Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

On Single-Index Coefficient Regression Models

Yingcun Xia & W. K. Li

To cite this article: Yingcun Xia & W. K. Li (1999) On Single-Index Coefficient Regression Models, Journal of the American Statistical Association, 94:448, 1275-1285

To link to this article: http://dx.doi.org/10.1080/01621459.1999.10473880

Published online: 17 Feb 2012.

On Single-Index Coefficient Regression Models

Yingcun XIA and W. K. LI

In this article we investigate a class of single-index coefficient regression models under dependence. This includes many existing models, such as the smooth transition threshold autoregressive (STAR) model of Chan and Tong, the functional-coefficient autoregressive (FAR) model of Chen and Tsay, and the single-index model of Ichimura. Compared to the varying-coefficient model of Hastie and Tibshirani, our model can avoid the curse of dimensionality in multivariate nonparametric estimations. Another advantage of this model is that a threshold variable is chosen automatically. An estimation method is proposed, and the corresponding estimators are shown to be consistent and asymptotically normal. Some simulations and applications are also reported.

KEY WORDS: Kernel smoothing; Nonparametric time series; Single-index model; Strongly mixing; Varying-coefficient model.

1. INTRODUCTION

Hastie and Tibshirani (1993) proposed a varying-coefficient model,

$$y = \phi_0(Z) + \phi_1(Z)x_1 + \cdots + \phi_p(Z)x_p + \varepsilon, \qquad (1)$$

where $\phi_i(\cdot), i = 0, \ldots, p$, are unknown coefficient functions; y is the response variable; Z is a random q-covariate and X is a random p-covariate. Following Tong (1990), Z may be called threshold variables. Model (1) and its variants have gained much attention in the literature and have been intensively investigated (see, e.g., Chen and Tsay 1993; Hastie and Tibshirani 1993). Hastie and Tibshirani (1993) gave many examples of (1) and proposed some methods to estimate the coefficient functions. In the case of $q = 1$, Xia and Li (1999) made a complete study of the model under dependence. They used the local linear smoother to estimate the coefficient functions, proved their uniform convergence with logarithm rate, and obtained their asymptotic distribution as Gaussian processes. Although these results can be extended to the case $q > 1$, an unrealistically large sample size is needed in practice to obtain an appropriate estimate of $\phi_i(\cdot), i = 0, 1, \ldots, p$. This is the so-called "curse of dimensionality" in estimating multivariate nonparametric regression functions.

Single-index models, or projection pursuit regression, have proven to be an efficient way of coping with the high-dimensional problem in nonparametric regressions (see, e.g., Hall 1989; Ichimura 1993). The idea is restricting the general multivariate regression function to a special form. The single-indexing specification of Ichimura (1993) has proven to be very efficient in practice. It is natural to introduce this idea to (1) in specifying the coefficient functions. More generally, we suggest using a nonlinear function to "single index" the threshold variables. Therefore, we propose the following model:

$$y = \phi_0(g(\theta_0, Z)) + \phi_1(g(\theta_0, Z))x_1 + \cdots + \phi_p(g(\theta_0, Z))x_p + \varepsilon, \qquad (2)$$

where $g(\theta, \mathcal{Z})$: $\Re^{k+q} \mapsto \Re$ is known up to a parametric vector $\theta \in \Theta, \Theta \subset \Re^k$ is usually a convex subset, and $\mathcal{Z} \in \Re^q$. This kind of single indexing was first considered by Ichimura (1993). A special case is when $g(\theta_0, \mathcal{Z}) = \theta_0^T \mathcal{Z}$ with $\|\theta_0\| = 1$, which may be seen as a generalized case of the linear single-index model (see Härdle, Hall, and Ichimura 1993 for details). Ichimura (1993) gave some other forms of the single-indexing function $g(\theta, \mathcal{Z})$.

In this article we consider (2) under dependence. Suppose that $\{(Z_t, X_t, y_t), t = 1, 2, \ldots, n\}$ is a random sample with $Z_t = (z_{t1}, \ldots, z_{tq})^T$ and $X_t = (x_{t1}, \ldots, x_{tp})^T$. We allow the sample to be a sequence of dependent data. Therefore, (2) may be seen as a nonparametric time series model if we take $x_{ti} = y_{t-i}, i = 1, \ldots, p$, and $z_{tj} = y_{t-j}, j = 1, \ldots, q$. Model (2) thus includes many existing nonlinear times series models; examples include the exponential autoregressive (EXPAR) model of Haggan and Ozaki (1981), the smooth transition threshold autoregressive (STAR) model of Chan and Tong (1986), and the functional-coefficient autoregressive (FAR) model of Chen and Tsay (1993). These models have been found to be very useful in nonlinear time series modeling. For these models, an important issue is to find a proper threshold variable (see, e.g., Chen 1995; Tong 1990). Chen (1995) proposed a method for searching for a threshold variable for an open-loop threshold autoregressive model. Another advantage of (2) is that the threshold variable $g(\theta, Z_t)$ can be chosen automatically by estimating θ.

Next we discuss the geometric ergodicity of (2) when it is used as a nonparametric time series model. Suppose that $\phi_i(\cdot), i = 1, \ldots, p$, are measurable functions and that $\phi_i(\cdot)$ can be written as $\phi_i(\cdot) = \alpha_i(\cdot) + \beta_i(\cdot)$ with $\beta_i(\mathcal{Z})\|\mathcal{Z}\|$ bounded on $\Re^q$ and $|\alpha_i(\cdot)| \le c_i$ 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. Furthermore, if the density function of ε_t is positive everywhere and $\lim_{|u|\to\infty} \sup |\phi_0(u)/u| \to 0$, then the Markov chain $\{y_t\}$ is geometrically ergodic (see Chen and Tsay 1993). More generally, we assume herein that $\{(Z_t, X_t, y_t)\}$ is a strongly mixing sequence, which is a very common dependence assumption. By the results of Pham (1986), a geometrically ergodic time series is a strongly mixing se-

Yingcun Xia is postdoctoral fellow and W. K. Li is Professor, Department of Statistics, The University of Hong Kong, Pokfulam Road, Hong Kong (E-mail: *hrntlwk@hku.hk*). The research of Li was partially supported by the Hong Kong Research Grant Council.

Journal of the American Statistical Association
December 1999, Vol. 94, No. 448, Theory and Methods

quence. Therefore, the strongly mixing assumption is suitable for (2) as a time series model under the aforementioned conditions.

In this article we obtain the estimators of $\phi_i(\cdot), i = 0, 1, \ldots, p$, by kernel smoothing and the estimator of θ by the least squares method. We prove the asymptotic normality of the estimator of θ and the consistency of the estimators for $\phi_i(\cdot)$ with the optimal uniform convergence rate. Some simulations and applications to real datasets are considered. We find that the single-index coefficient regression models are more suitable to the datasets under consideration than models proposed earlier by other authors. The rest of this article is organized as follows. Section 2 states the estimation method and the main results. Section 3 presents some simulations and applications to real data. The Appendix provides proofs.

2. METHOD OF ESTIMATION AND MAIN RESULTS

Throughout this article, we assume that $\{(Z_t, X_t, y_t)\}$ is strictly stationary. This is implied by the geometric ergodicity if we adopt an appropriate initial distribution. Let $f(\mathcal{Z}, \mathcal{X}, y)$ denote the joint density function of (Z_t, X_t, y_t), $f(\mathcal{Z}, \mathcal{X})$ denote the density function of (Z_t, X_t) and $f(\mathcal{Z})$ denote the density function of Z_t if they do not cause any confusion. Define $\phi_{0\theta}(z), \phi_{1\theta}(z), \ldots, \phi_{p\theta}(z)$ as the solution to the minimizer

$$\min_{\phi_{0\theta},\ldots,\phi_{p\theta}} E\{(y_t - \phi_{0\theta}x_{t0} - \cdots - \phi_{p\theta}x_{tp})^2 \,|\, g(\theta, Z_t) = z\}, \quad (3)$$

where we take $x_{t0} \equiv 1$ for convenience of exposition and redefine $X_t = (x_{t0}, x_{t1}, \ldots, x_{tp})^T$. Let $\Phi_\theta(z) = (\phi_{0\theta}(z), \phi_{1\theta}(z), \ldots, \phi_{p\theta}(z))^T$, $\mathcal{W}(z,\theta) = E(X_t X_t^T | g(\theta, Z_t) = z)$, and $\mathcal{V}(z,\theta) = E(X_t y_t | g(\theta, Z_t) = z)$. We can see that the solution to (3) is

$$\Phi_\theta(z) = [\mathcal{W}(z,\theta)]^{-1}\mathcal{V}(z,\theta),$$

provided that $[\mathcal{W}(z,\theta)]^{-1}$ exists. Particularly, $\phi_0(z), \phi_1(z), \ldots, \phi_p(z)$ are just the solutions to (3) at $\theta = \theta_0$ if we further assume that $E(\varepsilon_t | Z_t, X_t) = 0$ a.s. Therefore, $\Phi_{\theta_0}(z) = (\phi_0(z), \phi_1(z), \ldots, \phi_p(z))^T$.

For the sequence of observations $\{(Z_t, X_t, y_t), t = 1, \ldots, n\}$, following the idea of kernel smoothing, the estimator of $\phi_{i\theta}(z)$ is the solution of a_i to the minimizer

$$\min_{a_0, a_1, \ldots, a_p} \sum_{t=1}^{n} \left[y_t - \sum_{i=0}^{p} a_i x_{ti} \right]^2 K_h(g(\theta, Z_t) - z), \quad (4)$$

where $K(\cdot)$ is a kernel function, $K_h(\cdot) = K(\cdot/h)$, and $h \in \mathcal{H}$ is the bandwidth. Let $\mathcal{W}_n(z,\theta,h) = (nh)^{-1}\sum_{t=1}^n K_h(g(\theta, Z_t) - z) X_t X_t^T$ and $\mathcal{V}_n(z,\theta,h) = (nh)^{-1}\sum_{t=1}^n K_h(g(\theta, Z_t) - z) X_t y_t$. The solution to (4) [i.e., the estimator of $\Phi_\theta(z)$] is

$$\hat{\Phi}_\theta(z) = [\mathcal{W}_n(z,\theta,h)]^{-1}\mathcal{V}_n(z,\theta,h). \quad (5)$$

In the case where $[\mathcal{W}_n(z,\theta,h)]^{-1}$ does not exist or is too large, we may consider only a subset $\mathcal{A}$ of $\Re^q$ such that $\mathcal{W}_n(z,\theta,h)$ tends to a positive definite matrix for $z \in \{g(\theta, \mathcal{Z})\colon \mathcal{Z} \in \mathcal{A}\}$. In fact, we show in the next section that $\mathcal{A}$ is nonempty under certain conditions.

