Journal of Multivariate Analysis **82**, 166–188 (2002)
doi:10.1006/jmva.2001.2012

Local Polynomial Fitting in Semivarying Coefficient Model

Wenyang Zhang

The London School of Economics & Political Science, London, United Kingdom

and

Sik-Yum Lee and Xinyuan Song

The Chinese University of Hong Kong, Shatin, N.T. Hong Kong

Received September 29, 1999; published online November 13, 2001

smoothness, the model can easily be estimated via simple local regression. This leads to the one-step estimation procedure. In this paper, we consider a semivarying coefficient model which is an extension of the varying coefficient model, which is called the semivarying-coefficient model. Procedures for estimation of the linear part and the nonparametric part are developed and their associated statistical properties are studied. The proposed methods are illustrated by some simulation studies and a real example. © 2001 Elsevier Science (USA)

AMS 2000 subject classifications: 62G05; 65G08; 62G20.

Key words and phrases: semivarying-coefficient models; varying-coefficient models; local polynomial fit; one-step method; two-step method; optimal rate of convergence; mean squared errors.

1. INTRODUCTION

In recent years, many useful nonparametric techniques have been developed to relax traditional parametric models and to exploit possible hidden structure. For an introduction to these methods, see the books by Hastie and Tibshirani (1990), Green and Silverman (1994), Wand and Jones (1995) and Fan and Gijbels (1996), among others. In the literature, there are many powerful approaches incorporated to avoid the so-called "curse of dimensionality." Examples include additive modeling (Breiman and Friedman, 1985; Hastie and Tibshirani 1990), low-dimensional interaction modeling (Friedman 1991; Gu and Wahba, 1993; Stone *et al.* 1997), multiple-index models (Härdle and Stoker 1989), partially linear models

0047-259X/01 $35.00

(Wahba 1984; Green and Silverman 1994) and their hybrids (Carroll *et al.* 1997). An important alternative to the additive and other models is the varying-coefficient model (Hastie and Tibshirani, 1993), in which the coefficients of the linear models are replaced by smoothing nonparametric functions and hence the regression coefficients are allowed to vary as functions of other factors. The varying-coefficient model is defined by the following linear model:

$$Y = \sum_{j=1}^{p} a_j(U)\, X_j + \varepsilon, \tag{1.1}$$

for given covariates $(U, X_1, \ldots, X_p)^T$ and response variable Y with $E(\varepsilon \mid U, X_1, \ldots, X_p) = 0$, and $\mathrm{Var}(\varepsilon \mid U, X_1, \ldots, X_p) = \sigma^2(U)$. By selecting $X_1 \equiv 1$, the model allows varying intercept in the model. Due to the generality of the function $a_j(U)$, the modeling bias can be reduced significantly and the "curse of dimensionality" can be avoided. Moreover, it is well-recognized that (Hastie and Tibshirani, 1993 and its discussion) the model has extremely wide applications. For example, see Hoover *et al.* (1998) for novel applications of the model to longitudinal data; Fan, Yao and Cai (2000) for applications in ecologic data analysis; Chen and Tsay (1993), and Cai, Fan and Yao (2000) for nonlinear time series applications.

Assuming the coefficient functions $a_j(U)$ possess about the same degree of smoothness, Hastie and Tibshirani (1993) proposed an one-step 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). Under the condition that the coefficient functions $a_j(\cdot)$, $j = 1, \ldots, p$, possess different degree of smoothness, Fan and Zhang (1999) showed that the one-step method is not optimal, and they proposed a two-step method to repair this drawback. The "two-step" idea has been applied to other problems, such as the functional data analysis (Fan and Zhang, 1998).

In practice, investigators often want to know if a covariate affects the response or if the coefficients are really varying, see Fan and Zhang (2000). This amounts to test if the whole function is zero or constant, namely, testing the null hypothesis H_0 that $a_j(U) = a_j$, for some j. The model under H_0 will be called a semivarying coefficients model, which is defined by the following linear model

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

for given covariates $(U, X_1, \ldots, X_p, X_{p+1}, \ldots, X_{p+m})^T$ and response variable Y with $E(\varepsilon \mid U, X_1, \ldots, X_p, X_{p+1}, \ldots, X_{p+m}) = 0$, and $\mathrm{Var}(\varepsilon \mid U, X_1, \ldots, X_p, X_{p+1}, \ldots, X_{p+m}) = \sigma^2(U)$. This model consists of a nonparametric part that involves coefficient functions $a_j(U)$, $j = 1, \ldots, p$ and a linear part that involves constant coefficients a_j, $j = p+1, ..., p+m$. The classical semiparametric model can be regarded as its special case with $a_2(U) \equiv \cdots \equiv a_p(U) \equiv 0$ and $X_1 \equiv 1$. On the other hand, if the constant coefficient a_j is viewed as a function, the semivarying coefficient model can be regarded as a special case of the varying coefficient model. In this context, since the coefficient functions clearly admit different degree of smoothness, a two-step method should be used to obtain the estimators for $a_j(\cdot)$, $j = 1, ..., p+m$. This approach requires to estimate $p+m$ functions. The associated estimation problem is clearly more complicated than the one that just requires to estimate p functions and m constant coefficients with a semivarying coefficient model. Moreover, it is expected that the variance of the estimator produced by this approach is larger.

To establish a test statistic for H_0, we require to develop an estimation method for the semiparametric coefficient model. In this paper, such an estimation procedure is developed. The main ideas of this procedure are as follows: The coefficients in the linear part of the model defined in (1.2) are first estimated via local polynomial fitting with a small bandwidth. This provides a class of raw estimators for each coefficient in the linear part. The final estimates of these coefficients are obtained by taking the average of these raw estimators. After replacing the coefficients in the linear part by the obtained estimators, (1.2) is simplified to a varying coefficient model. Then, the method given in Fan and Zhang (1999) is employed to produce the estimators of the coefficient functions in nonparametric part. For this procedure, the asymptotic conditional bias and variance will be derived to give some idea about the mean square error of the estimator. Moreover, we will show that the optimal rate of the convergence of the linear part estimator is almost the same as the least squares estimator for the classical linear model, and the performance of the nonparametric part estimator is asymptotically identical to the estimator under the ideal situation where the unknown constants of the linear part are known.

2. ESTIMATION PROCEDURE AND ITS PROPERTIES

Results presented in this paper are developed under some technical conditions which are presented in the Appendix. Also, the following notations are used: Let $a_j^{(i)}$ denote the ith derivative of $a_j(\cdot)$, $f(u)$ be the marginal density of U, $r_{i,j}(u) = E(X_i X_j \mid U = u)$,

$$\mu_i = \int t^i K(t)\, dt, \qquad \boldsymbol{\alpha}_q = (\mu_{q+1}, ..., \mu_{2q+1})^T, \qquad \boldsymbol{\mu}_q = (\mu_0, ..., \mu_q)^T \qquad \text{and}$$

$$v_i = \int t^i K^2(t)\, dt.$$

Let Γ_q be a $(q+1)\times(q+1)$ matrix with the (i, j)th element μ_{i+j}, $\tilde{\Gamma}_q$ be the matrix similar to Γ_q except replacing μ_i by ν_i, $a_{j,i}^{(k)} = a_j^{(k)}(U_i)$; and $\mathscr{D} = (U_1, \ldots, U_n, X_{11}, \ldots, X_{1n}, \ldots, X_{p1}, \ldots, X_{pn})^T$; and $e_{k,j}$ be the unit vector of length j with 1 at position k.

2.1. *Estimation Procedure for the Linear Part*

Our estimation procedure consists of the following two steps.

In the first step, for $j = 1, \ldots, p$, and each point u, we approximate the function $a_j(U)$ locally based on the idea of local polynomial modeling as

$$a_j(U) \approx \sum_{k=0}^{q} \frac{1}{k!} a_j^{(k)}(u)(U-u)^k, \qquad j = 1, \ldots, p, \tag{2.1}$$

for U in a neighborhood of u. Then, consider the following local least-squares problem: Minimize

$$\sum_{i=1}^{n} \left\{ Y_i - \sum_{j=1}^{p} \sum_{l=0}^{q} \beta_{j,l}(U_i-u)^l X_{ij} - \sum_{j=1+p}^{p+m} a_j X_{ij} \right\}^2 K_{h_1}(U_i-u), \tag{2.2}$$

with respect to $\beta_{j,l}$ and a_j, for a given kernel function K and a bandwidth h_1. The solution of (2.2) gives an initial estimator $\hat{a}_{1,j}(u)$ of a_j. Note that this estimator depends on u. Let

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

and

$$W = \operatorname{diag}(K_{h_1}(U_1-u), \ldots, K_{h_1}(U_n-u)), \qquad K_h(\cdot) = K(\cdot/h)/h.$$

When $u = U_i$, we denote $\mathbf{X}$ and W by X_i and W_i, respectively.

