Journal of Multivariate Analysis **74**, 116–134 (2000)
doi:10.1006/jmva.1999.1883, available online at http://www.idealibrary.com on IDEAL®

Variable Bandwidth Selection in Varying-Coefficient Models

Wenyang Zhang

Institute of Mathematics & Statistics, University of Kent at Canterbury, Canterbury, England

and

Sik-Yum Lee

Department of Statistics, The Chinese University of Hong Kong, Hong Kong

Received June 24, 1998

The varying-coefficient model is an attractive alternative to the additive and other models. One important method in estimating the coefficient functions in this model is the local polynomial fitting approach. In this approach, the choice of bandwidth is crucial. If the unknown curve is spatial homogeneous, a constant bandwidth is sufficient. However, for estimating curves with a more complicated structure, a variable bandwidth is needed. The present article focuses on a variable bandwidth selection procedure, and provides the conditional bias and the conditional variance of the estimator, the convergence rate of the bandwidth, and the asymptotic distribution of its error relative to the theoretical optimal variable bandwidth. © 2000 Academic Press

AMS 1991 subject classifications: primary 62G05; secondary 62G08, or 62G20.

Key words and phrases: varying-coefficient models, local polynomial fitting, data-driven bandwidth selection, relative error, asymptotic normality, assessment of conditional bias and variance.

1. INTRODUCTION

In recent years, various nonparametric techniques have been developed to increase the flexibility of the regression modeling. Examples include the approach of penalized least squares (Wahba, 1990) and the local polynomial modeling (Fan and Gijbels, 1996). Among these methods, the local polynomial approach has been shown to be an attractive method from both theoretical and practical points of view; see, for example, Stone (1977), Cleveland (1979), Fan and Gijbels (1992), Fan (1993), Lu (1996), and Ruppert and Wand (1994). In this procedure, the choice of bandwidth is crucial in the analysis. In the literature, there are many proposals for

0047-259X/00 $35.00

selecting the bandwidth; examples include the cross-validation technique (Bowman, 1984; Scott and Terrell, 1987; Vieu, 1991; Hall and Johnstone, 1992; Fan *et al.*, 1996a) and the plug-in approach (Woodroofe, 1970; Sheather and Jones, 1991; Jones *et al.* 1996).

Based on the model

$$Y = a(U) + \sigma(U)\,\varepsilon, \tag{1.1}$$

and in the context of local polynomial fitting, Fan and Gijbels (1995) developed a procedure for bandwidth selection which can be applied to both constant and variable bandwidth selections in a wide variety of situations. Further theoretical foundation of their procedure has been provided by Fan *et al.* (1996b). An important extension of the model (1.1) as well as other linear models is the varying-coefficient model (Hastie and Tibshirani, 1993)

$$Y = \sum_{j=1}^{p} a_j(U)\, X_j + \varepsilon, \tag{1.2}$$

for given covariates $(U, X_1, ..., X_p)^T$ and response variable Y with

$$E(\varepsilon \mid U, X_1, ..., X_p) = 0, \qquad \text{and} \qquad \mathrm{Var}(\varepsilon \mid U, X_1, ..., X_p) = \sigma^2(U).$$

It is well recognized that (see Hastie and Tibshirani, 1993, and its discussion) this model has extremely wide applications. For example, see Hoover *et al.* (1997), Brumback and Rice (1998), Wu *et al.* (1998), and Fan and Zhang (1998) for application to longitudinal data, and Chen and Tsay (1993) and Cai *et al.* (1998) for application to nonlinear time series. Assuming the coefficient functions possess about the same degree of smoothness, Hastie and Tibshirani (1993) proposed an estimate for $a_j(U)$ via the dynamic linear model (West *et al.*, 1985; West and Harrison, 1989) and the approach of penalized least squares (Wahba, 1990). A two-step method was proposed by Fan and Zhang (1997) to analyze the model in which the coefficient functions admit different degrees of smoothness.

Inspired by the work of Fan and Zhang (1997), we propose in this paper an estimation method based on local polynomial fitting for analyzing varying coefficient models with coefficient functions that possess about the same degree of smoothness. Asymptotic expressions for the conditional bias and conditional variance of the estimators are derived. Like other methods in nonparametric estimation, the selection of bandwidth in the kernel function is crucial. In general, if the unknown coefficient functions are spatial homogeneous, a constant bandwidth is sufficient; however, for estimating coefficient functions with more complex structures, variable bandwidth is needed. In this paper, we present a variable bandwidth selection procedure.

Moreover, we establish the rate of convergence of the bandwidth selector and provide the asymptotic distribution of its error relative to the theoretical optimal variable bandwidth. These results are extensions of Fan *et al.*'s (1996b) work that associated with Fan and Gijbels' (1995) procedure for model (1.1) to the more general varying-coefficient models as defined in (1.2). Since it is assumed that the coefficient functions possess about the same degree of smoothness, we use the same bandwidth for estimating all the coefficient functions.

The paper is organized as follows. In Section 2, we briefly discuss the estimation method and its associated bandwidth selection procedure. The main asymptotic results are provided in Section 3. Some illustrative examples on the bandwidth selection procedure are given in Section 4, and technical proofs are given in Section 5.

2. ESTIMATION METHODS AND BANDWIDTH SELECTION PROCEDURE

Throughout this article, we assume that the coefficient functions $a_j(\cdot)$, $j=1, ..., p$, in model (1.2) possess about the same degrees of smoothness. Consider identically and independent distributed (i.i.d.) random observations $\{(U_i, X_{i1}, ..., X_{ip}, Y_i),\ i=1, ..., n\}$. Based on the arguments given in Fan (1992), we adopt local polynomials of odd order q to estimate the functions $a_j(\cdot)$, $j=1, ..., p$. For each given point u_0, we approximate the function locally as

$$a_j(u) \approx \sum_{l=0}^{q} \beta_{j,l}(u-u_0)^l, \tag{2.1}$$

for u in a neighborhood of u_0; and consider the following local least-squares problem: Minimize

$$\sum_{i=1}^{n} \left\{ Y_i - \sum_{j=1}^{p} \sum_{l=0}^{q} \beta_{j,l}(U_i-u_0)^l X_{ij} \right\}^2 K_h(U_i-u_0), \tag{2.2}$$

for a given kernel function K and bandwidth h, where $K_h(\cdot)=K(\cdot/h)/h$. Let

$$Y=(Y_1, ..., Y_n)^T, \qquad W=\operatorname{diag}(K_h(U_1-u_0), ..., K_h(U_n-u_0)),$$

and

$$\mathbf{X}_q=\begin{pmatrix} X_{11} & \cdots & X_{11}(U_1-u_0)^q & \cdots & X_{1p} & \cdots & X_{1p}(U_1-u_0)^q \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{n1}(U_n-u_0)^q & \cdots & X_{np} & \cdots & X_{np}(U_n-u_0)^q \end{pmatrix}.$$

Moreover, let $\beta=(\beta_{1,0}, ..., \beta_{1,q}, ..., \beta_{p,0}, ..., \beta_{p,q})^T$; the solution of the least-squares problem (2.2) gives the following one-step estimator $\hat{\beta}$ of β:

$$\hat{\beta}=(\hat{\beta}_{1,0}, ..., \hat{\beta}_{1,q}, ..., \hat{\beta}_{p,0}, ..., \hat{\beta}_{p,q})^T=(\mathbf{X}_q^T W\mathbf{X}_q)^{-1}\mathbf{X}_q^T WY. \quad (2.3)$$

