presentation, we will only consider model (1) in the theorems and their proof. The extension to the above more general model is trivial.

We propose the following local linear estimator to estimate the functions

$$\mathbf{f}(x) = [f_1(x), \ldots, f_p(x)]'$$

nonparametrically. Let

$$\hat{\mathbf{f}}(x) = \arg\min_{\boldsymbol{\beta}} \sum_{t=l+1}^{n} (X_t - \mathbf{X}_t'\boldsymbol{\beta})^2 K_h(X_{t-d} - x) \tag{2}$$

where $\mathbf{X}_t = [X_{t-1}, \ldots, X_{t-p}]'$ and $K_h(u) = h^{-1}K(u/h)$ where K is a kernel function, h is the bandwidth and $l = \max\{d, p\}$. It is easily seen that

$$\hat{\mathbf{f}}(x) = (\mathbf{X}'\mathbf{W}_x\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_x\mathbf{Y}$$

© Blackwell Publishers Ltd 2001

where $\mathbf{X} = [\mathbf{X}_{l+1} : \cdots : \mathbf{X}_n]'$, $\mathbf{Y} = [X_{l+1}, \ldots, X_n]'$ and $\mathbf{W}_x$ is a diagonal matrix with the diagonal elements being $K_h(X_{t-d} - x)$ for $t = l+1, \ldots, n$.

The asymptotic properties of the above estimator can be summarized in the following theorem. The theorem concerns only the continuous points. We will study the case of discontinuous coefficient functions in Section 4.

Define $\mu_2 = \int u^2 K(u)\mathrm{d}u$ and $K_2^2 = \int K^2(u)\mathrm{d}u$. Let $p_{i,j,d}$ be the joint stationary density of the triplet $(X_{t-i}, X_{t-j}, X_{t-d})$ and $p(x)$ be the stationary marginal density of X_t.

THEOREM 1. *Let x be a continuous point of the coefficient functions $f_1, \ldots, f_p$. Under assumptions A1 to A8 listed in the Appendix, we have*

$$n^{2/5}(\hat{\mathbf{f}}(x) - \mathbf{f}(x) - \beta^2\mathbf{b}(x)) \xrightarrow{\mathrm{D}} \mathrm{N}_p(0, \beta^{-1}\sigma^2 K_2^2 \mathbf{A}^{-1}(x))$$

where $\mathbf{A}(x) = p(x)E[\mathbf{X}_t\mathbf{X}_t'|X_{t-d} = x]$ and $\mathbf{b}(x) = \mu_2\mathbf{A}^{-1}(x)\mathbf{B}(x)$ where $\mathbf{B}(x)$ is a vector with ith element being

$$\sum_{j=1}^{p}\int uv\left\{\frac{1}{2}f_j''(x)p_{i,j,d}(u, v, x) + f_j'(x)p_{i,j,d}'(u, v, x)\right\}\mathrm{d}u\,\mathrm{d}v$$

with $f_j'(x)$ and $f_j''(x)$ being the first and second derivative of $f_j(x)$ respectively, and $p_{i,j,d}'$ being the partial derivative with respect to the third argument.

The proof of the above theorem is given in the Appendix. We note that the asymptotic result is similar to that of kernel estimation of a response curve. It can be easily extended to resemble that of local polynomial estimation of a response curve using the estimator

$$\hat{\mathbf{f}}(x) = \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} \sum_{t=l+1}^{n}(X_t - \mathbf{X}_t'\boldsymbol{\beta} - \mathbf{X}_t'\boldsymbol{\gamma}(X_{t-d} - x))K_h(X_{t-d} - x)$$

This estimator should entertain many nice properties of the local polynomial estimator, though the derivation of the asymptotic distribution becomes more complicated and tedious. In this paper, we restrict ourselves to the estimator in (2).

The theorem shows that the estimator has the rate of convergence of one-dimensional smoothing. As a consequence, estimation of the response surface $E[X_t|\mathbf{X}_t = \mathbf{x}]$ will have the same rate of convergence, hence does not suffer the curse of dimensionality as that in direct p-dimensional estimation of the surface. This advantage is due to the special structure of the model, which serves as a dimension reduction tool. Specifically, let $\mathbf{x} = [x_1, \ldots, x_p]'$. The conditional mean function

$$m(\mathbf{x}) = E[X_t|\mathbf{X}_t = \mathbf{x}] = \boldsymbol{x}'\mathbf{f}(x_d)$$

can be estimated by $\hat{m}(\mathbf{x}) = \mathbf{x}'\hat{\mathbf{f}}(x_d)$ and we have

© Blackwell Publishers Ltd 2001

COROLLARY 1. *Under the conditions of Theorem 1, we have*

$$n^{2/5}(\hat{m}(\mathbf{x}) - m(\mathbf{x}) - \beta^2\mathbf{x}'\mathbf{b}(x_d)) \xrightarrow{D} \mathrm{N}(0, \beta^{-1}\sigma^2 K_2^2 \mathbf{x}'\mathbf{A}^{-1}(x_d)\mathbf{x})$$

Automatic bandwidth selection procedure is always one of the key ingredients in practical implementation of nonparametric procedures. There are many approaches such as the cross-validation (CV) approach of Härdle and Vieu (1992) and Cheng and Tong (1992) in time series, the plug-in approach of Sheather (1983, 1986), Ruppert *et al.* (1995), Park and Marron (1990) and many others in regression. It is somewhat difficult to use the plug-in approach here since the bias term involves the partial derivative of a three-dimensional density, which is not easy to estimate. Hence, we suggest to use the CV procedure through the response surface estimation. Specifically, define

$$\mathrm{CV}(h) = \sum_{i=l+1}^{n} \left(X_i - \sum_{j=1}^{p} \hat{f}_{h,j}^{(-i)}(X_{i-d})X_{i-j} \right)^2 w(X_{i-d})$$

where $\hat{f}_{h,j}^{(-i)}$, $j = 1, \ldots, p$ are the leave-one estimator, i.e. the estimator defined in (2) but without the $t = i$ term in the summation. The weight function w has a compact support.

Consider the following second-order EXPAR model (3),

$$\text{EXPAR(2):} \quad X_t = (0.5 - 1.1\mathrm{e}^{-50X_{t-1}^2})X_{t-1} + (0.3 - 0.5\mathrm{e}^{-50X_{t-1}^2})X_{t-2} + 0.2\varepsilon_t \tag{3}$$

Figure 1 shows the estimation results of a simulated series from (3) with

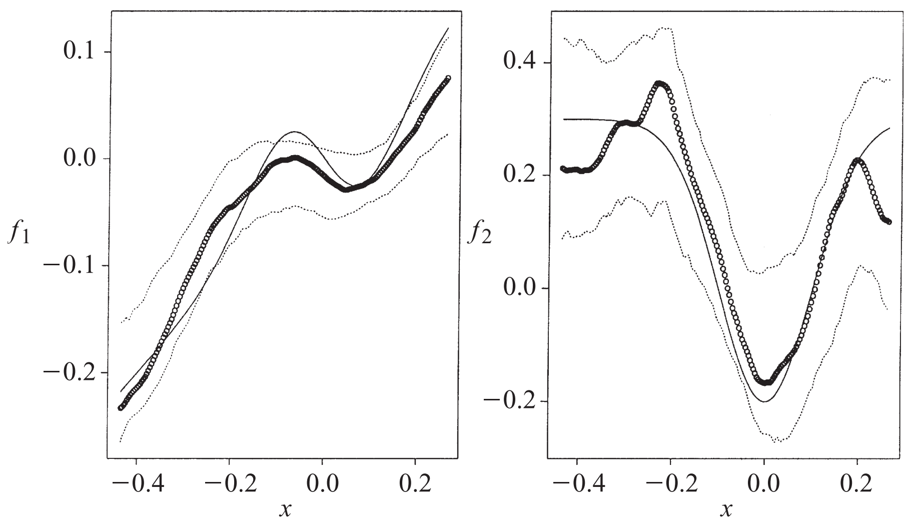

FIGURE 1. Function estimation of a simulated series from model (3): the dots are the estimated functions; the solid lines are the true functions; and the dashed lines are the pointwise confidence band

© Blackwell Publishers Ltd 2001

$\varepsilon_t \sim N(0, 1)$ and 400 samples using the optimal CV bandwidth with the quartic kernel function

$$K(u) = 0.9375(1 - u^2)^2 \mathrm{I}(|u| \leqslant 1)$$

Since X_{t-1} is both the threshold variable and one of the AR lags, we treat this model as

$$X_t = f_0(X_{t-1}) + f_1(X_{t-1})X_{t-2}$$

The dots in Figure 1 are the estimated functions, the solid lines are the true functions, and the dashed lines are the bootstrap 95% pointwise confidence band. The detailed description of the bootstrap confidence band is given in Section 4.

