Statistica Sinica **9**(1999), 735-757

ON THE ESTIMATION AND TESTING OF FUNCTIONAL-COEFFICIENT LINEAR MODELS

Yingcun Xia and W. K. Li

Jinan University and the University of Hong Kong

Abstract: In this paper we investigate the estimation and testing of the functional coefficient linear models under dependence, which includes the functional coefficient autoregressive model of Chen and Tsay (1993). We use local linear smoothing to estimate the coefficient functions of a functional-coefficient linear model, prove their uniform consistency, and derive their asymptotic distributions in terms of Gaussian processes. From these distributions we can obtain some tests about coefficient functions and the model. Some simulations and a study of real data are reported.

Key words and phrases: FAR model, local linear smoother, nonparametric regression, strongly mixing sequence, Wiener process.

1. Introduction

Nonparametric regression analysis has gained much attention in the last decades, due primarily to the fact that it provides a versatile method of exploring a general relationship between variables. It gives predictions of observations without reference to a fixed parametric model. Properties of such analysis have been discussed extensively. However for the case of multi-dimensional predictor regressions, there is a major problem with most approaches in use in that they usually require unrealistically large sample sizes. Because of this, one suggestion is to bring back a measure of parametric models, and some semiparametric models, such as partial linear models and additive models, have been proposed in the literature, see for example Härdle (1990).

Hastie and Tibshirani (1993) proposed the varying-coefficient model. This model has gained much attention and has been found useful in applications. Chen and Tsay (1993) introduced the model to time series analysis and proposed the so called functional-coefficient autoregressive (FAR) model defined as follows. Suppose $\{\mathbf{x}_t\}$ is the time series under consideration. Then $\mathbf{x}_t$ satisfies

$$\mathbf{x}_t = \theta_1(X^*_{t-d})\mathbf{x}_{t-1} + \theta_2(X^*_{t-d})\mathbf{x}_{t-2} + \cdots + \theta_p(X^*_{t-d})\mathbf{x}_{t-p} + \varepsilon_t, \qquad (1.1)$$

where p is a positive integer, ε_t is a sequence of i.i.d. random variables with mean 0, variance σ^2 and independent of $\mathbf{x}_{t-i}$ for any $i > 0$; $X^*_{t-d} = (\mathbf{x}_{t-i_1}, \ldots, \mathbf{x}_{t-i_k})^T$, $i_j > 0, j = 1, \ldots, k$, where A^T denotes the transpose of matrix A; the $\theta_i(X)$,

$X = (x^{(1)}, \ldots, x^{(k)})$, $i = 1, \ldots, p$, are measurable functions from $\mathbb{R}^k$ to $\mathbb{R}$ and $\theta_i(X) = \alpha_i(X) + \beta_i(X)$ with $\beta_i(X)x^{(i)}$ uniformly bounded and $|\alpha_i(X)| \le c_i$ such that all the roots of $\lambda^p - c_1\lambda^{p-1} - c_2\lambda^{p-2} - \cdots - c_p = 0$ are inside the unit circle. It is easy to see that many existing nonlinear AR models are special cases of FAR models. These include the threshold autoregressive model of Tong (1983, 1990) and the exponential autoregressive model of Haggan and Ozaki (1981). Another advantage of model (1.1) is that it is not necessary to choose a special form of coefficient functions as Tong (1983) and Haggan and Ozaki (1981) did. Some probability properties of the FAR model have been given by Chen and Tsay (1993). The arranged local regression (ALR) procedure is proposed, as they call it, to estimate the coefficient functions $\theta_i(X)$, $i = 1, \ldots, p$. Under some assumptions, they proved that the estimator of $\theta_i(X)$ by the ALR is mean square consistent. They also gave a method to build FAR models.

In this paper, we investigate a more general functional-coefficient linear (FL) model

$$\mathbf{y}_t = \theta_0(Z_t) + \theta_1(Z_t)\mathbf{x}_{t1} + \theta_2(Z_t)\mathbf{x}_{t2} + \cdots + \theta_p(Z_t)\mathbf{x}_{tp} + \varepsilon_t \ , \tag{1.2}$$

where Z_t is a $k-$vector random variable and $\theta_0(X), \ldots, \theta_p(X)$ are some unknown functions. If we take $\mathbf{y}_t = \mathbf{x}_t$, $Z_t = X^*_{t-d}$, and $\mathbf{x}_{ti} = \mathbf{x}_{t-i}$, $i = 1, \ldots, p$, then model (1.2) becomes model (1.1). We assume that $\{(\mathbf{y}_t, Z_t^T, \mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp})\}$ is a strongly mixing sequence. From Theorem 1.2 of Chen and Tsay (1993), if the density function of ε_t is positive everywhere, then the Markov chain $\{\mathbf{x}_t\}$ generated by model (1.1) is geometrically ergodic, which implies that model (1.1) is strongly mixing with a mixing coefficient of geometric rate (cf. Bradley (1986)). Therefore the FAR satisfies the strongly mixing assumption.

As far as estimation is concerned, kernel estimation has been proven to be a very useful method in dealing with nonparametric statistical problems (Fan and Gijbels (1996)). Fan (1993) further proved that the local linear smoother has better properties than the Nadaraya-Watson kernel estimation. In this paper, we will obtain estimators of $\theta_i(\cdot), i = 0, \ldots, p$, by local linear smoothing, and prove that they are uniformly consistent with the optimal rate. We further obtain an estimator of σ^2 and its asymptotic property. Our results are stronger than those of Chen and Tsay (1993). Based on the theory of Bickel and Rosenblatt (1973), we will also obtain the asymptotic distributions of the estimators, and use them to test hypotheses of the form

$$H_{0i} : \theta_i(x) \equiv \phi_i(x, \gamma) \qquad x \in [a, b], \tag{1.3}$$

where $\phi_i(x, \gamma)$ is a known real function with parameter $\gamma \in \Gamma$, $i = 0, \ldots, p$, and $\Gamma \subset \mathbb{R}^q$, for some positive integer q, is the parameter space. For instance,

$\phi_i(x,\gamma) = \gamma_1 + \gamma_2\Phi((x-\gamma_3)^2/\gamma_4)$ in the smooth transition threshold autoregressive model proposed by Chan and Tong (1986), where $\Phi(x)$ is the standard normal distribution function. A special case of the test is $\theta_i(x) \equiv 0$, which is of importance in building a FL model. By these results, we can make statistical inference about the FL model similar to linear regression models. Furthermore, these results can be applied to the FAR model of Chen and Tsay (1993).

The rest of this paper is organized as follows. Section 2 states the estimation of the model and some assumptions for the following discussion. Section 3 proves the uniform convergence rate of the estimates while Section 4 derives the asymptotic distributions of the estimators. All proofs are relegated to the appendix. In Section 5, we report some results on Monte Carlo simulations and an application to real data.

2. The Estimation of FL Models and Assumptions

For simplicity, we consider only the case $Z_t = \mathbf{z}_t$, where $\mathbf{z}_t$ is a univariate random variable. Without loss of generality we write (1.2) as

$$\mathbf{y}_t = \theta_1(\mathbf{z}_t)\mathbf{x}_{t1} + \theta_2(\mathbf{z}_t)\mathbf{x}_{t2} + \cdots + \theta_p(\mathbf{z}_t)\mathbf{x}_{tp} + \varepsilon_t. \tag{2.1}$$

If we take $\mathbf{x}_{t1} \equiv 1$, then (2.1) becomes (1.2). If $\mathbf{z}_t = \mathbf{x}_{tj}$ for some j we may write $\theta_j(\mathbf{z}_t)\mathbf{x}_{tj}$ as $\theta_j(\mathbf{z}_t)\cdot 1$. We first make an assumption about $\theta_i(x)$.
(A1) The coefficient functions $\theta_i(x)$, $i = 1,\ldots,p$, are bounded and have bounded second derivatives.

Suppose that $\{(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp}),\ t = 1,\ldots,n\}$ is a sequence of observations from model (2.1). By a Taylor expansion, $\theta_i(\mathbf{z}_t) = \theta_i(x) + \theta'(x)(\mathbf{z}_t - x) + \frac{1}{2}\theta_i''(\mathbf{z}_{ti})(\mathbf{z}_t - x)^2$, $i = 1,\ldots,p$, where $\mathbf{z}_{ti}$ is a point between $\mathbf{z}_t$ and x. From (2.1) we have

$$\mathbf{y}_t = \sum_{i=1}^{p}\Big[\theta_i(x) + \theta_i'(x)(\mathbf{z}_t - x)\Big]\mathbf{x}_{ti} + \frac{1}{2}\sum_{i=1}^{p}\theta_i''(\mathbf{z}_{ti})\mathbf{x}_{ti}(\mathbf{z}_t - x)^2 + \varepsilon_t.$$

Following the idea of locally linear smoothing, the estimator of $\theta_i(x)$ is the solution of a_i to the following minimizer:

$$\min_{a_i,b_i,\ i=1,\ldots,p}\sum_{t=1}^{n}\Big\{\mathbf{y}_t - \sum_{i=1}^{p}\Big[a_i + b_i(\mathbf{z}_t - x)\Big]\mathbf{x}_{ti}\Big\}^2 K\Big(\frac{\mathbf{z}_t - x}{h}\Big), \tag{2.2}$$

or

$$\min_{a_i,b_i,\ i=1,\ldots,p}\sum_{t=1}^{n}\Big[\mathcal{E}_t + \sum_{i=1}^{p}(a_i - \theta_i(x))\mathbf{x}_{ti} + \sum_{i=1}^{p}(b_i - \theta_i'(x))\mathbf{x}_{ti}(\mathbf{z}_t - x)\Big]^2 K\Big(\frac{\mathbf{z}_t - x}{h}\Big), \tag{2.3}$$

where $K(\cdot)$ is a kernel function, h is the bandwidth and $\mathcal{E}_t = \sum_{i=1}^p \theta_i''(\mathbf{z}_{ti})\mathbf{x}_{ti}(\mathbf{z}_t - x)^2/2 + \varepsilon_t$. Let $w_t = K((\mathbf{z}_t - x)/h)$, $w_{1t} = K((\mathbf{z}_t - x)/h)(\mathbf{z}_t - x)/h$, $w_{2t} = K((\mathbf{z}_t - x)/h)((\mathbf{z}_t - x)/h)^2$, $w_{4t} = K((\mathbf{z}_t - x)/h)((\mathbf{z}_t - x)/h)^4$ and

$$S_t = (\mathbf{x}_{t2}, \ldots, \mathbf{x}_{tp}), \quad T_t = (\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp}) \ ,$$
$$U_t = (S_t^T \quad (\mathbf{z}_t - x)T_t^T)^T, \quad V_t = (T_t^T \quad (\mathbf{z}_t - x)T_t^T)^T.$$