We use the normalized weighted residual sum of squares from the local polynomial of order q fit to estimate $\sigma^2=\sigma^2(u_0)$ as

$$\hat{\sigma}^2=\hat{\sigma}^2(u_0)=\frac{1}{\operatorname{tr}\{W-(\mathbf{X}_q^T W\mathbf{X}_q)^{-1}\mathbf{X}_q^T W^2\mathbf{X}_q\}}\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 K_h(U_i-u_0), \quad (2.4)$$

where $\hat{Y}=(\hat{Y}_1, ..., \hat{Y}_n)^T=\mathbf{X}_q(\mathbf{X}_q^T W\mathbf{X}_q)^{-1}\mathbf{X}_q^T WY$. Let

$$\hat{\mathbf{a}}(\cdot)=(\hat{a}_1(\cdot), ..., \hat{a}_p(\cdot))^T, \qquad \mathbf{a}(\cdot)=(a_1(\cdot), ..., a_p(\cdot))^T.$$

Obviously, the local polynomial regression estimate of $\mathbf{a}=\mathbf{a}(u_0)$ is given by

$$\hat{\mathbf{a}}=\hat{\mathbf{a}}(u_0)=(I_p\otimes e_{1,q})(\mathbf{X}_q^T W\mathbf{X}_q)^{-1}\mathbf{X}_q^T WY, \quad (2.5)$$

where $\otimes$ denotes the Kronecker product and $e_{k,q}$ denotes the unit vector of length $q+1$ with 1 at position k.

To introduce the variable bandwidth selection procedure, the following notations are required:

$$\mu_i=\int t^i K(t)\,dt, \qquad \mathbf{u}_i=(\mu_{q+i}, ..., \mu_{2q+i})^T, \qquad \text{and} \qquad \nu_i=\int t^i K^2(t)\,dt.$$

Note that $\mu_i=0$, and $\nu_i=0$ when the kernel function $K(\cdot)$ is symmetric and i is odd. Let Γ_q be a $(q+1)\times(q+1)$ matrix with elements

$$\tau_{ij}=\begin{cases}0 & i+j=\text{odd}\\ \mu_{i+j} & i+j=\text{even}\end{cases} \qquad \text{for} \quad i, j=0, ..., q,$$

Let $\tilde{\Gamma}_q$ be the matrix similar to Γ_q with μ_i replacing by ν_i, and let $\mathscr{D}$ be the observed covariates vector

$$\mathscr{D}=(U_1, ..., U_n, X_{11}, ..., X_{1n}, ..., X_{p1}, ..., X_{pn})^T.$$

Let $r_{ij}(u)=E(X_iX_j \mid U=u)$, $r_{ij}=r_{ij}(u_0)$, for $i, j=1, ..., p$, $\Omega(u)$ and Ω are matrices with their (i, j)th elements equal to $r_{ij}(u)$ and r_{ij}, respectively. Moreover, let $\operatorname{bias}(\hat{a}_i(u)\mid\mathscr{D})$ be the conditional bias of $\hat{a}_i(u)$ given $\mathscr{D}$,

$$\mathbf{b}(u)=\operatorname{bias}(\hat{\mathbf{a}}(u)\mid\mathscr{D})=(\operatorname{bias}(\hat{a}_1(u)\mid\mathscr{D}), ..., \operatorname{bias}(\hat{a}_p(u)\mid\mathscr{D}))^T, \qquad \text{and}$$

$$\mathbf{b}=\mathbf{b}(u_0).$$

Based on the definition of the varying-coefficient model defined in (1.2), we define the mean squared error

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{a}}(\cdot)) &= E\left(E\left[\left\{\sum_{j=1}^{p}(\hat{a}_j(U)-a_j(U))\,X_j\right\}^2 \,\middle|\, \mathscr{D}\right]\right)\\ &= E\left\{\sum_{j=1}^{p}(\hat{a}_j(U)-a_j(U))\,X_j\right\}^2 \qquad (2.6)\end{aligned}$$

as a criterion to depict the error of the estimators $\hat{a}_j$, $j=1, ..., p$, where the first expectation is taken over $U, X_1, ..., X_p$ which are random variables that are independent of the observed sample. It can be shown that

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{a}}(\cdot)) &= \mathrm{tr}\, E\{\Omega(U)\, E((\hat{\mathbf{a}}(U)-\mathbf{a}(U))(\hat{\mathbf{a}}(U)-\mathbf{a}(U))^T \mid U)\}\\ &= \mathrm{tr}\, E[E\{\Omega(U)(\hat{\mathbf{a}}(U)-\mathbf{a}(U))(\hat{\mathbf{a}}(U)-\mathbf{a}(U))^T \mid U, \mathscr{D}\}]\\ &= E\{\mathbf{b}^T(U)\,\Omega(U)\,\mathbf{b}(U)+\mathrm{tr}(\Omega(U)\,\mathrm{Cov}(\hat{\mathbf{a}}(U)\mid U, \mathscr{D}))\}.\end{aligned}$$

Define

$$\mathrm{MSE}(\hat{\mathbf{a}}(u)\mid \mathscr{D}) = \mathbf{b}^T(u)\,\Omega(u)\,\mathbf{b}(u)+\mathrm{tr}(\Omega(u)\,\mathrm{Cov}(\hat{\mathbf{a}}(u)\mid \mathscr{D})). \qquad (2.7)$$

Note that

$$\mathrm{Cov}(\hat{\mathbf{a}}(U)\mid U, \mathscr{D}) = \mathrm{Cov}(\hat{\mathbf{a}}(u)\mid \mathscr{D})|_{u=U},$$

hence

$$\mathrm{MSE}(\hat{\mathbf{a}}(\cdot)) = E[E\{\mathrm{MSE}(\hat{\mathbf{a}}(u)\mid \mathscr{D})|_{u=U}\mid \mathscr{D}\}].$$

The variable bandwidth h_{opt} that minimizes the $\mathrm{MSE}(\hat{\mathbf{a}}\mid \mathscr{D})$ is called the theoretical optimal variable bandwidth at u_0. Since the quantity $\mathrm{MSE}(\hat{\mathbf{a}}\mid \mathscr{D})$ depends on some unknown quantities, it is impossible to find the theoretical optimal variable bandwidth. To cope with this problem, the plug-in approach (Ruppert *et al.*, 1995) first derives the minimizer of the asymptotic expression of $\mathrm{MSE}(\hat{a}\mid D)$, namely, the asymptotic optimal variable bandwidth; then replaces the unknown parameters by their estimators. A disadvantage of this approach is that it depends heavily on the asymptotic expressions. The following more reasonable approach is to minimize a good estimator of $\mathrm{MSE}(\hat{\mathbf{a}}\mid \mathscr{D})$ and take the minimizer as the variable bandwidth. The quantity $\mathrm{MSE}(\hat{\mathbf{a}}\mid \mathscr{D})$ will be estimated as below.