Figure 2 shows the CV curves of five series generated from the above process. We can see that the procedure is reasonably robust.

3. TESTING LINEARITY IN FAR MODELS

When all the coefficient functions are constant functions, an FAR model becomes a linear AR model. In this section, we develop a testing procedure to determine if the underlying process is linear.

Let $\hat{\mathbf{f}}(x)$ be the estimators in (2) and $\hat{\boldsymbol{\phi}} = [\hat{\phi}_1, \ldots, \hat{\phi}_p]$ be the Yule–Walker estimator of a linear AR(p) model. We define statistic T to test linearity:

$$T = \frac{1}{n} \sum_{t=l+1}^{n} \mathbf{d}_t' \mathbf{d}_t w(X_{t-d}) \tag{4}$$

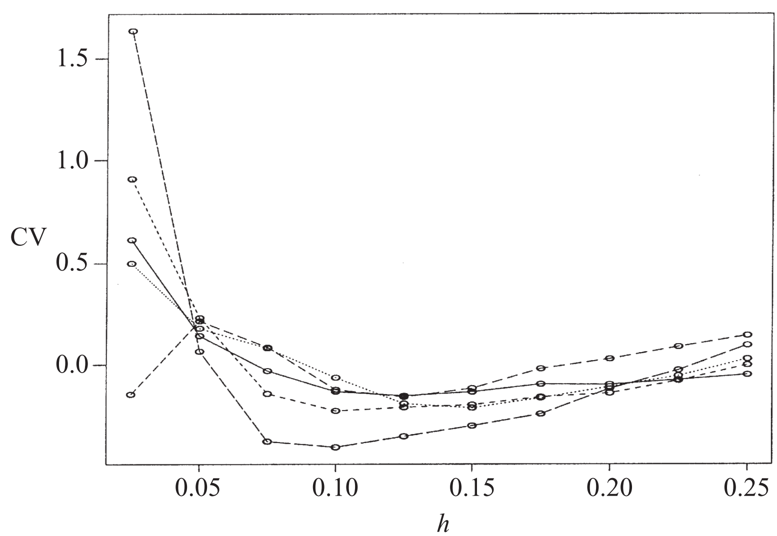

FIGURE 2. CV curves for five simulated series from model (3)

© Blackwell Publishers Ltd 2001

where

$$\mathbf{d}_t = (\mathbf{X}'\mathbf{W}_t\mathbf{X})(\hat{\mathbf{f}}(X_{t-d}) - \hat{\boldsymbol{\phi}})$$

with $\mathbf{W}_t$ being a diagonal matrix with diagonal element being $K_h(X_{i-d} - X_{t-d})$ for $i = l+1, \ldots, n$. The weight function w has a compact support, designed to reduce the boundary effects on the test statistic. Note that, traditionally, one would use $\mathbf{d}_t = \hat{\mathbf{f}}(X_{t-d}) - \hat{\boldsymbol{\phi}}$ in (4). The use of $\mathbf{d}_t$, in its current form, is purely for the simplicity and weaker conditions in obtaining the asymptotic distribution of T. We have the following theorem for the asymptotic distribution of the test statistic T.

THEOREM 2. *Under conditions A2, B1 and B2 listed in the Appendix, and the null hypothesis that $f_j(x) = \phi_j$, for $j = 1, \ldots, p$, with all the roots of $z^p - \phi_1 z^{p-1} - \cdots - \phi_p = 0$ inside the unit circle, we have*

$$nh^{1/2}T \xrightarrow{\mathrm{D}} \mathrm{N}(h^{-1/2}a_0, s_0^2)$$

where

$$a_0 = 2K_2^2\sigma^2 E\left[\sum_{k=1}^{p} X_{t-k}^2 p(X_{t-d})w(X_{t-d})\right]$$

and

$$s_0^2 = \sigma^4 \int K(u)K(v)K(u-z)K(v-z)\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}z \int s^4(x)p^2(x)w^2(x)\,\mathrm{d}x$$

where

$$s^2(x) = E\left[\left(\sum_{k=1}^{p} X_{t-k}\right)^2 \middle| X_{t-d} = x\right]$$

THEOREM 3. *Under the conditions A1–A4, B1–B2 listed in the Appendix, and the alternative hypothesis that at least one of the coefficient functions $f_j(x)$ are not constant, then*

$$T \xrightarrow{\mathrm{D}} \mathrm{N}(a_1, s_1^2/n)$$

where

$$a_1 = E[(\mathbf{f}(X_t) - \boldsymbol{\phi})'\mathbf{A}(X_t)\mathbf{A}(X_t)(\mathbf{f}(X_t) - \boldsymbol{\phi})w(X_t)]$$

and

$$s_1^2 = \mathrm{Var}[(\mathbf{f}(X_t) - \boldsymbol{\phi})'\mathbf{A}(X_t)\mathbf{A}(X_t)(\mathbf{f}(X_t) - \boldsymbol{\phi})w(X_t)]$$

where

$$\mathbf{A}(x) = p(x)E[\mathbf{X}_t\mathbf{X}_t'|X_{t-d} = x]$$

© Blackwell Publishers Ltd 2001

and $\boldsymbol{\phi}$ is the coefficients of the best linear prediction of X_t given $X_{t-1}, \ldots, X_{t-p}$.

The theorem shows that as $nh \to \infty$, T goes to zero in probability under the null hypothesis. Hence, a large value of the statistic indicates departure from linearity. It also shows that, under the null hypothesis, $nh^{1/2}T$ is asymptotic normal with finite variance, but the mean goes to infinity. This type of results were observed by Härdle and Mammen (1993), Hjellvik *et al*. (1997) in similar problems. The proof of the theorems basically follows similar proofs in Yoshihara (1976) and Hjellvik *et al*. (1997). First, we obtain the Hoeffding's decomposition of the test statistic. Then, a martingale central limit theorem is used on the resulting U-statistic. The proof is tedious, and hence is not presented here.

Although Theorem 2 can be used to obtain asymptotic level of the test statistic, it is noted by many researchers, e.g. Skaug and Tjøstheim (1993) and Hjellvik and Tjøstheim (1995, 1996) that, in finite samples, the asymptotic level does not perform well in most cases. Hence, for practical purposes, we suggest to use bootstrap procedures. Specifically, we first obtain residuals

$$\hat{e}_t = X_t - \sum_{i=1}^{p} \hat{\phi}_i X_{t-i}$$

where $\hat{\phi}_i$, $i = 1, \ldots, p$ are the Yule–Walker estimates of a linear AR(p) model fitted to the data. Then, we create bootstrap versions of the process

$$X_t^* = \sum_{i=1}^{p} \hat{\phi}_i X_{t-i}^* + e_t^*$$

for $t = l + 1, \ldots, n$, where e_t^* are independently sampled from $\{\hat{e}_{l+1}, \ldots, \hat{e}_n\}$ with replacement and $X_t^* = X_t$ for $t = 1, \ldots, l$. Then, a bootstrap value of the test statistic T^* is obtained by replacing X_t^* in the place of X_t in calculating the test statistic (4). Then the bootstrap null distribution can be obtained.