The solution to (2.2), i.e. the estimators of $\theta_i(x)$ and $\theta_i'(x)$, $i = 1, \ldots, p$, is

$$(\hat\theta_1(x), \ldots, \hat\theta_p(x) \quad \hat\theta_1'(x), \ldots, \hat\theta_p'(x))^T = (\sum_{t=1}^n w_t V_t V_t^T)^{-1} \sum_{t=1}^n w_t V_t \mathbf{y}_t. \tag{2.4}$$

In case that $(\sum_{t=1}^n w_t V_t V_t^T)^{-1}$ does not exist, we may substitute $(\sum_{t=1}^n w_t V_t V_t^T + n^{-2} I_{2p})^{-1}$ in practice, where I_k is the $k \times k$ unit matrix. However, as we will show in Section 3, $(\sum_{t=1}^n w_t V_t V_t^T)^{-1}$ exists $a.s.$ as $n \longrightarrow \infty$.

In order to investigate the asymptotic properties, we introduce another expression for $\hat\theta_i(x)$, $i = 1, \ldots, p$. Here we only take $\hat\theta_1(x)$ for example. By Lemma 3 of Lai, Robbins and Wei (1979) and (2.3), we have

$$\begin{aligned}
\hat\theta_1(x) &= \theta_1(x) + \frac{\sum_{t=1}^n w_t(\mathbf{x}_{t1} - J_n^T H_n^{-1} U_t)\mathcal{E}_t}{\sum_{t=1}^n w_t(\mathbf{x}_{t1} - J_n^T H_n^{-1} U_t)^2} \\
&= \theta_1(x) + \frac{\sum_{i=1}^p \sum_{t=1}^n w_t(\mathbf{x}_{t1} - J_n^T H_n^{-1} U_t)\theta_i''(\mathbf{z}_{ti})\mathbf{x}_{ti}(\mathbf{z}_t - x)^2}{2\sum_{t=1}^n w_t(\mathbf{x}_{t1} - J_n^T H_n^{-1} U_t)^2} \\
&\quad + \frac{\sum_{t=1}^n w_t(\mathbf{x}_{t1} - J_n^T H_n^{-1} U_t)\varepsilon_t}{\sum_{t=1}^n w_t(\mathbf{x}_{t1} - J_n^T H_n^{-1} U_t)^2},
\end{aligned} \tag{2.5}$$

where

$$H_n = \sum_{t=1}^n w_t U_t U_t^T = \begin{pmatrix} \sum_{t=1}^n w_t S_t S_t^T & h\sum_{t=1}^n w_{1t} S_t T_t^T \\ & h^2 \sum_{t=1}^n w_{2t} T_t T_t^T \end{pmatrix} \triangleq \begin{pmatrix} P_n & hR_n \\ & h^2 Q_n \end{pmatrix},$$
$$J_n = \sum_{t=1}^n w_t \mathbf{x}_{t1} U_t = \begin{pmatrix} \sum_{t=1}^n w_t \mathbf{x}_{t1} S_t \\ h\sum_{t=1}^n w_{1t}\mathbf{x}_{t1} T_t \end{pmatrix} \triangleq \begin{pmatrix} A_n \\ hB_n \end{pmatrix}.$$

Let $\Gamma_n = \sum_{t=1}^n w_t V_t V_t^T$. Then

$$\Gamma_n^{-1}(1,1) = \left[\sum_{t=1}^n w_t(\mathbf{x}_{t1} - J_n^T H_n^{-1} U_t)^2\right]^{-1}, \tag{2.6}$$

where $\Gamma_n^{-1}(i,i)$ is the (i,i)th entry of Γ_n^{-1}, $i = 1, \ldots, 2p$. Equations (2.5) and (2.6) are very useful for future proofs and calculations. Throughout this paper, we will omit the lower-left part of a symmetric matrix for brevity.

For convenience in the rest of the paper, we make the following assumptions.

(A2) $\{(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp})\}$ is a strictly stationary and strongly mixing sequence with mixing coefficient $\alpha(k) = O(\varrho^k)$ for some $0 < \varrho < 1$.
(A3) For each t, $E(\varepsilon_t | \mathbf{z}_t, \mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp}) = 0$ *a.s.*, $E\varepsilon_t^2 = \sigma^2$ and $E\varepsilon_t^3 < \infty$.
(A4) The density function of $\mathbf{z}_1$, say $f(x)$, has a bounded continuous derivative.
(A5) The conditional densities $f_{\mathbf{z}_1|(\mathbf{y}_1,\mathbf{x}_{1i})}(z_1|y_1, x_{1i})$, $f_{(\mathbf{z}_1,\mathbf{z}_l)|(\mathbf{y}_1,\mathbf{y}_l,\mathbf{x}_{1i},\mathbf{x}_{li})}(z_1, z_l|y_1, y_l, x_{1i}, x_{li})$, $f_{\mathbf{z}_1|(\mathbf{x}_{1i},\ \mathbf{x}_{1j})}(z_1|x_{1i},\ x_{1j})$ and $f_{(\mathbf{z}_1,\mathbf{z}_l)|(\mathbf{x}_{1i},\mathbf{x}_{1j},\mathbf{x}_{li},\mathbf{x}_{lj})}(z_1,\ z_l|x_{1i},\ x_{1j},\ x_{li},\ x_{lj})$ are bounded for all $l \geq 1$ and $i, j = 1, \ldots, p$.
(A6) $E(\mathbf{x}_{ti}^6) < \infty$ and the conditional expectation $v_{ij}(x) = E(\mathbf{x}_{ti}\mathbf{x}_{tj} \mid \mathbf{z}_t = x)$ is bounded and has a bounded derivative, $i, j = 1, \ldots, p$.
(A7) The kernel function $K(x)$ is a density function with a compact support $[-\delta_0,\ \delta_0]$ and bounded derivative such that $K(-\delta_0) = K(\delta_0) = 0$, $\int yK(y)dy = 0$ and $\int y^2 K(y)dy = 1$.

Assumption (A2) is made only for the purpose of simplicity, it can be weakened to $\alpha(k) = O(k^{-\iota})$ for some $\iota > 0$. However, many time series models satisfy assumption (A2)—examples are the nonparametric ARCH models (Masry and Tjøstheim (1995)) and model (1.1) if the density of ε_t is positive everywhere (Chen and Tsay (1993)). Assumption (A3) is a common one for time series models and $E\varepsilon_t^3 < \infty$ can be changed to $E|\varepsilon_t|^{r_0} < \infty$ for some $r_0 > 2$. Similarly, $E(\mathbf{x}_{1i}^6)$ in (A6) can be changed to $E(\mathbf{x}_{1i}^r)$ for some $r > 4$. However, these changes may make the following arguments more complicated. We shall not pursue this matter further. Assumption (A5) is a common one for dependent data, see Masry and Tjøstheim (1995). Other kernel functions can be used in (A7) but result in slightly more tedious arguments. The other assumptions are those in related papers, such as Bickel and Rosenblatt (1973), Chen and Tsay (1993) and Härdle (1989).

Let $\phi_k = \int y^k K(y)dy,\ k = 0, 1, \ldots.$ Then $\phi_0 = \phi_2 = 1$ and $\phi_1 = 0$ from (A7). Let $M, c, c_i, i = 0, 1, \ldots,$ be constants which may have different values at different places. By $C_n(x) = \bar{O}(a_n, D)$, we mean $\sup_{x \in D} |C_n(x)| = O(a_n)$ *a.s.*, where $D \subset \mathbb{R}$. We write $\mathbf{1}_k$ to denote the $k-$vector of ones.

3. Consistency of Estimators

In nonparametric regression we are mainly interested in the parts with dense observations, i.e. where $f(x) > 0$. This can be judged using the histograms of $\{\mathbf{z}_t, t = 1, \ldots, n\}$ in practice. Denote the minimum eigenvalue of a matrix A by $\lambda_{\min}(A)$. To ensure that $\lambda_{\min}((v_{ij}(x))_{i,j=1,\ldots,p}) > 0$ as needed for the estimation of $\theta_i(x)$, we may check $\lambda_{\min}(\sum_{\mathbf{z}_t \in x \pm \delta} T_t T_t^T) > 0$, where $T_t = (\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp})$ and δ is a small number proportional to h. For any $\epsilon > 0$, define $\mathbb{R}_\epsilon = \{x : f(x)\lambda_{\min}((v_{ij}(x))_{i,j=1,\ldots,p}) > \epsilon\}$. In this section, we prove that the estimator $\hat{\theta}_i(x)$ converges to $\theta_i(x)$ a.s. uniformly on $\mathbb{R}_\epsilon$ and obtain the corresponding convergence rate. We also obtain the consistency rate of $\hat{\sigma}^2$, the estimator of σ^2.

The following theorem tells us the uniformly consistency rate for the estimators of coefficient functions.

Theorem 3.1. *If the assumptions* (A1)-(A7) *hold and* $h = O(n^{-\rho})$ *for some* $0 < \rho < 1/3$, *then* $\hat{\theta}_i(x)$ *is a uniformly consistent estimator of* $\theta_i(x)$ *on* $\mathbb{R}_\epsilon$, *and*

$$|\hat{\theta}_i(x) - \theta_i(x)| = \bar{O}(\delta_{n2}, \mathbb{R}_\epsilon), \qquad i = 1, \dots, p,$$

where $\delta_{n2} = h^2 + (\log n/(nh))^{1/2}$.

From the results of Stute (1982), the above convergence rate is optimal. We can also prove that the estimator $\hat{\theta}'_i(x)$ is uniformly consistent but that is not our concern here.