For each $j = p+1, \ldots, p+m$, a total of n initial estimators of a_j can be obtained as below by taking u equal to each of $U_1, \ldots, U_n$,

$$\hat{a}_{1,j}(U_i) = e_{pq+j,\ell}^T (\mathbf{V}_i^T W_i \mathbf{V}_i)^{-1} \mathbf{V}_i^T W_i Y, \qquad i = 1, \ldots, n,$$

where $\ell = p(q+1)+m$,

$$\mathbf{V}_i = (\mathbf{X}_i \mathbf{T}), \qquad \text{and} \qquad \mathbf{T} = \begin{pmatrix} X_{1(p+1)} & \cdots & X_{1(p+m)} \\ \vdots & \ddots & \vdots \\ X_{n(p+1)} & \cdots & X_{n(p+m)} \end{pmatrix}.$$

A procedure to select the bandwidth is required in the local least squares problem (2.2). The unknown coefficient functions $a_1(\cdot), \ldots, a_p(\cdot)$ demand a small bandwidth, while the unknown constants $a_{p+1}, \ldots, a_{p+m}$ demand a large bandwidth. So, if we wish to reduce the MSE corresponding to the estimators of $a_1(\cdot), \ldots, a_p(\cdot)$ by selecting a small bandwidth, the MSE corresponding to the estimators of $a_{p+1}, \ldots, a_{p+m}$ will be increased. No matter how to select the bandwidth h_1, we cannot obtain good estimators for both the unknown coefficient functions and the unknown constants. Therefore, $\hat{a}_{1,j}(\cdot)$ is only taken as an initial estimator of a_j.

As it will be shown in Theorem 1, the conditional bias of our final estimator is $O_P(h_1^{q+1})$, while the conditional variance is $O_P(n^{-1})$, as long as $nh_1/\log h_1 \to \infty$. This motivates us to choose the bandwidth h_1 as small as possible, subject to the constraint $nh_1/\log h_1 \to \infty$. On the other hand, since the sample size is finite in practice, large variance of the initial estimator should be avoided. Thus, we suggest the following bandwidth selection procedure in practical implementation: using the cross-validation method (see, Hoover *et al.*, 1998) to select a bandwidth h; then, take $h_1 = 0.5h$ (say) as the final bandwidth for the first step.

In the second step, the main objective is to reduce the variance of initial estimator which has small bias but large variance. One natural approach is to take the average of $\hat{a}_{1,j}(U_i)$ over $i = 1, \ldots, n$. This leads to the following final estimator $\hat{a}_j$ of a_j:

$$\hat{a}_j = \frac{1}{n} \sum_{i=1}^{n} \hat{a}_{1,j}(U_i) = \frac{1}{n} \sum_{i=1}^{n} e_{pq+j,\ell}^T (\mathbf{V}_i^T W_i \mathbf{V}_i)^{-1} \mathbf{V}_i^T W_i Y, \qquad j = p+1, \ldots, p+m. \tag{2.3}$$

Let

$$\begin{aligned} \Omega_{1,2}(u) &= E((X_1, \ldots, X_p)^T (X_{p+1}, \ldots, X_{p+m}) \mid U = u), \\ \Omega(u) &= E((X_1, \ldots, X_p)^T (X_1, \ldots, X_p) \mid U = u) \\ \mathbf{A}(u) &= \Omega_{2,2}(u) - \boldsymbol{\mu}_q^T \Gamma^{-1} \boldsymbol{\mu}_q \Omega_{1,2}^T(u)\, \Omega^{-1}(u)\, \Omega_{1,2}(u), \\ \text{and} \quad \Omega_{2,2}(u) &= E((X_{p+1}, \ldots, X_{p+m})^T (X_{p+1}, \ldots, X_{p+m}) \mid U = u). \end{aligned}$$

The following theorem gives the asymptotic bias and variance of the final estimator $\hat{a}_j$:

THEOREM 1. *Under the conditions $nh_1/\log h_1 \to \infty$ and $h_1 \to 0$, we have*

$$bias(\hat{a}_j \mid \mathscr{D}) = M_1 h_1^{q+1}(1 + o_P(1)), \qquad Var(\hat{a}_j \mid \mathscr{D}) = M_2 n^{-1}(1 + o_P(1)),$$

where

$$M_1 = \frac{(\mu_{q+1} - \boldsymbol{\mu}_q^T \Gamma_q^{-1} \boldsymbol{\alpha}_q)}{(q+1)!} e_{j-p,m}^T \mathbf{A}^{-1}(U)\, \Omega_{1,2}^T(U) \begin{pmatrix} a_1^{(q+1)}(U) \\ \vdots \\ a_p^{(q+1)}(U) \end{pmatrix}$$

and

$$\begin{aligned} M_2 = E\{&e_{j-p,m}^T \mathbf{A}^{-1}(U) \\ &\times [\Omega_{2,2} + \{(\boldsymbol{\mu}_q^T \Gamma_q^{-1} \boldsymbol{\mu}_q)^2 - 2(\boldsymbol{\mu}_q^T \Gamma_q^{-1} \boldsymbol{\mu}_q)\}\, \Omega_{1,2}^T(U)\, \Omega^{-1}(U)\, \Omega_{1,2}(U)] \\ &\times \mathbf{A}^{-1}(U)\, e_{j-p,m} \sigma^2(U)\}. \end{aligned}$$

Proofs of this theorem and the other theorems are given in the Appendix. It follows from Theorem 1 that the conditional bias of our estimator is of order $O_P(h_1^{q+1})$, while the conditional variance is of order $O_P(n^{-1})$. By taking h_1 as small as possible subject to the constraint $nh_1/\log h_1 \to \infty$, the bias can be very close to zero, while the variance is of order $O_P(n^{-1})$. Actually, if we choose h_1 such that $n^{1/2} h_1^{(q+1)} \to 0$, our estimator has a root-n convergent rate. In another words, the asymptotic performance of our estimator for the linear part of the semivarying coefficient model is the same as the least squares estimator for the classical linear model. We will see from our simulation study that whether the coefficient functions in the nonparametric part possess the same degree of smoothness does not affect the linear part estimation.

2.2. *A One-Stage Estimation Procedure for the Nonparametric Part*

We first consider an one-stage procedure in estimating the coefficient functions in the nonparametric part. Similar to the first step in the procedure for estimating the linear part, we approximate $a_j(U)$ locally as in (2.1), for $j = 1, \ldots, p$. By minimizing (2.2) with respect to $\beta_{j,l}$ and a_j, we obtain the one-stage estimator of $a_j(u)$, as follows,

$$\hat{a}_{v,o,j}(u) = e_{(q+1)(j-1)+1,\ell}^T (\mathbf{V}^T W_o \mathbf{V})^{-1}\, \mathbf{V}^T W_o Y, \qquad j = 1, \ldots, p, \qquad (2.4)$$

where $\mathbf{V}$ is the matrix $\mathbf{V}_i$ with $U_i = u$, W_o is W with $h_1 = h_o$.

In practical implementation of this one-stage procedure, the bandwidth h_o can be chosen by the cross-validation method (Hoover *et al.*, 1998). The conditional bias and the conditional variance of this estimator are given by the following theorem:

THEOREM 2. *Under the conditions $nh_o \to \infty$ and $h_o \to 0$, the conditional bias of $\hat{a}_{v,o,j}(u)$ is*

$$\mathit{bias}(\hat{a}_{v,o,j}(u) \mid \mathcal{D}) = \frac{h_o^{q+1}}{(q+1)!} e_{1,q+1}^T \Gamma_q^{-1} \boldsymbol{\alpha}_q a_j^{(q+1)}(u) + h_o^{q+1} R_1 + o_P(h_o^{q+1});$$

and the conditional variance is

$$Var(\hat{a}_{v,o,j}(u) \mid \mathscr{D}) = \frac{\sigma^2(u)}{nh_o f(u)} e^T_{1,q+1} \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} e_{1,q+1} e^T_{j,p} \Omega^{-1}(u)\, e_{j,p} + \frac{\sigma^2(u)}{nh_o f(u)} R_2 + o_P((nh_o)^{-1}),$$

where

$$R_1 = \frac{e^T_{1,q+1} \Gamma^{-1} \boldsymbol{\mu}_q (\boldsymbol{\mu}_q \Gamma^{-1} \boldsymbol{\alpha}_q - \mu_{q+1})}{(q+1)!} e^T_{j,p} \Omega^{-1}(u)\, \Omega_{1,2}(u)\, \mathbf{A}^{-1}(u) \times \Omega^T_{1,2}(u) \begin{pmatrix} a_1^{(q+1)}(u) \\ \vdots \\ a_p^{(q+1)}(u) \end{pmatrix},$$