In Table 1, we present a small scale simulation for checking the performance of the proposed tests. In addition to model (3), we include five other models:

TABLE I

PERCENTAGE OF REJECTION OF THE LINEARITY TEST

α	AR(2)	AR(1)	EXPAR(2)	TAR	STAR	EXPAR(1)
0.10	10	4	96	100	94	40
0.05	8	4	94	100	90	24
0.01	2	2	84	90	74	10

© Blackwell Publishers Ltd 2001

$$\text{AR(2):} \qquad X_t = 0.6X_{t-1} - 0.3X_{t-2} + \varepsilon_t \tag{5}$$

$$\text{AR(1):} \qquad X_t = 0.5X_{t-1} + \varepsilon_t \tag{6}$$

$$\begin{aligned}\text{TAR:} \qquad X_t &= (0.4 - 1.0I(X_{t-1} > 0))X_{t-1} \\ &\quad + (-0.8 + 1.0I(X_{t-1} > 0))X_{t-2} + \varepsilon_t\end{aligned} \tag{7}$$

$$\begin{aligned}\text{STAR:} \qquad X_t &= \left(0.5 - \frac{1.1}{1 + \mathrm{e}^{-2X_{t-1}}}\right)X_{t-1} \\ &\quad + \left(0.3 - \frac{0.5}{1 + \mathrm{e}^{-2X_{t-1}}}\right)X_{t-2} + \varepsilon_t\end{aligned} \tag{8}$$

$$\text{EXPAR(1):} \qquad X_t = 0.5 - 1.1\mathrm{e}^{-50X_{t-1}^2} + 0.2\varepsilon_t \tag{9}$$

For each model, we generated 50 series of size 400. We used 50 bootstrap replications to obtain the bootstrap null distribution. For each model, we used a common bandwidth obtained by averaging five CV bandwidths of five simulated samples. Table 1 presents the percentage of rejection of the null hypothesis under three different α levels. From the table, we can see that the proposed testing procedure works reasonably well.

4. TESTING THRESHOLD TYPE OF DISCONTINUITY

First, we define local linear estimates of the coefficient functions using one-sided kernels. Let

$$\mathbf{f}^+(x) = \lim_{\delta \to 0+} \mathbf{f}(x + \delta) \qquad \text{and} \qquad \mathbf{f}^-(x) = \lim_{\delta \to 0-} \mathbf{f}(x + \delta)$$

Define

$$\hat{\mathbf{f}}^+(x) = \arg\min_{\boldsymbol{\beta}} \sum_{t=l+1}^{n} (X_t - \mathbf{X}_t'\boldsymbol{\beta})^2 K_h^+(X_{t-d} - x)$$

and

$$\hat{\mathbf{f}}^-(x) = \arg\min_{\boldsymbol{\beta}} \sum_{t=l+1}^{n} (X_t - \mathbf{X}_t'\boldsymbol{\beta})^2 K_h^-(X_{t-d} - x)$$

where

$$K_h^+(u) = 2K_h(u)I(u \geqslant 0)$$

and

$$K_h^-(u) = 2K_h(u)I(u \leqslant 0)$$

© Blackwell Publishers Ltd 2001

where K is a symmetric kernel function with bounded support and $\int K(u)\mathrm{d}u = 1$. Let

$$\mu_1^+ = \int uK^+(u)\mathrm{d}u \qquad \text{and} \qquad \mu_1^- = \int uK^-(u)\mathrm{d}u$$

We have the following theorem:

THEOREM 4. *Under conditions A1–A4 and C1–C2 listed in the Appendix, we have*

(i) $\hat{\mathbf{f}}^+(x)$ *and* $\hat{\mathbf{f}}^-(x)$ *are asymptotically uncorrelated.*
(ii)

$$n^{1/3}(\hat{\mathbf{f}}^+(x) - \mathbf{f}^+(x) - \beta\mathbf{b}^+(x)) \xrightarrow{\mathrm{D}} \mathrm{N}(0, \beta^{-1}\sigma^2 K_2^2 \mathbf{A}^{-1}(x))$$

and

$$n^{1/3}(\hat{\mathbf{f}}^-(x) - \mathbf{f}^-(x) - \beta\mathbf{b}^-(x)) \xrightarrow{\mathrm{D}} \mathrm{N}(0, \beta^{-1}\sigma^2 K_2^2 \mathbf{A}^{-1}(x))$$

where

$$\mathbf{A}(x) = p(x)E[\mathbf{X}_t\mathbf{X}_t'|X_{t-d} = x]$$

$$\mathbf{b}^+(x) = \mathbf{A}^{-1}(x)\mathbf{B}^+(x)$$

$$\mathbf{b}^-(x) = \mathbf{A}^{-1}(x)\mathbf{B}^-(x)$$

with $\mathbf{B}^+(x)$ *and* $\mathbf{B}^-(x)$ *being vectors with ith element being*

$$\mu_1^+ \sum_{j=1}^{p} f_j'^{(+)}(x)E[X_{t-i}X_{t-j}|X_{t-d} = x] \qquad \text{and}$$

$$\mu_1^- \sum_{j=1}^{p} f_j'^{(-)}(x)E[X_{t-i}X_{t-j}|X_{t-d} = x]$$

respectively, where $f_j'^{(+)}$ *and* $f_j'^{(-)}$ *are the left and right derivatives of the function* f_j *at point* x.

The proof of the theorem is similar to that of Theorem 1. A brief discussion is given in the Appendix. Note that the convergence rate is lower than that of the two-sided estimator. Similar results were obtained by Cline and Hart (1991) for density estimation. Müller (1992) proposed similar procedures and investigated their asymptotic properties in a regression setting.

In Figure 3, we present the estimated coefficient functions using one-sided and two-sided kernels from a simulated TAR(2) series of (7). The sample size used is 400. Again, the quartic kernel is used, with bandwidth $h = 1.5$ for one-sided kernels and $h = 0.75$ for the two-sided kernel. We can see that, away

© Blackwell Publishers Ltd 2001

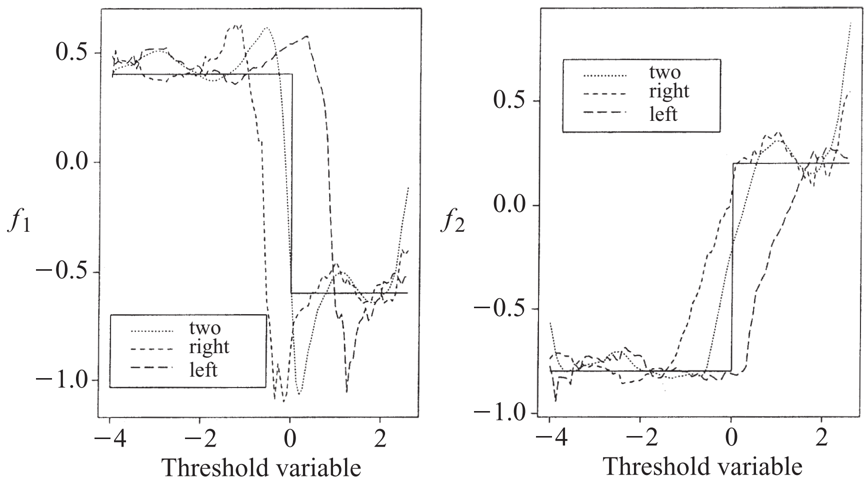

FIGURE 3. Function estimates using two-sided, left and right sided kernels from a simulated series of model (7). The solid lines are the true functions

from the discontinuous point ($x = 0$), both one-sided and two-sided estimates work well. Note that, in the TAR case, away from the discontinuous point, the functions are constant, hence there is no bias in those estimates. Thus there is not much difference between one-sided and two-sided estimates. Around the discontinuous point, $\hat{\mathbf{f}}^{+}$ is consistent right of the point, $\hat{\mathbf{f}}^{-}$ is consistent left of the point and the two-sided estimate is not consistent. In Figure 4, we plotted