The conditional bias $\mathbf{b}$ in MSE$(\hat{\mathbf{a}} \mid \mathscr{D})$ is equal to $(I_p \otimes e_{1,q}^T)(\mathbf{X}_q^T W \mathbf{X}_q)^{-1} \mathbf{X}_q^T W \ell$, where $\ell = (\ell_1, ..., \ell_n)^T$ with

$$\ell_i = \sum_{j=1}^{p} \left(a_j(U_i) - \sum_{k=0}^{q} \beta_{j,k} (U_i - u_0)^k \right) X_{ij}.$$

Based on the Taylor expansion of order m, the conditional bias can be approximated by $(I_p \otimes e_{1,q}^T)(\mathbf{X}_q^T W \mathbf{X}_q)^{-1} \mathbf{X}_q^T W \eta$, where η is a n vector with ith element equal to

$$\sum_{j=1}^{p} \sum_{k=1}^{m} \beta_{j,q+k} (U_i - u_0)^{q+k} X_{ij}.$$

For convenience, we take $m = 2$; then $(I_p \otimes e_{1,q}^T)(\mathbf{X}_q^T W \mathbf{X}_q)^{-1} \mathbf{X}_q^T W \eta$ is simplified as

$$(I_p \otimes e_{1,q}^T)(\mathbf{X}_q^T W \mathbf{X}_q)^{-1} \mathbf{X}_q^T W \mathbf{X}_q^* \mathbf{d},$$

where

$$\mathbf{d}_i = (\beta_{1,q+i}, ..., \beta_{p,q+i})^T, \qquad \text{for} \quad i = 1, 2, ...;$$
$$\mathbf{d} = (\mathbf{d}_1^T \otimes (1, 0) + \mathbf{d}_2^T \otimes (0, 1))^T,$$

and

$$\mathbf{X}_q^* = \begin{pmatrix} X_{11}(U_1 - u_0)^{q+1} & X_{11}(U_1 - u_0)^{q+2} & \cdots & & \\ \vdots & \vdots & \ddots & & \\ X_{n1}(U_n - u_0)^{q+1} & X_{n1}(U_n - u_0)^{q+2} & \cdots & & \\ & & & X_{1p}(U_1 - u_0)^{q+1} & X_{1p}(U_1 - u_0)^{q+2} \\ & & & \vdots & \vdots \\ & & & X_{np}(U_n - u_0)^{q+1} & X_{np}(U_n - u_0)^{q+2} \end{pmatrix}.$$

The quantity $\mathbf{d}$ can be estimated by using a local polynomial regression of order g $(g > q)$ with a bandwidth h_*, namely

$$\hat{\mathbf{d}} = (I_p \otimes (e_{q+2,g}, e_{q+3,g})^T)(\mathbf{X}_g^T W_* \mathbf{X}_g)^{-1} \mathbf{X}_g^T W_* Y,$$

where $W_* = \text{diag}(K_{h_*}(U_1 - u_0), ..., K_{h_*}(U_n - u_0))$. The initial bandwidth h_* can be obtained by the minimizer of some residual squares criterion (RSC) as given in Zhang and Lee (1998); see also Fan and Gijbels (1995).

The conditional covariance is given by

$$\text{Cov}(\hat{\mathbf{a}} \mid \mathscr{D}) = (I_p \otimes e_{1,q}^T)(\mathbf{X}_q^T W \mathbf{X}_q)^{-1} \times (\mathbf{X}_q^T W \Psi W \mathbf{X}_q)(\mathbf{X}_q^T W \mathbf{X}_q)^{-1} (I_p \otimes e_{1,q}),$$

where $\Psi = \text{diag}(\sigma^2(U_1), ..., \sigma^2(U_n))$. We can approximate it by using the local homoscedasticity as follows:

$$(I_p \otimes e_{1,q}^T)(\mathbf{X}_q^T W \mathbf{X}_q)^{-1} (\mathbf{X}_q^T W^2 \mathbf{X}_q)(\mathbf{X}_q^T W \mathbf{X}_q)^{-1} (I_p \otimes e_{1,q})\, \sigma^2.$$

The unknown parameter σ^2 can be estimated by the normalized weighted residual sum of squares from a gth-order polynomial fit as below,

$$\hat{\sigma}^2 = \frac{1}{\text{tr}(W_*) - \text{tr}((\mathbf{X}_g^T W_* \mathbf{X}_g)^{-1} \mathbf{X}_g^T W_*^2 \mathbf{X}_g)} \sum_{i=1}^{n} (Y_i - \hat{Y}_{*i})^2 K_{h_*}(U_i - u_0),$$

where

$$\hat{Y}_* = (\hat{Y}_{*1}, ..., \hat{Y}_{*n})^T = \mathbf{X}_g(\mathbf{X}_g^T W_* \mathbf{X}_g)^{-1} \mathbf{X}_g^T W_* Y.$$

The estimate of the element r_{ij} in Ω will be obtained based on $(U_l, X_{li} X_{lj})$, $l = 1, ..., n$, using the local polynomial fit of order g with bandwidth $h_{0*} = O_P(h_*)$ as below,

$$\hat{r}_{ij} = e_{1,g}^T (\mathbf{V}^T W_{0*} \mathbf{V})^{-1} \mathbf{V}^T W_{0*} Z_{ij},$$

where $W_{0*} = \text{diag}(K_{h_{0*}}(U_1 - u_0), ..., K_{h_{0*}}(U_n - u_0))$ and

$$\mathbf{V} = \begin{pmatrix} 1 & \cdots & (U_1 - u_0)^g \\ \vdots & \ddots & \vdots \\ 1 & \cdots & (U_n - u_0)^g \end{pmatrix}, \qquad Z_{ij} = \begin{pmatrix} X_{1i} X_{1j} \\ \vdots \\ X_{ni} X_{nj} \end{pmatrix}.$$

By Theorem 3.1 of Fan and Gijbels (1996), it can be shown that

$$\text{bias}(\hat{r}_{ij} \mid U_1, ..., U_n) = O_P(h_*^{g+1}), \qquad \text{Var}(\hat{r}_{ij} \mid U_1, ..., U_n) = O_P\left(\frac{1}{nh_*}\right). \tag{2.8}$$

Let $\hat{\Omega}$ be an estimator of Ω with elements $\hat{r}_{ij}$; we obtain the following estimate of $\text{MSE}(\hat{\mathbf{a}} \mid \mathscr{D})$:

$$\begin{aligned} \widehat{\text{MSE}}(\hat{\mathbf{a}} \mid \mathscr{D}) = {} & \hat{\mathbf{d}}^T \mathbf{X}_q^{*T} W \mathbf{X}_q (\mathbf{X}_q^T W \mathbf{X}_q)^{-1} \\ & \times (\hat{\Omega} \otimes e_{1,q} e_{1,q}^T)(\mathbf{X}_q^T W \mathbf{X}_q)^{-1} \mathbf{X}_q^T W \mathbf{X}_q^* \hat{\mathbf{d}} \\ & + \text{tr}\{(\mathbf{X}_q^T W \mathbf{X}_q)^{-1} (\mathbf{X}_q^T W^2 \mathbf{X}_q)(\mathbf{X}_q^T W \mathbf{X}_q)^{-1} \\ & \times (\hat{\Omega} \otimes e_{1,q} e_{1,q}^T)\, \hat{\sigma}^2\}. \end{aligned} \tag{2.9}$$

Finally, we select the variable bandwidth

$$\hat{h}_{opt} = \arg\min_{h} \widehat{\mathrm{MSE}}(\hat{\mathbf{a}} \mid \mathscr{D})$$

as our final bandwidth to be used in the analysis with polynomial of order q. Hereafter, we call this variable bandwidth the estimated optimal variable bandwidth.

3. MAIN ASYMPTOTIC RESULTS

We first impose the following technical conditions:

(1) $EX_j^{2s} < \infty$, for $s > 2$, $j = 1, ..., p$.