Let $\hat{\Phi}_{\theta t}(z) = (\hat{\phi}_{0\theta t}(z), \ldots, \hat{\phi}_{p\theta t}(z))^T$ be the estimator from (5) using data $\{(Z_s, X_s, y_s)\colon s \neq t\}$. Define

$$\hat{S}(\boldsymbol{\theta}, h) = \sum_{Z_t \in \mathcal{A}} [y_t - \hat{\Phi}_{\theta t}^T(g(\boldsymbol{\theta}, Z_t))X_t]^2.$$

We estimate the parameter vector $\boldsymbol{\theta}$ and the bandwidth h by minimizing $\hat{S}(\boldsymbol{\theta}, h)$. Suppose that $(\hat{\theta}, \hat{h})$ is the pair of solutions. We estimate the variance of ε_t as

$$\hat{\sigma}^2 = \frac{1}{\#\{Z_t \in \mathcal{A}\}}\hat{S}(\hat{\theta}, \hat{h}), \quad (6)$$

where $\#\{Z_t \in \mathcal{A}\}$ denotes the number of elements in $\{Z_t \in \mathcal{A}\}$. Because we estimate $\hat{\Phi}_{\theta t}(\cdot)$ without using the observation $(Z_t, X_t, y_t), \sigma^2$ is usually overestimated by (6); that is, $\hat{\sigma}^2 > \sigma^2$ usually (see eq. 2.5 of Härdle, Hall, and Marron 1988). However, it thus can be used to assess the goodness of fit of the model from an unfavorable point of view.

We expect that $\hat{\theta} \to \boldsymbol{\theta}_0$ and $\hat{h}/h_0 \to 1$ a.s., where h_0 is the theoretically optimal bandwidth that minimizes

$$J(h) = \int\int_{\mathcal{A}} E\{[\hat{\Phi}_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, \mathcal{Z})) - \Phi_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, \mathcal{Z}))]^T \mathcal{X}\}^2 \times f(\mathcal{Z}, \mathcal{X})\, d\mathcal{Z}\, d\mathcal{X}.$$

It is not difficult to see that $h_0 \propto n^{-1/5}$ (cf. Härdle and Vieu 1992), where $\propto$ means that both sides have the same order. Next, we consider only the case

$$\min_{\theta \in \Theta_n, h \in \mathcal{H}_n} \hat{S}(\boldsymbol{\theta}, h),$$

where $\Theta_n = \{\boldsymbol{\theta}\colon \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le C_0 n^{-\delta}\}$ with $3/10 < \delta < 1/2$, $\mathcal{H}_n = \{h\colon C_1 n^{-1/5} \le h \le C_2 n^{-1/5}\}$ for some constants C_0 and $C_1 < C_2$. Note that by Corollary 1 (which follows); $\|\hat{\theta} - \boldsymbol{\theta}_0\| = O_p(n^{-1/2})$, and so $\hat{\theta}$ is well inside the cone Θ_n. Therefore, our restriction for $\boldsymbol{\theta}$ to the cone Θ_n does not exclude any minima of interest and is made without loss of generality. Because $\hat{h}$ is expected to be close to $h_0 \propto n^{-1/5}$, we should look for a minimum of $\hat{S}(\boldsymbol{\theta}, h)$ that involves h approximately equal to a constant multiple of $n^{-1/5}$. Similar restrictions were also made by Hall (1989). Härdle et al. (1993) went so far as to consider only the cone Θ_n with $\delta = 1/2$. These regions are imposed only for the theoretical proof. In practice, we need only assume that $\hat{S}(\boldsymbol{\theta}, h)$ has a single minimum point over $\Theta \times \mathcal{H}$.

To discuss the asymptotic properties, we need the following assumptions. Assume that $\mathcal{A} \subset \Re^q$ is the union of a number of open convex sets such that $\min_{\mathcal{Z} \in \mathcal{A}} \det[\mathcal{W}(g(\boldsymbol{\theta}_0, \mathcal{Z}), \boldsymbol{\theta}_0) f(\mathcal{Z})] > 0$. We concentrate our attention in this region. Given $\rho > 0$, let $\mathcal{A}^\rho$ denote the set of all points in $\Re^q$ with distance no further than ρ from $\mathcal{A}$. Let $\mathcal{U} = \{z = g(\boldsymbol{\theta}_0, \mathcal{Z})\colon \mathcal{Z} \in \mathcal{A}\}, \mathcal{U}^\rho = \{z = g(\boldsymbol{\theta}_0, \mathcal{Z})\colon \mathcal{Z} \in \mathcal{A}^\rho\}$, $z_{\theta t} = g(\boldsymbol{\theta}, Z_t)$, $z_t = g(\boldsymbol{\theta}_0, Z_t)$, and $f(z, \boldsymbol{\theta})$ denote the density function of $z_{\theta t}$.

(C1) $\phi_{i\boldsymbol{\theta}}(z), i = 0, \ldots, p, \mathcal{W}(z, \boldsymbol{\theta}), \mathcal{V}(z, \boldsymbol{\theta})$, and $f(z, \theta)$ have bounded, continuous second-order derivatives with respect to $z \in \mathcal{U}^\rho$ for all $\boldsymbol{\theta} \in \Theta$.

(C2) $\{(Z_t, X_t, y_t)\}$ is a strictly stationary and strongly mixing sequence with mixing coefficient $\alpha(l) = O(\rho^l)$ for some $0 < \rho < 1$.
(C3) $c < f(\mathcal{Z}) < C$ for some positive constants c and C and has bounded second-order derivatives in $\mathcal{A}^\rho$.
(C4) $g(\boldsymbol{\theta}, \mathcal{Z})$ has bounded derivatives in $\Theta \times \mathcal{A}^\rho$.
(C5) The conditional density functions $f_{z_1|y_1,X_1}(z|y,\mathcal{X})$, $f_{z_1,z_l|y_1,y_l,X_1,X_l}(z_1,z_l|y_1,y_l,\mathcal{X}_1,\mathcal{X}_l)$ are bounded for all $l \geq 1$.
(C6) $E(\varepsilon_t|Z_t,X_t) = 0$ a.s., $E(\varepsilon_t^2) = \sigma^2, E\|X_t\|^l < \infty$, and $E|y_t|^l < \infty$ for all $l > 0$.
(C7) $K(z)$ is supported on the interval $(-1, 1)$ and is a symmetric probability density function with a bounded derivative. Furthermore, the Fourier transformation of $K(z)$ is absolutely integrable.

(C1) is needed to use the symmetric kernel in (C7) to achieve an order of bias $O(h^2)$ for the kernel smoothing. Higher-order bias can be obtained by imposing stronger conditions on the smoothness of $\phi_i(z)$ and conditions on the kernel function (see, e.g., Fan and Gijbels 1996). (C2) is made only for the purpose of simplicity; it can be weakened to $\alpha(l) = O(l^{-\kappa})$ for some $\kappa > 0$. However, many time series models satisfy (C2). Examples are the nonparametric ARCH models (Masry and Tjøstheim 1995) and (2) as a time series model under certain conditions. (C3) is made to estimate the coefficient functions in the region $\mathcal{A}^\rho$. (C4) is used to obtain the asymptotic properties for the nonlinear least squares estimator of $\boldsymbol{\theta}$. (C5) is a commonly used assumption for dependent data (see, e.g., Masry and Tjøstheim 1995). In (C6), $E(\varepsilon_t|Z_t, X_t) = 0$ a.s. ensures the consistency of the estimators. The moments may exist only up to a sufficiently large l. This is assumed so that the Chebyshev inequality can be applied. The second part of (C7) is needed to use the moment inequality of Kim and Cox (1995). Similar assumptions have been made by others, including Kim and Cox (1995), Härdle et al. (1993), and Ichimura (1993).

Define

$$\mu(\mathcal{Z}|\boldsymbol{\theta}) = E\left(\frac{\partial g(\boldsymbol{\theta}, Z_t)}{\partial \boldsymbol{\theta}}\bigg| g(\boldsymbol{\theta}, Z_t) = g(\boldsymbol{\theta}, \mathcal{Z})\right),$$

$$k_1 = \int z^2 K(z)\, dz,$$

$$k_2 = \int K^2(z)\, dz,$$

$$V_n = \sum_{Z_t \in \mathcal{A}} \left[\frac{\partial g(\boldsymbol{\theta}_0, Z_t)}{\partial \boldsymbol{\theta}} - \mu(Z_t|\boldsymbol{\theta}_0)\right] \dot{\Phi}_{\boldsymbol{\theta}_0}^T(g(\boldsymbol{\theta}_0, Z_t)) X_t \varepsilon_t,$$

$$\tilde{V}_0 = \frac{1}{n}\operatorname{var}(V_n),$$

$$a_1 = \sigma^2 k_2 \int_{\mathcal{A}} [f(g(\boldsymbol{\theta}_0, \mathcal{Z}), \boldsymbol{\theta}_0)]^{-1} f(\mathcal{Z})\, d\mathcal{Z},$$

and

$$a_2 = \frac{1}{4}k_1^2 \int\int_{\mathcal{A}} [\ddot{\Phi}_{\boldsymbol{\theta}_0}^T(g(\boldsymbol{\theta}_0, \mathcal{Z}))\mathcal{X}]^2 f(\mathcal{Z}, \mathcal{X})\, d\mathcal{Z}\, d\mathcal{X},$$

where $\dot{\Phi}_{\boldsymbol{\theta}_0}(z) = \partial\Phi_{\boldsymbol{\theta}_0}(z)/\partial z$ and $\ddot{\Phi}_{\boldsymbol{\theta}_0}(z) = \partial\dot{\Phi}_{\boldsymbol{\theta}_0}(z)/\partial z$. In this notation, $J(h) \cong a_1 h^{-1} + a_2 n h^4$ and thus $h_0 \cong (a_1/4a_2 n)^{1/5}$, where $\cong$ means that the quotient of both sides tends to 1 as $n \to \infty$ (see Härdle et al. 1993). Let

$$\tilde{S}(\boldsymbol{\theta}) = \sum_{Z_t \in \mathcal{A}} [y_t - \Phi_{\boldsymbol{\theta}}^T(g(\boldsymbol{\theta}, Z_t))X_t]^2$$

and

$$T(h) = \sum_{Z_t \in \mathcal{A}} \{[\hat{\Phi}_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, Z_t)) - \Phi_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, Z_t))]^T X_t\}^2.$$

We have the following decomposition about the quantity $\hat{S}(\boldsymbol{\theta}, h)$.