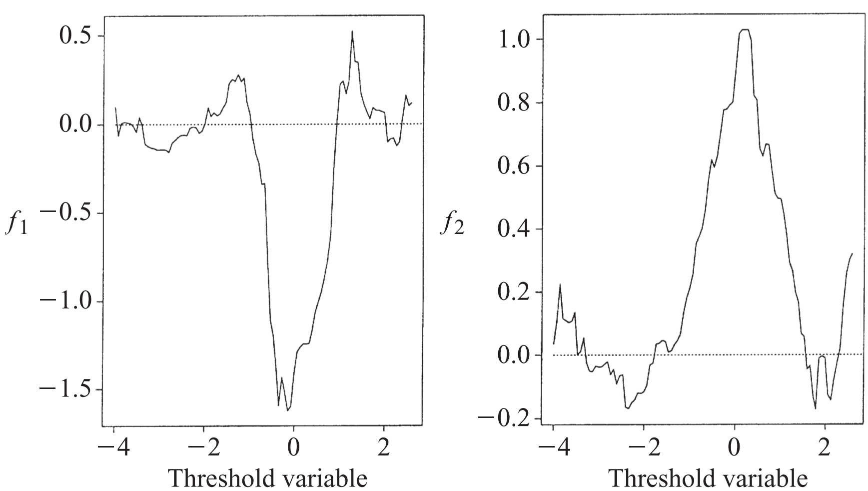

FIGURE 4. The difference of function estimates using left- and right-side kernels from a simulated series of model (7)

© Blackwell Publishers Ltd 2001

$\hat{f}_1^+(x) - \hat{f}_1^-(x)$ and $\hat{f}_2^+(x) - \hat{f}_2^-(x)$. We can see that around the discontinuity point $x = 0$, the differences between the two estimated functions are the largest. Hence, we suggest to use statistic s to test threshold type of nonlinearity:

$$S = \sup_{x \in \mathscr{D}} \max_{j=1,\ldots,p} |\hat{f}_j^+(x) - \hat{\mathrm{f}}_j^-(x)|$$

where $\mathscr{D}$ is a compact interval of interest. Since one-sided kernels have severe boundary effects, in our simulation studies we choose $\mathscr{D}$ be the interval between 20 and 80 percentile of the data range, for samples of size of 400.

THEOREM 5. *Under conditions A1–A3, C1–C4 and the null hypothesis that the functions are all continuous, we have*

$$S \xrightarrow{\mathrm{P}} 0$$

and, if there is a discontinuity point in $\mathscr{D}$, then

$$S \xrightarrow{\mathrm{P}} \max_{j=1,\ldots,p} \sup_{x \in \mathscr{D}} |f_j^+(x) - f_j^-(x)| > 0$$

The above theorem shows that, under the null hypothesis, the test statistic goes to zero in probability. Hence, a large value of the statistic indicates tendency of departing from the null hypothesis. If the null hypothesis is rejected, the function

$$\max_{j=1,\ldots,p} |\hat{f}_j^+(x) - \hat{f}_j^-(x)|$$

can be used to estimate the location of the threshold. Note that, in practice, we have to use a grid in $\mathscr{D}$ to calculate the statistic S and the grid must be finer than at least half of the bandwidth to obtain reasonable results.

The asymptotic distribution of the test statistic is very difficult to obtain and may not be useful in practice. So, again, we use bootstrap approaches. However, there are several difficulties. Note that, to construct bootstrapped version of the data under the null hypothesis (that the functions are continuous in the interval of consideration), we must estimate the coefficient functions nonparametrically, e.g. using two-sided kernels as in Section 2. However, it is a local estimator, and hence cannot be used outside the data range. In addition, it suffers from the boundary effects. So, if ones tries to construct a bootstrap version of the time series using

$$X_t^* = \sum_{i=1}^{p} \hat{f}_i(X_{t-d}^*) X_{t-i}^* + \varepsilon_t^*$$

where ε_t^* is sampled from the residuals

$$\hat{\varepsilon}_t = X_t - \sum_{i=1}^{p} \hat{f}_i(X_{t-d}) X_{t-i}$$

© Blackwell Publishers Ltd 2001

not only is the residual distribution not correct (due to the boundary effects), but also the generated threshold variable X^*_{t-d} may be out of the range of reliable estimate of f_i. To overcome these two difficulties, we convert the problem to a regression setting. Specifically, we fix the original design matrix using the original data and bootstrap only the response. This is slightly different from our time series setting but, under strong mixing condition, the effect will be minimal. In addition, to reduce the boundary effect, we only bootstrap the observation with X_{t-d} within 10 and 90 percentile of the data range. For data outside the range, we always use the original observation. We also only resample from the residuals obtained within the same region.

We performed a small-scale simulation to check the performance of the proposed testing method. In addition to models (5), (3), (8) and (7), we also tried four more models:

$$\text{STAR-2:} \qquad X_t = \left(0.5 - \frac{1.1}{1 + \mathrm{e}^{X_{t-1}}}\right)X_{t-1} + \left(0.3 - \frac{0.5}{1 + \mathrm{e}^{X_{t-1}}}\right)X_{t-2} + \varepsilon_t \quad (10)$$

$$\text{TAR-2:} \qquad X_t = (0.4 - 0.5I(X_{t-1} > 0))X_{t-1} + (-0.8 + 0.5I(X_{t-1} > 0))X_{t-2} + \varepsilon_t \quad (11)$$

$$\text{TSIN:} \qquad X_t = (1 - 2I(X_{t-1} > 0))\cos(0.5\pi X_{t-1})X_{t-1} + (-0.5 + I(X_{t-1} > 0))\cos(\pi x_{t-1})x_{t-2} + 0.2\varepsilon_t \quad (12)$$

$$\text{TSIN-2:} \qquad X_t = (0.5 - I(X_{t-1} > 0))\cos(0.5\pi X_{t-1})X_{t-1} + (-0.2 + 0.4I(X_{t-1} > 0))\cos(\pi X_{t-1})X_{t-2} + 0.2\varepsilon_t \quad (13)$$

Table 2 shows the percentage of rejections of the null hypothesis that the coefficient functions are continuous, in 50 simulated samples, each of sample size 400. The differences between TAR and TAR-2, TSIN and TSIN-2 are the jump size. We can see that, with smaller jump size (TAR2 and TSIN-2), the power of the test is smaller, as expected.

TABLE II
PERCENTAGE OF REJECTIONS OF THE CONTINUITY TEST

α	AR(2)	EXPAR(2)	STAR	STAR-2	TAR	TAR-2	TSIN	TSIN-2
0.10	14	6	6	2	96	30	48	40
0.05	6	2	4	2	92	24	38	34
0.01	2	0	2	0	84	12	22	18

© Blackwell Publishers Ltd 2001

5. THE CHICKENPOX EXAMPLE

In this section, we analyze the monthly record of chickenpox cases in New York city from January 1928 to June 1972 (Sugihara and May, 1990). Following Chen and Tsay (1993), we make a log transformation $X_t = \log(Y_t + 743)$ to stabilize the variability. The display of the data is shown in Figure 5.