$$\begin{aligned} R_2 = {} & (2e^T_{1,q+1} \Gamma_q^{-1} \boldsymbol{\mu}_q \boldsymbol{\mu}_q^T \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} e_{1,q+1} - 2e^T_{1,q+1} \Gamma_q^{-1} \boldsymbol{\mu}_q e^T_{1,q+1} \Gamma_q^{-1} \mathbf{v}_q) \\ & \times e^T_{j,p} \Omega^{-1}(u)\, \Omega_{1,2}(u)\, \mathbf{A}^{-1}(u)\, \Omega^T_{1,2}(u)\, \Omega^{-1}(u)\, e_{j,p} \\ & + (e^T_{1,q+1} \Gamma_q^{-1} \boldsymbol{\mu}_q)^2 (\boldsymbol{\mu}_q^T \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} \boldsymbol{\mu}_q - 2\mathbf{v}_q^T \Gamma_q^{-1} \boldsymbol{\mu}_q) \\ & \times e^T_{j,p} \Omega^{-1}(u)\, \Omega_{1,2}(u)\, \mathbf{A}^{-1}(u)\, \Omega^T_{1,2}(u)\, \Omega^{-1}(u)\, \Omega_{1,2}(u) \\ & \times \mathbf{A}^{-1} \Omega^T_{1,2}(u)\, \Omega^{-1}(u)\, e_{j,p} \\ & + v_0 (e^T_{1,q+1} \Gamma_q^{-1} \boldsymbol{\mu}_q)^2 e^T_{j,p} \Omega^{-1}(u)\, \Omega_{1,2}(u)\, \mathbf{A}^{-1}(u)\, \Omega_{2,2}(u) \\ & \times \mathbf{A}^{-1}(u)\, \Omega^T_{1,2} \Omega^{-1}(u)\, e_{j,p}, \end{aligned}$$

with $\mathbf{v}_q = (v_0, \ldots, v_q)^T$.

Now, consider the ideal situation where $a_{p+1}, \ldots, a_{p+m}$ are known. In this situation, the model is simplified to the usual varying coefficient model (1.1). From Fan and Zhang (1999), the coefficient functions can be estimated by their one-step estimator with bandwidth h_o,

$$\hat{a}_{o,j}(u) = e^T_{(j-1)(q+1)+1,\, p(q+1)} (\mathbf{X}^T W_o \mathbf{X})^{-1} \mathbf{X}^T W_o \tilde{Y}, \qquad j = 1, \ldots, p, \tag{2.5}$$

where $\tilde{Y} = (\tilde{Y}_1, \ldots, \tilde{Y}_n)$ with $\tilde{Y}_i = Y_i - \sum_{j=p+1}^{p+m} a_j X_{ij}$. The resulting estimator is called the ideal estimator, whose bias and variance are given by the following theorem.

THEOREM 3. *If* $nh_o \to \infty$ *and* $h_o \to 0$, *then*

$$bias(\hat{a}_{o,j}(u) \mid \mathscr{D}) = \frac{1}{(q+1)!} e^T_{1,q+1} \Gamma_q^{-1} \boldsymbol{\alpha}_q a_j^{(q+1)}(u)\, h_o^{q+1} (1 + o_P(1))$$

and asymptotic conditional variance

$$Var(\hat{a}_{o,j}(u)\mid\mathscr{D})=\frac{\sigma^2(u)}{nh_o f(u)}\,e^T_{1,q+1}\Gamma_q^{-1}\tilde{\Gamma}_q\Gamma_q^{-1}e_{1,q+1}e^T_{j,p}\Omega^{-1}(u)\,e_{j,p}(1+o_P(1)).$$

This is Theorem 1 in Zhang and Lee (2000).

From Theorem 2 and Theorem 3, we find that for $j=1, \ldots, p$, the conditional bias of $\hat{a}_{v,o,j}(u)$ is equal to the conditional bias of the ideal estimator plus another term. Similar phenomenon is also observed for the conditional variance.

Similar to the varying coefficient model, this estimator is not optimal for coefficient functions in the nonparametric part admit different degree of smoothness, see the Example 3 in next section. To improve the performance, an estimation procedure that utilize the "two-step" idea is proposed. For brevity, we will first present our method under the assumption that the coefficient functions admit the same degree of smoothness; and then explain how to extend it to the more complicated cases.

2.3. *A Two-Stage Estimation Procedure for the Nonparametric Part*

In the first stage of the proposed two-stage procedure, the estimation procedure described in subsection 2.1 is used to obtain $\hat{a}_j$ of a_j in the linear part, for $j=p+1, \ldots, p+m$; see equation (2.3).

In the second stage, replacing the unknown constant coefficient a_j in the semivarying coefficient model by its estimator $\hat{a}_j$, we obtain the following model

$$Y-\sum_{j=p+1}^{p+m}\hat{a}_jX_j=\sum_{j=1}^{p}a_j(U)\,X_j+\varepsilon. \tag{2.6}$$

This is a varying coefficient model. When the coefficient functions admit the same degree of smoothness, the final estimator of $a_j(u)$ is given by (2.5) with $\tilde{Y}_i=Y_i-\sum_{k=p+1}^{p+m}\hat{a}_kX_{ik}$ and $h_o=h_3$, namely, the estimator $\hat{a}_{TS,j}(u)$ of $a_j(u)$, for $j=1, \ldots, p$, is

$$\hat{a}_{TS,j}(u)=e^T_{(q+1)(j-1)+1,\,p(q+1)}(\mathbf{X}^TW_{TS}\mathbf{X})^{-1}$$
$$\times\mathbf{X}^TW_{TS}\left(Y_1-\sum_{k=p+1}^{p+m}\hat{a}_kX_{1k}, \ldots, Y_n-\sum_{k=p+1}^{p+m}\hat{a}_kX_{nk}\right)^T,$$

where

$$W_{TS}=\mathrm{diag}(K_{h_3}(U_1-u), \ldots, K_{h_3}(U_n-u)).$$

Let

$$H_{TS} = I_n - \frac{1}{n} \mathbf{T}(\mathbf{0}\ I_m) \sum_{i=1}^{n} (\mathbf{V}_i^T W_i \mathbf{V}_i)^{-1} \mathbf{V}_i^T W_i,$$

where $\mathbf{0}$ is a $m \times p$ matrix with all elements equal to zero. We have

$$a_{TS,j}(u) = e^T_{(j-1)(q+1)+1,\, p(q+1)} (\mathbf{X}^T W_{TS} \mathbf{X})^{-1} \mathbf{X}^T W_{TS} H_{TS} Y, \qquad j = 1, ..., p. \tag{2.7}$$

Thus, this estimator is a linear estimator.

In the implementation of our estimation procedure, the bandwidth h_3 in the second stage can be chosen by some data-driven method, see Zhang and Lee (1998); or by the cross-validation method, see Hoover *et al.* (1998). The asymptotic conditional bias and conditional variance of the two-step estimator is given by the following theorem.

THEOREM 4. *Under the conditions* $nh_1/\log h_1 \to \infty$, $h_3 \to 0$ *and* $h_1 = o(h_3)$, *the conditional bias of* $\hat{a}_{TS,j}$, $j = 1, ..., p$, *is*

$$bias(\hat{a}_{TS,j}(u) \mid \mathscr{D}) = \frac{1}{(q+1)!} e^T_{1,q+1} \Gamma_q^{-1} \boldsymbol{\alpha}_q a_j^{(q+1)}(u)\, h_3^{q+1} (1 + o_P(1)),$$

and the conditional variance is

$$Var(\hat{a}_{TS,j}(u) \mid \mathscr{D}) = \frac{\sigma^2(u)}{nh_3 f(u)} e^T_{1,q+1} \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} e_{1,q+1} e^T_{j,p} \Omega^{-1}(u)\, e_{j,p} (1 + o_P(1)).$$

Under the situation that the coefficient functions admit about the same degree of smoothness, Theorems 4 and 3 indicate that our two-stage estimator $\hat{a}_{TS,j}(u)$ has same asymptotic conditional bias and variance as the ideal estimator. Hence, the performance of our two-stage estimator and the ideal estimator is asymptotically identical.

Our proposed two-stage procedure can be modified to cope with the case where $a_1(\cdot), ..., a_p(\cdot)$ admit different degrees of smoothness. In particular, consider the situation where $a_p(\cdot)$ is smoother than $a_1(\cdot), ..., a_{p-1}(\cdot)$. In the implementation of our two-stage procedure for this situation, the first stage is the same as before. In the second stage, the one-step method as described in Fan and Zhang (1999) is replaced by their two-step method in estimating the model (2.6). Now the ideal estimator for the ideal situation, where $a_{p+1}, ..., a_{p+m}$ are known, can be obtained similarly via Fan and Zhang's two-step method. Using the same technique as in the proof of Theorem 4, we can show that the modified two-step estimator has the same asymptotic conditional bias and variance as the ideal estimator. Hence, the two-stage procedure is better than the one-stage method for the semivarying coefficient model with coefficient functions in the nonparametric part that admit different degrees of smoothness.

From Theorems 2, 3 and 4, we can see that the convergence rates of one-stage, two-stage and ideal estimator are the same under the situation where coefficient functions in the nonparametric part have the same smoothness. Moreover, we will see from Examples 1 and 2 in the next section that the finite sample performances of $\hat{a}_{v,o,j}(u)$, $\hat{a}_{TS,j}(u)$ and ideal estimator are the same. So, if we have the prior information that the coefficient functions in nonparametric part have the same smoothness, the two-stage procedure is not necessary. But in practice, we have not such information, hence the two-stage procedure will be recommended.