(2) Let $a_j^{(i)}$ denote the ith derivative of $a_j(\cdot)$; $a_j^{(q+3)}(\cdot)$ is continuous in a neighborhood of u_0, for $j = 1, ..., p$. Further, assume $a_j^{(q+1)}(u_0) \neq 0$, for $j = 1, ..., p$.

(3) The marginal density $f(u)$ of U has a continuous second derivative in some neighborhood of u_0 and $f(u_0) \neq 0$.

(4) The functions $r_{ij}(\cdot)$ and $\sigma^2(\cdot)$ have bounded second derivatives in a neighborhood of u_0.

(5) The function $K(t)$ is a symmetric density function with a compact support.

(6) $2/(2q+3) > (g-q)/(2g+3)$ and $g > q$.

The asymptotic expansions for the conditional bias and variance are given via the following theorem. The proof of this theorem, which will be given in the next section, is based on an extension of the arguments in Fan and Zhang (1997).

THEOREM 1. *Under conditions* (1)–(5), *if* $n^{-b} < h < n^{-a}$ *for* $0 < a < b < 1$, *then the conditional bias and covariance of* $\hat{\mathbf{a}}$ *have the following expansions uniformly for* $h \in [n^{-b}, n^{-a}]$,

$$\mathrm{bias}(\hat{\mathbf{a}} \mid \mathscr{D}) = e_{1,q}^T \Gamma_q^{-1} \mathbf{u}_1 \mathbf{d}_1 h^{q+1} \left(1 + O_P\left(h^2 + \frac{\log n}{\sqrt{nh}}\right)\right),$$

and

$$\mathrm{Cov}(\hat{\mathbf{a}} \mid \mathscr{D}) = \frac{\sigma^2}{nhf(u_0)} (e_{1,q}^T \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} e_{1,q})\, \Omega^{-1} \left(1 + O_P\left(h^2 + \frac{\log n}{\sqrt{nh}}\right)\right).$$

Remark. From Theorem 1, we see that the order of the convergence is $O_P(h^2 + \log n/\sqrt{nh})$. This result is crucial for getting the asymptotic distribution of the estimated optimal variable bandwidth. Fan and Zhang (1997) obtained some expressions for the asymptotic bias and variance, but they did not give the convergence rate. Using the technique of the proof of Theorem 1, we can get

$$\text{bias}(\hat{\boldsymbol{\beta}} \mid \mathscr{D}) = \mathbf{d}_1 \otimes \mathbf{G}_q^{-1} \Gamma_q^{-1} \mathbf{u}_1 h^{q+1}(1 + o_P(1)),$$

and

$$\text{Cov}(\hat{\boldsymbol{\beta}} \mid \mathscr{D}) = \frac{\sigma^2}{nhf(u_0)} \Omega^{-1} \otimes (\mathbf{G}_q^{-1} \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} \mathbf{G}_q^{-1})(1 + o_P(1)),$$

where $\mathbf{G}_q = \text{diag}(1, h, ..., h^q)$. This provides the asymptotic bias and the asymptotic variance of the estimator of the derivative of $\mathbf{a}$. Moreover, from the proof of Theorem 1, we can see that for the even derivative of $\mathbf{a}$, its asymptotic bias and asymptotic variance still have the convergence rate $O_P(h^2 + \log n/\sqrt{nh})$.

To derive the rate of convergence for bandwidth selection, we need the following lemma of Fan *et al.* (1996b):

LEMMA 1. *Suppose that a function* $M(h)$ *has the asymptotic expansion*

$$M(h) = ch^{2(p+1-v)} \left(1 + O_P\left(h^2 + \frac{\log n}{\sqrt{nh}}\right)\right) + \frac{\alpha}{nh^{2v+1}} \left(1 + O_P\left(h^2 + \frac{\log n}{\sqrt{nh}}\right)\right),$$

uniformly in $h \in [n^{-b}, n^{-a}]$. *Let* h_{min} *be the minimizer of* $M(h)$. *Then*

$$h_{min} = \left(\frac{(2v+1)\,\alpha}{2(p+1-v)\,cn}\right)^{1/(2p+3)} (1 + O_P(n^{-2/(2p+3)} \log n)),$$

and

$$M(h_{min}) = c^{1-s}\alpha^s(2p+3)(2v+1)^{-(1-s)} \times [2(p+1-v)]^{-s} n^{-s}(1 + O_P(n^{-2/(2p+3)} \log n)),$$

with $s = 2(p+1-v)/(2p+3)$, *provided that* $c, \alpha > 0$. *If* $p > 1$, *the* $\log n$ *factor does not have to appear in the "* O_P*-terms."*

Combining Theorem 1 *and* (2.7), *we obtain*

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{a}} \mid \mathscr{D}) = {} & (e_{1,q}^T \Gamma_q^{-1} \mathbf{u}_1)^2\, \mathbf{d}_1^T \Omega \mathbf{d}_1 h^{2q+2} (1 + O_P(h^2 + \log n/\sqrt{nh})) \\ & + p\sigma^2 (nhf(u_0))^{-1}\, (e_{1,q}^T \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} e_{1,q}) \\ & \times (1 + O_P(h^2 + \log n/\sqrt{nh})).\end{aligned}$$

From the above result and Lemma 1, *the following theorem is valid.*

THEOREM 2. *Under conditions* (1)–(5),

$$\frac{h_{opt} - h_{a,\,opt}}{h_{a,\,opt}} = O_P(n^{-2/(2q+3)} \log n),$$

where

$$h_{a,\,opt} = \left(\frac{p\sigma^2 e_{1,q}^T \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} e_{1,q}}{2(q+1)(e_{1,q}^T \Gamma_q^{-1} \mathbf{u}_1)^2\, \mathbf{d}_1^T \Omega \mathbf{d}_1 f(u_0)\, n} \right)^{1/(2q+3)}.$$

It is clear $h_{a,\,opt}$ *is the minimizer of*

$$(e_{1,q}^T \Gamma_q^{-1} \mathbf{u}_1)^2\, \mathbf{d}_1^T \Omega \mathbf{d}_1 h^{2q+2} + \frac{p\sigma^2}{nhf(u_0)} (e_{1,q}^T \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} e_{1,q}).$$

We call $h_{a,\,opt}$ the asymptotic optimal variable bandwidth. From Theorem 2, we see that the rate of the relative difference between the asymptotic optimal variable bandwidth and the optimal variable bandwidth converges to zero.

Now, we consider the estimated optimal variable bandwidth $\hat{h}_{opt}$. Let

$$\hat{\mathbf{d}}_1 = (I_p \otimes e_{q+2,\,g}^T)(\mathbf{X}_g^T W_* \mathbf{X}_g)^{-1} \mathbf{X}_g^T W_* Y$$

be an estimator of $\mathbf{d}_1$ obtained by using local polynomial of order g modeling with bandwidth h_*. Using arguments similar to Theorem 1, we have

$$\begin{aligned}\widehat{\mathrm{MSE}}(\hat{\mathbf{a}} \mid \mathscr{D}) = {} & (e_{1,q}^T \Gamma_q^{-1} \mathbf{u}_1)^2\, \hat{\mathbf{d}}_1^T \hat{\Omega} \hat{\mathbf{d}}_1 h^{2q+2} (1 + O_P(h^2 + \log n/\sqrt{nh})) \\ & + \hat{\sigma}^2 (nhf(u_0))^{-1}\, (e_{1,q}^T \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} e_{1,q}) \\ & \times \mathrm{tr}\{\hat{\Omega} \Omega^{-1}\} (1 + O_P(h^2 + \log n/\sqrt{nh})),\end{aligned}$$