In Chen and Tsay (1993), the threshold lag X_{t-12} is identified. This is a natural choice since the data demonstrated nonlinear seasonal pattern. Using CV and post-sample forecast, we identify the AR lags 1, 3, 9, 24 are significant which resulted in the following FAR model

$$X_t = f_1(X_{t-12})X_{t-1} + f_3(X_{t-12})X_{t-3} + f_9(X_{t-12})X_{t-9} + f_{24}(X_{t-12})X_{t-24} + \varepsilon_t \tag{14}$$

To select the optimal bandwidth, the CV criterion specified in Section 2 is employed. Figure 6 plots the CV(h) versus the bandwidth h, which shows that the optimal bandwidth is $\hat{h} = 0.27$.

The linearity test proposed in Section 3 is applied to the dataset using the optimal bandwidth. It rejects the linearity overwhelmingly with bootstrap p-value less than 0.01. On the other hand, the continuity test in Section 4 fails to reject the continuity assumption, with p-value at 0.54. Hence, we estimate model (14) with a two-sided kernel with $\hat{h} = 0.27$, using the estimation procedure in section 2. Figure 7 shows the estimated functions with the pointwise confidence band.

To evaluate the performance of this nonparametric model, we compare the

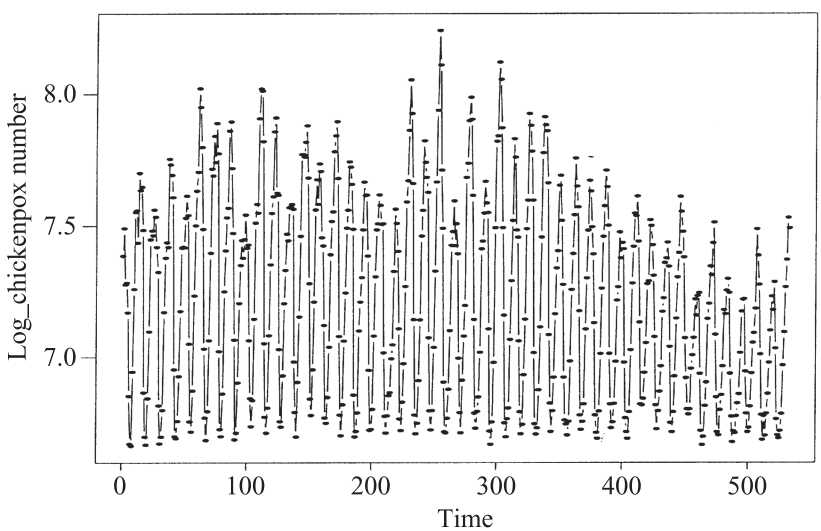

FIGURE 5. Log transformed chickenpox series

© Blackwell Publishers Ltd 2001

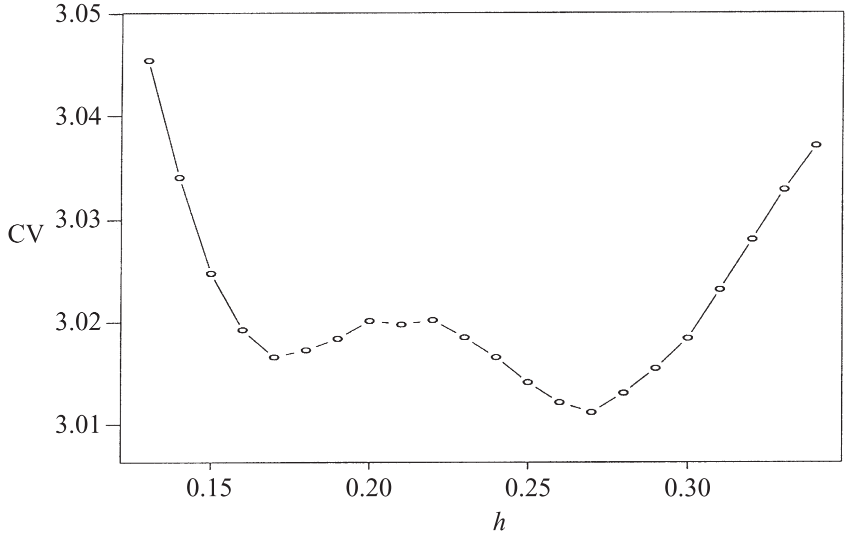

FIGURE 6. Plot of CV versus bandwidth of the chickenpox example

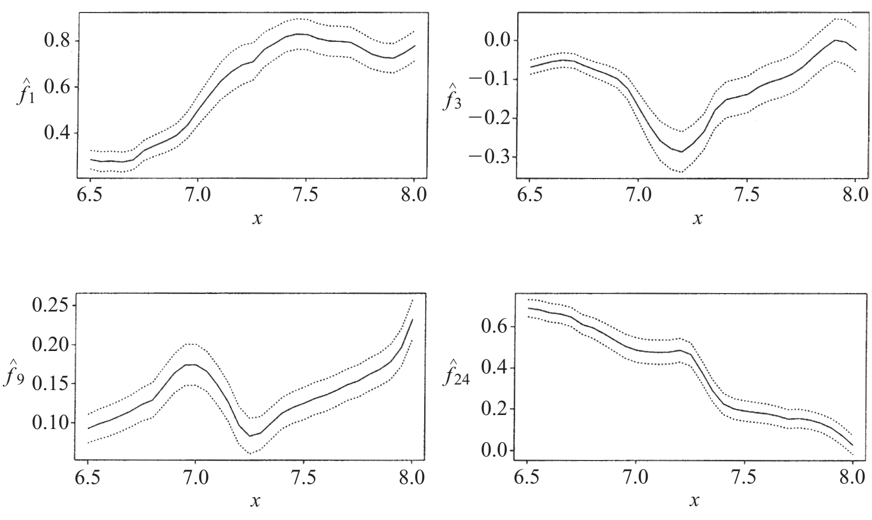

FIGURE 7. Estimated functions with pointwise confidence band of the chickenpox example

out-sample multi-step forecasting performance with the parametric threshold AR (TAR) model found in Chen and Tsay (1993). Their model assumed discontinuity of the coefficient functions and used three threshold regimes, with total 16 AR coefficient and two threshold values. They have shown that their model is significantly better than the standard linear AR model and the seasonal

© Blackwell Publishers Ltd 2001

TABLE III

MULTI-STEP PREDICTION PERFORMANCE COMPARISON AMONG THE FAR MODEL, THE TAR MODEL AND THE SEASONAL AR MODEL

lead time	1	2	3	4	5	6	7	8	9	10	11
FAR vs TAR	0.101	0.142	0.082	0.063	0.066	0.073	0.052	0.026	0.001	0.016	0.031
FAR vs SAR	−0.027	0.091	0.245	0.392	0.496	0.550	0.579	0.590	0.592	0.609	0.613
TAR vs SAR	−0.143	0.059	0.177	0.352	0.461	0.515	0.556	0.580	0.592	0.602	0.600

AR (SAR) model in multi-step forecasts. Here, we repeat their experiment, using the first 450 observations to estimate the model and calculating 1–11-step ahead post-sample forecasting error using the last 70 observations. Table 3 shows the forecasting performance comparison among the nonparametric functional coefficient AR model, the parametric threshold AR (TAR) model and a seasonal AR model (SAR). The mean squared forecasting error improvement is calculated as follows:

$$\text{FAR vs TAR:} \quad \frac{\text{MSE(TAR)-MSE(FAR)}}{\text{MSE(TAR)}}$$

$$\text{FAR vs SAR:} \quad \frac{\text{MSE(SAR)-MSE(FAR)}}{\text{MSE(SAR)}}$$

$$\text{TAR vs SAR:} \quad \frac{\text{MSE(SAR)-MSE(TAR)}}{\text{MSE(SAR)}}$$

From Table 3, we can see that in short-term forecasting, the nonparametric FAR model assuming continuity outperforms the threshold AR model between 6% and 14%, but with limited improvement in the longer-term forecasting. Note that the TAR model significantly outperforms the seasonal AR in longer-term forecasting, but underperforms in short-term forecasting. On the other hand, the FAR model consistently outperforms the seasonal AR model, in both short-term and longer-term forecasting, except the one-step ahead forecast in which the difference is minimal. This example demonstrated the flexibility and superiority of this class of models, and the importance of linearity and continuity tests.