3. SIMULATIONS AND EXAMPLE

The following three examples will be used to illustrate the empirical performance of our method.

EXAMPLE 1. $Y = \sin(2\pi U)\, X_1 + \cos(2\pi U)\, X_2 + X_3 + \varepsilon$,

EXAMPLE 2. $$Y = \sin(2\pi U)\, X_1 + (3.5[\exp\{-(4U-1)^2\} + \exp\{-(4U-3)^2\}] - 1.5)\, X_2 + X_3 + \varepsilon,$$

EXAMPLE 3. $Y = \sin(6\pi U)\, X_1 + \sin(2\pi U)\, X_2 + X_3 + \varepsilon$,

where U follows a uniform distribution on $[0, 1]$; X_1, X_2 and X_3 are normally distributed. The correlation coefficients between X_1 and X_2, X_1 and X_3, X_2 and X_3 are $2^{-1/2}$, 2^{-1} and $2^{-1/2}$ respectively. Further, the marginal distribution of X_1, X_2 and X_3 is the standard normal and ε, U and (X_1, X_2, X_3) are independent. The random variable ε follows a normal distribution with mean zero and variance σ^2. For these examples, the σ^2 is chosen so that the noise to signal ratio is about 1:5. Obviously, the functional coefficients in Examples 1 and 2 possess the same degree of smoothness; but the functional coefficients in Example 3 admit different degrees of smoothness.

For each of the above examples, we conducted 100 replications with sample size $n = 250$ and 500. Furthermore, the local linear fit ($q = 1$) for the coefficient functions is considered, and the kernel function is taken to be the Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$. The MISE for estimating the functional coefficient and the mean squared errors (MSE) for estimating the constant coefficient are recorded.

Figure 1 depicts the MSE of the estimator $\hat{a}_j$ of the constant coefficient in the linear part as a function of bandwidth h_1. It indicates that the

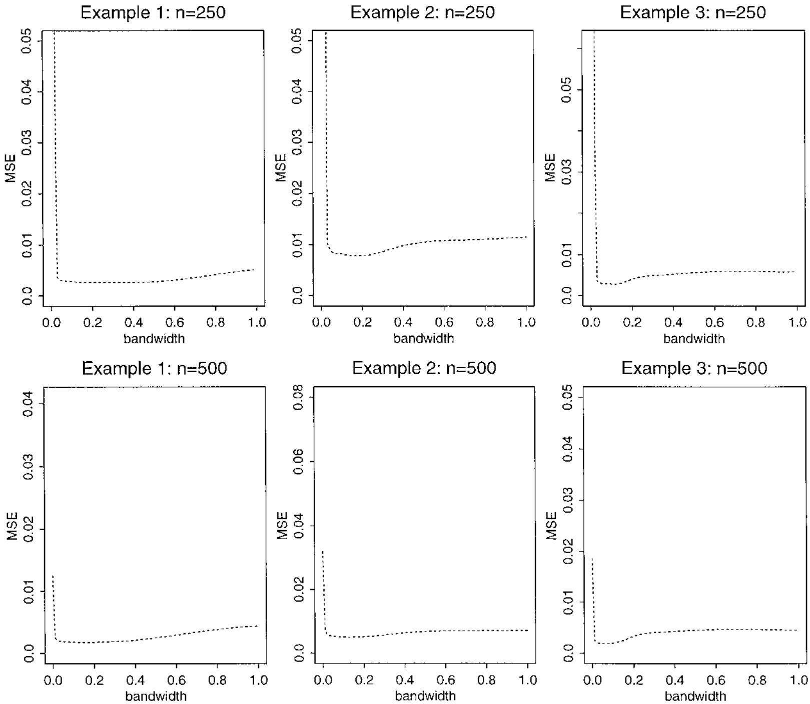


FIG. 1. Relation of MSE and the bandwidth in the linear part estimation.

bandwidth selection procedure for estimating the constant coefficient is less crucial as long as it is not too small. This agrees with Theorem 1.

Based on a sample size $n = 250$, $h_1 = 0.08$, and 100 replications, the mean and standard error of the estimator of the constant coefficient are summarized in Table I. From Table I, it can be seen that our estimators in all the examples are very close to the true values. Since the functional coefficients in Example 3 admit different degrees of smoothness, we may conclude that the estimation of the linear part is not affected by the smoothness of the coefficient functions in the nonparametric part.

TABLE I

The Means and Standard Errors of the Constant Coefficients Estimators

	Example 1	Example 2	Example 3
Mean	1.0013	1.0017	1.0011
Standard errors	0.0540	0.0627	0.0538

To give some idea on the performance of the estimation methods for the nonparametric part, we will demonstrate that the performances of the one-stage, two-stage and the ideal estimators are almost the same in Examples 1 and 2; while the two-stage estimator is better than the one-stage estimator in Example 3.

First, we discuss Examples 1 and 2. For the ideal estimator and the one-stage estimator given in (2.4), the values of the MISE are plotted against bandwidth. For the two-stage procedure, the small initial bandwidth h_1 was taken to 0.08 and 0.05 when $n = 250$ and 500, respectively. The MISE for the two-stage estimator was then computed as a function of h_3 and plotted against h_3. These plots are presented in Figs. 2 and 3. From these figures, we can see that the MISE curves of the one-stage estimator (solid curve), the two-stage estimator (short-dashed curve) and the ideal estimator (long-dashed curve) are almost coincident. This indicates that the empirical performances of these estimators are almost identical. It is in line with our asymptotic results in Theorems 2, 3 and 4.

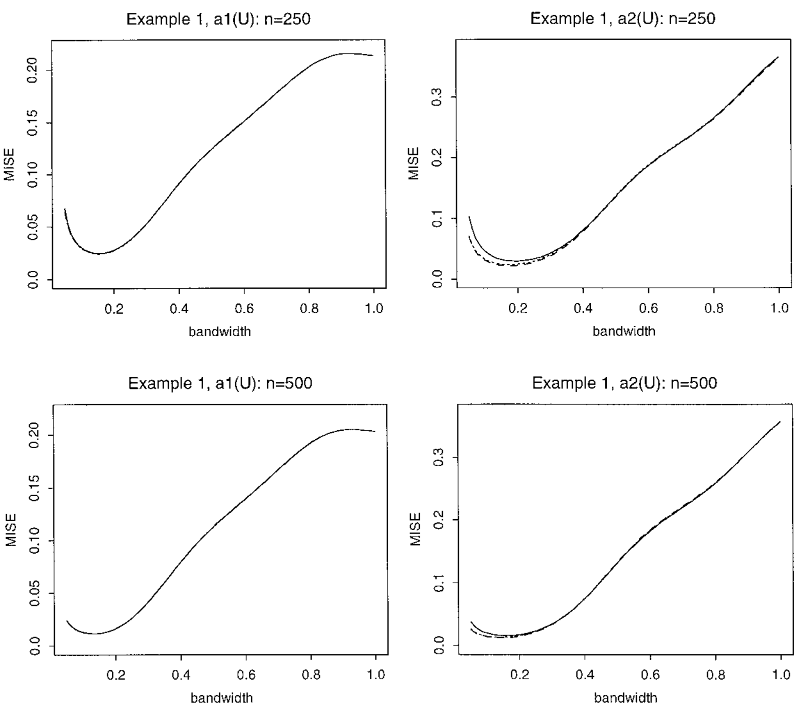


FIG. 2. Comparisons of the performance between the one-step (solid-curve), the two-step (short-dashed curve) and the ideal estimators (long-dashed curve).