uniformly for $h \in [n^{-b}, n^{-a}]$ with $0 < a < b < 1$. From Lemma 1, we have

$$\frac{\hat{h}_{opt} - \hat{h}_{a,\,opt}}{\hat{h}_{a,\,opt}} = O_P(n^{-2/(2q+3)} \log n), \tag{3.1}$$

where

$$\hat{h}_{a,\,opt} = \left(\frac{\hat{\sigma}^2 e_{1,q}^T \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} e_{1,q} \operatorname{tr}\{\hat{\Omega}\Omega^{-1}\}}{2(q+1)(e_{1,q}^T \Gamma_q^{-1} \mathbf{u}_1)^2 \hat{\mathbf{d}}_1^T \hat{\Omega} \hat{\mathbf{d}}_1 f(u_0)\, n} \right)^{1/(2q+3)},$$

the minimizer of

$$(e_{1,q}^T \Gamma_q^{-1} \mathbf{u}_1)^2 \hat{\mathbf{d}}_1^T \hat{\Omega} \hat{\mathbf{d}}_1 h^{2q+2} + \frac{\hat{\sigma}^2 \operatorname{tr}\{\hat{\Omega}\Omega^{-1}\}}{nhf(u_0)} (e_{1,q}^T \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} e_{1,q}).$$

It follows from Theorem 2 and (3.1) that the asymptotic result for the error of the estimated optimal variable bandwidth $\hat{h}_{opt}$ relative to the theoretical optimal variable bandwidth h_{opt} can be established by the connection between $h_{a,\,opt}$ and $\hat{h}_{a,\,opt}$.

The following theorem provides the asymptotic distribution of the error of the estimated optimal variable bandwidth $\hat{h}_{opt}$ relative to the theoretical optimal variable bandwidth h_{opt}.

THEOREM 3. *Under conditions* (1)–(6) *and* $\sigma^2(u) = \sigma^2$ *in a neighborhood of* u_0, *if* $h_* = O(n^{-1/(2g+3)})$, *then the asymptotic distribution of*

$$\sqrt{nh_*^{2q+3} f(u_0)} \left\{ \frac{\hat{h}_{opt} - h_{opt}}{h_{opt}} + \frac{2\mathbf{d}_1^T \Omega \mathbf{d}_{g-q+1} e_{q+2,g}^T \Gamma_g^{-1} \mathbf{u}_{g-q+1} h_*^{g-q}}{(2q+3)\, \mathbf{d}_1^T \Omega \mathbf{d}_1} \right\}$$

is

$$N\left(0,\, \frac{4\sigma^2 e_{q+2,g}^T \Gamma_g^{-1} \tilde{\Gamma}_g \Gamma_g^{-1} e_{q+2,g}}{(2q+3)^2\, \mathbf{d}_1^T \Omega \mathbf{d}_1}\right).$$

Remark. From Theorem 3 we can see that the error of the estimated optimal variable bandwidth $\hat{h}_{opt}$ relative to the theoretical optimal variable bandwidth h_{opt} is of order $O_P(n^{-(g-q)/(2g+3)})$. Moreover, the asymptotic bias and the asymptotic variance of the estimated optimal variable bandwidth are given by

$$\left\{ -\frac{2\mathbf{d}_1^T \Omega \mathbf{d}_{g-q+1} e_{q+2,g}^T \Gamma_g^{-1} \mathbf{u}_{g-q+1} h_{opt} h_*^{g-q}}{(2q+3)\, \mathbf{d}_1^T \Omega \mathbf{d}_1} \right\} (1 + o(1)),$$

and

$$\frac{4\sigma^2 e_{q+2,g}^T \Gamma_g^{-1} \tilde{\Gamma}_g \Gamma_g^{-1} e_{q+2,g} h_{opt}^2}{(2q+3)^2\, \mathbf{d}_1^T \Omega \mathbf{d}_1 f(u_0)\, h_*^{2q+3} n} (1 + o(1)),$$

respectively, where the Γ_g and $\tilde{\Gamma}_g$ are similarly defined as Γ_q and $\tilde{\Gamma}_q$.

4. SOME EXAMPLES

The following three examples are used to illustrate the empirical performance of our bandwidth selection method,

Example 1: $Y = X_1 \cos(3U) + 2X_2 \exp\{-16U^2\} + \varepsilon,$

Example 2: $Y = X_1 \cos(3U) + 2X_2 \exp\{-16U^2\} \sin(2U) + \varepsilon,$

Example 3: $Y = X_1 \sin(U) + 2X_2 \exp\{-16U^2\} + \varepsilon,$

where X_1 and X_2 are normally distributed with correlation coefficient $2^{-1/2}$, $E(X_1) = E(X_2) = 0$ and $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = 1$; U follows a uniform distribution on $[-2, 2]$, ε is distributed as normal with mean zero and variance σ^2; ε, U, and (X_1, X_2) are independent. For each example, a sample size $n = 200$ was considered. In general, the noise-to-signal ratio $\sigma^2/\mathrm{Var}\{E(Y \mid U, X_1, X_2)\}$ indicates the difficulty of the estimation problem: the bigger this ratio the more difficult the problem. For these examples, we choose σ^2 such that the noise-to-signal ratio is about 1 : 5, namely

$$\sigma^2 = 0.2\,\mathrm{Var}\{E(Y \mid U, X_1, X_2)\}.$$

Moreover, the local linear fit ($q = 1$) for the regression curve estimation is considered, and the kernel function is taken to be the Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$. In our bandwidth selection procedure, the initial bandwidth h_* is chosen by minimizing the residual squares criterion as given in Zhang and Lee (1998), and h_{0*} is chosen to be the same as h_*.

Based on 100 replications, the MSE $(\hat{a}(\cdot))$, see (2.6), of the estimated regression function with (i) the variable bandwidth obtained by our method, (ii) the theoretical optimal constant bandwidth, and (iii) the bandwidth obtained by the cross-validation method (see Hoover *et al.*, 1997) are computed and they are respectively denoted as MSE1, MSE2, and MSE3. In these three examples, the theoretical optimal constant bandwidths are all equal to 0.32. Table I describes the gain of our variable bandwidth over the theoretical optimal constant bandwidth and the bandwidth obtained via the cross-validation method. From this table, we

TABLE I

Comparison between the Bandwidth Selection Rules

	(MSE3-MSE1)/MSE1	(MSE2-MSE1)/MSE1
Example 1	0.288644	0.089782
Example 2	0.138218	0.010000
Example 3	0.408973	0.202803

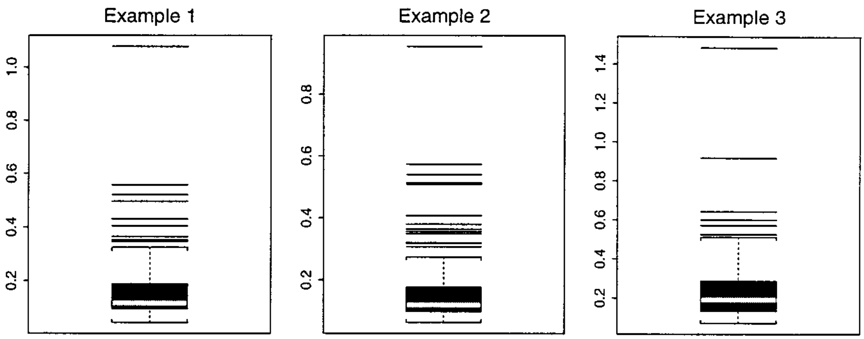

FIG. 1. The box-plots of deviations.

see that the empirical performance of our variable bandwidth is the best. Compared to the cross-validation method and the theoretical optimal constant bandwidth, the MSE of our procedure are reduced respectively by about (29%, 14%, 41 %) and (9%, 1%, 20 %) in Examples 1, 2, and 3.