APPENDIX

First we list all the necessary assumptions for the theorems:

A1 The process is geometrically ergodic. A set of sufficient ergodic conditions for the FAR model can be found in Chen and Tsay (1993) and Cline and Pu (1995).

© Blackwell Publishers Ltd 2001

A2 The kernel function K is a positive, compactly supported bounded function, with

$$\int K(u)\mathrm{d}u = 1$$

$$\int uK(u)\mathrm{d}u = 0$$

and

$$|K(x_1) - K(x_2)| < c|x_1 - x_2|$$

for all x_1 and x_2 in its support.

A3 The density of the stationary distribution exists and is bounded.

A4 The matrix

$$\mathbf{A}(y) = p(y)E[\mathbf{X}_t\mathbf{X}_t'|X_{t-d} = y]$$

is of full rank. $\mathbf{A}(y)$ and $\mathbf{A}^{-1}(y)$ are bounded element-wise in a neighbourhood of x.

A5 Let $p_{i,j,d}$ be the joint density of $(X_{t-i}, X_{t-j}, X_{t-d})$. We assume that $p_{i,j,d}$ has Holder continuous first partial derivative with respect to the third argument.

A6 The second derivative of the coefficient functions exists and is Holder continuous.

A7 The term

$$\int uv\left\{\frac{1}{2}f_j''(y)p_{i,j,d}(u, v, y) + f_j'(y)p_{i,j,d}'(u, v, y)\right\}\mathrm{d}u\,\mathrm{d}v$$

is bounded in a neighbourhood of x for all $1 \leqslant i,\ j \leqslant p$.

A8 $h = \beta n^{-1/5}$ for $\beta > 0$.

A9 $h = \beta n^{-1/5}$ for $\beta > 0$.

B1 The joint density of distinct elements of $\{X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4}, X_{t_5}, X_{t_6}, X_{t_7}, X_{t_8}, X_{t_9}\}$ is continuous and bounded by a constant independent of t_i, for $i = 1, \ldots, 9$.

B2 As $n \to \infty$, then $h \to 0$ and

$$\frac{nh^{(2+4\delta)/(1+\delta)}}{\log n} \to \infty$$

C1 $f_j^{\prime(+)}$ and $f_j^{\prime(-)}$ exist and are Holder continuous in $(x, x+\delta)$ and $(x-\delta, x)$, respectively.

C2 $h = \beta n^{-1/3}$ for $\beta > 0$.

C3 The matrix

$$\mathbf{A}(y) = p(y)E[\mathbf{X}_t\mathbf{X}_t'|X_{t-d} = y]$$

is of full rank. $\mathbf{A}(y)$ and $\mathbf{A}^{-1}(y)$ are bounded uniformly in a compact interval $\mathscr{D}$ of interest.

C4 $f_j^{\prime(+)}$ and $f_j^{\prime(-)}$ exist and are Holder continuous in $\mathscr{D}$.

We need the following lemmas.

LEMMA 1 *(Liptser and Shirjaev, 1980, Corollary 6)*
Denote $\mathscr{F}_k$ be a σ-field. Let, for every $n > 0$, the sequence $(\eta_{nk}, \mathscr{F}_k)$ be a square integrable martingale difference, i.e.

$$E(\eta_{nk}|\mathscr{F}_{k-1}) = 0 \qquad \text{and} \qquad E(\eta_{nk}^2) < \infty, \qquad \text{for } 1 \leqslant k \leqslant n$$

and let

© Blackwell Publishers Ltd 2001

$$\sum_{i=1}^{n} E(\eta_{nk}^2) = 1 \qquad \text{for any } n \geqslant n_0 \geqslant 0$$

The conditions

$$\sum_{k=1}^{n} E(\eta_{nk}^2 | \mathscr{F}_{k-1}) \xrightarrow{\mathrm{P}} 1 \qquad \text{as} \quad n \to \infty$$

$$\sum_{k=1}^{n} E(\eta_{nk}^2 I(|\eta_{nk}| > \varepsilon | \mathscr{F}_{k-1}) \xrightarrow{\mathrm{P}} 0 \qquad \text{as} \quad n \to \infty$$

for $\varepsilon > 0$, are necessary and sufficient for convergence

$$\sum_{k=1}^{n} \eta_{nk} \xrightarrow{\mathrm{D}} \mathrm{N}(0,\ 1)$$

LEMMA 2. *Let p be the stationary density of X_t. Under condition (A1) (geometric ergodic), we have*

$$n^{-1} \sum_{t=l+1}^{n} X_{t-i} X_{t-j} K_h(X_{t-d} - x) - E[X_{t-i} X_{t-j} | X_{t-d} = x] p(x) = o_p(1)$$

and

$$n^{-1} \sum_{t=l+1}^{n} E[X_{t-i} X_{t-j} K_h(X_{t-d} - x)] - E[X_{t-i} X_{t-j} | X_{t-d} = x] p(x) = o(1)$$

PROOF. By an ergodic theorem, we have

$$n^{-1} \sum_{t=l+1}^{n} X_{t-i} X_{t-j} K_h(X_{t-d} - x) - E[X_{t-i} X_{t-j} K_h(X_{t-d} - x)] = o_p(1)$$

Let $p_{i,j,d}$ be the joint density of $(X_{t-i},\ X_{t-j},\ X_{t-d})$. We have

$$\begin{aligned} E[X_{t-i} X_{t-j} K_h(X_{t-d} - x)] &= \int uvK(w) p_{i,j,d}(u,\ v,\ x + hw)\, \mathrm{d}u\, \mathrm{d}v\, \mathrm{d}w \\ &= \int uv p_{i,j,d}(u,\ v,\ x)\, \mathrm{d}u\, \mathrm{d}v\, \mathrm{d}w (1 + o(1)) \\ &= p(x) E[X_{t-i} X_{t-j} | X_{t-d} = x](1 + o(1)) \end{aligned}$$

PROOF OF THEOREM 1. Let $n^* = n - l$ and $\boldsymbol{\varepsilon} = [\varepsilon_{t+1},\ \ldots,\ \varepsilon_n]'$. Then

$$\hat{\mathbf{f}}(x) - \mathbf{f}(x) = (\mathbf{X}'\mathbf{W}_x\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_x[\mathbf{Y} - \boldsymbol{\varepsilon} - \mathbf{X}\mathbf{f}(x)] + (\mathbf{X}'\mathbf{W}_x\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_x\boldsymbol{\varepsilon} \equiv I_1 + I_2$$

First, we work with $(n^*)^{-1}\mathbf{X}'\mathbf{W}_x\mathbf{X}$. The $(i,\ j)$th element of $(n^*)^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}$ is

$$\frac{1}{n^*} \sum_{t=l+1}^{n} X_{t-i} X_{t-j} K_h(X_{t-d} - x) = p(x) E[X_{t-i} X_{t-j} | X_{t-d} = x](1 + o_p(1))$$

by Lemma 2. Hence

$$(n^*)^{-1}\mathbf{X}'\mathbf{W}_x\mathbf{X} = \mathbf{A}(x)(1 + o_p(1))$$

where

© Blackwell Publishers Ltd 2001

$$\mathbf{A}(x) = p(x)E[\mathbf{X}_t\mathbf{X}_t'|X_{t-d} = x]$$

Next, the ith element of $(n^*)^{-1}\mathbf{X}'\mathbf{W}_x(\mathbf{Y} - \boldsymbol{\varepsilon} - \mathbf{Xf}(x))$ is

$$\begin{aligned}
&\frac{1}{n^*}\sum_{t=l+1}^{n} X_{t-i}K_h(X_{t-d} - x)\sum_{j=1}^{p}\{f_j(X_{t-d}) - f_j(x)\}X_{t-j} \\
&= \sum_{j=1}^{p}\frac{1}{n^*}\sum_{t=l+1}^{n} X_{t-i}X_{t-j}K_h(X_{t-d} - x)\{f_j(X_{t-d}) - f_j(x)\} \\
&= \sum_{j=1}^{p}\int uvK(w)\{f_j(x + hw) - f_j(x)\}p_{i,j,d}(u, v, x + hw)\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w(1 + o_p(1)) \\
&= \mu_2 h^2\sum_{j=1}^{p}\int uv\left\{\frac{1}{2}f''(x)p_{i,j,d}(u, v, x) + f'(x)p'_{i,j,d}(u, v, x)\right\}\mathrm{d}u\,\mathrm{d}v(1 + o_p(1)) \\
&= \mu_2 h^2 B_i(x)(1 + o_p(1))
\end{aligned}$$