Theorem 1. Under the preceding conditions, we may write

$$\hat{S}(\boldsymbol{\theta}, h) = \tilde{S}(\boldsymbol{\theta}) + T(h) + R_1(\boldsymbol{\theta}, h) + R_2(h), \tag{7}$$

where $R_2(h)$ does not depend on $\boldsymbol{\theta}$, and

$$\sup_{\boldsymbol{\theta} \in \Theta_n, h \in \mathcal{H}_n} |R_1(\boldsymbol{\theta}, h)| = O(n^{1/5-\tau}\log n),$$

$$\sup_{h \in \mathcal{H}_n} |R_2(h)| = O(n^{1/5-\tau}) \quad \text{a.s.,} \tag{8}$$

with $0 < \tau < 1/10$. Furthermore,

$$\tilde{S}(\boldsymbol{\theta}) = n[\tilde{V}_0^{1/2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - n^{-1/2}\sigma\boldsymbol{\xi}]^T \times [\tilde{V}_0^{1/2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - n^{-1/2}\sigma\boldsymbol{\xi}] + R_3 + R_4(\boldsymbol{\theta}) \tag{9}$$

and

$$T(h) = a_1 h^{-1} + a_2 n h^4 + R_5(h), \tag{10}$$

where $\boldsymbol{\xi}$ is an asymptotically standard normal random k-vector such that $V_n = n^{1/2}\sigma\tilde{V}_0^{1/2}\boldsymbol{\xi}$, R_3 depends on neither $\boldsymbol{\theta}$ nor h, and

$$\sup_{\boldsymbol{\theta} \in \Theta_n} |R_4(\boldsymbol{\theta})| = o(1), \qquad \sup_{h \in \mathcal{H}_n} |R_5(h)| = o(n^{1/5}) \quad \text{a.s.}$$

Theorem 1 is a direct extension of the work of Härdle et al. (1993) to a more complicated model (2) under the strongly mixing situation. However, its proof requires several rather technical lemmas. These results are derived for a larger parameter cone Θ_n than that of Härdle et al. (1993). By Theorem 1, we can investigate properties of the estimator of h assuming that $\boldsymbol{\theta}_0$ is known. The problem is then equivalent to the case of $q = 1$. Similarly, we can investigate the estimator of $\boldsymbol{\theta}$ assuming that h_0 is known. From Theorem 1, we further have the following results about the estimators of $\boldsymbol{\theta}$ and h.

Corollary 1. Under the assumptions of Theorem 1, and for each t, ε_t is independent of $\{(Z_s, X_s): s \le t\}$. Then

$$\hat{\theta} - \boldsymbol{\theta}_0 = n^{-1/2}\sigma(V_0^-)^{1/2}\boldsymbol{\xi} + o_p(n^{-1/2})$$

and

$$\hat{h} = h_0 + o(n^{-1/5}) \quad \text{a.s.},$$

where

$$V_0 = \int\int_{\mathcal{A}} \left[\frac{\partial g(\theta_0, \mathcal{Z})}{\partial \boldsymbol{\theta}} - \mu(\mathcal{Z}|\boldsymbol{\theta}_0)\right] \times \left[\frac{\partial g(\boldsymbol{\theta}_0, \mathcal{Z})}{\partial \boldsymbol{\theta}} - \mu(\mathcal{Z}|\boldsymbol{\theta}_0)\right]^T \times [\dot{\Phi}^T_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, \mathcal{Z}))\mathcal{X}]^2 f(\mathcal{Z}, \mathcal{X})\, d\mathcal{Z}\, d\mathcal{X},$$

V_0^- denotes a generalized inverse of V_0, and $\boldsymbol{\xi}$ is as defined in Theorem 1, with $\tilde{V}_0$ replaced by V_0.

As for the estimators of $\Phi_{\boldsymbol{\theta}_0}(z) = (\phi_0(z), \phi_1(z), \ldots, \phi_p(z))^T$ and σ^2, we have the following convergence rates.

Theorem 2. Under the assumptions of Theorem 1, we have

$$\sup_{z\in\mathcal{U}} \|\hat{\Phi}_{\hat{\theta}}(z) - \Phi_{\boldsymbol{\theta}_0}(z)\| = O\left(\left(\frac{\log n}{nh}\right)^{1/2}\right) \quad \text{a.s.}$$

and

$$\hat{\sigma}^2 - \sigma^2 = O\left(\left(\frac{\log\log n}{n}\right)^{1/2}\right) \quad \text{a.s.}$$

Note that the convergence rate of $\hat{\Phi}_{\hat{\theta}}(z)$ is uniform over $\mathcal{U}$. From the results of Stute (1982) and the law of iterated logarithm, we can see that the foregoing convergence rates are optimal. More complicated arguments allow the results to hold for local linear smoothing, which provides more accurate estimators under more relaxed conditions (see Fan and Gijbels 1996). We have the following asymptotic normality result about the pointwise properties of $\hat{\Phi}_{\hat{\theta}}(z)$.

Theorem 3. Under the assumptions of Corollary 1, for any $z \in \mathcal{U}$, we have

$$\sqrt{nh}\{\hat{\Phi}_{\hat{\theta}}(z) - \Phi_{\boldsymbol{\theta}_0}(z) - B(z)h^2\} \xrightarrow{D} \mathrm{N}(0, [\mathcal{W}(z, \boldsymbol{\theta}_0)]^{-1}k_2\sigma^2),$$

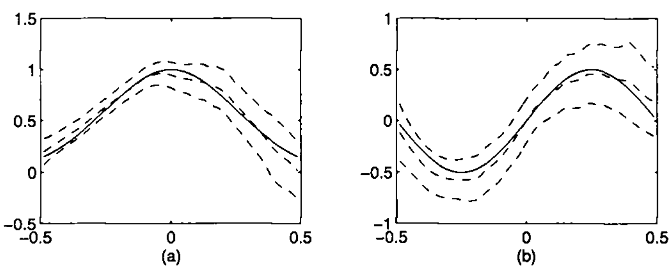

Figure 1. Simulation Results for Example 1. The solid curves in (a) and (b) denote the real coefficient functions $\phi_0(z)$ and $\phi_1(z)$. The dashed curves denote the estimated coefficient functions and pointwise approximate 95% confidence limits.

where

$$B(z) = \frac{1}{2}\, k_1\left[2(\mathcal{W}(z, \boldsymbol{\theta}_0))^{-1}\frac{\partial}{\partial z}\, \mathcal{W}(z, \boldsymbol{\theta}_0)\dot{\Phi}_{\boldsymbol{\theta}_0}(z) + 2(f(z, \boldsymbol{\theta}_0))^{-1}\frac{\partial}{\partial z}\, f(z, \boldsymbol{\theta}_0)\dot{\Phi}_{\boldsymbol{\theta}_0}(z) + \ddot{\Phi}_{\boldsymbol{\theta}_0}(z)\right]h^2.$$

3. SIMULATIONS AND APPLICATIONS

In this section we first use simulations to support the claim that the estimation method for the model works well. We then apply the model and the method to some real datasets. The single-index coefficient regression model was found to give a good fit to these datasets. In the following examples, we use the Epanechnikov kernel.

Example 1. We first consider the following single-index coefficient regression model:

$$y_t = \exp(-8(.8y_{t-1} + .6y_{t-2} - .6)^2) + .5\sin(2\pi(.8y_{t-1} + .6y_{t-2} - .6))y_{t-1} + .1\varepsilon_t,$$

where the ε_t's are independent standard normal random variables. It corresponds to the model

$$y_t = \phi_0(g_1(\boldsymbol{\theta}, y_{t-1}, y_{t-2})) + \phi_1(g_1(\boldsymbol{\theta}, y_{t-1}, y_{t-2}))y_{t-1} + .1\varepsilon_t.$$

Here $g_1(\boldsymbol{\theta}, y_{t-1}, y_{t-2}) = \cos(\boldsymbol{\theta})y_{t-1} + \sin(\boldsymbol{\theta})y_{t-2}$ with $\boldsymbol{\theta} = .6435$. For sample sizes $n = 50, 100$, and 200, 500 independent realizations were simulated. Take $\mathcal{A}$ such that $(.8, .6)\mathcal{Z} \in [-.5, .5]$. We minimize $\hat{S}(\boldsymbol{\theta}, h)$ within $\boldsymbol{\theta} \in [.2, 1.3]$ and $h \in [.01, .2]$. Table 1 gives the estimates of $h, \boldsymbol{\theta}$, and σ^2.

Table 1 confirms our theoretical results. Stable estimates of the parameters θ and σ^2 were obtained even when the sample size was as small as 50. Therefore, reliable estimates of coefficient functions may be obtained using our method. Figure 1 shows the estimated coefficient functions from a typical simulated dataset of size $n = 200$. We can see that the estimated functions are close to the real ones. From Theorem 3, the approximate pointwise confidence interval can also be drawn as in Figure 1 (see Xia and Li 1999 for details).

In the following examples, we take $\mathcal{A}$ sufficiently large to include all of the observations. The single-indexing specification of Ichimura (1993) has proven to be very efficient in practice. From (1) and (2), we make the same kind of specification on the coefficient functions. To illustrate the efficiency of this specification in our model, we give the following example.

Example 2. Consider the Hénon map with dynamic noise,

$$y_t = 6.8 - .19y_{t-1}^2 + .28y_{t-2} + .2\xi_t, \tag{11}$$

where $\{\xi_t\}$ is a sequence of i.i.d. random variables with standard normal distribution truncated in the interval $[-12, 12]$. By simple calculations, the Hénon map (11) can

Table 1. Means and Standard Errors (in Parentheses) of Estimated h, θ, and σ^2 for Different Sample Size n

n	h	θ	σ^2
50	.0757(1.7303 × 10^{-2})	.6401(6.7229 × 10^{-2})	.0133(3.3911 × 10^{-3})
100	.0590(1.2665 × 10^{-2})	.6450(4.8006 × 10^{-2})	.0115(2.2147 × 10^{-3})
200	.0500(1.0278 × 10^{-2})	.6438(3.5133 × 10^{-2})	.0101(7.2621 × 10^{-4})

be written as

$$y_t = \tilde{\phi}_0(y_{t-2}, y_{t-3}) + \tilde{\phi}_1(y_{t-2}, y_{t-3})y_{t-2} + \tilde{\phi}_2(y_{t-2}, y_{t-3})y_{t-3} + \eta_t, \quad (12)$$

where $E(\eta_t|y_{t-2}, y_{t-3}) = 0$ a.s. Model (12) is a varying-coefficient model defined in (1).