To give more ideas on the empirical performance of our method, we use

$$\int \left(\frac{\hat{h}_{opt}(u) - h_{opt}(u)}{h_{opt}(u)} \right)^2 du \tag{4.1}$$

to describe the deviation of the variable bandwidth obtained by our method, $\hat{h}_{opt}(u)$, and the theoretical optimal variable bandwidth, $h_{opt}(u)$. Based on 100 replications, the box plot of the deviations defined in (4.1) is presented in Fig. 1. From this figure, we see that the deviation is small.

5. PROOF OF THEOREMS

Proof of Theorem 1. Let Φ be $p \times p$ matrix with elements

$$\phi_{ij} = \left. \frac{d(r_{ij}(u) f(u))}{du} \right|_{u=u_0},$$

and let Γ_q^* be a $(q+1) \times (q+1)$ matrix with elements

$$\tau_{ij}^* = \begin{cases} 0 & i+j = \text{even} \\ \mu_{i+j+1} & i+j = \text{odd}. \end{cases}$$

We first note that if (ξ_i, η_i), $i = 1, ..., n$ are i.i.d. random observations from the population associated with (ξ, η), then

$$\frac{1}{nh^j} \sum_{i=1}^n \eta_i (\xi_i - x_0)^j K_h(\xi_i - x_0) = t(x_0)\, \mu_j + ht'(x_0)\, \mu_{j+1} + O_P\left(h^2 + \frac{\log n}{\sqrt{nh}}\right),$$

uniformly for $h \in [n^{-b}, n^{-a}]$ with $0 < a < b < 1$, where $t(\cdot) = f_\xi(\cdot)\, v(\cdot)$, and $f_\xi(\cdot)$ is the density function of ξ and $v(x) = E(\eta \mid \xi = x)$.

Based on this result and let $0 < a < b < 1$, it can be shown that

$$\frac{1}{n}(I_p \otimes \mathbf{G}_q)^{-1}(\mathbf{X}_q^T W \mathbf{X}_q)(I_p \otimes \mathbf{G}_q)^{-1}$$
$$= \{\Omega \otimes \Gamma_q f(u_0) + \Phi \otimes \Gamma_q^* h + O(h^2)\}\left(1 + O_P\left(\frac{\log n}{\sqrt{nh}}\right)\right), \quad (5.1)$$

and

$$n^{-1}(I_p \otimes \mathbf{G}_q)^{-1}\mathbf{X}_q^T W\ell = \{\Omega \otimes (\mathbf{u}_1, \mathbf{u}_2) f(u_0) + \Phi \otimes (\mathbf{u}_2, \mathbf{u}_3)\, h + O(h^2)\}$$
$$\times (I_p \otimes \mathbf{G}_1)\, \mathbf{d} h^{q+1}(1 + O_P(\log n/\sqrt{nh}))$$

uniformly for $h \in [n^{-b}, n^{-a}]$. Using the fact $(A + hB)^{-1} = A^{-1} - hA^{-1}BA^{-1} + O(h^2)$, we get

$$\{n^{-1}(I_p \otimes \mathbf{G}_q)^{-1}(\mathbf{X}_q^T W \mathbf{X}_q)(I_p \otimes \mathbf{G}_q)^{-1}\}^{-1}$$
$$= \{f(u_0)^{-1}(\Omega^{-1} \otimes \Gamma_q^{-1}) - hf(u_0)^{-2}(\Omega^{-1}\Phi\Omega^{-1})$$
$$\otimes (\Gamma_q^{-1}\Gamma_q^*\Gamma_q^{-1}) + O(h^2)\}$$
$$\times (1 + O_P(\log n/\sqrt{nh})), \quad (5.2)$$

uniformly for $h \in [n^{-b}, n^{-a}]$. Hence, $\text{bias}(\hat{\mathbf{a}} \mid \mathscr{D})$ is equal to

$$[I_p \otimes (e_{1,q}^T\Gamma_q^{-1}\mathbf{u}_1, e_{1,q}^T\Gamma_q^{-1}\mathbf{u}_2) - hf(u_0)^{-1}$$
$$\times \{(\Omega^{-1}\Phi) \otimes (e_{1,q}^T\Gamma_q^{-1}\Gamma_q^*\Gamma_q^{-1}\mathbf{u}_1, e_{1,q}^T\Gamma_q^{-1}\Gamma_q^*\Gamma_q^{-1}\mathbf{u}_2)\}$$
$$+ hf(u_0)^{-1}\{(\Omega^{-1}\Phi) \otimes (e_{1,q}^T\Gamma_q^{-1}\mathbf{u}_2, e_{1,q}^T\Gamma_q^{-1}\mathbf{u}_3)\} + O(h^2)]$$
$$\times (I_p \otimes \mathbf{G}_1)\, \mathbf{d} h^{q+1}(1 + O_P(\log n/\sqrt{nh})),$$

uniformly for $h \in [n^{-b}, n^{-a}]$. Based on similar reasonings in Fan *et al.* (1996b), it can be shown that $e_{1,q}^T\Gamma_q^{-1}\mathbf{u}_2 = 0$, and $e_{1,q}^T\Gamma_q^{-1}\Gamma_q^*\Gamma_q^{-1}\mathbf{u}_1 = 0$. Hence

$$\text{bias}(\hat{\mathbf{a}} \mid \mathscr{D}) = e_{1,q}^T\Gamma_q^{-1}\mathbf{u}_1\mathbf{d}_1 h^{q+1}(1 + O_P(h^2 + \log n/\sqrt{nh})).$$

Next, we consider $\text{Cov}(\hat{\mathbf{a}} \mid \mathscr{D})$. Let $\tilde{\Gamma}_q^*$ be the matrix similar to Γ_q^* except replacing μ_i by v_i, and let Φ^* be $p \times p$ matrix with element $d(\sigma^2(u) f(u) r_{ij}(u))/du|_{u=u_0}$. Using arguments similar to those used in getting (5.1), we have

$$hn^{-1}(I_p \otimes \mathbf{G}_q)^{-1}(\mathbf{X}_q^T W\Psi W\mathbf{X}_q)(I_p \otimes \mathbf{G}_q)^{-1}$$
$$= \{\Omega \otimes \tilde{\Gamma}_q f(u_0)\,\sigma^2 + \Phi^* \otimes \tilde{\Gamma}_q^*\, h + O(h^2)\}(1 + O_P(\log n/\sqrt{nh}))$$

uniformly for $h \in [n^{-b}, n^{-a}]$. Combining this result and (5.2), we get

$$\mathrm{Cov}(\hat{\mathbf{a}} \mid \mathscr{D}) = \sigma^2(nhf(u_0))^{-1}(e_{q,1}^T \Gamma_q^{-1}\tilde{\Gamma}_q\Gamma_q^{-1}e_{q,1})\,\Omega^{-1}$$
$$\times (1 + O_P(h^2 + \log n/\sqrt{nh}))$$

uniformly for $h \in [n^{-b}, n^{-a}]$. This establishes the results of Theorem 1.