The second equality is the result of the ergodic theorem. Let $\mathbf{B}(x) = [B_1(x), \ldots, B_p(x)]'$, then

$$I_1 = \mu_2 h^2\mathbf{A}^{-1}(x)\mathbf{B}(x)(1 + o_p(1))$$

Now we work with I_2. The ith element of $(n^*)^{-1}\mathbf{X}'W_x\boldsymbol{\varepsilon}$ is

$$e_i = (n^*)^{-1}\sum_{t=l+1}^{n} X_{t-i}K_h(X_{t-d} - x)\varepsilon_t$$

We show that e_i is asymptotically normal by checking all the conditions of Lemma 1. First, since ε_t is independent of X_s for all $s < t$, we have $E(e_i) = 0$. Standard calculation yields

$$\begin{aligned}
s_i^2 = \mathrm{Var}(\varepsilon_i) &= \frac{1}{n^{*2}}\sigma^2\sum_{t=l+1}^{n} E[X_{t-i}^2K_h^2(X_{t-d} - x)] \\
&= \frac{1}{n^*h}\sigma^2K_2^2E[X_{t-i}^2|X_{t-d} = x]p(x)(1 + o(1))
\end{aligned}$$

where $p_{i,d}$ is the joint density of X_{t-i} and X_{t-d}. Define

$$\eta_t = \frac{1}{n^*}\frac{X_{t-i}K_h(X_{t-d} - x)}{s_i}\varepsilon_t$$

Note that η_t actually depends on i. For brevity, we here work with a fixed i and suppress the index i on η. Let $\mathscr{F}_t$ be the σ-field generated by $[X_1, \ldots, X_t]$. Since ε_t is independent of X_s for all $s < t$, we have

© Blackwell Publishers Ltd 2001

$$E[\eta_t|\mathscr{F}_{t-1}] = 0$$

$$E[\eta_t^2] = E\left[\frac{X_{t-i}^2 K_h^2(X_{t-d}-x)}{n^{*2} s_i^2}\varepsilon_t^2\right] < \infty$$

$$\sum_{t=l+1}^{n} E[\eta_t^2] = \frac{\sum_{t=l+1}^{n} E[X_{t-i}^2 K_h^2(X_{t-d}-x)\varepsilon_t^2]}{n^{*2} s_i^2} = 1$$

$$\sum_{t=l+1}^{n} E[\eta_t^2|\mathscr{F}_{t-1}] = \frac{\sum_{t=l+1}^{n} X_{t-i}^2 K_h^2(X_{t-d}-x)\sigma^2}{n^{*2} s_i^2} \to 1 \qquad \text{as} \quad n\to\infty$$

Finally, for any $\epsilon > 0$, we want to show

$$\sum_{t=l+1}^{n} E[\eta_t^2 I(|\eta_t| > \epsilon)|\mathscr{F}_{t-1}] = \sum_{t=l+1}^{n} \frac{X_{t-i}^2 K_h^2(X_{t-d}-x) E(\varepsilon_t^2 I(|\eta_t| > \epsilon)|\mathscr{F}_{t-1})}{n^{*2} s_i^2(x)} = o(1)$$

For some constants C_1 and C, we have

$$\begin{aligned} E[\varepsilon_t^2 I(|\eta_t| > \epsilon)] &\leqslant \{E(\varepsilon_t^4) E[I(|\eta_t| > \epsilon)]\}^{1/2} \\ &\leqslant C_1\left[P\left(|\varepsilon_t| > \frac{n^* s_i \epsilon}{|X_{t-i}| K_h(X_{t-d}-x)}\right)\right]^{1/2} \\ &\leqslant C_1 \frac{\sigma |X_{t-i}| K_h(X_{t-d}-x)}{n^* s_i \epsilon} \end{aligned}$$

Hence,

$$\left|\sum_{t=l+1}^{n} E[\eta_t^2 I(|\eta_t| > \epsilon)|\mathscr{F}_{t-1}]\right| \leqslant C \frac{\sum_{t=l+1}^{n} |X_{t-i}|^3 K_h^3(X_{t-d}-x)\sigma}{n^3 s_i^3 \epsilon} = o_p(1)$$

Then by Lemma 1, we have

$$\frac{1}{s_i}\left[\frac{1}{n^*}\sum_{t=l+1}^{n} X_{t-i} K_h(X_{t-d}-x)\varepsilon_t\right] \xrightarrow{\mathscr{D}} \mathrm{N}(0,\, 1)$$

Also, note that

$$\begin{aligned} &\frac{1}{n^{*2}} E\left[\left(\sum_{t=l+1}^{n} X_{t-i} K_h(X_{t-d}-x)\varepsilon_t\right)\left(\sum_{t=l+1}^{n} X_{t-j} K_h(X_{t-d}-x)\varepsilon_t\right)\right] \\ &= \frac{1}{nh^*}\sigma^2 K_2^2 E[X_{t-i} X_{t-j}|X_{t-d} = x]\, p(x)(1 + o(1)) \end{aligned}$$

By a Cramer–Wold device, it is easy to show that

$$(n^* h)\left(\frac{1}{n^*}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{\varepsilon}\right) \xrightarrow{\mathrm{D}} \mathrm{N}_p(0,\, \sigma^2 K_2^2 \mathbf{A}(x))$$

© Blackwell Publishers Ltd 2001

Hence

$$(n^* h)I_2 \xrightarrow{\mathrm{D}} \mathrm{N}_p(0, \sigma^2 K_2^2 \mathbf{A}^{-1}(x))$$

Let $h = \beta n^{-1/5}$, the theorem then follows.

PROOF OF THEOREM 4. The proof is essentially the same as that of Theorem 1 with slight differences in the bias calculation since $\int uK^+(u)\mathrm{d}u = \mu_1^+ \neq 0$. The asymptotic uncorrelation of $\hat{\mathbf{f}}^+(x)$ and $\hat{\mathbf{f}}^-(x)$ is due to the fact that the covariance of the ith element of $(n^*)^{-1}\mathbf{X}'\mathbf{W}_x^+\boldsymbol{\varepsilon}$ and the jth element of $(n^*)^{-1}\mathbf{X}'\mathbf{W}_x^-\boldsymbol{\varepsilon}$ is

$$\frac{1}{n^{*2}} \sum_{t_1=l+1}^{n} \sum_{t_2=l+1}^{n} E[X_{t_1-i}X_{t_2-j}K_h^+(X_{t_1-d} - x)K_h^-(X_{t_2-d} - x)\varepsilon_{t_1}\varepsilon_{t_2}]$$

Note that for $t_1 < t_2$, ε_{t_2} is independent of the rest of the terms. Since $E(\varepsilon_{t_2}) = 0$, all the terms with $t_1 \neq t_2$ are zero. On the other hand, for $t_1 = t_2$,

$$K_h^+(X_{t_1-d} - x)K_h^-(X_{t_1-d} - x) = 0$$

by definition. Hence $\hat{\boldsymbol{f}}^+(x)$ and $\hat{\boldsymbol{f}}^-(x)$ are asymptotically uncorrelated.