Next, we want to give an estimator of σ^2, which will be used in making statistical inference about model (1.1). From (2.1) we have

$$\mathbf{y}_t I(\mathbf{z}_t \in \mathbb{R}_\epsilon) = \theta_1(\mathbf{z}_t) I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\mathbf{x}_{t1} + \cdots + \theta_p(\mathbf{z}_t) I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\mathbf{x}_{tp} + I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\varepsilon_t.$$

The estimator of $\mathbf{y}_t$, say $\hat{\mathbf{y}}_t$, satisfies

$$\hat{\mathbf{y}}_t I(\mathbf{z}_t \in \mathbb{R}_\epsilon) = \hat{\theta}_1(\mathbf{z}_t) I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\mathbf{x}_{t1} + \cdots + \hat{\theta}_p(\mathbf{z}_t) I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\mathbf{x}_{tp}.$$

Here we may take $\hat{\theta}_i(x) \equiv 0$, if $x \notin \mathbb{R}_\epsilon$. Therefore from Theorem 3.1,

$$\begin{aligned} e_t &\stackrel{\Delta}{=} \mathbf{y}_t I(\mathbf{z}_t \in \mathbb{R}_\epsilon) - \hat{\mathbf{y}}_t I(\mathbf{z}_t \in \mathbb{R}_\epsilon) \\ &= \sum_{i=1}^p (\theta_i(\mathbf{z}_t) - \hat{\theta}_i(\mathbf{z}_t)) I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\mathbf{x}_{ti} + I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\varepsilon_t \\ &= \sum_{i=1}^p O(\delta_{n2})\mathbf{x}_{ti} + I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\varepsilon_t \quad a.s. \\ e_t^2 &= \sum_{i=1}^p \Big[O(\delta_{n2}^2)\mathbf{x}_{ti}^2 + O(\delta_{n2})\mathbf{x}_{ti} I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\varepsilon_t\Big] + I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\varepsilon_t^2 \quad a.s. \end{aligned}$$

By the LIL for strongly mixing sequence (cf. Rio (1995)),

$$\begin{aligned} &|\frac{1}{n}\sum_{t=1}^n \mathbf{x}_{ti}^2| < M\ , \qquad \frac{1}{n}\sum_{t=1}^n |\mathbf{x}_{ti} I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\varepsilon_t| < M \quad a.s., \\ &|\frac{1}{n}\sum_{t=1}^n I(\mathbf{z}_t \in \mathbb{R}_\epsilon)\varepsilon_t^2 - \sigma^2 P(\mathbf{z}_1 \in \mathbb{R}_\epsilon)| = O((\log\log n/n)^{\frac{1}{2}}) \quad a.s. \end{aligned}$$

and

$$|\frac{1}{n}\sum_{t=1}^n I(\mathbf{z}_t \in \mathbb{R}_\epsilon) - P(\mathbf{z}_1 \in \mathbb{R}_\epsilon)| = O((\log\log n/n)^{\frac{1}{2}}) \quad a.s. \tag{3.1}$$

Hence,

$$|\frac{1}{n}\sum_{t=1}^{n} e_t^2 - \sigma^2 P(\mathbf{z}_1 \in \mathbb{R}_\epsilon)| = O(\delta_{n2} + (\log\log n/n)^{\frac{1}{2}}) \quad a.s. \tag{3.2}$$

Let $\#\mathbf{A}$ denote the number of elements in set $\mathbf{A}$. Then $\#\{\mathbf{z}_t \in \mathbb{R}_\epsilon, \quad t = 1,\ldots,n\} = \sum_{t=1}^n I(\mathbf{z}_t \in \mathbb{R}_\epsilon)$. Thus from (3.1) and (3.2), we obtain an estimator of σ^2 as

$$\hat{\sigma}^2 = \frac{1}{\#\{\mathbf{z}_t \in \mathbb{R}_\epsilon, \quad t = 1,\ldots,n\}} \sum_{\mathbf{z}_t \in \mathbb{R}_\epsilon} (\mathbf{y}_t - \hat{\mathbf{y}}_t)^2.$$

Theorem 3.2. *If* (A1)-(A7) *hold and* $h = O(n^{-\rho})$ *for some* $0 < \rho < 1/3$, *then*

$$|\hat{\sigma}^2 - \sigma^2| = O(\delta_{n2}) \quad a.s.$$

4. Distributions of Estimators

For ease of exposition, we consider only an interval $[a,\ b]$ such that $\min_{x\in[a,b]}(\lambda_{\min}((v_{ij}(x))_{i,j=1,\ldots,p})f(x)) > 0$. If (A4) and (A6) hold, there exists $[a_0, b_0]$ such that $c_0 = \min\{a-a_0, b_0-b\} > 0$ and $\min_{x\in[a_0,b_0]}(\lambda_{\min}((v_{ij}(x))_{i,j=1,\ldots,p})f(x)) > 0$. We further assume

(A8) For each t, ε_t is independent of $\{\mathbf{z}_s, \mathbf{x}_{s1},\ldots,\mathbf{x}_{sp}:\ s \le t\}$ and $E|\varepsilon_t|^{r_1} < \infty$, $E|\mathbf{x}_{ti}|^{r_1} < \infty,\ i = 1,\ldots,p$, for some $r_1 > 8$.

Let

$$Y_{0n}(x) = h^{-\frac{1}{2}} \int_{a_0}^{b_0} K\Big(\frac{s-x}{h}\Big) d\mathcal{W}(s-a_0),$$
$$M_{ni}(x) = \hat{\sigma}^{-1}(\Gamma_n^{-1}(i,i))^{-\frac{1}{2}}(\hat{\theta}_i(x) - \theta_i(x)), \quad i = 1,\ldots,p,$$

where $\mathcal{W}(x)$ is a standard Wiener process and $\Gamma_n^{-1}(i,i)$ is defined in (2.6). We have the following asymptotic distribution for the estimators of coefficient functions.

Theorem 4.1. *If assumptions* (A1)-(A8) *hold and* $h = O(n^{-\rho})$ *for some* $1/5 < \rho < 1/3$, *then* $\sup_{a\le x\le b}|M_{ni}(x)|$ *and* $\sup_{a\le x\le b}|Y_{0n}(x)|$ *have the same asymptotic distribution.*

Notice that the bandwidth can not take its optimal value $h_0 \propto n^{-1/5}$. This is because of the bias of the estimates $\hat{\theta}_i(x)$. To see it more clearly, we further assume that $\theta_i(x)$, $i = 1,\ldots,p$, have bounded third order derivatives. Then by (2.3) and the results in Section 3, we can show that

$$\hat{\theta}_1(x) = \theta_1(x) + \sum_{i=1}^{p} \frac{\sum_{t=1}^n w_t(\mathbf{x}_{t1} - J_n^T H_n^{-1} U_t)\mathbf{x}_{ti}(\mathbf{z}_t - x)^2}{2\sum_{t=1}^n w_t(\mathbf{x}_{t1} - J_n^T H_n^{-1} U_t)^2} \theta_i''(x)$$

$$+\frac{\sum_{t=1}^n w_t(\mathbf{x}_{t1}-J_n^T H_n^{-1}U_t)\varepsilon_t}{\sum_{t=1}^n w_t(\mathbf{x}_{t1}-J_n^T H_n^{-1}U_t)^2}$$
$$+\sum_{i=1}^p \frac{\sum_{t=1}^n w_t(\mathbf{x}_{t1}-J_n^T H_n^{-1}U_t)\mathbf{x}_{ti}(\mathbf{z}_t-x)^3}{6\sum_{t=1}^n w_t(\mathbf{x}_{t1}-J_n^T H_n^{-1}U_t)^2}\theta_i'''(\mathbf{z}_{ti}^*)$$
$$=\theta_1(x)+B_1(x)h^2+\frac{1}{nh}C_1(x)\sum_{t=1}^n w_t V_t\varepsilon_t+O(\delta_{n1}\delta_{n2}) \quad a.s., \qquad (4.1)$$

where $B_1(x)$ is a continuous function, $C_1(x)$ is a continuous vector, $\delta_{1n}=h+(\log n/(nh))^{1/2}$ and $\mathbf{z}_{ti}^*$ is a point between $\mathbf{z}_t$ and x. Similarly, we have $\hat\theta_i(x)=\theta_i(x)+B_i(x)h^2+C_i(x)\sum_{t=1}^n w_tV_t\varepsilon_t/(nh)+O(\delta_{n1}\delta_{n2})$ $a.s.$, $i=2,\ldots,p$. Thus

$$E\hat\theta_i(x)=\theta_i(x)+B_i(x)h^2+O(\delta_{n1}\delta_{n2}).$$

The main bias term is $B_i(x)h^2$, which restricts the value of h in Theorem 4.1.

To allow the optimal bandwidth h_0 be used, we need to remove the bias term $B_i(x)h^2$ in the estimator of $\theta_i(x)$. Similar to (2.2), we consider

$$\min_{f_i,e_i,d_i,c_i,\ i=1,\ldots,p}\sum_{t=1}^n\Big\{\mathbf{y}_t-\sum_{i=1}^p\Big[f_i+e_i(\mathbf{z}_t-x)+d_i(\mathbf{z}_t-x)^2+c_i(\mathbf{z}_t-x)^3\Big]\mathbf{x}_{ti}\Big\}^2K\Big(\frac{\mathbf{z}_t-x}{h'}\Big), \qquad (4.2)$$

where h' is another bandwidth. The solution of d_i to the above problem is the estimator of $\theta_i''(x)$, say $\hat\theta_i''(x)$. We obtain the estimator of the bias as

$$\hat B_1(x)=\sum_{i=1}^p\frac{\sum_{t=1}^n w_t(\mathbf{x}_{t1}-J_n^TH_n^{-1}U_t)\mathbf{x}_{ti}(\mathbf{z}_t-x)^2}{2\sum_{t=1}^n w_t(\mathbf{x}_{t1}-J_n^TH_n^{-1}U_t)^2}\hat\theta_i''(x).$$

Similarly, we can get the other bias estimator of $\theta_i(x)$, say $\hat B_i(x)$, $i=2,\ldots,p$. If we replace $\hat\theta_i(x)$ with $\hat\theta_i(x)-\hat B_i(x)h^2$, then Theorem 4.1 and Corollaries 4.1 and 4.2 below hold with $h\propto n^{1/5}$. See Xia (1998).