Next, we want to see the performance of a wrongly identified single-index coefficient regression model fitted to the simulated Hénon map data. We make a two-step ahead prediction of y_t using y_{t-2} and y_{t-3} by the following model:

$$y_t = \phi_0(g_2(\boldsymbol{\theta}, y_{t-2}, y_{t-3})) + \phi_1(g_2(\boldsymbol{\theta}, y_{t-2}, y_{t-3}))y_{t-2} + \phi_2(g_2(\boldsymbol{\theta}, y_{t-2}, y_{t-3}))y_{t-3} + \varepsilon_t, \quad (13)$$

where $g_2(\boldsymbol{\theta}, y_{t-2}, y_{t-3}) = \cos(\boldsymbol{\theta})y_{t-2} + \sin(\boldsymbol{\theta})y_{t-3}$. Keep in mind that the Hénon map (11) cannot be written in the form of (13) or (2). A sample of 200 observations was generated from the Hénon map (11). The first 150 observations are used to estimate (13), and the last 50 are retained for predictions. We estimate $\boldsymbol{\theta}$ to be 2.9322, and thus $g_2(2.9322, y_{t-2}, y_{t-3}) = -.9781y_{t-2} + .2079y_{t-3}, \hat{h} = .8$, and $\hat{\sigma}^2 = .2009$. Figure 2 shows the estimated coefficient functions. These coefficient functions vary greatly over the interval $[-10, 10]$, which means that we can not use the fixed-coefficient autoregression (AR) model to fit the data. Figure 3 shows that (13) still makes some good predictions to the retained data. In contrast, an AR(3) model performs rather badly. On the other hand, it is impossible for us to use the observations to estimate (12) or (1) due to the sparseness of data on the two-dimensional space.

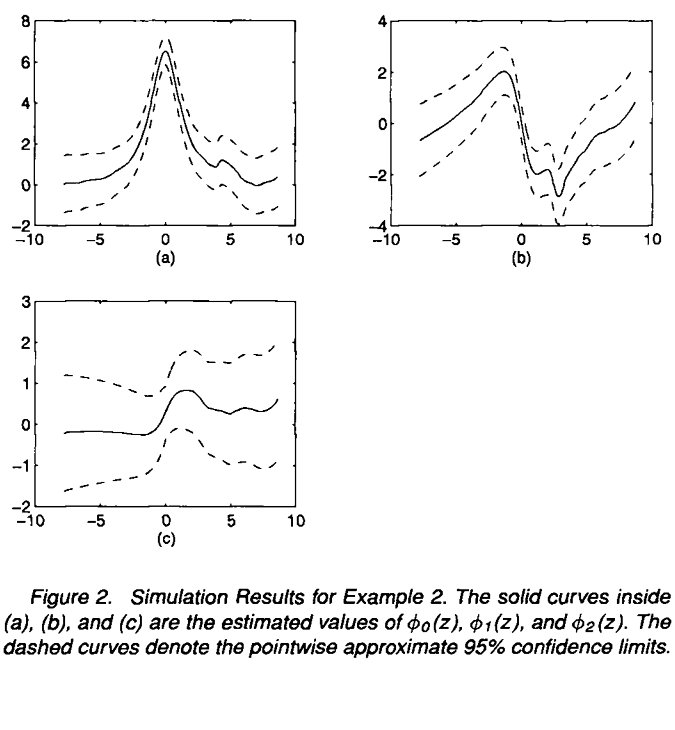

Figure 2. Simulation Results for Example 2. The solid curves inside (a), (b), and (c) are the estimated values of $\phi_0(z)$, $\phi_1(z)$, and $\phi_2(z)$. The dashed curves denote the pointwise approximate 95% confidence limits.

Example 3. This example considers the Canadian lynx data for 1821–1934. Haggan and Ozaki (1981) used the EXPAR model to fit these data. Following them, we make the same transformation of data y_t = log(lynx number in year $(1820 + t)$). We fit the transformed data by the model

$$y_t = \phi_0(g_3(\boldsymbol{\theta}, y_{t-1}, y_{t-2})) + \phi_1(g_3(\boldsymbol{\theta}, y_{t-1}, y_{t-2}))y_{t-1} + \phi_2(g_3(\boldsymbol{\theta}, y_{t-1}, y_{t-2}))y_{t-2} + \varepsilon_t,$$

where $g_3(\boldsymbol{\theta}, y_{t-1}, y_{t-2}) = \cos(\boldsymbol{\theta})y_{t-1} + \sin(\boldsymbol{\theta})y_{t-2}$. We estimate the parameters as

$$\hat{\theta} = .6374, \qquad \hat{h} = .4,$$

and

$$\hat{\sigma}^2 = .0463(.0401).$$

The value in the parentheses is the estimate of σ^2 with $\hat{\Phi}_{\theta t}(\cdot)$ replaced by $\hat{\Phi}_{\theta}(\cdot)$ in (6) using the same bandwidth $\hat{h}$ and orientation $\hat{\theta}$. (see Cheng and Tong 1993). These values are much smaller than Tong's nonlinear model (1990, p. 410), which has a residual variance of .0496. In our model the best threshold variable is the linear combination $-.5878y_{t-1} + .8090y_{t-2}$. This is rather different from what Haggan and Ozaki (1981) and Tong (1990) have proposed, where y_{t-2} or y_{t-1} is used as the threshold variable. The estimated coefficient functions are shown in Figure 4.

Example 4. In this example we consider the annual sunspot number for the period 1700–1979. These data have been analyzed by various authors. Chen and Tsay (1993) and Tong (1990) proposed two different threshold variables. Here we want to make a brief comparison between them under the following model. Under the same transformation of data $y_t = 2\{[1 + (\text{sunspot number in year } (1699 + t))]^{1/2} - 1\}$ as Tong (1990) used, we fit the following model to the data:

$$\begin{aligned} y_t = {} & \phi_0(g_4(\boldsymbol{\theta}, y_{t-3}, y_{t-8})) + \phi_1(g_4(\boldsymbol{\theta}, y_{t-3}, y_{t-8}))y_{t-1} \\ & + \phi_2(g_4(\boldsymbol{\theta}, y_{t-3}, y_{t-8}))y_{t-2} \\ & + \phi_3(g_4(\boldsymbol{\theta}, y_{t-3}, y_{t-8}))y_{t-3} \\ & + \phi_4(g_4(\boldsymbol{\theta}, y_{t-3}, y_{t-8}))y_{t-8} + \varepsilon_t, \end{aligned} \quad (14)$$

where, as in Example 3, $g_4(\boldsymbol{\theta}, y_{t-3}, y_{t-8}) = \cos(\boldsymbol{\theta})y_{t-3} + \sin(\theta)y_{t-8}$. Our model is similar to that of Chen and Tsay (1993). Tong (1990) took y_{t-8} as the threshold variable, whereas Chen and Tsay (1993) took y_{t-3} as the threshold variable. For these threshold variables, the results of estimation for model (14) are

$$y_{t-3}\ (\boldsymbol{\theta} = 0){:}\ \hat{h} = 1.9, \qquad \hat{\sigma}^2 = 4.12 \quad (15)$$

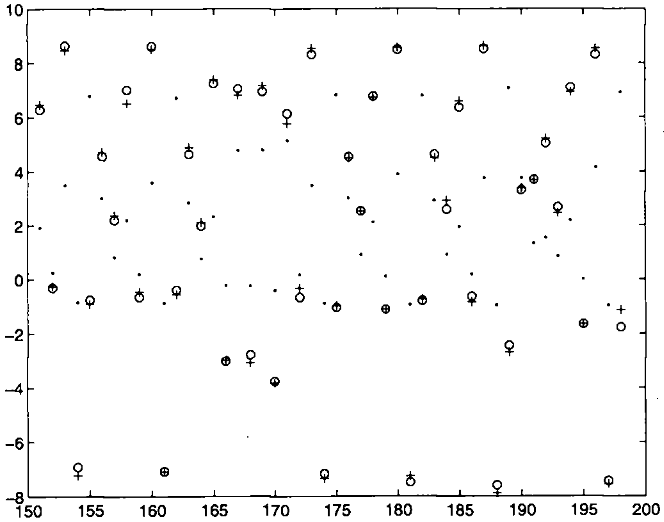

Figure 3. Simulation Results for Example 2. +, observations; o, predicted values using (13); ·, predicted values using the AR(3) model $y_t = \beta_0 + \beta_1 y_{t-2} + \beta_2 y_{t-3} + \varepsilon_t$.

and

$$y_{t-8}\ (\boldsymbol{\theta} = \pi/2):\ \hat{h} = 2.0, \qquad \hat{\sigma}^2 = 4.47. \tag{16}$$

Now we estimate the threshold variable directly; the estimated parameters are

$$\hat{\boldsymbol{\theta}} = -0.1256, \qquad \hat{h} = 1.7,$$

and

$$\hat{\sigma}^2 = 3.72(3.14). \tag{17}$$

Therefore, the threshold variable is $.9921y_{t-3} - .1253y_{t-8}$. Comparing (15) and (16) with (17), we find that y_{t-3} seems to be more suitable than y_{t-8} as the threshold variable for (14). The value 3.14 in the parentheses is the residual variance using all observations, as explained in Example 3. Compared to the results of Chen and Tsay (1993), our result lends some support to their nonlinear model.

Table 2. Multistep Predictions of Annual Sunspot Numbers for 1980–1987 and Prediction Errors (in Parentheses)

Year	Observation	TAR prediction	FAR prediction	Our prediction
1980	154.7	160.1 (−5.4)	168.5 (−13.8)	156.8 (−2.1)
1981	140.5	141.8 (−1.3)	140.6 (−0.1)	138.8 (1.7)
1982	115.9	96.4 (19.5)	105.9 (10.0)	113.3 (2.6)
1983	66.6	61.8 (4.8)	70.9 (−4.3)	69.0 (−2.4)
1984	45.9	31.0 (14.9)	42.1 (3.8)	43.6 (2.3)
1985	17.9	18.1 (−0.2)	22.5 (−4.6)	25.5 (−7.6)
1986	13.4	18.9 (−5.5)	12.1 (1.3)	9.2 (4.2)
1987	29.2	29.9 (−0.7)	7.5 (21.7)	16.0 (13.2)

Finally, we use the model to predict the sunspot numbers for 1980–1987. The results, along with the results of the TAR model of Tong (1990) and the FAR model of Chen and Tsay (1993), are shown in Table 2. The mean squared errors for the TAR model predictions, the FAR model predictions, and our predictions are 85.87, 102.14, and 34.44. In this sense, our model outperforms both the TAR and FAR models. (See also Chen and Tsay 1993 for a detailed explanation about the large prediction errors for the year 1987 using the FAR model and our model.)