Proof of Theorem 3. First, note that

$$h_{a,\,opt} = O(n^{-1/(2q+3)}) \qquad \text{and} \qquad \hat{h}_{a,\,opt} = O_P(n^{-1/(2q+3)}).$$

This result together with Eq. (3.1) and Theorem 2 leads to

$$h_{opt} = O(n^{-1/(2q+3)}) \qquad \text{and} \qquad \hat{h}_{opt} = O_P(n^{-1/(2q+3)}).$$

Consequently,

$$\frac{\hat{h}_{opt} - h_{opt}}{h_{opt}} = \frac{\hat{h}_{a,\,opt} - h_{a,\,opt}}{h_{a,\,opt}} + O_P(n^{-2(2q+3)^{-1}}\log n).$$

Further, note that

$$\frac{\hat{h}_{a,\,opt} - h_{a,\,opt}}{h_{a,\,opt}} = \frac{\hat{h}_{a,\,opt}}{h_{a,\,opt}} - 1 = \left(\frac{\mathrm{tr}(\hat{\Omega}\Omega^{-1})\,\hat{\sigma}^2}{p\sigma^2}\right)^{1/(2q+3)} \left(\frac{\mathbf{d}_1^T\Omega\mathbf{d}_1}{\hat{\mathbf{d}}_1^T\hat{\Omega}\hat{\mathbf{d}}_1}\right)^{1/(2q+3)} - 1 \tag{5.3}$$

and

$$\left(\frac{\mathrm{tr}(\hat{\Omega}\Omega^{-1})\,\hat{\sigma}^2}{p\sigma^2}\right)^{1/(2q+3)} \left(\frac{\mathbf{d}_1^T\,\Omega\mathbf{d}_1}{\hat{\mathbf{d}}_1^T\hat{\Omega}\hat{\mathbf{d}}_1}\right)^{1/(2q+3)} - 1$$
$$= \left(\frac{\mathbf{d}_1^T\,\Omega\mathbf{d}_1}{\hat{\mathbf{d}}_1^T\hat{\Omega}\hat{\mathbf{d}}_1}\right)^{1/(2q+3)} \left\{\left(\frac{\mathrm{tr}(\hat{\Omega}\Omega^{-1})\,\hat{\sigma}^2}{p\sigma^2}\right)^{1/(2q+3)} - \left(\frac{\hat{\mathbf{d}}_1^T\hat{\Omega}\hat{\mathbf{d}}_1}{\mathbf{d}_1^T\,\Omega\mathbf{d}_1}\right)^{1/(2q+3)}\right\}. \tag{5.4}$$

From (2.8), we have

$$\hat{\Omega} = \Omega\left(1 + O_P\left(h_*^{g+1} + \frac{1}{\sqrt{nh_*}}\right)\right), \tag{5.5}$$

which gives

$$\frac{\operatorname{tr}(\hat{\Omega}\Omega^{-1})\,\hat{\sigma}^2}{p\sigma^2}=\left(\frac{\hat{\sigma}^2}{\sigma^2}\right)\left(1+O_P\left(h_*^{g+1}+\frac{1}{\sqrt{nh_*}}\right)\right).$$

From Theorem 2 in Zhang and Lee (1998),

$$\hat{\sigma}^2=\sigma^2\left(1+O_P\left(h_*^{2g+2}+\frac{1}{\sqrt{nh_*}}\right)\right).$$

So,

$$\left(\frac{\operatorname{tr}(\hat{\Omega}\Omega^{-1})\,\hat{\sigma}^2}{p\sigma^2}\right)^{1/(2q+3)}=1+O_P\left(h_*^{g+1}+h_*^{2g+2}+\frac{1}{\sqrt{nh_*}}\right)$$
$$=1+O_P\left(h_*^{g+1}+\frac{1}{\sqrt{nh_*}}\right). \tag{5.6}$$

Using the Taylar expansion and (5.5), we obtain

$$\left(\frac{\hat{\mathbf{d}}_1^T\hat{\Omega}\hat{\mathbf{d}}_1}{\mathbf{d}_1^T\Omega\mathbf{d}_1}\right)^{1/(2q+3)}=\left(1+\frac{2\mathbf{d}_1^T\Omega(\hat{\mathbf{d}}_1-\mathbf{d}_1)}{(2q+3)\,\mathbf{d}_1^T\Omega\mathbf{d}_1}+O_P(\|\hat{\mathbf{d}}_1-\mathbf{d}_1\|^2)\right)$$
$$\times\left(1+O_P\left(h_*^{g+1}+\frac{1}{\sqrt{nh_*}}\right)\right), \tag{5.7}$$

where $\|\hat{\mathbf{d}}_1-\mathbf{d}_1\|^2=(\hat{\mathbf{d}}_1-\mathbf{d}_1)^T(\hat{\mathbf{d}}_1-\mathbf{d}_1)$.

Now, we prove that $\sqrt{nh_*^{2q+3}f(u_0)}\,(\hat{\mathbf{d}}_1-\mathbf{d}_1)$ is asymptotic normal. Let $\boldsymbol{\varepsilon}=(\varepsilon_1, ..., \varepsilon_n)^T$, it can be shown that

$$\sqrt{nh_*^{2q+3}f(u_0)}\,(\hat{\mathbf{d}}_1-\mathbf{d}_1)$$
$$=\sqrt{nh_*^{2q+3}f(u_0)}\,(I_p\otimes e_{q+2,\,g}^T)(\mathbf{X}_g^TW_*\mathbf{X}_g)^{-1}\mathbf{X}_g^TW_*\boldsymbol{\varepsilon}$$
$$+\sqrt{nh_*^{2q+3}f(u_0)}\{(I_p\otimes e_{q+2,\,g}^T)(\mathbf{X}_g^TW_*\mathbf{X}_g)^{-1}$$
$$\times\mathbf{X}_g^TW_*E(Y\mid\mathscr{D})-\mathbf{d}_1\}$$

and

$$\sqrt{nh_*^{2q+3}f(u_0)}\,(I_p\otimes e_{q+2,\,g}^T)(\mathbf{X}_g^TW_*\mathbf{X}_g)^{-1}\mathbf{X}_g^TW_*\boldsymbol{\varepsilon}$$
$$=\{\Omega^{-1}\otimes(e_{q+2,\,g}^T\Gamma_g^{-1})\}\,f(u_0)^{-1}\sqrt{nh_*f(u_0)}\,n^{-1}(I_p\otimes\mathbf{G}_g^{-1})$$
$$\times\mathbf{X}_g^TW_*\boldsymbol{\varepsilon}(1+o_P(1)).$$

From the central limit theorem,

$$\sqrt{nh_* f(u_0)}\, n^{-1}(I_p \otimes \mathbf{G}_g^{-1})\, \mathbf{X}_g^T W_* \boldsymbol{\varepsilon} \xrightarrow{D} N(0, (\Omega \otimes \tilde{\Gamma}_g)\, \sigma^2 f(u_0)^2).$$

So,

$$\sqrt{nh_*^{2q+3} f(u_0)}\, (I_p \otimes e_{q+2,g}^T)(\mathbf{X}_g^T W_* \mathbf{X}_g)^{-1} \mathbf{X}_g^T W_* \boldsymbol{\varepsilon} \xrightarrow{D} N(0, e_{q+2,g}^T \Gamma_g^{-1} \tilde{\Gamma}_g \Gamma_g^{-1} e_{q+2,g} \Omega^{-1} \sigma^2).$$