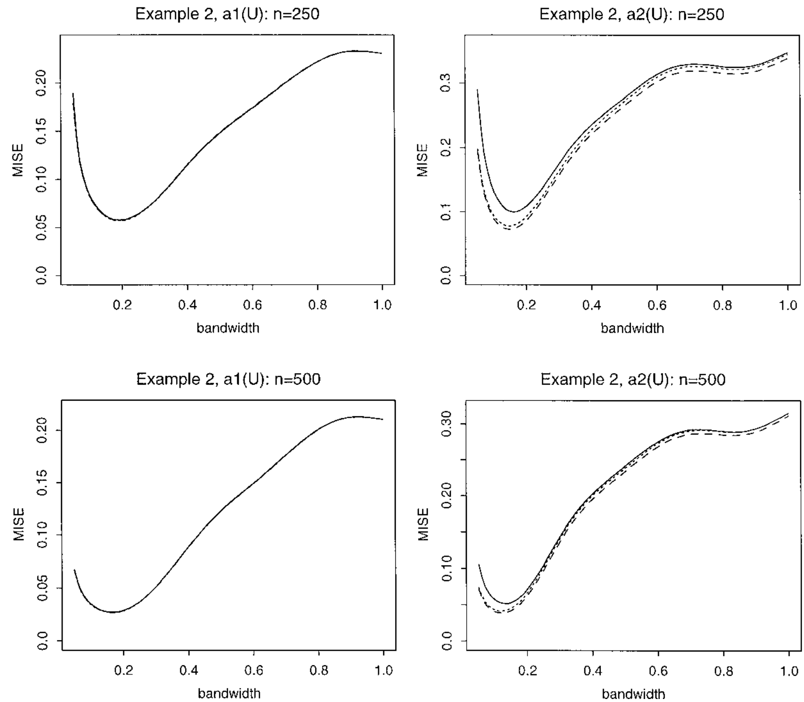


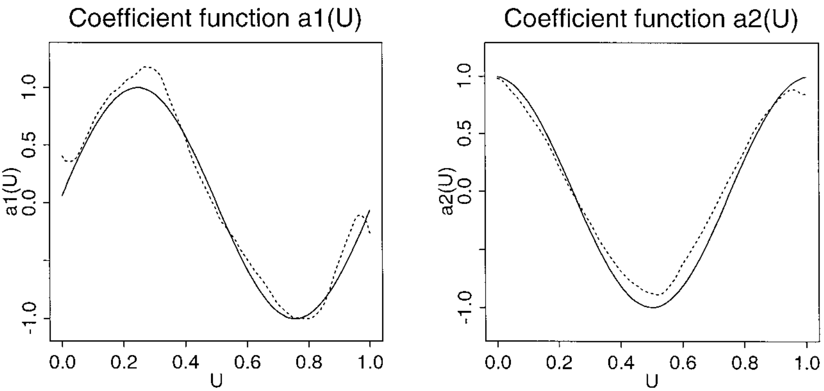
FIG. 3. Comparisons of the performance between the one-step (solid curve), the two-step (short-dashed curve) and the ideal estimators (long-dashed curve).

To provide more information on the performance of the two-stage estimator, we select among the 100 replications the one that the estimator has the median performance, in the combination with $h_1 = 0.05$ and $n = 500$. Figure 4 depicts the true functions and the estimated functions. It can be seen that the estimated functions are quite accurate.

In Example 3, we only discuss the estimation of $a_2(u) = \sin(2\pi u)$, which fluctuates less than $a_1(u) = \sin(6\pi u)$. The modified two-stage procedure as discussed at the end of subsection 2.3 was used to obtain the two-stage estimator. For completeness, the two stages of the whole procedure are described as below: (1) Replace the constant coefficient by its estimator given in (2.3) with bandwidth h_1, and obtain the usual varying coefficient model (2.6). (2) Apply Fan and Zhang's two-step method to (2.6). Denote the bandwidth in the first step of their two-step method by h_2 and the second step h_3.

The bandwidths $\{h_1, h_2\}$ for $n = 250$ and 500 were taken to be $\{0.08, 0.06\}$ and $\{0.05, 0.04\}$, respectively. Then we computed the MISE for the two-step estimators as a function of h_3. The first two graphs in

Example 1



Example 2

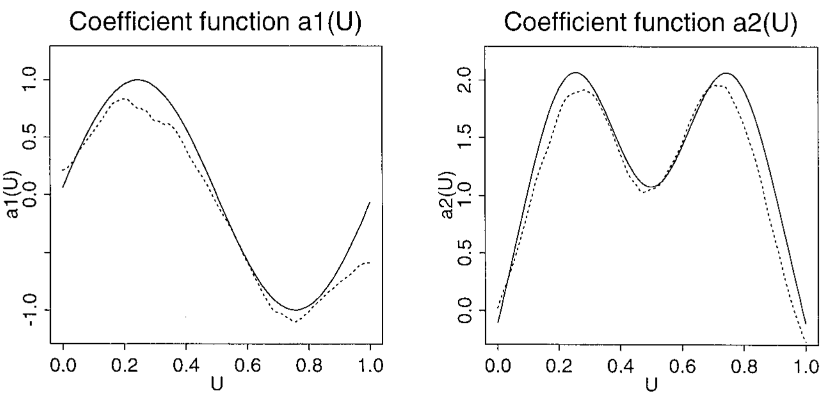


FIG. 4. The solid curve is the true coefficient function, the dashed curve is its two-step estimator.

Fig. 5 report the results for both the one-stage and the two-stage estimators. They indicate that the improvement of the two-stage estimator is quite substantial. Further, for the two-stage estimator, the MISE curve is flatter than that for the one-stage method. This in turn suggests that the bandwidth for the two-stage estimator is less crucial than that for the one-stage procedure. For the combination with $h_1 = 0.05$, $h_2 = 0.04$ and $n = 500$, we select the replication that the two-step estimator has the median performance for further illustration. The third graph in Fig. 5

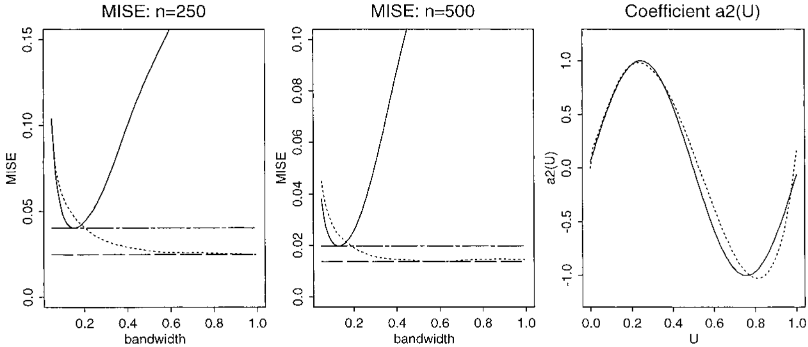


FIG. 5. The first two graphs give comparisons of the performance between the one-step estimator (solid curve) and the two-step estimator (short-dashed curve). The third graph depicts the two-step estimator (dashed curve) and the true function (solid curve).

depicts the true function (solid curve) and the estimated function. It seems that the estimated function is again quite accurate.

We now illustrate the methodology via an application to an environmental data set. This data consists of a collection of daily measurements of chemical pollutants and other environmental factors in Hong Kong between January 1, 1994 and December 31, 1995; see Fan and Zhang (1999). Our main interests are to study the association between the levels of chemical pollutants and the number of daily total hospital admissions for circulation and respiration, and to examine the extent to which the association varies with time. We consider the relation among the number of daily hospital admission (Y) and the level of pollutant Sulphur Dioxide X_2 (in $\mu g/m^3$), the level of pollutant Nitrogen Dioxide X_3 (in $\mu g/m^3$), and the level of dust X_4 (in $\mu g/m^3$).

We took $X_1 = 1$ for the intercept term, and $U = t$. To avoid the extreme effect by boundary, we used the data from Feb. 20, 1994 to Nov. 10, 1995. On the basis of our main objective for examining the association between the levels of chemical pollutants, and the analysis given in Fan and Zhang (2000) that the hypothesis "$a_4(\cdot) =$ a constant" is not significant, we analyzed the data with coefficient functions $a_2(t)$ and $a_3(t)$ but treating the coefficient corresponding to the level of dust $a_4(\cdot)$ as a constant. So, the model

$$Y = a_1(t) + a_2(t)\, X_2 + a_3(t)\, X_3 + a_4 X_4 + \varepsilon$$

was used. As we argued before, estimators obtained from the proposed semivarying coefficient model are more accurate because their variances are reduced. The estimator of the constant coefficient is 0.136. For the nonparametric part, the two-step method was used to estimate the coefficient

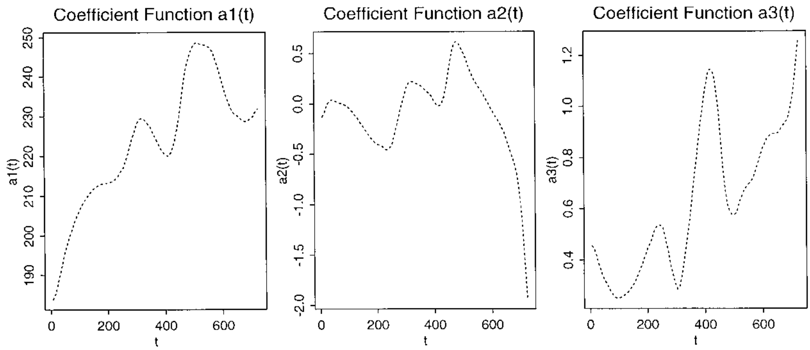


FIG. 6. The estimated coefficient functions.

function $a_j(\cdot)$, $j = 1, 2, 3$. The bandwidth $h_1 = 0.03 * 629$ (three percents of the whole interval) was chosen. In the second step, the bandwidth h_3 was equal to 30% of the interval length. Figure 6 depicts the estimated coefficient functions. It indicates that there is some strong time effect, and the time trend $a_1(t)$ is increasing instead of seasonal.