Similar to the asymptotic normality of the coefficient estimators in linear models, Theorem 4.1 is a basic asymptotic result for statistical inference about the coefficient functions in FL models. From Theorem 2 and Theorem 3.1 of Bickel and Rosenblatt (1973), we have

Corollary 4.1. *Under the assumptions of Theorem* 4.1*, we have*

$$P\Big((-2\log h)^{\frac12}(d^{-1}\sup_{a\le x\le b}|M_{ni}(x)|-\mu_n)<z\Big)\longrightarrow e^{-2e^{-z}}, \quad i=1,\ldots,p,$$

where $d=\sqrt{\kappa_2(b-a)}$ *and*

$$\mu_n=(-2\log h)^{\frac12}+\frac{1}{(-2\log h)^{\frac12}}(\log\frac{\kappa_1^{\frac12}}{2\pi\kappa_2^{\frac12}}),$$

with $\kappa_2 = \int K^2(x)dx$ *and* $\kappa_1 = \int (K'(x))^2 dx$.

We can now derive some statistical inference about FL models in a global sense, i.e. statistical inferences based on a global measure of how good $\hat{\theta}_i(x)$ is an estimator of $\theta_i(x)$ on $[a,b]$. Similar to Härdle (1989) and Eubank and Speckman (1993), we can construct confidence bands for the coefficient function $\theta_i(x)$. We have

$$\lim_{n\to+\infty} P\Big(\hat{\theta}_i(x) - d\hat{\sigma}(\Gamma_n^{-1}(i,i))^{\frac{1}{2}}[\frac{z}{(-2\log h)^{\frac{1}{2}}} + \mu_n] \le \theta_i(x) \le \hat{\theta}_i(x) + d\hat{\sigma}(\Gamma_n^{-1}(i,i))^{\frac{1}{2}}[\frac{z}{(-2\log h)^{\frac{1}{2}}} + \mu_n], \quad \text{for all } a \le x \le b\Big) = e^{-2e^{-z}}. \tag{4.3}$$

Similar to linear models, we can also test hypotheses of the type given by (1.3). We first use some standard nonlinear time series estimation methods to estimate the parameter γ under $H_{0i},\ i = 0,\ldots,p$. Denote the estimator by $\hat{\gamma}$. It is known that under some regular conditions, $\hat{\gamma}$ is root-n consistent. See Gallant (1987). Let $M_{ni}(x,\hat{\gamma}) = \hat{\sigma}^{-1}(\Gamma_n^{-1}(i,i))^{-\frac{1}{2}}(\hat{\theta}_i(x) - \phi_i(x,\hat{\gamma}))$. By Corollary 4.1, we have

Corollary 4.2. *Suppose* $H_{0i} : \theta_i(x) \equiv \phi_i(x,\gamma_0)$ *holds for some* $\gamma_0 \in \Gamma$ *and all* $x \in [a,b]$*;* $\phi_i(x,\gamma)$ *has a bounded derivative in* $\gamma \in \Gamma$ *and* $\hat{\gamma} - \gamma_0 = O_p(n^{-1/2})$*. Under the assumptions of Theorem* 4.1*, we have*

$$P\Big((-2\log h)^{\frac{1}{2}}(d^{-1}\sup_{a\le x\le b}|M_{ni}(x,\hat{\gamma})| - \mu_n) < z\Big) \longrightarrow e^{-2e^{-z}}, \quad i = 1,\ldots,p.$$

Using Corollary 4.2, we can obtain the critical value given the significance level and make our decision. We give some examples in the next section.

5. Simulations and Application to Real Data

In this section, we illustrate our modeling methodology by using both simulated and real data. For the simulated data, we see how well the testing method works. Similar work has been done by Härdle (1989) and Eubank and Speckman (1993) in much simpler cases. For the real data set, we concentrate on the building of a FL model.

5.1. Selection of bandwidth

An important issue for kernel estimation is the choice of bandwidths. There are many methods to do this. Here we combine cross-validation with the plug-in method.

First, we use the cross-validation method to select the bandwidth h' for the estimation of $\theta''_i(x)$ in the bias term $B_i(x)$. See the discussion after Theorem 4.1. Let $\hat{\vartheta}_i(x)$ be the solution of f_i in (4.2). Then $\hat{\vartheta}_i(x)$ is the local third order polynomial estimator of $\theta_i(x)$. Let $\hat{\vartheta}_{i(t)}(x)$ be the estimator of $\theta_i(x)$ from (4.2) using data $\{(\mathbf{y}_s, \mathbf{z}_s, \mathbf{x}_{s1}, \ldots, \mathbf{x}_{sp}) : s \neq t\}$. Define

$$CV(h) = \sum_{a \le \mathbf{z}_t \le b} \Big[\mathbf{y}_t - \hat{\vartheta}_{1(t)}(\mathbf{z}_t)\mathbf{x}_{t1} - \cdots - \hat{\vartheta}_{p(t)}(\mathbf{z}_t)\mathbf{x}_{tp}\Big]^2 G(\mathbf{z}_t),$$

where $G(\cdot)$ is a weight function. The cross-validation bandwidth for $\{\hat{\vartheta}_i(x),\ i = 1, \ldots, p\}$ is then

$$\hat{h}'_0 = \arg\min_{h'} CV(h). \tag{5.1}$$

If we further assume that all the moments of ε_t, $\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp}$ and $\mathbf{y}_t$ exist, then we can show, following the same steps of Xia and Li (1997), that

$$\frac{\hat{h}'_0 - h'_o}{h'_o} = o_p(1),$$

where h'_o is the ideal bandwidth for the third order polynomial smoothers $\{\hat{\vartheta}_i(x), i = 1, \ldots, p\}$ in the sense of mean integrated squared error (MISE), where

$$MISE = E\int_a^b \Big[\sum_{i=1}^{p}(\hat{\vartheta}_i(z) - \theta_i(z))\mathbf{x}_{ti}\Big]^2 G(z)f(z)dz. \tag{5.2}$$

Notice that $\hat{h}'_0$ is not the asymptotic ideal bandwidth for $\{\hat{\theta}''_i(x),\ i = 1, \ldots, p\}$. But the ideal bandwidth $\hat{h}'$ for $\{\hat{\theta}''_i(x),\ i = 1, \ldots, p\}$ can be obtained as

$$\hat{h}' = adj_{2,3}\hat{h}'_0,$$

where $adj_{2,3}$ is an adjustment constant which depends only on the kernel function. For example, if we use the Gaussian kernel, $adj_{2,3} = 0.8665$. See Fan and Gijbels (1996, p.67) about this.

For the local linear smoother $\{\hat{\theta}_i(x),\ i = 1, \ldots, p\}$ of (2.2), by some algebraic calculation the ideal bandwidth in the sense of MISE (with $\hat{\vartheta}_i(x)$ replaced by $\hat{\theta}_i(x)$ in (5.2)) is

$$h_o = \Big(\frac{\kappa_2\sigma^2}{n\int E(\sum_{i=1}^{p}\theta''_i(z)\mathbf{x}_{ti})^2 G(z)f(z)dz}\Big)^{1/5}.$$

The idea of the plug-in method is to replace the unknown quantities in h_o with their estimators. Since we have already estimated these quantities, the plug-in method is appealing to us. The estimated optimal bandwidth for $\{\hat{\theta}_i(x),\ i = 1, \ldots, p\}$ is then

$$\hat{h} = \Big(\frac{\kappa_2 \hat{\sigma}^2}{\sum_{a \le \mathbf{z}_t \le b} (\sum_{i=1}^{p} \hat{\theta}_i''(\mathbf{z}_t)\mathbf{x}_{ti})^2 G(\mathbf{z}_t)}\Big)^{1/5}. \tag{5.3}$$

5.2. Simulation

For all the data analyzed below, we use the Epanechnikov kernel: $K(x) = 3(1 - x^2/5)I(x^2 \le 5)/(4\sqrt{5})$. Thus $\int_{-\sqrt{5}}^{+\sqrt{5}} K^2(x)dx = 3\sqrt{5}/25$, $\int_{-\sqrt{5}}^{+\sqrt{5}} (K'(x))^2 dx = 3\sqrt{5}/50$. The bandwidth is chosen by (5.1) and (5.3) with $G(x) \equiv 1$.

Consider the following model

$$\mathbf{y}_t = \theta_1(\mathbf{y}_{t-4}) + \theta_2(\mathbf{y}_{t-4})\mathbf{y}_{t-2} + 0.05\varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} N(0,1). \tag{5.4}$$

We want to test hypotheses

$$H_{10}: \ \theta_1(x) = 0.8e^{-\gamma_1(x-0.5)^2},$$
$$H_{20}: \ \theta_2(x) = 0.6\sin(\gamma_2\pi(x-0.5)^2),$$

where $x \in [0,\ 1]$ and $\gamma_1 = 2$ and $\gamma_2 = 4$. These coefficient functions are shown in Figure 1.

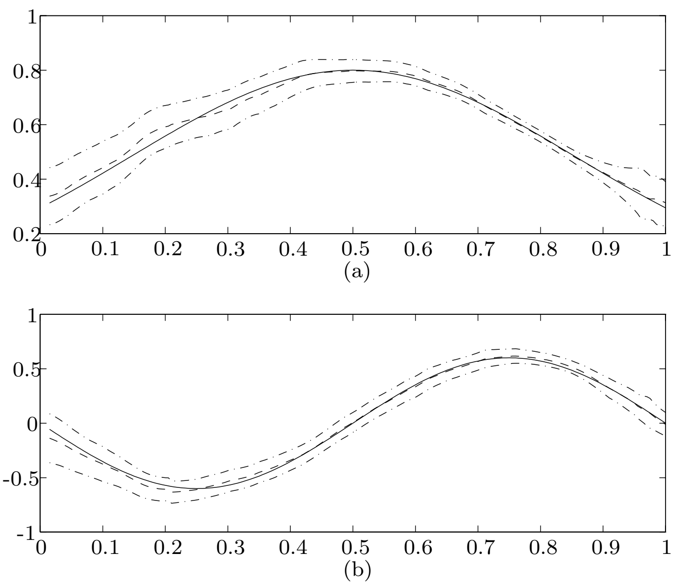

Figure 1. (a) and (b) corresponds to coefficient functions $\theta_1(x)$ and $\theta_2(x)$ respectively. ‘ —-’: real coefficients; ‘- - -’: estimated coefficient; ‘-.-.’: 90% confidence bands.