Example 5. We now consider a regression model with dependent observations. The experimental setup involves a person riding a computer-controlled cycle ergometer for the purpose of studying health problems. The data were collected over a period of 10 minutes on four variables: work load (WORK), total volume of inspired and expired flows (VENT), and O_2 and CO_2 breathed out. There are 144 observations for each series. These series have been found to follow some nonlinear time series (Tong 1990). Here we consider the relation between CO_2 breathed out with other variables. Following Tong (1990), let $y_t = \sqrt{CO_2}/10, x_{t1} = \sqrt{WORK}/10, x_{t2} = VENT/10$, and $x_{t3} = \sqrt{O_2}/10$. We fit the relation with the single-index coefficient regression

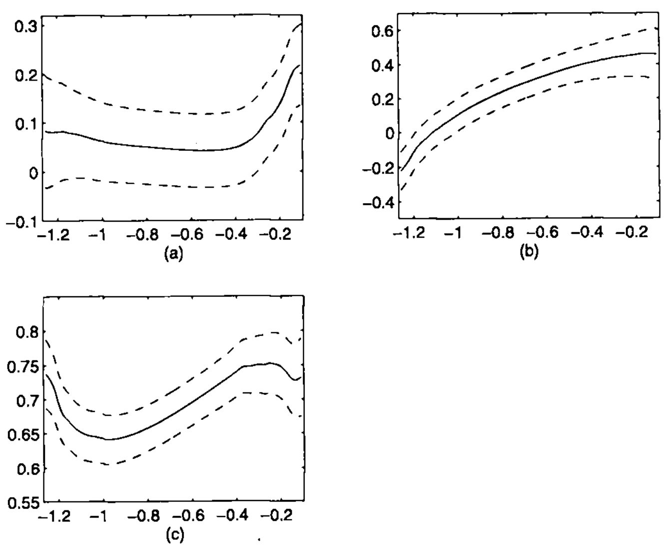

Figure 4. Results for Example 3. The solid curves inside (a), (b), and (c) are the estimated values of $\phi_0(z)$, $\phi_1(z)$, and $\phi_2(z)$. The dashed curves denote the pointwise approximate 95% confidence limits.

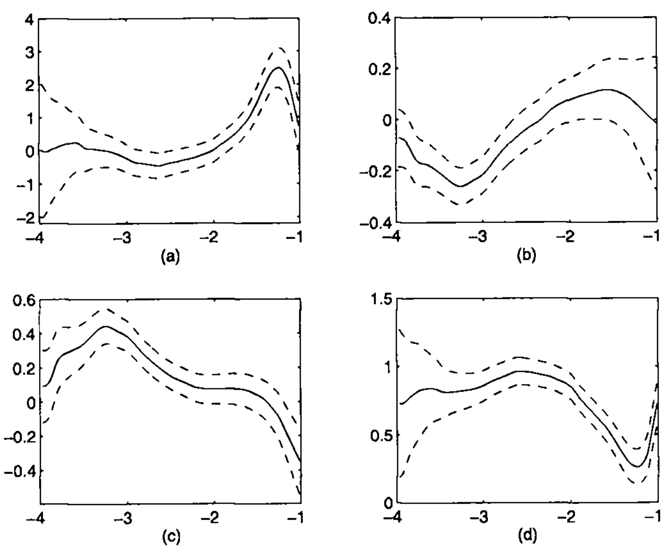

Figure 5. Results for Example 5. The solid curves inside (a)–(d) are the estimated values of $\phi_0(z)$, $\phi_1(z)$, $\phi_2(z)$, and $\phi_3(z)$. The dashed curves denote the pointwise approximate 95% confidence limits.

model

$$
\begin{aligned}
y_t = {} & \phi_0(g_5(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, x_{t1}, x_{t2}, x_{t3})) \\
& + \phi_1(g_5(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, x_{t1}, x_{t2}, x_{t3}))x_{t1} \\
& + \phi_2(g_5(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, x_{t1}, x_{t2}, x_{t3}))x_{t2} \\
& + \phi_3(g_5(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, x_{t1}, x_{t2}, x_{t3}))x_{t3} + \varepsilon_t,
\end{aligned}
$$

where, using polar coordinates, we let $g_5(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, x_{t1}, x_{t2}, x_{t3}) = \cos(\boldsymbol{\theta}_1)x_{t1} + \sin(\boldsymbol{\theta}_1)\cos(\boldsymbol{\theta}_2)x_{t2} + \sin(\boldsymbol{\theta}_1)\sin(\boldsymbol{\theta}_2)x_{t3}$. The estimates were found to be

$$\hat{\boldsymbol{\theta}}_1 = 1.0472, \qquad \hat{\boldsymbol{\theta}}_2 = 3.7699, \qquad \hat{h} = .28,$$

and

$$\hat{\sigma}^2 = .0085(.0070).$$

The variance of the fitted error using the linear regression model, $y_t = c_0 + c_1x_{t1} + c_2x_{t2} + c_3x_{t3} + \varepsilon_t$, is .0115. The single-index model gives a great improvement to the goodness of fit. Figure 5 shows the estimated coefficient functions, which also suggest that the constant coefficient regression model is not adequate for this dataset.

4. CONCLUSION

In this article we proposed the single-index coefficient regression model and its nonparametric estimation method. The result is also suitable for time series analysis. The model can achieve three goals. First, it can avoid the "curse of dimensionality" in nonparametric estimations. Second, it can be used to choose the threshold variable by estimating the single-index parameters. Finally, it gives an efficient approximate to the unknown multivariate regression functions. The result is supported by our simulations and real data studies. Both theory and empirical study suggest that the single-index coefficient regression models in conjunction with the proposed estimation method could be useful in both nonlinear regression and nonlinear time series analyses.

APPENDIX: PROOFS

We now give the proofs of the results in Section 2. We use $A_n(\mathcal{Z}, \boldsymbol{\theta}, h) = \bar{O}(a_n, \mathcal{A}, \Theta, \mathcal{H})$ to denote $\sup_{\mathcal{Z}\in\mathcal{A},\boldsymbol{\theta}\in\Theta,h\in\mathcal{H}} |A_n(\mathcal{Z},\boldsymbol{\theta},h)| = O(a_n)$ a.s. If $A_n(z,\boldsymbol{\theta})$ is a matrix, then $A_n(\mathcal{Z},\boldsymbol{\theta},h) = \bar{O}(a_n,\mathcal{A},\Theta,\mathcal{H})$ means that all of its elements are $\bar{O}(a_n,\mathcal{A},\Theta,\mathcal{H})$. Similar notations hold for $B_n(\boldsymbol{\theta},h) = \bar{O}(a_n,\Theta,\mathcal{H})$ and $C_n(\boldsymbol{\theta}) = \bar{O}(a_n,\Theta)$. For simplicity, we assume that all of the observations $Z_t \in \mathcal{A}$. Otherwise, we take into account only those Z_t inside $\mathcal{A}$ (see Härdle et al. 1993). By continuity, we can show that for all those $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$, $\min_{z\in\mathcal{U}}\{f(z,\boldsymbol{\theta})\det[\mathcal{W}(z,\boldsymbol{\theta})]\} > 0$. Therefore, for simplicity, we may assume that this property holds for all $\boldsymbol{\theta} \in \Theta$. Let C be a constant below; it may have different values at different places.

Lemma A.1

Suppose that $\varphi(\mathcal{Z},\boldsymbol{\theta})$ is a bounded function and has bounded derivatives with respect to $\boldsymbol{\theta} \in \Theta$ uniformly for $\mathcal{Z} \in \mathcal{A}$. $\{(Z_t,\xi_t)\}$ is a strongly mixing and strictly stationary sequence with mixing coefficient as in (C2) and $E(\xi_t|Z_t) = 0$ a.s., $E|\xi_t|^l < \infty$ for all $l > 1$. Then

$$\sum_{t=1}^{n} \xi_t[\varphi(Z_t,\boldsymbol{\theta}) - \varphi(Z_t,\boldsymbol{\theta}_0)] = \bar{O}((n^{1-2\delta}\log n)^{1/2}, \Theta_n).$$

Using the moment inequality for strongly mixing sequences and the "continuity argument" on Θ_n (cf. Kim and Cox 1995), Lemma A.1 is easily established.

Lemma A.2

Suppose that $\varphi(g(\boldsymbol{\theta},\mathcal{Z}))$ is a bounded function and has bounded derivatives with respect to $\boldsymbol{\theta}$, and that $\{(Z_t,\xi_t)\}$ is as defined in Lemma A.1. The conditional densities $f_{z_1|\xi_1}(z|v)$ and $f_{z_1,z_l|\xi_1,\xi_l}(z_1,z_l|v_1,v_l)$ are bounded for all $l > 1$, where $z_t = g(\boldsymbol{\theta}_0, Z_t)$. Let $U_t(\boldsymbol{\theta},\mathcal{Z},h) = K_h(g(\boldsymbol{\theta},Z_t) - g(\boldsymbol{\theta},\mathcal{Z}))\varphi(g(\boldsymbol{\theta},Z_t))\xi_t$. If (C4), (C5), and (C7) hold, then

$$\sum_{t=1}^{n}[U_t(\boldsymbol{\theta},\mathcal{Z},h) - E(U_t(\boldsymbol{\theta},\mathcal{Z},h))] = \bar{O}((n^{4/5}\log n)^{1/2}, \mathcal{A}, \Theta, \mathcal{H}_n)$$

and

$$
\begin{aligned}
&\sum_{t=1}^{n}\{U_t(\boldsymbol{\theta},\mathcal{Z},h) - U_t(\boldsymbol{\theta}_0,\mathcal{Z},h) - E[U_t(\boldsymbol{\theta},\mathcal{Z},h) - U_t(\boldsymbol{\theta}_0,\mathcal{Z},h)]\} \\
&\qquad = \bar{O}(n^{2/5-\delta}(\log n)^{1/2}, \mathcal{A}, \Theta_n, \mathcal{H}_n).
\end{aligned}
$$

Proof

Here we prove only the second equation. Let $\tilde{T}(n)$ be any positive function of n, which may tend to infinity as $n \to \infty$, and

$$\bar{\xi}_t = \xi_t I(|\xi_t| < \tilde{T}(n)),$$

$$\bar{U}_t(\boldsymbol{\theta},\mathcal{Z},h) = K_h(g(\boldsymbol{\theta},Z_t) - g(\boldsymbol{\theta},\mathcal{Z}))\varphi_{\theta}(g(\boldsymbol{\theta},Z_t))\bar{\xi}_t,$$

and

$$\varsigma_n(\boldsymbol{\theta},\mathcal{Z},h) = \sum_{t=1}^{n}[\bar{U}_t(\boldsymbol{\theta},\mathcal{Z},h) - \bar{U}_t(\boldsymbol{\theta}_0,\mathcal{Z},h)].$$