4. FINAL REMARKS

The varying and the semivarying coefficient models have close relationship with each other. Based on our motivations mentioned before, it is advantageous to use the semivarying coefficient model for situations where we have a clear model of interest in mind; that is, we have the prior intention or knowledge on which independent variable X_j is associated with a coefficient function and which is associated with a constant coefficient. For exploratory analysis, we need to decide which coefficients are functions and which are constants. This is essentially a model selection problem. One procedure to handle this problem is sketched as follows. First, we use one-step method for the varying coefficient model (see, Fan and Zhang, 1999) to obtain initial estimates $\hat{a}_{1,j}(\cdot)$ of the coefficient functions. Then, based on $\{(U_i, \hat{a}_{1,j}(U_i)), i = 1, \ldots, n\}$, we decide whether $a_j(\cdot)$ has a constant trend via some univariate model selection procedures (see, Linhart and Zucchini, 1986). Clearly, more theoretical and computational results on the model selection problem are needed to be developed in the future.

APPENDIX

The following technical conditions will be assumed to develop the asymptotic results:

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

(2) $a_j^{(q+1)}(\cdot)$ is continuous in a neighborhood of u, for $j = 1, ..., p-1$. Further, assume $a_j^{(q+1)}(u) \neq 0$, for $j = 1, ..., p-1$.

(3) The function $a_p(\cdot)$ has a continuous $(q+3)$th derivative in a neighborhood of u.

(4) The marginal density $f(u)$ of U has a continuous derivative and $f(u) \neq 0$. In addition, $f(u)$ has a compact support.

(5) For any i, j, $r_{ij}(\cdot)$ is continuous in a neighborhood of u.

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

To prove Theorem 1–Theorem 4, the following lemma which follows immediately from a result in Mack and Silverman (1982), is required:

LEMMA 1. *Let $(X_1, Y_1), ..., (X_n, Y_n)$ be i.i.d random vectors, where the Y_i's are scalar random variables. Assume further that $E|y|^s < \infty$ and $\sup_x \int |y|^s f(x, y)\, dy < \infty$, where f denotes the joint density of (X, Y). Let K be a bounded positive function with a bounded support and satisfying a Lipschitz condition. Then*

$$\sup_{x \in D} \left| \frac{1}{n} \sum_{i=1}^{n} [K_h(X_i - x)\, Y_i - E\{K_h(X_i - x)\, Y_i\}] \right| = O_P\left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} \right),$$

provided that $n^{2\varepsilon - 1} h \to \infty$ for some $\varepsilon < 1 - s^{-1}$.

Proof of Theorem 1. First, we compute the bias of $\hat{a}_j$, $j = p+1, ..., p+m$ under the condition $nh_1 / \log h_1 \to \infty$, we have

$$\begin{aligned}
\operatorname{bias}(\hat{a}_j \mid \mathscr{D}) &= E\{e^T_{pq+j,\ell}(\mathbf{V}_1^T W_1 \mathbf{V}_1)^{-1}\, \mathbf{V}_1^T W_1 Y \mid \mathscr{D}\} - a_j \\
&= \frac{h_1^{q+1}}{(q+1)!}\, e^T_{pq+j,\ell} \begin{pmatrix} \Omega(U) \otimes \Gamma_q & \Omega_{1,2}(U) \otimes \boldsymbol{\mu}_q \\ \Omega^T_{1,2}(U) \otimes \boldsymbol{\mu}_q^T & \Omega_{2,2}(U) \end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix} \Omega(U) \otimes \boldsymbol{\alpha}_q \\ \mu_{q+1} \Omega^T_{1,2}(U) \end{pmatrix} \begin{pmatrix} a_1^{(q+1)}(U) \\ \vdots \\ a_p^{(q+1)}(U) \end{pmatrix} (1 + o_P(1)) \\
&= \frac{h_1^{q+1}(\mu_{q+1} - \boldsymbol{\mu}_q^T \Gamma_q^{-1} \boldsymbol{\alpha}_q)}{(q+1)!}\, e^T_{j-p,m} \mathbf{A}^{-1}(U)\, \Omega^T_{1,2}(U) \begin{pmatrix} a_1^{(q+1)}(U) \\ \vdots \\ a_p^{(q+1)}(U) \end{pmatrix} (1 + o_P(1)).
\end{aligned}$$

The conditional variance of $\hat{a}_j$ is

$$\operatorname{Var}(\hat{a}_j \mid \mathscr{D}) = \frac{1}{n^2}(1, \ldots, 1)\begin{pmatrix} e^T_{pq+j,\ell}(\mathbf{V}_1^T W_1 \mathbf{V}_1)^{-1}\,\mathbf{V}_1^T W_1 \\ \vdots \\ e^T_{pq+j,\ell}(\mathbf{V}_n^T W_n \mathbf{V}_n)^{-1}\,\mathbf{V}_n^T W_n \end{pmatrix}\Psi$$

$$\times \begin{pmatrix} e^T_{pq+j,\ell}(\mathbf{V}_1^T W_1 \mathbf{V}_1)^{-1}\,\mathbf{V}_1^T W_1 \\ \vdots \\ e^T_{pq+j,\ell}(\mathbf{V}_n^T W_n \mathbf{V}_n)^{-1}\,\mathbf{V}_n^T W_n \end{pmatrix}^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} e^T_{pq+j,\ell}(\mathbf{V}_i^T W_i \mathbf{V}_i)^{-1}\,\mathbf{V}_i^T W_i \Psi W_j \mathbf{V}_j (\mathbf{V}_j^T W_j \mathbf{V}_j)^{-1}\, e_{pq+j,\ell}$$

$$= \frac{1}{n}E\left\{ e^T_{pq+j,\ell}\begin{pmatrix} \Omega(U)\otimes\Gamma_q & \Omega_{1,2}(U)\otimes\boldsymbol{\mu}_q \\ \Omega^T_{1,2}(U)\otimes\boldsymbol{\mu}_q^T & \Omega_{2,2}(U) \end{pmatrix}^{-1}\right.$$

$$\times \begin{pmatrix} \Omega(U)\otimes\boldsymbol{\mu}_q\boldsymbol{\mu}_q^T & \Omega_{1,2}(U)\otimes\boldsymbol{\mu}_q \\ \Omega^T_{1,2}(U)\otimes\boldsymbol{\mu}_q^T & \Omega_{2,2}(U) \end{pmatrix}$$

$$\left.\times \begin{pmatrix} \Omega(U)\otimes\Gamma_q & \Omega_{1,2}(U)\otimes\boldsymbol{\mu}_q \\ \Omega^T_{1,2}(U)\otimes\boldsymbol{\mu}_q^T & \Omega_{2,2}(U) \end{pmatrix}^{-1} e_{pq+j,\ell}\sigma^2(U)\right\}(1+o_P(1))$$

$$= \frac{1}{n}E\left\{ e^T_{pq+j,\ell}\begin{pmatrix} \Omega(U)\otimes\Gamma_q & \Omega_{1,2}(U)\otimes\boldsymbol{\mu}_q \\ \Omega^T_{1,2}(U)\otimes\boldsymbol{\mu}_q^T & \Omega_{2,2}(U) \end{pmatrix}^{-1} e_{pq+j,\ell}\sigma^2(U)\right.$$

$$+ e^T_{pq+j,\ell}\begin{pmatrix} \Omega(U)\otimes\Gamma_q & \Omega_{1,2}(U)\otimes\boldsymbol{\mu}_q \\ \Omega^T_{1,2}(U)\otimes\boldsymbol{\mu}_q^T & \Omega_{2,2}(U) \end{pmatrix}^{-1}\begin{pmatrix} \Omega(U)\otimes(\boldsymbol{\mu}_q\boldsymbol{\mu}_q^T-\Gamma_q) & 0 \\ 0 & 0 \end{pmatrix}$$

$$\left.\times \begin{pmatrix} \Omega(U)\otimes\Gamma_q & \Omega_{1,2}(U)\otimes\boldsymbol{\mu}_q \\ \Omega^T_{1,2}(U)\otimes\boldsymbol{\mu}_q^T & \Omega_{2,2}(U) \end{pmatrix}^{-1} e_{pq+j,\ell}\sigma^2(U)\right\}(1+o_P(1))$$

$$= \frac{1}{n}E\{e^T_{j-p,m}\mathbf{A}^{-1}(U)[\,\Omega_{2,2} + \{(\boldsymbol{\mu}_q^T\Gamma_q^{-1}\boldsymbol{\mu}_q)^2 - 2(\boldsymbol{\mu}_q^T\Gamma_q^{-1}\boldsymbol{\mu}_q)\}$$

$$\times\,\Omega^T_{1,2}(U)\,\Omega^{-1}(U)\,\Omega_{1,2}(U)]\times\mathbf{A}^{-1}(U)\,e_{j-p,m}\sigma^2(U)\}(1+o_P(1)).$$

Proof of Theorem 2. It follows from standard argument and calculation.