Take the sample size n to be 300, 500 and 800 respectively. We repeat each realization 500 times using a different seed. The parameters γ_1 and γ_2 are estimated as in Gallant (1987). The empirical sizes and powers of the tests are shown in Table 1. For the power study, the true model was simulated by adding the term $0.01\sin(32\pi x)$ to both $\theta_1(x)$ and $\theta_2(x)$ above. We observe that in Table 1 the empirical powers are high and the empirical sizes are close to the nominal significance level with all sample sizes. We can also construct confidence bands for the coefficients using (4.3). Figure 1 shows the 90% confidence bands from a typical data set with size $n = 300$.

Table 1. Empirical sizes and powers of the tests about model (5.4) for different sample size n

			$n = 300$	$n = 500$	$n = 800$
0.2 nominal	$\theta_1(x)$	size	0.226	0.218	0.214
significance		power	0.652	0.838	0.946
level	$\theta_2(x)$	size	0.258	0.242	0.224
		power	0.932	0.988	1.000
0.1 nominal	$\theta_1(x)$	size	0.120	0.116	0.116
significance		power	0.606	0.714	0.920
level	$\theta_2(x)$	size	0.144	0.122	0.120
		power	0.880	0.982	1.000
0.05 nominal	$\theta_1(x)$	size	0.054	0.048	0.044
significance		power	0.514	0.648	0.872
level	$\theta_2(x)$	size	0.066	0.058	0.056
		power	0.856	0.928	1.000

5.3. Application to the Australian blowfly data

We turn to the study of Australian blowfly data. There are $n = 361$ observations with outliers at $t = 11$ and $t = 48$ (cf. Tsay (1988)). Following Chan and Tong (1986), we take a log transformation of the data, $\mathbf{y}_t = \log_{10}($ blowfly population $)$, and consider the following model

$$\mathbf{y}_t = b_0(\mathbf{y}_{t-8}) + b_1(\mathbf{y}_{t-8})\mathbf{y}_{t-1} + \cdots + b_7(\mathbf{y}_{t-8})\mathbf{y}_{t-7} + \varepsilon_t \ . \qquad (5.5)$$

The estimates of $b_i(x),\ i = 0, \ldots, 7$, are displayed in Figure 2(a), and $\hat{\sigma}^2 = 0.0178$.

We test the null hypothesis

$$H_{0i}: \quad b_i(x) \equiv 0, \qquad x \in [2.0, 4.0].$$

$i = 0, \ldots, 7$. Calculate

$$M_{ni} = (-2\log h)^{1/2}\Big[\hat{\sigma}(\psi(b-a))^{-1/2} \sup_{2\le x\le 4} |(\Gamma_n^{-1}(i,i))^{-\frac{1}{2}}\hat{b}_i(x)| - \mu_n\Big],$$

where $b - a = 2.0$. The test procedures are given in Table 2.

Table 2. The values of M_{ni}, their corresponding p-values and the variance estimate of ε_t

	step 1		step 2		step 3	
	M_{ni}	p-value	M_{ni}	p-value	M_{ni}	p-value
$b_0(x)$	14.78	0.001	37.38	0.000	42.63	0.000
$b_1(x)$	54.15	0.000	54.36	0.000	55.80	0.000
$b_2(x)$	6.99	0.002	6.64	0.003	16.84	0.001
$b_3(x)$	7.53	0.009	7.09	0.002		
$b_4(x)$	5.41	0.002	5.21	0.011		
$b_5(x)$	3.19	0.079	4.71	0.018		
$b_6(x)$	1.60	0.332				
$b_7(x)$	1.71	0.304				
$\hat{\sigma}^2$	0.0178		0.0181		0.0192	

The p-values in Table 2 are defined as $1-P(M_{ni} < *)$, $i = 1, \ldots, 7$, (for example, $1 - P(M_{n1} < 14.78) = 0.001$). By removing $b_6(x)$ and $b_7(x)$ according to their probability of significance, Step 2 arrives at a better model for the data. However, following Tong (1990, p.337), we can also consider a model with a lower order (step 3). Similar to linear regression, we write the FAR regression as

$$\underset{}{\hat{\mathbf{y}}_t} = \underset{(0.000)}{\hat{b}_0(\mathbf{y}_{t-8})} + \underset{(0.000)}{\hat{b}_1(\mathbf{y}_{t-8})}\mathbf{y}_{t-1} + \underset{(0.003)}{\hat{b}_2(\mathbf{y}_{t-8})}\mathbf{y}_{t-2}\ , \quad \hat{\sigma}^2 = 0.0192\ . \tag{5.6}$$

The values in parentheses are the corresponding $p-$values of the tests of $b_i(x) \equiv 0$, $i = 1, 2, 3$. The coefficient functions, the real data and the predicted values are shown in Figure 2(b). There we can see why the threshold value is chosen to be about 3.00 by Chan and Tong (1986) and Tong (1990), the coefficient functions change directions about this point.

As Chen and Tsay pointed out, FAR models have widespread applications. Related work on model identification and building has been considered by many authors. For example, Chen (1995) gives a procedure to select the threshold variable. Together with the results in this paper, one can make statistical inferences about FL models much as can be done in linear models.

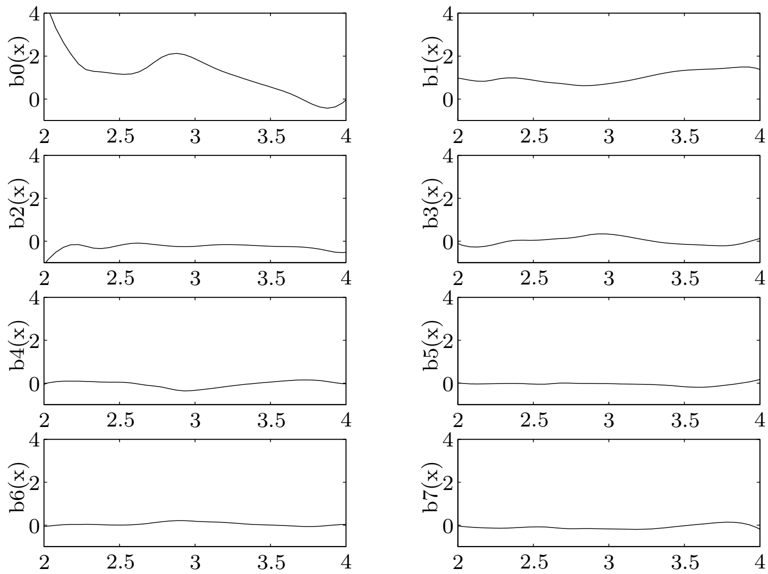

Figure 2(a). Estimates of coefficient functions of model (5.5).

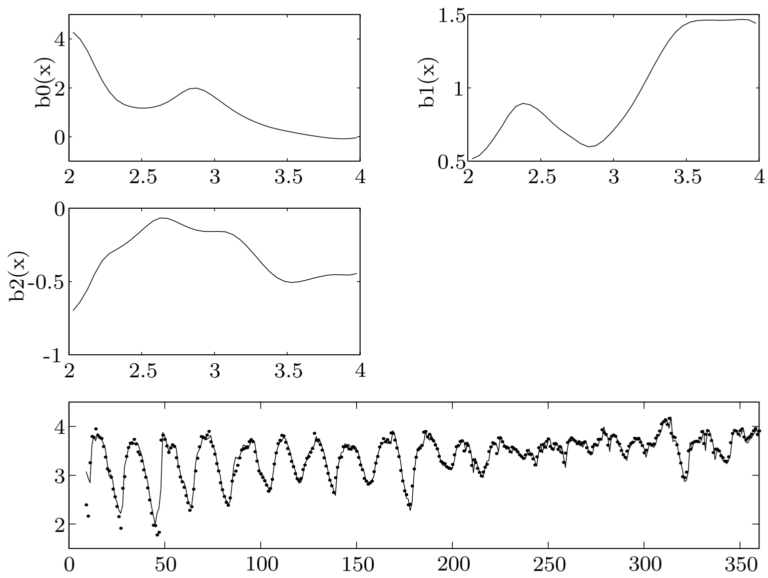

Figure 2(b). Estimates of coefficient functions of model (5.6) and the observations vs estimated values, · observations, – estimated values.

Acknowledgement

The authors are most grateful to the referees for suggesting improvements over an earlier version of this paper. W.K. Li's research was supported by the Hong Kong Research Grant Council.

Appendix. Proofs

In this section, we present some basic results about kernel smoothing under dependence.

Theorem A1. *Suppose that* $\{(\chi_t, \ \xi_t)\}$ *is a strictly stationary and strongly mixing sequence such that*

(i) *The mixing coefficient* $\alpha(k) = O(\varrho^k)$ *for some* $0 < \varrho < 1$*;*

(ii) $E(\xi_t|\chi_t = x)$ *is bounded;*

(iii) *The conditional densities* $f_{(\chi_1|\xi_1)}(u_1|v_1)$ *and* $f_{(\chi_1,\chi_l|\xi_1,\xi_l)}(u_1, u_l|v_1, v_l)$ *are bounded for all* $l > 1$;

(iv) $E|\xi_t|^\nu < \infty$ *for some* $\nu > 2$*;*

(v) *the density function* $f(x)$ *of* χ_t *is bounded.*

If $\phi(x)$ *is any bounded measurable function and* $U(x)$ *is any integrable function with bounded derivative and compact support, then*

(1) $\sum_{t=1}^n \Big[U((\chi_t - x)/a_n)\phi(\chi_t)\xi_t - E(U((\chi_t - x)/a_n)\phi(\chi_t)\xi_t)\Big] = \bar{O}((na_n \log n)^{1/2}, \mathbb{R})$,

(2) $n^{-1} \sum_{t=1}^n I(\chi_t < x)\xi_t - \int_{-\infty}^x E(\xi_t|\chi_t = s)f(s)ds = \bar{O}((n^{-1}\log n)^{1/2}, \ \mathbb{R})$,

(3) $\sup_{x,|\delta|<a_n} \Big|\sum_{t=1}^n I(x - \delta < \chi_t < x + \delta)\xi_t - \int_{x-\delta}^{x+\delta} E(\xi_t|\chi_t = s)f(s)ds\Big| = O((na_n \log n)^{1/2}) \quad a.s,$

where $a_n = O(n^{-\epsilon_1})$ *for some* $0 < \epsilon_1 < 1 - 2/\nu$.