Further, it can be shown that

$$\sqrt{nh_*^{2q+3} f(u_0)}\, \{(I_p \otimes e_{q+2,g}^T)(\mathbf{X}_g^T W_* \mathbf{X}_g)^{-1} \mathbf{X}_g^T W_* E(Y \mid \mathscr{D}) - \mathbf{d}_1\} = \sqrt{nh_*^{2q+3} f(u_0)}\, (e_{q+2,g}^T \Gamma_g^{-1} \mathbf{u}_{g-q+1})\, \mathbf{d}_{g-q+1} h_*^{g-q} (1 + o_P(1)),$$

hence,

$$\sqrt{nh_*^{2q+3} f(u_0)}\, \{\hat{\mathbf{d}}_1 - \mathbf{d}_1 - (e_{q+2,g}^T \Gamma_g^{-1} \mathbf{u}_{g-q+1})\, \mathbf{d}_{g-q+1} h_*^{g-q}\} \xrightarrow{D} (0, e_{q+2,g}^T \Gamma_g^{-1} \tilde{\Gamma}_g \Gamma_g^{-1} e_{q+2,g} \Omega^{-1} \sigma^2).$$

Further, it follows from the above result and (5.7) that the asymptotic distribution of

$$\sqrt{nh_*^{2q+3} f(u_0)} \times \left\{ \left(\frac{\hat{\mathbf{d}}_1^T \hat{\Omega} \hat{\mathbf{d}}_1}{\mathbf{d}_1^T \Omega \mathbf{d}_1} \right)^{1/(2q+3)} - 1 - \frac{2\mathbf{d}_1^T \Omega \mathbf{d}_{g-q+1} (e_{q+2,g}^T \Gamma_g^{-1} \mathbf{u}_{g-q+1})\, h_*^{g-q}}{(2q+3)\, \mathbf{d}_1^T \Omega \mathbf{d}_1} \right\}$$

is

$$N\left(0, \frac{4\sigma^2 e_{q+2,g}^T \Gamma_g^{-1} \tilde{\Gamma}_g \Gamma_g^{-1} e_{q+2,g}}{(2q+3)^2\, \mathbf{d}_1^T \Omega \mathbf{d}_1}\right).$$

Finally, it follows from (5.3), (5.4), and (5.6) that the asymptotic distribution of

$$\sqrt{nh_*^{2q+3} f(u_0)} \left\{ \frac{\hat{h}_{opt} - h_{opt}}{h_{opt}} + \frac{2\mathbf{d}_1^T \Omega \mathbf{d}_{g-q+1} (e_{q+2,g}^T \Gamma_g^{-1} \mathbf{u}_{g-q+1})\, h_*^{g-q}}{(2q+3)\, \mathbf{d}_1^T \Omega \mathbf{d}_1} \right\}$$

is

$$N\left(0, \frac{4\sigma^2 e_{q+2,g}^T \Gamma_g^{-1} \tilde{\Gamma}_g \Gamma_g^{-1} e_{q+2,g}}{(2q+3)^2\, \mathbf{d}_1^T \Omega \mathbf{d}_1}\right).$$

ACKNOWLEDGMENTS

The authors are grateful to the associate editor and two referees for their helpful comments, and to Esther L. S. Tam for typing the manuscript.

REFERENCES

A. W. Bowman, An alternative method of cross-validation for the smoothing of density estimates, *Biometrika* **71** (1984), 353–360.

B. A. Brumback and J. A. Rice, Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion), *J. Amer. Statist. Assoc.* **93** (1998), 961–994.

Z. Cai, J. Fan, and Q. Yao, Functional-coefficient regression models for nonlinear time series, *J. Amer. Statist. Assoc.*, 1998, to appear.

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

W. S. Cleveland, Robust locally weighted regression and smoothing scatterplots, *J. Amer. Statist. Assoc.* **74** (1979), 829–836.

J. Fan, Design-adaptive nonparametric regression, *J. Amer. Statist. Assoc.* **87** (1992), 998–1004.

J. Fan, Local linear regression smoothers and their minmax efficiency, *Ann. Statist.* **21** (1993), 196–216.

J. Fan and I. Gijbels, Variable bandwidth and local linear regression smoothers, *Ann. Statist.* **20** (1992), 2008–2036.

J. Fan and I. Gijbels, Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation, *J. Roy. Statist. Soc. Ser. B.* **57** (1995), 371–394.

J. Fan and I. Gijbels, "Local Polynomial Modeling and Its Applications," Chapman & Hall, London, 1996.

J. Fan and W. Zhang, Statistical estimation in varying coefficient models, 1997, submitted for publication.

J. Fan and J. Zhang, Functional linear models for longitudinal data, *J. Roy. Statist. Soc. Ser. B.*, 1998, to appear.

J. Fan, P. Hall, M. Martin, and P. Patil, On the local smoothing of nonparametric curve estimators, *J. Amer. Statist. Assoc.* **91** (1996a), 258–266.

J. Fan, I. Gijbels, T.-C. Hu, and L.-S. Huang, An asymptotic study of variable bandwidth selection for local polynomial regression with application to density estimation, *Statistica Sinica* **6** (1996b), 113–127.

P. Hall and I. Johnstone, Empirical functionals and efficient smoothing parameter selection (with discussion), *J. Roy. Statist. Soc. Ser. B.* **54** (1992), 475–530.

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

D. R. Hoover, J. A. Rice, C. O. Wu, and L. P. Yang, Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika*, 1997, to appear.

M. C. Jones, J. S. Marron, and S. J. Sheather, A brief survey of bandwidth selection for density estimation, *J. Amer. Statist. Assoc.* **91** (1996), 401–407.

Zhan-Qian Lu, Multivariate locally weighted polynomial fitting and partial derivatives estimation, *J. Multivariate* **59** (1996), 187–205.

D. Ruppert and M. P. Wand, Multivariate weighted least squares regression, *Ann. Statist.* **22** (1994), 1346–1370.

D. Ruppert, S. J. Sheather, and M. P. Wand, An effective bandwidth selector for local least squares regression, *J. Amer. Statist. Assoc.* **90** (1995), 1257–1270.

D. W. Scott and G. R. Terrell, Biased and unbiased cross-validation in density estimation, *J. Amer. Statist. Assoc.* **82** (1987), 1131–1146.

S. J. Sheather and M. C. Jones, A reliable data-based bandwidth selection method for kernel density estimation, *J. Roy. Statist. Soc. Ser. B.* **53** (1991), 683–690.

C. J. Stone, Consistent nonparametric regression, *Ann. Statist.* **5** (1977), 595–645.

P. Vieu, Nonparametric regression: Optimal local bandwidth choice, *J. Roy. Statist. Soc. Ser. B.* **53** (1991), 453–464.

G. Wahba, "Spline Models for Observing Data," SIAM, Philadelphia, 1990.

M. West, P. J. Harrison, and H. S. Migon, Dynamic generalized linear models and Bayesian forecasting (with discussion), *J. Amer. Statist. Assoc.* **80** (1985), 73–97.

M. West and P. J. Harrison, "Bayesian Forecasting and Dynamic Models," Springer, New York, 1989.

C. O. Wu, Chin-Tsang Chiang, and D. R. Hoover, Asymptotic confidence regions for smoothing of a varying-coefficient model with longitudinal data, *J. Amer. Statist. Assoc.* **93** (1998), 1388–1402.

M. Woodroofe, On choosing a delta-sequence, *Ann. Math. Statist.* **41** (1970), 1665–1671.

W. Zhang and S. Y. Lee, On local polynomial fitting of varying coefficient models, 1998, submitted for publication.