From the proof of lemma 4.4 of Masry and Tjøstheim (1995), we need only to prove that

$$\sup_{\boldsymbol{\theta}\in\Theta_n,\mathcal{Z}\in\mathcal{A},h\in\mathcal{H}_n} \operatorname{var}(\varsigma_n(\boldsymbol{\theta},\mathcal{Z},h)) = O(n^{4/5-2\delta}). \tag{A.1}$$

Next we prove (A.1) under more relaxed conditions than those in Lemma A.2. Suppose that the lth ($l > 2$) moment of X_t and y_t exist. It is easy to see that for any $l' > l$,

$$\sum_{n=1}^{\infty} n^{l'}[\alpha(n)]^{1-2/l} < \infty. \tag{A.2}$$

By stationarity, we have

$$
\begin{aligned}
\operatorname{var}[\varsigma_n(\boldsymbol{\theta},\mathcal{Z},h)] &= n\operatorname{var}[\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)] \\
&\quad + 2\sum_{t=2}^{n}(n-t)\operatorname{cov}[\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h), \bar{U}_t(\boldsymbol{\theta},\mathcal{Z},h)] \\
&\triangleq R_1 + R_2.
\end{aligned}
$$

Let $f(\mathcal{Z},v)$ be the density function of (Z_t,ξ_t); then

$$E|\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_1(\boldsymbol{\theta}_0,\mathcal{Z},h)|^l$$
$$=\int |K_h(g(\boldsymbol{\theta},\mathcal{Z}')-g(\boldsymbol{\theta},\mathcal{Z}))\varphi(g(\boldsymbol{\theta}_0,\mathcal{Z}')) - K_h(g(\boldsymbol{\theta}_0,\mathcal{Z}')-g(\boldsymbol{\theta}_0,\mathcal{Z}))\varphi(g(\boldsymbol{\theta}_0,\mathcal{Z}'))|^l|v|^l \times I(|v|\le T(n))f(\mathcal{Z}',v)\,d\mathcal{Z}'\,dv$$
$$\le C\int |K_h(g(\boldsymbol{\theta},\mathcal{Z}')-g(\boldsymbol{\theta},\mathcal{Z})) - K_h(g(\boldsymbol{\theta}_0,\mathcal{Z}')-g(\boldsymbol{\theta},\mathcal{Z}))|^l|v|^l f(\mathcal{Z}',v)\,d\mathcal{Z}'\,dv.$$

Note that for bounded $\mathcal{Z}$ and $\mathcal{Z}'$,

$$K\left(\frac{g(\boldsymbol{\theta},\mathcal{Z}')-g(\boldsymbol{\theta},\mathcal{Z})}{h}\right)-K\left(\frac{g(\boldsymbol{\theta}_0,\mathcal{Z}')-g(\boldsymbol{\theta}_0,\mathcal{Z})}{h}\right)$$
$$=K'\left(\frac{g(\boldsymbol{\theta}_0,\mathcal{Z}')-g(\boldsymbol{\theta}_0,\mathcal{Z})}{h}\right)\frac{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T}{h}\frac{\partial g(\boldsymbol{\theta}_0,\mathcal{Z})}{\partial\boldsymbol{\theta}} + O(n^{-2\delta}h^{-2}),$$

whereas

$$\int\left|K'\left(\frac{g(\boldsymbol{\theta}_0,\mathcal{Z}')-g(\boldsymbol{\theta}_0,\mathcal{Z})}{h}\right)\times\frac{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T}{h}\frac{\partial g(\boldsymbol{\theta}_0,\mathcal{Z})}{\partial\boldsymbol{\theta}}\right|^l|v|^l f(\mathcal{Z},v)\,d\mathcal{Z}\,dv$$
$$\le Cn^{-\delta l}\int\left|K'\left(\frac{z'-z}{h}\right)\right|^l f_{z_1|\xi_1}(z|v)\,dz'|v|^l f_{\xi_t}(v)\,dv$$
$$=O(n^{-\delta l}h).$$

We have

$$E|\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_1(\boldsymbol{\theta}_0,\mathcal{Z},h)|^l=O(n^{-\delta l}h),\quad l=2,3,\ldots \tag{A.3}$$

By the Minkovski inequality, we have

$$[E|\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_1(\boldsymbol{\theta}_0,\mathcal{Z},h)-E\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)+E\bar{U}_1(\boldsymbol{\theta}_0,\mathcal{Z},h)|^l]^{1/l} \le Cn^{-\delta}h^{1/l},\qquad l=2,3,\ldots \tag{A.4}$$

Similarly, it is easy to check that

$$\text{cov}[\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_1(\boldsymbol{\theta}_0,\mathcal{Z},h), \bar{U}_t(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_t(\boldsymbol{\theta}_0,\mathcal{Z},h)]=O(n^{-2\delta}h^2). \tag{A.5}$$

From (A.3), we have

$$R_1\le nE[\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_1(\boldsymbol{\theta}_0,\mathcal{Z},h)]^2=O(n^{1-2\delta}h)=O(n^{4/5-2\delta}).$$

For R_2, we have

$$R_2\le n\sum_{t=2}^{n}|\text{cov}[\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_1(\boldsymbol{\theta}_0,\mathcal{Z},h), \bar{U}_t(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_t(\boldsymbol{\theta}_0,\mathcal{Z},h)]|$$
$$=n\sum_{t=2}^{\tilde{n}}|\text{cov}[\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_1(\boldsymbol{\theta}_0,\mathcal{Z},h), \bar{U}_t(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_t(\boldsymbol{\theta}_0,\mathcal{Z},h)]|$$
$$+n\sum_{t=\tilde{n}+1}^{n}|\text{cov}[\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_1(\boldsymbol{\theta}_0,\mathcal{Z},h), \bar{U}_t(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_t(\boldsymbol{\theta}_0,\mathcal{Z},h)]|$$
$$\triangleq R_{21}+R_{22},$$

where $\tilde{n}$ is the integer part of $h^{-(1-2/l)/l'}$. By (A.5) and noting that $\tilde{n}h=O(1)$, it follows that

$$R_{21}\le Cnn^{-2\delta}h^2\tilde{n}=O(n^{4/5-2\delta}).$$

For R_{22}, we use Davydov's lemma (Hall and Heyde 1980, p. 278, cor. A2) and (A.4),

$$\text{cov}[\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_1(\boldsymbol{\theta}_0,\mathcal{Z},h),\bar{U}_t(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_t(\boldsymbol{\theta}_0,\mathcal{Z},h)]$$
$$\le 8[\alpha(t)]^{1-2/l}[E|\bar{U}_1(\boldsymbol{\theta},\mathcal{Z},h)-\bar{U}_1(\boldsymbol{\theta}_0,\mathcal{Z},h)|^l]^{2/l}$$
$$\le Cn^{-2\delta}h^{2/l}[\alpha(t)]^{1-2/l}.$$

Note that $\tilde{n}\cong h^{-(1-1/l)/l'}$; from (A.2),

$$R_{22}\le Cn^{1-2\delta}h^{2/l}\sum_{t=\tilde{n}+1}^{n}[\alpha(t)]^{1-2/l}$$
$$\le\frac{Cn^{1-2\delta}h}{h^{1-2/l}\tilde{n}^{l'}}\sum_{t=\tilde{n}+1}^{\infty}t^{l'}[\alpha(t)]^{1-2/l}=O(n^{4/5-2\delta}).$$

Consequently, (A.1) follows.

Lemma A.3

Suppose that $\{(\chi_t,\xi_t,Z_t)\}$ is a strongly mixing sequence with mixing coefficient of geometric rate as in (C2) and that ξ_t and Z_t satisfy the conditions in Lemma A.2, and that $r_n(x)$ is a measurable function such that $E(r_n(\chi_t)|Z_t)=0$ a.s. and $Er_n^l(\chi_t)\le M_l\delta_n^l$ for all $l>0$, where M_l does not depend on n. If (C7) holds, then

$$\sum_{t=1}^{n}r_n(\chi_t)\sum_{s=1}^{n}[K_h(g(\boldsymbol{\theta},Z_s)-g(\boldsymbol{\theta},Z_t))\xi_s-K_{1h}(g(\boldsymbol{\theta},Z_t),\boldsymbol{\theta})] = \bar{O}(n^{1-\tau}\delta_n,\Theta,\mathcal{H}_n)$$

and

$$\sum_{t=1}^{n}r_n(\chi_t)\sum_{s=1}^{n}[K_h(g(\boldsymbol{\theta},Z_s)-g(\boldsymbol{\theta},Z_t))\xi_s - K_h(g(\boldsymbol{\theta}_0,Z_s)-g(\boldsymbol{\theta}_0,Z_t))\xi_s - K_{1h}(g(\boldsymbol{\theta},Z_t),\boldsymbol{\theta})+K_{1h}(g(\boldsymbol{\theta}_0,Z_t),\boldsymbol{\theta}_0)]$$
$$=\bar{O}(n^{1-\delta-\tau}\delta_n,\Theta_n,\mathcal{H}_n),$$

where $K_{1h}(z,\boldsymbol{\theta})=E[K_h(g(\boldsymbol{\theta},Z_s)-z)\xi_s]$ and $0<\tau<1/10$.

Proof

We prove only the first equation as an example. Let $\eta>0$ be given for now; it will be determined later. Because $\Theta_n\times\mathcal{H}_n$ is a compact subset, we can find $n'=O(n^{\eta+k/2+1/5})$ balls in $\Re^{k+1}$ centered at $(\boldsymbol{\theta}_j,h_j)$ with radius less than $n^{-\eta}$ such that these balls cover $\Theta_n\times\mathcal{H}_n$.