Theorem A1(1) can be proved as in Masry and Tjøstheim (1995), while A1(2) and A1(3) were proved by Xia and Zhou (1997). From Theorem A1, we get the following lemma by taking $U(x) = K(x)x^k$ and $a_n = h$.

Lemma A2. *Suppose* (A2)-(A7) *hold. If* $h = O(n^{-\rho})$ *for some* $0 < \rho < 1/3$*, then*

$$\frac{1}{nh}\sum_{t=1}^n w_{kt}\mathbf{x}_{ti}\mathbf{x}_{tj} = \phi_k v_{ij}(x)f(x) + \bar{O}(\delta_{n1}, \mathbb{R}),$$
$$\frac{1}{nh}\sum_{t=1}^n w_{kt}\mathbf{x}_{ti}\varepsilon_t = \bar{O}(\delta_{n\infty}, \mathbb{R}), \quad i, j = 1, \ldots, p, \quad k = 0, 1, 2, 4,$$

where $w_{kt} = K((\mathbf{z}_t - x)/h)((\mathbf{z}_t - x)/h)^k$*,* $\delta_{n1} = h + (\log n/(nh))^{1/2}$*,* $\delta_{n\infty} = (\log n/(nh))^{1/2}$ *and* $\phi_k = \int u^k K(u)du$.

Let $A(x) = (v_{12}(x), v_{13}(x), \ldots, v_{1p}(x))^T$, $P(x) = (v_{ij}(x))_{i,j=2,\ldots,p}$ and $Q(x) = (v_{ij}(x))_{i,j=1,\ldots,p}$. From Lemma A2, we can easily show that (notations have been defined between (2.2) and (2.6))

$$\frac{1}{nh}A_n = A(x)f(x) + \bar{O}(\delta_{n1}, \mathbb{R})\mathbf{1}_{p-1}, \quad \frac{1}{nh}B_n = \bar{O}(1, \mathbb{R})\mathbf{1}_p,$$

$$\frac{1}{nh}R_n = \bar{O}(1, \mathbb{R})\mathbf{1}_{p-1}, \qquad \frac{1}{nh}Q_n = Q(x)f(x) + \bar{O}(\delta_{n1}, \mathbb{R})\mathbf{1}_p\mathbf{1}_p^T,$$
$$\frac{1}{nh}P_n = P(x)f(x) + \bar{O}(\delta_{n1}, \mathbb{R})\mathbf{1}_{p-1}\mathbf{1}_{p-1}^T. \tag{A.1}$$

Furthermore, since $f(x)\lambda_{\min}(P(x)) \geq f(x)\lambda_{\min}(Q(x)) > \epsilon$ as $x \in \mathbb{R}_\epsilon$, we have

$$(\frac{1}{nh}P_n)^{-1} = (P(x))^{-1}f^{-1}(x) + \bar{O}(\delta_{n1}, \mathbb{R}_\epsilon)\mathbf{1}_{p-1}\mathbf{1}_{p-1}^T \ ,$$
$$(\frac{1}{nh}Q_n)^{-1} = (Q(x))^{-1}f^{-1}(x) + \bar{O}(\delta_{n1}, \mathbb{R}_\epsilon)\mathbf{1}_p\mathbf{1}_p^T. \tag{A.2}$$

It is known that for any symmetric matrices A and D such that appropriate inverses exist,

$$\begin{pmatrix} A & B \\ & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + FE^{-1}F^T & -FE^{-1} \\ & E^{-1} \end{pmatrix},$$

where $E = D - B^T A^{-1} B$, $F = A^{-1}B$. Therefore,

$$H_n^{-1} = \begin{pmatrix} P_n^{-1} + P_n^{-1}R_n\mathcal{H}_n^{-1}R_n^TP_n^{-1} & -hP_n^{-1}R_n\mathcal{H}_n^{-1} \\ & h^{-2}\mathcal{H}_n^{-1} \end{pmatrix},$$

where $\mathcal{H}_n = Q_n - R_n^T P_n^{-1} R_n$, and

$$J_n^T H_n^{-1} = (A_n^T P_n^{-1} + A_n^T P_n^{-1} R_n \mathcal{H}_n^{-1} R_n P_n^{-1} - B_n^T \mathcal{H}_n^{-1} R_n^T P_n^{-1},$$
$$-h^{-1}A_n^T P_n^{-1} R_n \mathcal{H}_n^{-1} + h^{-1} B_n \mathcal{H}_n^{-1}), \tag{A.3}$$
$$J_n^T H_n^{-1} U_t = A_n P_n^{-1} S_t + A_n^T P_n^{-1} R_n \mathcal{H}_n^{-1} R_n^T P_n^{-1} S_t - B_n^T \mathcal{H}_n^{-1} R_n^T P_n^{-1} S_t$$
$$-h^{-1}(\mathbf{z}_t - x)(A_n^T P_n^{-1} R_n \mathcal{H}_n^{-1} - B_n^T \mathcal{H}_n^{-1})T_n. \tag{A.4}$$

It follows from (A.1)-(A.4) that

$$J_n^T H_n^{-1} = (A^T(x)(P(x))^{-1} + \bar{O}(\delta_{n1}, \mathbb{R}_\epsilon)\mathbf{1}_{p-1}^T, \quad \bar{O}(\delta_{n1}, \mathbb{R}_\epsilon)\mathbf{1}_p^T), \tag{A.5}$$
$$J_n^T H_n^{-1} J_n = A^T(x)(P(x))^{-1}A(x)f(x) + \bar{O}(\delta_{n1}, \mathbb{R}_\epsilon). \tag{A.6}$$

Thus

$$\frac{1}{nh}\sum_{t=1}^{n} w_t(\mathbf{x}_{t1} - J_n^T H_n^{-1} U_t)^2 = \frac{1}{nh}\sum_{t=1}^{n} w_t \mathbf{x}_{t1}^2 - J_n^T H_n^{-1} J_n$$
$$= v_{11}(x)f(x) - A^T(x)(P(x))^{-1}A(x)f(x) + \bar{O}(\delta_{n1}, \mathbb{R}_\epsilon)$$
$$= f(x)det(Q(x))/det(P(x)) + \bar{O}(\delta_{n1}, \mathbb{R}_\epsilon). \tag{A.7}$$

This result means that $\hat{\theta}_1(x)$ makes sense at (2.5). In notations defined between (2.2) and (2.6), we have the following result.

Lemma A3. *Assume* (A2)-(A7) *hold and* $h = O(n^{-\rho})$ *for some* $0 < \rho < 1/3$. *Then* $det(\sum_{t=1}^n w_t V_t\ V_t^T)$ *tends to infinity a.s. uniformly in* $x \in \mathbb{R}_\epsilon$ *as* $n \to \infty$.

Proof. Decompose

$$\sum_{t=1}^n w_t V_t V_t^T = \begin{pmatrix} \sum_{t=1}^n w_t T_t T_t^T & h\sum_{t=1}^n w_{1t} T_t T_t^T \\ & h^2 \sum_{t=1}^n w_{2t} T_t T_t^T \end{pmatrix} = \begin{pmatrix} \sqrt{nh} I_p & 0 \\ & \sqrt{nh^3} I_p \end{pmatrix}$$
$$\times \begin{pmatrix} (nh)^{-1} \sum_{t=1}^n w_t T_t T_t^T & (nh)^{-1} \sum_{t=1}^n w_{1t} T_t T_t^T \\ & (nh)^{-1} \sum_{t=1}^n w_{2t} T_t T_t^T \end{pmatrix} \begin{pmatrix} \sqrt{nh} I_p & 0 \\ & \sqrt{nh^3} I_p \end{pmatrix}.$$

By Lemma A2, we have

$$\frac{1}{nh}\sum_{t=1}^n w_t T_t T_t^T = Q(x) f(x) + \bar{O}(\delta_{n1}, \mathbb{R}_\epsilon) \mathbf{1}_p \mathbf{1}_p^T,$$
$$\frac{1}{nh}\sum_{t=1}^n w_{1t} T_t T_t^T = \bar{O}(\delta_{n1}, \mathbb{R}_\epsilon) \mathbf{1}_p \mathbf{1}_p^T,$$
$$\frac{1}{nh}\sum_{t=1}^n w_{2t} T_t T_t^T = Q(x) f(x) + \bar{O}(\delta_{n1}, \mathbb{R}_\epsilon) \mathbf{1}_p \mathbf{1}_p^T.$$

Notice that $nh^3 \to +\infty$ and Lemma A3 follows.

Lemma A3 ensures that the estimator of (2.4) makes sense.

Theorem A2. *Let* $\{(\chi_t, \xi_t)\}$ *be defined as in Theorem* A1. *Suppose* $E(\xi_t^2|\chi_t = s)$ *is bounded and* $E|\xi_t|^{8+\iota} < \infty$ *for some* $\iota > 0$; $\{\nu_t\}$ *is a sequence of i.i.d. r.v's with* $E\nu_t = 0$, $E\nu_t^2 = \sigma^2$ *and* $E|\nu_t|^{8+\iota} < \infty$. *For each* t, *suppose* ν_t *is independent of* $\{(\chi_s,\ \xi_s),\ s \le t\}$. *Then on a possibly enlarged probability space, there exists a sequence of standard Wiener process* $\mathcal{W}_n(x),\ x > 0$, *such that*

$$\sum_{t=1}^n I(\chi_t < x)\xi_t \nu_t - \sqrt{n}\sigma \mathcal{W}_n(\int_{-\infty}^x E(\xi_t^2|\chi_t = s) f(s) ds) = \bar{O}(n^{1/4}(\log n)^{\frac{3}{4}},\ [a,\ b]).$$

Note that the process $\sum_{t=1}^n I(\chi_t < x)\xi_t \nu_t,\ x \in \mathbb{R}$, is a hybrid of an empirical and a partial-sum process, and is of some interest in itself. Theorem A2 can be proved using the Skorohod embedding of multivariate random variables (Kiefer (1972)). The proof is lengthy, details can be obtained from the authors.