PROOF OF THEOREM 5. With the result of theorem 4, it is easy to show, by construction of finite open intervals, that for a compact interval,

$$\sup_{x\in\mathscr{D}} |\hat{f}_j^+(x) - f_j^+(x)| \xrightarrow{\mathrm{P}} 0 \qquad \text{and} \qquad \sup_{x\in\mathscr{D}} |\hat{f}_j^-(x) - f_j^-(x)| \xrightarrow{\mathrm{P}} 0$$

for all $j = 1, \ldots, p$. Hence

$$\begin{aligned} \Big|\sup_{x\in\mathscr{D}} |\hat{f}_j^+(x) - \hat{f}_j^-(x)| - \sup_{x\in\mathscr{D}} |f_j^+(x) - f_j^-(x)|\Big| &\leqslant \sup_{x\in\mathscr{D}} |\hat{f}_j^+(x) - f_j^+(x)| + \sup_{x\in\mathscr{D}} |\hat{f}_j^-(x) - f_j^-(x)| \\ &\xrightarrow{\mathrm{P}} 0 \end{aligned}$$

Thus,

$$\sup_{x\in\mathscr{D}} |\hat{f}_j^+(x) - \hat{f}_j^-(x)| \xrightarrow{\mathrm{P}} \sup_{x\in\mathscr{D}} |f_j^+(x) - f_j^-(x)|$$

The theorem then follows.

ACKNOWLEDGMENT

Rong Chen's research was supported in part by NSF grant DMS 9626113 and Sonderforschungsbereich 373 'Quantifikation und Simulation Ökonomischer Prozesse', Hombold-Universitate zu Berlin. Lon-Mu Liu's research was support in part by grants from the UIC Center for Research in Information Management (CRIM) and Scientific Computing Associates Corp. The authors thank the editor, an associate editor and two anonymous referees for their encouragement and helpful comments.

NOTES

1. Rong Chen and Lon-Mu Liu are both Professor, Department of Information and Decision Sciences (M/C 294), University of Illinois at Chicago, Chicago, IL 60607

© Blackwell Publishers Ltd 2001

REFERENCES

CHAN, K. S. and TONG, H. (1986) On estimating thresholds in autoregressive models. *J. Time Series Anal.* 7, 179–90.
CHEN, R. and TSAY, R. S. (1993) Functional-coefficient autoregressive models. *J. Am. Statist. Assoc.* 88, 298–308.
CHENG, B. and TONG, H. (1992) On consistent non-parametric order determination and chaos (with discussion). *J. R. Statist. Soc. B* 54, 451–74.
CLEVELAND, W. S. and DEVLIN, S. J. (1988) Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Statist. Assoc.* 83, 596–610.
CLINE, D. B. H. and HART, J. D. (1991) Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics* 22, 69–84.
CLINE, D. B. H. and PU, H. H. (1995) Geometric ergodicity of nonparametric nonlinear time series, *Technical report. Texas: Dept. of Statistics, Texas A & M University.*
FAN, J. and GIJBELS, I. (1996) *Local Polynomial Modeling and Its Applications*. UK: Chapman & Hall.
GRANGER, C. and TERÄSVIRTA, T. (1993) *Modeling Nonlinear Economic Relationships*. Oxford: Oxford University Press.
GYÖRFI, L., HÄRDLE, W., SARDA, P. and VIEU, P. (1989) Nonparametric Curve Estimation from Time Series. Lecture Notes in Statistics 60. Heidelberg: Springer-Verlag.
HAGGAN, V. and OZAKI, T. (1981) Modeling nonlinear vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* 68, 189–96.
HÄRDLE, W. and MAMMEN, E. (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics*. 21, 1926–47.
HÄRDLE, W. and VIEU, P. (1992) Kernel regression smoothing of time series. *J. Time Series Anal.* 13, 209–32.
HÄRDLE, W., LÜTKEPOHL, H. and CHEN, R. (1997) A review of nonparametric time series analysis. *International Statistical Review* 65, 49–72.
HART, J. D. (1996) Some automated methods of smoothing time-dependent data. *J. Nonparametric Statistics* 6, 115–42.
HASTIE, T. J. and TIBSHIRANI, R. J. (1993) Varying-coefficient models. *J. R. Statist. Soc. B.* 55, 757–96.
HJELLVIK, V. and TJØSTHEIM, D. (1995) Nonparametric tests of linearity for time series. *Biometrika* 82, 351–68.
HJELLVIK, V. and TJØSTHEIM, D. (1996) Nonparametric statistics for testing of linearity and serial independence. *J. Nonparametric Statistics* 6, 223–51.
HJELLVIK, V., YAO, Q. and TJØSTHEIM, D. (1997) Linearity testing using local polynomial approximation, *Technical report*. Norway: University of Bergen.
KEENAN, D. M. (1985) A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika* 72, 39–44.
LIPTSER, R. S. and SHIRJAEV, A. N. (1980) A functional central limit theorem for martingales. *Theory of Probability and its Applications* 25, 667–88.
LUUKKONEN, R. P., SAIKKONEN, P. and TERÄSVIRTA, T. (1988) Testing linearity against smooth transition autoregression. *Biometrika* 75, 491–9.
MÜLLER, H. G. (1992) Change-points in nonparametric regression analysis. *Annals of Statistics* 20, 737–61.
PARK, B. U. and MARRON, J. S. (1990) Comparison of data-driven bandwidth selectors. *J. Am. Statist. Assoc.* 85, 66–72.
PRIESTLEY, M. B. (1988) *Non-linear and Non-stationary Time Series Analysis*. New York: Academic Press.
RUPPERT, D., SHEATHER, S. J. and WAND, M. P. (1995) An efficient bandwidth selector for local least squares regression. *J. Am. Statist. Assoc.* 90, 1257–70.
SHEATHER, S. J. (1983) A data-based algorithm for choosing the window width when estimating the density at a point. *Computational Statistics and Data Analysis* 1, 229–38.
—— (1986) An improved data-based algorithm for choosing the window width when estimating the density at a point. *Computational Statistics and Data Analysis* 4, 61–5.
SKAUG, H. J. and TJØSTHEIM, D. (1993) Non-parametric tests of serial independence. In *Development in Time Series Analysis: in Honour Maurice B. Priestley* (ed. T. Subba Rao). London, New York: Chapman & Hall, 207–29.

© Blackwell Publishers Ltd 2001

SUGIHARA, G. and MAY, R. M. (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*. 344, 734–41.

TERÄSVIRTA, T. (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. *J. Am. Statist. Assoc.* 89, 208–18.

TJøSTHEIM, D. (1994) Nonlinear time series, a selective review. *Scand. J. Statist* 21, 97–130.

TONG, H. (1983) *Threshold Models in Nonlinear Time Series Analysis*. Lecture Notes in Statistics 21, Heidelberg: Springer.

—— (1990) *Nonlinear Time Series Analysis: A Dynamic Approach*. Oxford: Oxford University Press.

TSAY, R. S. (1986) Nonlinearity tests for time series. *Biometrika* 73, 461–6.

—— (1989) Testing and modeling threshold autoregressive processes. *J. Am. Statist. Assoc.* 84, 231–40.

TSYBAKOV, A. B. (1986) Robust reconstruction of functions by the local approximation method. *Problems of Information Transmission* 22, 133–46.

YOSHIHARA, K. (1976) Limiting behaviour of U-statistics for stationary absolutely regular process. *Z. Wahr. v. Gebiete* 35, 237–52.

© Blackwell Publishers Ltd 2001

Copyright © 2003 EBSCO Publishing