We may consider only the summation over the set of indices constrained by $|s-t|>N$ for N($<n$) chosen appropriately. Indeed, by Lemma A.2, we have

$$\sum_{|s-t|<N}r_n(\chi_t)[K_{h_j}(g(\boldsymbol{\theta}_j,Z_s)-g(\boldsymbol{\theta}_j,Z_t))\xi_s-K_{1h_j}(g(\boldsymbol{\theta}_j,Z_t),\boldsymbol{\theta}_j)] = O(n\delta_n(Nn^{-1/5}\log n)^{1/2}). \tag{A.6}$$

Therefore, the first equation holds for summation over $|s-t| \le N$ if we take $N = o(n^{\varepsilon'})$ with $\varepsilon' < 1/5 - 2\tau$. Let

$$R_j = \sum_{|s-t|>N} r_n(\chi_t)[K_{h_j}(g(\boldsymbol{\theta}_j, Z_s) - g(\boldsymbol{\theta}_j, Z_t))\xi_s - K_{1h_j}(g(\boldsymbol{\theta}_j, Z_t), \boldsymbol{\theta}_j)]/\delta_n$$

and

$$\beta_{1t}(u) = 2\pi r_n(\chi_t)\exp(-ig(\boldsymbol{\theta}_j, Z_t)u/h_j)/\delta_n$$

and

$$\beta_{2s}(u) = \exp(ig(\boldsymbol{\theta}_j, Z_t)u/h_j)\xi_s - \int \exp(ig(\boldsymbol{\theta}_j, \mathcal{Z})u/h_j)vf(\mathcal{Z}, v)\,d\mathcal{Z}\,dv,$$

(here i is the imaginary unit), and let $\psi(u)$ be the Fourier transformation of $K(z)$. We have

$$R_j = \int \sum_{|s-t|>N} \beta_{1t}(u)\beta_{2s}(u)\psi(u)\,du.$$

It is easy to see that

$$E\beta_{1t}(u) = E\beta_{2s}(u) = 0,$$
$$M_l = \max\{[E|\beta_{1t}(u)|^l]^{1/l}, [E|\beta_{2t}(u)|^l]^{1/l}\} < \infty.$$

M_l does not depend on n for any $l > 1$. It is easy to check that the other conditions of lemma 5.2 of Kim and Cox (1995) are satisfied. Noting that $h_j = O(n^{-1/5})$, for any $0 < \tau < 1/2$, we have

$$\begin{aligned} n^{\eta+k/2+1/5} & \max_{i\le n'} E|n^{-1+\tau}R_i|^{2l} \\ &\le Cn^{\eta+k/2+1/5}\sup_{i\le n'}\left[n^{2\tau l}\sum_{i=1+N/(4l)}^{\infty} i^{2l-1}\alpha(i) + n^{-2(1-\tau)l}\sum_{i=1}^{2l}(n^i N^{4l-i}h_j^{i/2})\right] \\ &\le Cn^{\eta+k/2+1/5}\sup_{i\le n'}\left[n^{2\tau l}\sum_{i=1+N/(4l)}^{\infty} i^{2l-1}\alpha(i) + n^{(8/5)l+2\tau l}\sum_{i=1}^{2l}(n^{-1}Nh_j^{-1/2})^{4l-i}\right]. \end{aligned}$$

The first term inside the brackets is negligible. The summation of the second term is dominated by $n^{-2l}N^{2l}h_j^{-l}$ as $i = 2l$ if $n^{-1}Nh_j^{-1/2} \to 0$, which is satisfied by taking $N = O(n^{\varepsilon''})$ with $\varepsilon'' = \varepsilon'/2$. We have

$$\begin{aligned} n^{\eta+k/2+1/5} & \sup_{j\le n'} E|n^{-1+\tau}R_j|^{2l} \\ &\le Cn^{\eta+k/2+1/5}\sup_{j\le n'} n^{-(2/5)l+2\tau l}N^{2l}h_j^{-l} \\ &\le Cn^{-\tau'}, \quad \text{for some} \quad \tau' > 1. \end{aligned}$$

The last inequality holds by taking l sufficiently large such that

$$\eta + k/2 + 1/5 + 2\varepsilon'' l + 2\tau l - l/5 < -1. \qquad \text{(A.7)}$$

Therefore, by the Chebyshev inequality, for any $\varepsilon > 0$,

$$\Pr(\max_{j\le n'}|n^{-1+\tau}R_j| > \varepsilon) \le Cn^{\eta+k/2+1/5}\max_{j\le n'}E|n^{-1+\tau}R_j|^{2l} \le Cn^{-\tau'}.$$

By the Borel–Cantelli lemma, we have

$$\max_{j\le n'}\sum_{t=1}^{n} r_n(\chi_t)\sum_{s=1}^{n}[K_{h_j}(g(\boldsymbol{\theta}_j, Z_s) - g(\boldsymbol{\theta}_j, Z_t))\xi_s - K_{1h_j}(g(\boldsymbol{\theta}_j, Z_t), \boldsymbol{\theta}_j)] = O(n^{1-\tau}\delta_n) \quad \text{a.s.}$$

By taking η sufficiently large and using the "continuity argument," it is easy to see that the first equation holds.

Let $\mathcal{W}(z, \boldsymbol{\theta}, h) = h^{-1}E\{K_h(g(\boldsymbol{\theta}, Z_t), z)X_tX_t^T\}$ and $\mathcal{V}(z, \boldsymbol{\theta}, h) = h^{-1}E\{K_h(g(\boldsymbol{\theta}, Z_t), z)X_ty_t\}$. By the continuity of $f(\mathcal{Z}, \mathcal{X}, y)$, we have

$$\begin{aligned} \mathcal{W}(z, \boldsymbol{\theta}, h) - \mathcal{W}(z, \boldsymbol{\theta}) &= O(h^2), \\ \mathcal{V}(z, \boldsymbol{\theta}, h) - \mathcal{V}(z, \boldsymbol{\theta}) &= O(h^2). \end{aligned} \qquad \text{(A.8)}$$

By Lemma A.2, we immediately have the following results about the estimator of the coefficient functions.

Lemma A.4

Suppose that (C1)–(C7) hold. Then

$$\mathcal{W}_n(z, \boldsymbol{\theta}, h) = \mathcal{W}(z, \boldsymbol{\theta})f(z, \boldsymbol{\theta}) + \bar{O}((n^{-4/5}\log n)^{1/2}, \mathcal{U}, \Theta, \mathcal{H}_n)$$

and

$$\mathcal{V}_n(z, \boldsymbol{\theta}, h) = \mathcal{V}(z, \boldsymbol{\theta})f(z, \boldsymbol{\theta}) + \bar{O}((n^{-4/5}\log n)^{1/2}, \mathcal{U}, \Theta, \mathcal{H}_n),$$

and thus

$$\hat{\Phi}_{\boldsymbol{\theta}}(z) = \Phi_{\boldsymbol{\theta}}(z) + \bar{O}((n^{-4/5}\log n)^{1/2}, \mathcal{U}, \Theta, \mathcal{H}_n).$$

Proof of Theorem 1

The proof of Theorem 1 is similar to the proof of Härdle et al. (1993). Here we only give a sketch of it. Define

$$D_t = [\hat{\Phi}_{\boldsymbol{\theta}_0 t}(g(\boldsymbol{\theta}_0, Z_t)) - \Phi_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, Z_t))]^T X_t,$$
$$d_t = [\Phi_{\boldsymbol{\theta}}(g(\boldsymbol{\theta}, Z_t)) - \Phi_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, Z_t))]^T X_t,$$

and

$$\Delta_t = \{\hat{\Phi}_{\boldsymbol{\theta}t}(g(\boldsymbol{\theta}, Z_t)) - \Phi_{\boldsymbol{\theta}}(g(\boldsymbol{\theta}, Z_t)) - [\hat{\Phi}_{\boldsymbol{\theta}_0 t}(g(\boldsymbol{\theta}_0, Z_t)) - \Phi_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, Z_t))]\}^T X_t.$$

In this notation,

$$\hat{S}(\boldsymbol{\theta}, h) - \tilde{S}(\boldsymbol{\theta}) = \sum_{t=1}^{n}(D_t^2 + \Delta_t^2) + 2\sum_{t=1}^{n}(D_t\Delta_t + D_td_t + \Delta_td_t - D_t\varepsilon_t - \Delta_t\varepsilon_t).$$

Corresponding to Theorem 1, let

$$R_1(\boldsymbol{\theta}, h) = \sum_{t=1}^{n}\Delta_t^2 + 2\sum_{t=1}^{n}(D_t\Delta_t + D_td_t + \Delta_td_t - \Delta_t\varepsilon_t)$$

and

$$R_2(h) = \sum_{t=1}^{n}D_t\varepsilon_t, \qquad T(h) = \sum_{t=1}^{n}D_t^2.$$

It follows that

$$\left|\hat{S}(\boldsymbol{\theta}, h) - \tilde{S}(\boldsymbol{\theta}) - \sum_{t=1}^{n}D_t^2 + 2\sum_{t=1}^{n}D_t\varepsilon_t\right| \le \sum_{t=1}^{n}\Delta_t^2 + 2\left(\sum_{t=1}^{n}\Delta_t^2\right)^{1/2}$$

$$\times \left[\left(\sum_{t=1}^{n} D_t^2 \right)^{1/2} + \left(\sum_{t=1}^{n} d_t^2 \right)^{1/2} \right]$$

$$+ 2 \left| \sum_{t=1}^{n} D_t d_t \right| + 2 \left| \sum_{t=1}^{n} \Delta_t \varepsilon_t \right| .$$

By Lemma A.4 and the law of large numbers for $\|X_t\|$ (Rio 1995), we have

$$T(h) = \sum_{t=1}^{n} D_t^2 = \bar{O}(n^{1/5} \log n, \mathcal{H}_n).$$

Similarly, we have

$$\sum_{t=1}^{n} d_t^2 = \bar{O}(n^{1-2\delta}, \Theta_n, \mathcal{H}_n).$$

From Lemma A.2 and some calculations,

$$\sum_{t=1}^{n} \Delta_t^2 = \bar{O}(n^{1/5-2\delta} \log n, \Theta_n, \mathcal{H}_n).$$

Hence

$$\sum_{t=1}^{n} \Delta_t^2 + 2 \left(\sum_{t=1}^{n} \Delta_t^2 \right)^{1/2} \left[\left(\sum_{t=1}^{n} D_t^2 \right)^{1/2} + \left(\sum_{t=1}^{n} d_t^2 \right)^{1/2} \right] = \bar{O}(n^{6/10-2\delta} \log n, \Theta_n, \mathcal{H}_n).$$

Write

$$\begin{aligned} D_t &= X_t^T [\mathcal{W}(g(\boldsymbol{\theta}_0, Z_t), \boldsymbol{\theta}_0)]^{-1} \\ &\quad \times [\mathcal{V}_n(g(\boldsymbol{\theta}_0, Z_t), \boldsymbol{\theta}_0, h) - \mathcal{V}(g(\boldsymbol{\theta}_0, Z_t), \boldsymbol{\theta}_0, h)] \\ &\quad + X_t^T [\mathcal{W}(g(\boldsymbol{\theta}_0, Z_t), \boldsymbol{\theta}_0)]^{-1} \\ &\quad \times [\mathcal{V}(g(\boldsymbol{\theta}_0, Z_t), \boldsymbol{\theta}_0, h) - \mathcal{V}(g(\boldsymbol{\theta}_0, Z_t), \boldsymbol{\theta}_0)] \\ &\quad + X_t^T \{[\mathcal{W}_n(g(\boldsymbol{\theta}_0, Z_t), \boldsymbol{\theta}_0)]^{-1} - [\mathcal{W}(g(\boldsymbol{\theta}_0, Z_t), \boldsymbol{\theta}_0)]^{-1}\} \\ &\quad \times \mathcal{V}_n(g(\boldsymbol{\theta}_0, Z_t), \boldsymbol{\theta}_0, h) \\ &\triangleq D_{1t} + D_{2t} + D_{3t}. \end{aligned}$$