Proof of Theorem 3.1. Here we only do the proof for $\hat{\theta}_1(x)$ as an example. By the Cauchy-Schwarz inequality, (A.7), and Lemma A2,

$$|\frac{1}{nh}\sum_{t=1}^n w_t(\mathbf{x}_{t1} - J_n^T H_n^{-1} U_t)\theta''(\mathbf{z}_{ti})(\mathbf{z}_t - x)^2 \mathbf{x}_{ti}|$$

$$
\begin{aligned}
&\le \Big\{\frac{1}{nh}\sum_{t=1}^{n} w_t(\mathbf{x}_{t1}-J_n^T H_n^{-1}U_t)^2 \frac{1}{nh}\sum_{t=1}^{n} w_t\theta_i''^2(\mathbf{z}_{ti})(\mathbf{z}_t-x)^4\mathbf{x}_{ti}^2\Big\}^{\frac{1}{2}}\\
&\le M\Big\{\frac{1}{nh}\sum_{t=1}^{n} w_t(\mathbf{x}_{t1}-J_n^T H_n^{-1}U_t)^2 h^4\frac{1}{nh}\sum_{t=1}^{n} w_{4t}\mathbf{x}_{ti}^2\Big\}^{\frac{1}{2}}\\
&= M\Big\{[v_{11}(x)f(x)-A(x)(P(x))^{-1}A^T(x)f(x)+\bar{O}(\delta_{n1},\mathbb{R}_\epsilon)]\\
&\qquad \times h^4[\phi_4 v_{11}(x)f(x)+\bar{O}(\delta_{n1},\mathbb{R}_\epsilon)]\Big\}^{\frac{1}{2}}\\
&= \bar{O}(h^2,\mathbb{R}_\epsilon), \qquad i=1,\ldots,p. \qquad\qquad (A.8)
\end{aligned}
$$

Thus from (A.7) and (A.8), the second term on the right hand side of (2.5) is $\bar{O}(h^2,\mathbb{R}_\epsilon)$. By Lemma A2, we have

$$
\frac{1}{nh}\sum_{t=1}^{n} w_t T_t\varepsilon_t = \bar{O}(\delta_{n\infty},\mathbb{R})\mathbf{1}_p, \quad \frac{1}{nh}\sum_{t=1}^{n} w_{1t} T_t\varepsilon_t = \bar{O}(\delta_{n\infty},\mathbb{R})\mathbf{1}_p.
$$

Therefore from (A.5),

$$
\begin{aligned}
&\sum_{t=1}^{n} w_t(\mathbf{x}_{t1}-J_n^T H_n^{-1}U_t)\varepsilon_t = (1,\ -J_n^T H_n^{-1})\sum_{t=1}^{n} w_t V_t\varepsilon_t\\
&= (1,\ -J_n^T H_n^{-1})(\sum_{t=1}^{n} w_t T_t^T\varepsilon_t,\ h\sum_{t=1}^{n} w_{1t}T_t^T\varepsilon_t)^T\\
&= (1,\ -A^T(x)(P(x))^{-1})\sum_{t=1}^{n} w_t T_t\varepsilon_t + \bar{O}(nh\delta_{n1}\delta_{n\infty},\mathbb{R}_\epsilon). \qquad (A.9)
\end{aligned}
$$

Combining (A.9) with (A.7) and using Lemma A2, the last term on the right hand side of (2.5) is equal to $\bar{O}(\delta_{n\infty},\mathbb{R}_\epsilon)$.

Proof of Theorem 4.1. We only prove the theorem for $M_{n1}(x)$, the other cases are similar. From (2.5) and (A.7)-(A.9), to derive the distribution of $\hat{\theta}_1(x)$ we only need to handle the term $\sum_{t=1}^n w_t(\mathbf{x}_{t1}-A^T(x)(P(x))^{-1}S_t)\varepsilon_t$. Let $\mathbf{u}_t = \mathbf{x}_{t1}-A^T(x)(P(x))^{-1}S_t$ and $\Lambda=[a_0,b_0]$. Write

$$
\sum_{t=1}^{n} w_t\mathbf{u}_t\varepsilon_t = \sum_{t=1}^{n} I(\mathbf{z}_t\in\Lambda)w_t\mathbf{u}_t\varepsilon_t + \sum_{t=1}^{n} I(\mathbf{z}_t\notin\Lambda)w_t\mathbf{u}_t\varepsilon_t \triangleq L_1+L_2. \quad (A.10)
$$

Noticing that $|\mathbf{z}_t-x|>c_0$ when $x\in[a,b]$ and $\mathbf{z}_t\notin\Lambda$, we have, (as $\frac{c_0}{h}>\delta_0$),

$$
\sup_{x\in[a,b]}\Big|I(\mathbf{z}_t\notin\Lambda)w_t\mathbf{u}_t\varepsilon_t\Big| \le K(\frac{c_0}{h})\,(|\mathbf{x}_t\varepsilon_t|+M\|S_t\varepsilon_t\|)=0\ a.s. \qquad (A.11)
$$

Therefore the term L_2 is negligible when n is sufficiently large. Let $\Lambda_i = [d_{n(i-1)},d_{ni})$, $i=1,\ldots,n-1$, $\Lambda_n=[d_{n(n-1)},\ b_0]$, where $d_{ni}=a_0+i(b_0-$

$a_0)/n,\ i = 0, \ldots, n$. Define $\tilde{\mathbf{z}}_t = d_{ni}$ if $\mathbf{z}_t \in \Lambda_i,\ i = 1, \ldots, n$, and $\tilde{\mathbf{z}}_t = 0$ otherwise. By definition, $|I(\mathbf{z}_t \in \Lambda)\mathbf{z}_t - \tilde{\mathbf{z}}_t| = O(1/n)$. Using the Strong Law of Large Numbers for $\{\mathbf{u}_t\varepsilon_t\}$, we have

$$\begin{aligned} L_1 &= \sum_{t=1}^{n} I(\mathbf{z}_t \in \Lambda)\Big[K\Big(\frac{\mathbf{z}_t - x}{h}\Big) - K\Big(\frac{\tilde{\mathbf{z}}_t - x}{h}\Big)\Big]\mathbf{u}_t\varepsilon_t + \sum_{t=1}^{n} I(\mathbf{z}_t \in \Lambda)K\Big(\frac{\tilde{\mathbf{z}}_t - x}{h}\Big)\mathbf{u}_t\varepsilon_t \\ &= O(\frac{1}{nh})\sum_{t=1}^{n}|\mathbf{u}_t\varepsilon_t| + \sum_{t=1}^{n} I(\mathbf{z}_t \in \Lambda)K\Big(\frac{\tilde{\mathbf{z}}_t - x}{h}\Big)\mathbf{u}_t\varepsilon_t \\ &= \bar{O}(\frac{1}{h}, [a,b]) + \sum_{t=1}^{n} I(\mathbf{z}_t \in \Lambda)K\Big(\frac{\tilde{\mathbf{z}}_t - x}{h}\Big)\mathbf{u}_t\varepsilon_t. \end{aligned}$$

Therefore

$$\frac{1}{\sqrt{nh}}L_1 = \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} K\Big(\frac{\tilde{\mathbf{z}}_t - x}{h}\Big)I(\mathbf{z}_t \in \Lambda)\mathbf{u}_t\varepsilon_t + \bar{O}(\frac{1}{\sqrt{nh^3}}, [a,b]), \tag{A.12}$$

while

$$\sum_{t=1}^{n} K\Big(\frac{\tilde{\mathbf{z}}_t - x}{h}\Big)I(\mathbf{z}_t \in \Lambda)\mathbf{u}_t\varepsilon_t = \sum_{i=1}^{n} K\Big(\frac{d_{ni} - x}{h}\Big)\sum_{t=1}^{n} I(\mathbf{z}_t \in \Lambda_i)\mathbf{u}_t\varepsilon_t. \tag{A.13}$$

Let $\varsigma_i = \sum_{j=1}^{i}\sum_{t=1}^{n} I(\mathbf{z}_t \in \Lambda_j)\mathbf{u}_t\varepsilon_t = \sum_{t=1}^{n} I(a_0 \le \mathbf{z}_t < d_{ni})\mathbf{u}_t\varepsilon_t$, $\varsigma_0 \equiv 0$. By Abel's summation we have

$$\begin{aligned} &\sum_{t=1}^{n} K\Big(\frac{\tilde{\mathbf{z}}_t - x}{h}\Big)I(\mathbf{z}_t \in \Lambda)\mathbf{u}_t\varepsilon_t \\ &= K\Big(\frac{b_0 - x}{h}\Big)\varsigma_n - \sum_{i=1}^{n-1}\Big[K\Big(\frac{d_{n(i+1)} - x}{h}\Big) - K\Big(\frac{d_{ni} - x}{h}\Big)\Big]\varsigma_i. \end{aligned} \tag{A.14}$$

Let $G(z) = \int_{a_0}^{z} E(\mathbf{u}_t^2|\mathbf{z}_t = s)f(s)ds,\ \ \tilde{\delta}_n = n^{\frac{1}{4}}(\log n)^{\frac{3}{4}}$. By Theorem A2 and the bounded variation of $K(x)$, if (A8) holds, then

$$|\varsigma_n - \sigma n^{1/2}\mathcal{W}(G(b_0))| = O(\tilde{\delta}_n) \quad a.s.$$

and

$$\begin{aligned} &|\sum_{i=1}^{n-1}\Big[K\Big(\frac{d_{n(i+1)} - x}{h}\Big) - K\Big(\frac{d_{ni} - x}{h}\Big)\Big][\varsigma_i - \sigma n^{\frac{1}{2}}\mathcal{W}(G(d_{ni}))]| \\ &\le \max_{1 \le i < n}|\varsigma_i - \sigma n^{\frac{1}{2}}\mathcal{W}(G(d_{ni}))|\sum_{i=1}^{n-1}\Big|K\Big(\frac{d_{n(i+1)} - x}{h}\Big) - K\Big(\frac{d_{ni} - x}{h}\Big)\Big| \\ &\le \bar{O}(\tilde{\delta}_n, [a,b]). \end{aligned}$$