Proof of Theorem 4. First, we compute the conditional bias of $\hat{a}_{TS,j}(u)$, $j = 1, \ldots, p$, as below.

$$
\begin{aligned}
&\operatorname{bias}(\hat{a}_{TS,j}(u) \mid \mathscr{D}) \\
&\quad = e^T_{(j-1)(q+1)+1,\, p(q+1)} (\mathbf{X}^T W_t \mathbf{X})^{-1} \mathbf{X}^T W_t \\
&\qquad \times \left(\sum_{k=1}^{p} a_k(U_1)\, X_{1k}, \ldots, \sum_{k=1}^{p} a_k(U_n)\, X_{nk} \right)^T - a_j(u) \\
&\qquad - \sum_{k=p+1}^{p+m} \operatorname{bias}(\hat{a}_k \mid \mathscr{D})\, e^T_{(j-1)(q+1)+1,\, p(q+1)} (\mathbf{X}^T W_t \mathbf{X})^{-1} \mathbf{X}^T W_t (X_{1k}, \ldots, X_{nk})^T \\
&\quad = \frac{1}{(q+1)!} e^T_{1,q+1} \Gamma_q^{-1} \boldsymbol{\alpha}_q a_j^{(q+1)}(u)\, h_3^{q+1} (1 + o_P(1)) \\
&\qquad - e^T_{1,q+1} \Gamma_q^{-1} \boldsymbol{\mu}_q e^T_{j,p} \Omega^{-1}(u)\, \Omega_{1,2}(u) \\
&\qquad \times (\operatorname{bias}(\hat{a}_{p+1} \mid \mathscr{D}), \ldots, \operatorname{bias}(\hat{a}_{p+m} \mid \mathscr{D}))^T (1 + o_P(1)) \\
&\quad = \frac{1}{(q+1)!} e^T_{1,q+1} \Gamma_q^{-1} \boldsymbol{\alpha}_q a_j^{(q+1)}(u)\, h_3^{q+1} (1 + o_P(1)) \\
&\qquad - \frac{h_1^{q+1} (\mu_{q+1} - \boldsymbol{\mu}_q^T \Gamma_q^{-1} \boldsymbol{\alpha}_q)}{(q+1)!} e^T_{1,q+1} \Gamma_q^{-1} \boldsymbol{\mu}_q \\
&\qquad \times e^T_{j,p} \Omega^{-1}(u)\, \Omega_{1,2}(u)\, \mathbf{A}^{-1}(U) \\
&\qquad \times \Omega^T_{1,2}(U) (a_1^{(q+1)}(U), \ldots, a_p^{(q+1)}(U))^T (1 + o_P(1)) \\
&\quad = \frac{1}{(q+1)!} e^T_{1,q+1} \Gamma_q^{-1} \boldsymbol{\alpha}_q a_j^{(q+1)}(u)\, h_3^{q+1} (1 + o_P(1)).
\end{aligned}
$$

In the following, we will discuss the conditional variance of $\hat{a}_{TS,j}(u)$. From (2.7), we have

$$
\begin{aligned}
&\operatorname{Var}(\hat{a}_{TS,j}(u) \mid \mathscr{D}) \\
&= e^T_{(j-1)(q+1)+1,\, p(q+1)} (\mathbf{X}^T W_t \mathbf{X})^{-1} \mathbf{X}^T W_t H_t \Psi H_t^T W_t \mathbf{X} (\mathbf{X}^T W_t \mathbf{X})^{-1} e_{(j-1)(q+1)+1,\, p(q+1)} \\
&= e^T_{(j-1)(q+1)+1,\, p(q+1)} (\mathbf{X}^T W_t \mathbf{X})^{-1} \mathbf{X}^T W_t \Psi W_t \mathbf{X} (\mathbf{X}^T W_t \mathbf{X})^{-1} e_{(j-1)(q+1)+1,\, p(q+1)} \\
&\quad - \frac{2}{n} e^T_{(j-1)(q+1)+1,\, p(q+1)} (\mathbf{X}^T W_t \mathbf{X})^{-1} \mathbf{X}^T W_t \mathbf{T} (\mathbf{0}\ I_m) \\
&\quad \times \sum_{i=1}^{n} (\mathbf{V}_i^T W_i \mathbf{V}_i)^{-1} \mathbf{V}_i^T W_i \Psi W_t \mathbf{X} (\mathbf{X}^T W_t \mathbf{X})^{-1} e_{(j-1)(q+1)+1,\, p(q+1)}
\end{aligned}
$$

$$+\frac{1}{n^2} e^T_{(j-1)(q+1)+1,\, p(q+1)} (\mathbf{X}^T W_t \mathbf{X})^{-1} \mathbf{X}^T W_t \mathbf{T}(\mathbf{0}\ I_m)$$

$$\times \sum_{i=1}^{n} (\mathbf{V}_i^T W_i \mathbf{V}_i)^{-1} \mathbf{V}_i^T W_i \Psi \sum_{i=1}^{n} W_i \mathbf{V}_i (\mathbf{V}_i^T W_i \mathbf{V}_i)^{-1} (\mathbf{0}\ I_m)^T \mathbf{T}^T W_t \mathbf{X}$$

$$\times (\mathbf{X}^T W_t \mathbf{X})^{-1} e_{(j-1)(q+1)+1,\, p(q+1)}$$

$$\triangleq \mathbf{J}_1 - 2\mathbf{J}_2 + \mathbf{J}_3.$$

Obviously,

$$\mathbf{J}_1 = \frac{\sigma^2(u)}{nh_3 f(u)} e^T_{1,q+1} \Gamma_q^{-1} \tilde{\Gamma}_q \Gamma_q^{-1} e_{1,q+1} e^T_{j,p} \Omega^{-1}(u)\, e_{j,p}(1 + o_P(1)).$$

Using Lemma 1, and by some calculation, it follows that under the conditions $nh_1/\log h_1 \to \infty$ and $h_1 = o(h_3)$,

$$\mathbf{J}_2 = \frac{(e^T_{1,q+1}\Gamma_q^{-1}\boldsymbol{\mu}_q)^2\, \sigma^2(u)}{n} e^T_{j,p} \Omega^{-1}(u)\, \Omega_{1,2}(u)(\mathbf{0}\ I_m)$$

$$\times \begin{pmatrix} \Omega(u) \otimes \Gamma_q & \Omega_{1,2}(u) \otimes \boldsymbol{\mu}_q \\ \Omega^T_{1,2}(u) \otimes \boldsymbol{\mu}_q^T & \Omega_{2,2}(u) \end{pmatrix}^{-1} \begin{pmatrix} e_{j,p} \otimes \boldsymbol{\mu}_q \\ \Omega^T_{1,2}(u)\, \Omega^{-1}(u)\, e_{j,p} \end{pmatrix} (1 + o_P(1))$$

$$= \frac{\sigma^2(u)}{n} (e^T_{1,q+1}\Gamma_q^{-1}\boldsymbol{\mu}_q)^2 (1 - \boldsymbol{\mu}_q^T \Gamma^{-1} \boldsymbol{\mu}_q)$$

$$\times e^T_{j,p} \Omega^{-1}(u)\, \Omega_{1,2}(u)\, \mathbf{A}^{-1}(u)\, \Omega^T_{1,2}(u)\, \Omega^{-1}(u)\, e_{j,p}(1 + o_P(1)).$$

Using Lemma 1, and the technique in the proof of Theorem 2 in Fan and Zhang (1999), we have

$$(\mathbf{0}\ I_m) \sum_{i=1}^{n} (\mathbf{V}_i^T W_i \mathbf{V}_i)^{-1} \mathbf{V}_i^T W_i \Psi \sum_{i=1}^{n} W_i \mathbf{V}_i (\mathbf{V}_i^T W_i \mathbf{V}_i)^{-1} (\mathbf{0}\ I_m)^T$$

$$= (\mathbf{0}\ I_m) \sum_{k=1}^{n} \sigma^2(U_k) \begin{pmatrix} \Omega_k \otimes \Gamma_q & \Omega_{1,2,k} \otimes \boldsymbol{\mu}_q \\ \Omega^T_{1,2,k} \otimes \boldsymbol{\mu}_q^T & \Omega_{2,2,k} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} (X_{k1}, \ldots, X_{kp})^T \otimes \boldsymbol{\mu}_q \\ (X_{k(p+1)}, \ldots, X_{k(p+m)})^T \end{pmatrix}$$

$$\times ((X_{k1}, \ldots, X_{kp}) \otimes \boldsymbol{\mu}_q^T, (X_{k(p+1)}, \ldots, X_{k(p+m)}))$$

$$\times \begin{pmatrix} \Omega_k \otimes \Gamma_q & \Omega_{1,2,k} \otimes \boldsymbol{\mu}_q \\ \Omega^T_{1,2,k} \otimes \boldsymbol{\mu}_q^T & \Omega_{2,2,k} \end{pmatrix}^{-1}$$