Let

$$d_t' = [\Phi_{\boldsymbol{\theta}}(g(\boldsymbol{\theta}, Z_t)) - \Phi_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, Z_t))]^T (X_t - E(X_t | g(\boldsymbol{\theta}_0, Z_t)))$$

and

$$d_t'' = [\Phi_{\boldsymbol{\theta}}(g(\boldsymbol{\theta}, Z_t)) - \Phi_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, Z_t))]^T E(X_t | g(\boldsymbol{\theta}_0, Z_t)).$$

By Lemma A.3,

$$\sum_{t=1}^{n} D_{1t} d_t' = \bar{O}(n^{1/5-\tau-\delta}, \Theta_n, \mathcal{H}_n).$$

By Lemmas A.1 and A.2 and (A.8), it follows that

$$\sum_{t=1}^{n} D_{2t} d_t' = \bar{O}(n^{3/10-\delta} (\log n)^{1/2}, \Theta_n, \mathcal{H}_n)$$

and

$$\sum_{t=1}^{n} D_{3t} d_t' = \bar{O}(n^{-1/5-\delta} \log n, \Theta_n, \mathcal{H}_n).$$

Hence

$$\sum_{t=1}^{n} D_t d_t' = \bar{O}(n^{3/10-\delta} (\log n)^{1/2}, \Theta_n, \mathcal{H}_n). \tag{A.9}$$

Note that (Härdle et al. 1993)

$$\Phi_{\boldsymbol{\theta}}(g(\boldsymbol{\theta}, \mathcal{Z})) = \Phi_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}, \mathcal{Z})) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \mu(\mathcal{Z} | \boldsymbol{\theta}_0) \times \dot{\Phi}_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, \mathcal{Z})) + O(n^{-1})$$

and

$$\Phi_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, \mathcal{Z})) = \Phi_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}, \mathcal{Z})) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \times \frac{\partial g(\boldsymbol{\theta}_0, \mathcal{Z})}{\partial \boldsymbol{\theta}} \dot{\Phi}_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, \mathcal{Z})) + O(n^{-1}).$$

Therefore,

$$d_t'' = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \left[\frac{\partial g(\boldsymbol{\theta}_0, Z_t)}{\partial \boldsymbol{\theta}} - \mu(Z_t | \boldsymbol{\theta}_0) \right] \times E(X_t | g(\boldsymbol{\theta}_0, Z_t)) \Phi_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0, Z_t)) + O(n^{-1}) \quad \text{a.s.}$$

By Lemma A.3 and noting that

$$E \left[\frac{\partial g(\boldsymbol{\theta}_0, Z_t)}{\partial \boldsymbol{\theta}} - \mu(Z_t | \boldsymbol{\theta}_0) \Big| g(\boldsymbol{\theta}_0, Z_t) \right] = 0 \quad \text{a.s.},$$

we have

$$\sum_{t=1}^{n} D_t d_t'' = \bar{O}(n^{1/5-\delta-\tau}, \Theta_n, \mathcal{H}_n). \tag{A.10}$$

It follows from (A.9) and (A.10) that

$$\sum_{t=1}^{n} D_t d_t = \bar{O}(n^{3/10-\delta} (\log n)^{1/2}, \Theta_n, \mathcal{H}_n).$$

Similarly, by Lemma A.3, we have

$$\sum_{t=1}^{n} \Delta_t \varepsilon_t = \bar{O}(n^{3/10-\delta} (\log n)^{1/2}, \Theta_n, \mathcal{H}_n))$$

and

$$\sum_{t=1}^{n} D_t \varepsilon_t = \bar{O}(n^{-\tau+1/5}, \mathcal{H}_n).$$

Therefore, we have established (7) and (8).

The proof of (9) is the same as for step (ix) of Härdle et al. (1993) by using the central limiting theorem for dependent data (Rio 1995). Following the procedure of Härdle and Vieu (1992), we can prove (10).

Proof of Theorem 2

The first part follows from Lemma A.4 and Corollary 1. For the second part, we have

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^{n} [y_t - \hat{\Phi}_{\hat{\theta}t}^T (g(\hat{\theta}, Z_t)) X_t]^2 \\ &= \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2 - \frac{2}{n} \sum_{t=1}^{n} [\hat{\Phi}_{\hat{\theta}t}(g(\hat{\theta}, Z_t)) - \Phi(g(\boldsymbol{\theta}_0, Z_t))]^T X_t \varepsilon_t \\ &\quad + \frac{1}{n} \sum_{t=1}^{n} \{[\hat{\Phi}_{\hat{\theta}t}(g(\hat{\theta}, Z_t)) - \Phi(g(\boldsymbol{\theta}_0, Z_t))]^T X_t\}^2. \end{aligned}$$

Therefore, the second part follows from the law of iterated logarithm of $\{\varepsilon_t^2\}$ and Lemmas A.3 and A.4.

Proof of Theorem 3

Note that for those Z_t such that $z_t = g(\boldsymbol{\theta}_0, Z_t)$ is close to z, we have

$$y_t = X_t^T \left\{ \Phi_{\boldsymbol{\theta}_0}(z) + \dot{\Phi}_{\boldsymbol{\theta}_0}(z)(z_t - z) + \frac{1}{2}\ddot{\Phi}_{\boldsymbol{\theta}_0}(z)(z_t - z)^2 \right\} + o_p([z_t - z]^2) + \varepsilon_t.$$

By some calculations, we have

$$(nh)^{-1} \sum_{t=1}^{n} K_h(z_t - z) X_t X_t^T = \mathcal{W}(z, \boldsymbol{\theta}_0) + O_p(h),$$

$$(nh)^{-1} \sum_{t=1}^{n} K_h(z_t - z) X_t X_t^T (z_t - z) = k_1 \frac{\partial}{\partial z} \mathcal{W}(z, \boldsymbol{\theta}_0) f(z, \boldsymbol{\theta}_0) h^2 + k_1 \mathcal{W}(z, \boldsymbol{\theta}_0) \frac{\partial}{\partial z} f(z, \boldsymbol{\theta}_0) h^2 + o_p(h^2),$$

and

$$(nh)^{-1} \sum_{t=1}^{n} K_h(z_t - z) X_t X_t^T (z_t - z)^2 = \frac{1}{2} k_1 \mathcal{W}(z, \boldsymbol{\theta}_0) f(z, \boldsymbol{\theta}_0) h^2 + o_p(h^2).$$

From the foregoing equations and (5), we have

$$\hat{\Phi}_{\boldsymbol{\theta}_0}(z) = \Phi_{\boldsymbol{\theta}_0}(z) + B(z)h^2 + [\mathcal{W}_n(z, \boldsymbol{\theta}_0, h)]^{-1} \sum_{t=1}^{n} K_h(z_t - z) X_t \varepsilon_t + o_p(h^2).$$

By theorem 4.4 of Masry and Tjøstheim (1995), we have

$$\sqrt{nh}\{\hat{\Phi}_{\boldsymbol{\theta}_0}(z) - \Phi_{\boldsymbol{\theta}_0}(z) - B(z)h^2\} \xrightarrow{D} \mathrm{N}(0, [\mathcal{W}(z, \boldsymbol{\theta}_0)]^{-1} k_2 \sigma^2). \quad \text{(A.11)}$$

Theorem 3 follows from (A.11) and Corollary 1.

[Received October 1997. Revised November 1998.]

REFERENCES

Chan, K. S., and Tong, H. (1986), "On Estimating Thresholds in Autoregressive Models," *Journal of Time Series Analysis*, 7, 179–190.

Chen, R. (1995), "Threshold Variable Selection in Open-Loop Threshold Autoregressive Models," *Journal of Time Series Analysis*, 16, 461–481.

Chen, R., and Tsay, S. (1993), "Functional-Coefficient Autoregressive Models," *Journal of the American Statistical Association*, 88, 298–308.

Cheng, B., and Tong, H. (1993), "On Residual Sums of Squares in Non-Parametric Autoregression," *Stochastic Processes and Their Applications*, 48, 157–174.

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

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

Hall, P. (1989), "On Projection Pursuit Regression," *The Annals of Statistics*, 17, 573–588.

Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, New York: Academic Press.

Härdle, W., Hall, P., and Ichimura, H. (1993), "Optimal Smoothing in Single-Index Models," *The Annals of Statistics*, 21, 157–178.

Härdle, W., Hall, P., and Marron, J. S. (1988), "How Far are Automatically Chosen Regression Smoothing Parameters From Their Optimum?," *Journal of the American Statistical Association*, 83, 86–101.

Härdle, W., and Vieu, P. (1992), "Kernel Regression Smoothing of Time Series," *Journal of Time Series Analysis*, 13, 209–224.

Hastie, T., and Tibshirani, R. (1993), "Varying-Coefficient Models" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 55, 757–796.

Huber, P. J. (1985), "Projection Pursuit" (with discussion), *The Annals of Statistics*, 13, 435–525.

Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics*, 58, 71–120.

Kim, T. Y., and Cox, D. D. (1995), "Asymptotic Behavior of Some Measures of Accuracy in Nonparametric Curve Estimation With Dependent Observations," *Journal of Multivariate Analysis*, 53, 67–93.

Masry, E., and Tjøstheim, D. (1995), "Nonparametric Estimation and Identification of Nonlinear ARCH Time Series: Strong Convergence and Asymptotic Normality," *Econometric Theory*, 11, 258–289.

Pham, D. T. (1986), "The Mixing Properties of Bilinear and Generalized Random Coefficient Autoregressive Models," *Stochastic Processes and Their Applications*, 23, 291–300.

Rio, E. (1995), "The Functional Law of the Iterated Logarithm for Stationary Strongly Mixing Sequences," *Annals of Probability*, 23, 1188–1203.

Stute, W. (1982), "A Law of the Logarithm for Kernel Density Estimators," *Annals of Probability*, 10, 414–422.

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

Xia, Y., and Li, W. K. (1999), "On the Estimation and Testing of Functional-Coefficient Linear Models," *Statistica Sinica*, 9, 735–757.