Consequently,

$$\sum_{t=1}^{n} K\Big(\frac{\tilde{\mathbf{z}}_t - x}{h}\Big) I(\mathbf{z}_t \in \Lambda)\mathbf{u}_t\varepsilon_t$$
$$= \sigma\sqrt{n}K\Big(\frac{b_0 - x}{h}\Big)W(G(b_0)) - \sigma\sqrt{n}\sum_{i=1}^{n-1}\Big[K\Big(\frac{d_{n(i+1)} - x}{h}\Big)$$
$$-K\Big(\frac{d_{ni} - x}{h}\Big)\Big]\mathcal{W}(G(d_{ni})) + \bar{O}(\tilde{\delta}_n, [a, b]). \tag{A.15}$$

For a Wiener process, it is known that (cf. Csörgő and Révész (1981), p.44)

$$\sup_{z\in\Lambda} |\mathcal{W}(G(z + \delta)) - \mathcal{W}(G(z))| = O((\delta \log(1/\delta))^{1/2}\ ) \quad a.s.$$

where δ is any small number. Using this property and the bounded variation of $K(x)$, we have

$$\sum_{i=1}^{n-1}\Big[K\Big(\frac{d_{n(i+1)} - x}{h}\Big) - K\Big(\frac{d_{ni} - x}{h}\Big)\Big]\mathcal{W}(G(d_{ni}))$$
$$= \int_{a_0}^{b_0} \mathcal{W}(G(s))dK\Big(\frac{s - x}{h}\Big) + \bar{O}(\Big(\frac{\log n}{n}\Big)^{\frac{1}{2}}, [a,\ b]). \tag{A.16}$$

From (A.12)-(A.16), we conclude that

$$\frac{1}{\sqrt{nh}}L_1 = \sigma h^{-\frac{1}{2}}\int_{a_0}^{b_0} K\Big(\frac{s - x}{h}\Big)d\mathcal{W}(G(s)) + \bar{O}(\frac{\tilde{\delta}_n}{\sqrt{nh}} + \frac{1}{\sqrt{nh^3}},\ [a, b]). \tag{A.17}$$

Hence from (2.5), (A.7), (A.8), (A.9), (A.17) and (A.10) we have

$$\frac{1}{\sqrt{nh}\sigma}\Big[\Gamma_n^{-1}(1, 1)\Big]^{-1}(\hat{\theta}_1(x) - \theta_1(x))$$
$$= h^{-\frac{1}{2}}\int_{a_0}^{b_0} K\Big(\frac{s - x}{h}\Big)d\mathcal{W}(G(s)) + \bar{O}(n^{\frac{1}{2}}h^{\frac{5}{2}} + n^{\frac{1}{2}}h^{\frac{1}{2}}\delta_{n1}\delta_{n\infty} + \frac{\tilde{\delta}_n}{\sqrt{nh}} + \frac{1}{\sqrt{nh^3}},\ [a, b])$$
$$\triangleq Y_n(x) + \bar{O}(\bar{\delta}, [a, b]). \tag{A.18}$$

If $h = O(n^{-\rho})$ for some $1/5 < \rho < 1/3$, then $\bar{\delta} \to 0$ as $n \to \infty$. Let $g(x, s) = [E(\mathbf{u}_t^2|\mathbf{z}_t = s)f(s)]^{\frac{1}{2}} = v_{11}(s)f(s) - A^T(x)(P(x))^{-1}A(s)f(s)$. From assumptions (A4) and (A6), we know that $g(x, s) > 0$, $(x, s) \in [a_0, b_0] \times [a_0, b_0]$, and has bounded derivative. For Wiener processes, it is easy to show by calculating their moments that

$$\mathcal{W}(G(y)) \overset{D}{=} \int_{a_0}^{y} g(x, s)d\mathcal{W}(s - a_0).$$

Thus

$$Y_n(x) \overset{D}{=} h^{-\frac{1}{2}} \int_{a_0}^{b_0} K\Big(\frac{s-x}{h}\Big) g(x,s) d\mathcal{W}(s-a_0). \tag{A.19}$$

Let $Y_{1n}(x) = Y_n(x)/g(x,x)$ and

$$Y_{0n}(x) = h^{-\frac{1}{2}} \int_{a_0}^{b_0} K\Big(\frac{s-x}{h}\Big) d\mathcal{W}(s-a_0).$$

A simple result about $Y_{0n}(x)$ is (cf. Xia (1998))

$$Y_{0n}(x) = \bar{O}((\log n)^{1/2},\ [a,\ b]). \tag{A.20}$$

Furthermore, following the method of Härdle (1989), we have

$$|Y_{1n}(x) - Y_{0n}(x)| = O_p(h^{\frac{1}{2}}), \tag{A.21}$$

uniformly for $x \in [a,\ b]$.

On the other hand, we have by definition,

$$(g(x,x))^2 = v_{11}(x)f(x) - A^T(x)(P(x))^{-1}A(x)f(x).$$

From (2.6) and (A.7) and noticing that $g(x,x)$ is bounded away from zero on $x \in \Lambda$, we have

$$|\sqrt{nh}(\Gamma_n^{-1}(1,1))^{\frac{1}{2}} - (g(x,x))^{-1}| = \bar{O}(\delta_{n1}, \Lambda). \tag{A.22}$$

Hence from (A.18), (A.20) and (A.22), we have

$$\begin{aligned} &|\sigma^{-1}(\Gamma_n^{-1}(1,1))^{-\frac{1}{2}}[\hat{\theta}_1(x) - \theta_1(x)] - Y_n(x)/g(x,x)| \\ &= \sigma^{-1}\Big|Y_n(x)\Big[\sqrt{nh}(\Gamma_n^{-1}(1,1))^{\frac{1}{2}} - (g(x,x))^{-1}\Big]\Big| + \bar{O}(\bar{\delta}, [a,b]) \\ &= \bar{O}(\bar{\delta} + \delta_{n1}(\log n)^{1/2}, [a,b]). \end{aligned} \tag{A.23}$$

From (A.22), (A.23), Theorems 3.1 and 3.2, we have

$$\begin{aligned} |M_{ni}(x) - Y_n(x)/g(x,x)| &\le |(\hat{\sigma}^{-1} - \sigma^{-1})(\Gamma_n^{-1}(1,1))^{-\frac{1}{2}}[\hat{\theta}_1(x) - \theta_1(x)]| \\ &\quad + |\sigma^{-1}(\Gamma_n^{-1}(1,1))^{-\frac{1}{2}}\Big[\hat{\theta}_1(x) - \theta_1(x)\Big] - Y_n(x)/g(x,x)| \\ &= \bar{O}(\sqrt{nh}\delta_{n2}^2 + \bar{\delta} + \delta_{n1}(\log n)^{1/2}, [a,b]). \end{aligned} \tag{A.24}$$

Hence, Theorem 4.1 follows from (A.21) and (A.24).

References

Bickel, P. L. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071-1095.

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In *Dependence in Probability and Statistics: A Survey of Recent Results* (Edited by E. Eberlein and M. S. Taqqu), 165-192. Birhauser, Boston.

Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *J. Time Ser. Anal.* **7**, 277-281.

Chen, R. (1995). Threshold variable selection in open-loop threshold autoregressive models. *J. Time Ser. Anal.* **16**, 461-481.

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

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88**, 1287-1301.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **2**, 196-216.

Fan, J. and Gijbels, J. (1996). *Local Polynomial Modeling and Its Applications.* Chapman and Hall, London.

Freedman, D. A. (1975). On tail probability for martingales. *Ann. Probab.* **3**, 100-118.

Gallant, A. R. (1987). *Nonlinear Statistics Models.* John Wiley, New York.

Haggan, V. and Ozaki, T. (1981). Modeling vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* **68**, 189-196.

Hanson, D. L. and Russo, P. (1983). Some results on increments of the Wiener processes with applications to lag sums of i.i.d. random variables. *Ann. Probab.* **11**, 428-436.

Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multivariate Anal.* **29**, 163-179.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Boston.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models (with discussion). *J. Roy. Statist. Soc. Ser. B* **55**, 757-796.

Kiefer, J. (1972). Skorohod embedding of multivariate RV's, and the sample DF. *Z. Wahrsch. verw. Geb.* **24**, 1-35.

Lai, T. L., Robbins, H. and Wei, C. Z. (1979). Strong consistency of least squares estimates in multiple regression II. *J. Multivariate Anal.* **9**, 343-361.

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

McLeish, D. L. (1975). A maximal inequality and dependent strong laws. *Ann. Probab.* **3**, 829-839.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* John Wiley, New York.

Rio, E. (1995). The functional law of the iterated logarithm for stationary strongly mixing sequences. *Ann. Probab.* **23**, 1188-1203.

Stute, W. (1982). A law of the logarithm for kernel density estimators. *Ann. Probab.* **10**, 414-422.

Tsay, R. S. (1988). Non-linear time series analysis of blowfly population. *J. Time Ser. Anal.* **9**, 247-263.

Tong, H. (1990). *Nonlinear Time Series Analysis: A Dynamic Approach.* Oxford University Press, London.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. Roy. Statist. Soc. Ser. B* **60**, 797-811.

Xia, Y. and Li, W. K. (1997). Asymptotic behavior of bandwidth selected by cross-validation method under dependence. Manuscript, The University of Hong Kong.

Xia, Y. and Zhou, Y. (1997). Strong uniform convergence rate for randomly weighted empirical processes and its applications under dependence. Manuscript, The University of Hong Kong.

Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong and College of Economics, Jinan University, Guangzhou, China.

E-mail: ycxia@hkusua.hku.hk

Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong.

E-mail: hrntlwk@hkucc.hku.hk

(Received August 1997; accepted September 1998)