$$
\begin{aligned}
&(\mathbf{0}\ I_m)^T \times (1+o_P(1)) \\
&\quad = nE\left\{(\mathbf{0}\ I_m)\begin{pmatrix} \Omega(U)\otimes\Gamma_q & \Omega_{1,2}(U)\otimes\boldsymbol{\mu}_q \\ \Omega_{1,2}^T(U)\otimes\boldsymbol{\mu}_q^T & \Omega_{2,2}(U)\end{pmatrix}^{-1}\right. \\
&\qquad \times\begin{pmatrix} \Omega(U)\otimes\boldsymbol{\mu}_q\boldsymbol{\mu}_q^T & \Omega_{1,2}(U)\otimes\boldsymbol{\mu}_q \\ \Omega_{1,2}^T(U)\otimes\boldsymbol{\mu}_q^T & \Omega_{2,2}(U)\end{pmatrix} \\
&\qquad \times\left.\begin{pmatrix} \Omega(U)\otimes\Gamma_q & \Omega_{1,2}(U)\otimes\boldsymbol{\mu}_q \\ \Omega_{1,2}^T(U)\otimes\boldsymbol{\mu}_q^T & \Omega_{2,2}(U)\end{pmatrix}^{-1}(\mathbf{0}\ I_m)^T\sigma^2(U)\right\}(1+o_P(1)) \\
&\quad = nE([\{(\boldsymbol{\mu}_q^T\Gamma_q^{-1}\boldsymbol{\mu}_q)^2-2(\boldsymbol{\mu}_q^T\Gamma_q^{-1}\boldsymbol{\mu}_q)\}\,\mathbf{A}^{-1}(U)\,\Omega_{1,2}^T(U)\,\Omega^{-1}(U)\,\Omega_{1,2}(U) \\
&\qquad \times\mathbf{A}^{-1}(U)+\mathbf{A}^{-1}(U)\,\Omega_{2,2}(U)\,\mathbf{A}^{-1}]\,\sigma^2(U))(1+o_P(1)),
\end{aligned}
$$

where $\Omega_{1,2,k}=\Omega_{1,2}(U_k)$, $\Omega_{2,2,k}=\Omega_{2,2}(U_k)$. So, we have

$$
\begin{aligned}
\mathbf{J}_3 &= \frac{(e_{1,q+1}^T\Gamma^{-1}\boldsymbol{\mu}_q)^2}{n}\,e_{j,p}^T\Omega^{-1}(u)\,\Omega_{1,2}(u) \\
&\quad\times E([\{(\boldsymbol{\mu}_q^T\Gamma_q^{-1}\boldsymbol{\mu}_q)^2-2(\boldsymbol{\mu}_q^T\Gamma_q^{-1}\boldsymbol{\mu}_q)\}\,\mathbf{A}^{-1}(U)\,\Omega_{1,2}^T(U)\,\Omega^{-1}(U)\,\Omega_{1,2}(U)\,\mathbf{A}^{-1}(U) \\
&\quad+\mathbf{A}^{-1}(U)\,\Omega_{2,2}(U)\,\mathbf{A}^{-1}]\,\sigma^2(U))\,\Omega_{1,2}^T(u)\,\Omega^{-1}(u)\,e_{j,p}(1+o_P(1)).
\end{aligned}
$$

Hence, the conditional variance of $\hat{a}_{TS,j}(u)$ is given by

$$
\begin{aligned}
\mathrm{Var}(\hat{a}_{TS,j}(u)\mid\mathscr{D}) &= \frac{\sigma^2(u)}{nh_3 f(u)}\,e_{1,q+1}^T\Gamma_q^{-1}\tilde{\Gamma}_q\Gamma_q^{-1}e_{1,q+1}e_{j,p}^T\Omega^{-1}(u)\,e_{j,p}(1+o_P(1)) \\
&\quad-\frac{2\sigma^2(u)}{n}(e_{1,q+1}^T\Gamma_q^{-1}\boldsymbol{\mu}_q)^2(1-\boldsymbol{\mu}_q^T\Gamma^{-1}\boldsymbol{\mu}_q) \\
&\quad\times e_{j,p}^T\Omega^{-1}(u)\,\Omega_{1,2}(u)\,\mathbf{A}^{-1}(u)\,\Omega_{1,2}^T(u)\,\Omega^{-1}(u)\,e_{j,p}(1+o_P(1)) \\
&\quad+\frac{(e_{1,q+1}^T\Gamma^{-1}\boldsymbol{\mu}_q)^2}{n}\,e_{j,p}^T\Omega^{-1}(u)\,\Omega_{1,2}(u)\,E([\{(\boldsymbol{\mu}_q^T\Gamma_q^{-1}\boldsymbol{\mu}_q)^2-2(\boldsymbol{\mu}_q^T\Gamma_q^{-1}\boldsymbol{\mu}_q)\} \\
&\quad\times\mathbf{A}^{-1}(U)\,\Omega_{1,2}^T(U)\,\Omega^{-1}(U)\,\Omega_{1,2}(U)\,\mathbf{A}^{-1}(U) \\
&\quad+\mathbf{A}^{-1}(U)\,\Omega_{2,2}(U)\,\mathbf{A}^{-1}]\,\sigma^2(U))\,\Omega_{1,2}^T(u)\,\Omega^{-1}(u)\,e_{j,p}(1+o_P(1)) \\
&= \frac{\sigma^2(u)}{nh_3 f(u)}\,e_{1,q+1}^T\Gamma_q^{-1}\tilde{\Gamma}_q\Gamma_q^{-1}e_{1,q+1}e_{j,p}^T\Omega^{-1}(u)\,e_{j,p}(1+o_P(1)).
\end{aligned}
$$

ACKNOWLEDGMENTS

The authors are grateful to the editors and two referees for their helpful comments. The assistance of Esther Tam in preparing the manuscript is appreciated.

REFERENCES

L. Breiman and J. H. Friedman, Estimating optimal transformations for multiple regression and correlation (with discussion), *J. Amer. Stat. Assoc.* **80** (1985), 580–619.

Z. Cai, J. Fan, and Q. Yao, Functional-coefficient regression models for nonlinear time series, *J. Amer. Statis. Assoc.* **98** (2000), 941–956.

R. J. Carroll, J. Fan, I. Gijbels, and M. P. Wand, Generalized partially linear single-index models, *J. Amer. Statist. Assoc.* **92** (1997), 477–489.

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.

W. S. Cleveland, E. Grosse, and W. M. Shyu, Local regression models, *in* "Statistical Models in S" (J. M. Chambers and T. J. Hastie, Eds.), pp. 309–376, Wadsworth & Brooks, Pacific Grove, 1991.

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

J. Fan and I. Gijbel, 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 and Hall, London, 1996.

J. Fan, Q. Yao and Z. Cai, Varying-coefficient linear models with unknown indices, manuscript submitted for publication, 2000.

J. Fan and J. Zhang, Comments on "Smoothing spline models for the analysis of nested and crossed samples of curves," *J. Amer. Statist. Assoc.* **93** (1998), 980–984.

J. Fan and W. Zhang, Statistical estimation in varying coefficient models, *Ann. Statist.* **27** (1999), 1491–1518.

J. Fan and W. Zhang, Simultaneous confidence bands are and hypothesis testing in varying-coefficient models, *Scand. J. Statist.* **27** (2000), 715–731.

J. H. Friedman, Multivariate adaptive regression splines (with discussion), *Ann. Statist.* **19** (1991), 1–141.

P. J. Green and B. W. Silverman, "Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach," Chapman and Hall, London, 1994.

C. Gu and G. Wahba, Smoothing spline ANOVA with component-wise Bayesian "confidence intervals," *J. Comput. Graph. Statist.* **2** (1993), 97–117.

W. Härdle and T. M. Stoker, Investigating smooth multiple regression by the method of average derivatives, *J. Amer. Statist. Assoc.* **84** (1989), 986–995.

T. J. Hastie and R. Tibshirani, "Generalized Additive Models," Chapman and Hall, London, 1990.

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* **85** (1998), 809–822.

H. Linhart and W. Zucchini, "Model Selection," Wiley, New York, 1986.

Y. P. Mack and B. W. Silverman, Weak and strong uniform consistency of kernel regression estimates, *Z. Wahr. Gebiete* **61** (1982), 405–415.

D. Ruppert, Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation, *J. Amer. Statist. Assoc.* **92** (1997), 1049–1062.

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.

C. J. Stone, M. Hansen, C. Kooperberg, and Y. K. Truong, Polynomial splines and their tensor products in extended linear modeling, *Ann. Statist.* **25** (1997), 1371–1470.

M. Stone, Cross-validatory choice and assessment of statistical predictions (with discussion), *J. Roy. Statist. Soc. Ser. B.* **36** (1974), 111–147.

G. Wahba, Partial spline models for semiparametric estimation of functions of several variables, *in* "Statistical Analysis of Time Series, Proceedings of the Japan U.S. Joint Seminar, Tokyo," pp. 319–329, Institute of Statistical Mathematics, Tokyo, 1984.

G. Wahba, "Spline Models for Observing Data," Society for Industrial and Applied Mathematics, Philadelphia, 1990.

M. P. Wand and M. C. Jones, "Kernel Smoothing," Chapman and Hall, London, 1995.

M. West and P. J. Harrison, "Bayesian Forecasting and Dynamic Models," Springer-Verlag, Berlin, 1989.

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.

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

W. Zhang and S. Y. Lee, Variable bandwidth selection in varying coefficient models, *J. Multivariate Anal.* **74** (2000), 116–134.