Ann. Inst. Statist. Math.
Vol. 57, No. 4, 637–653 (2005)

COMPARISONS BETWEEN SIMULTANEOUS AND COMPONENTWISE SPLINES FOR VARYING COEFFICIENT MODELS

CHIN-TSANG CHIANG

Department of Mathematics, National Taiwan University, Taipei 106, Taiwan, R.O.C.

(Received September 22, 2003; revised August 3, 2004)

Abstract. In this paper, we study the properties of the simultaneous and componentwise splines for the varying coefficient model with repeatedly measured (longitudinal) dependent variable and time invariant covariates. The proposed simultaneous smoothing spline estimators are mainly obtained from the penalized least squares with adjustment for the variations of covariates in the penalized terms. We do this mainly to avoid the penalized terms being influenced by the scales of the covariates and the random smoothing parameters appearing in the estimators, which complicates the derivation of the asymptotic properties of the estimators. It is shown in this study that our estimators have smaller variances than the componentwise ones. Through a Monte Carlo simulation and two empirical examples, the simultaneous smoothing splines are all found to be more accurate in the variances.

Key words and phrases: Componentwise smoothing splines, longitudinal data, mean squared error, penalized least squares, simultaneous smoothing splines, smoothing parameters, varying coefficient model.

1. Introduction

In biomedical and epidemiological studies, longitudinal data with time invariant covariates and repeatedly measured (longitudinal) dependent variable over time are frequently encountered. Generally speaking, this type of data is collected from n randomly selected subjects. For the i-th subject, let m_i, t_{ij}, Y_{ij}, and $\boldsymbol{X}_i^T = (X_{i0}, \ldots, X_{ik})$ with $X_{i0} = 1$, respectively denote the number of the repeated measurements, the time of the j-th measurement, the observed outcome at time t_{ij}, and the observed time invariant covariate vector. Here, the total number of observations is denoted by $N = \sum_{i=1}^n m_i$.

To model the relationship between the dependent variable $Y(t)$ and the time dependent or time invariant covariates $\boldsymbol{X}^T(t) = (X_0(t), \ldots, X_k(t))$ with $X_0(t) = 1$, Hoover *et al.* (1998) considered a more flexible varying coefficient model of Hastie and Tibshirani (1993)

$$Y(t) = \boldsymbol{X}^T(t)\boldsymbol{\beta}(t) + \varepsilon(t), \tag{1.1}$$

where $\boldsymbol{\beta}(t) = (\beta_0(t), \ldots, \beta_k(t))^T$ are smooth functions of t, and $\varepsilon(t)$ is a mean zero stochastic process and is independent of $\boldsymbol{X}(t)$. They also proposed a class of smoothing methods to estimate the coefficient curves. Based on model (1.1), Hoover *et al.* (1998), and Wu *et al.* (1998) have developed inferences for the kernel estimators. Under some specific designs, Fan and Zhang (2000) provided a simply implemented two-step smoothing alternative. When the covariates are time dependent, Wu *et al.* (2000) proposed a two-step smoothing method to avoid large biases appearing on the estimates. Using basis function approximations for estimating $\boldsymbol{\beta}(t)$, Huang *et al.* (2002) further developed a global smoothing procedure.

In this paper, we focus on the covariates which are invariant with respect to the time points. This is because the covariates of interest are often time invariant in clinical trials and other biomedical and epidemiological studies. Under this data setting, model (1.1) will be reduced to

$$Y(t) = \boldsymbol{X}^T \boldsymbol{\beta}(t) + \varepsilon(t), \tag{1.2}$$

where $\boldsymbol{X}^T = (X_0, \ldots, X_k)$ with $X_0 = 1$. Substituting $Y(t)$, $\boldsymbol{X}$ and t with observations Y_{ij}, $\boldsymbol{X}_i$ and t_{ij}, model (1.2) becomes

$$Y_{ij} = \boldsymbol{X}_i^T \boldsymbol{\beta}(t_{ij}) + \varepsilon(t_{ij}), \qquad i = 1, \ldots, n; \;\; j = 1, \ldots, m_i. \tag{1.3}$$

Based on (1.3), Wu and Chiang (2000), and Chiang *et al.* (2001) modified the methods of Hoover *et al.* (1998) into componentwise estimation methods to significantly simplify the computations. They also derived the asymptotic properties of the estimators through the explicit expressions of their asymptotic risk representations. Meanwhile, through a Monte Carlo simulation, the sample variances of their estimators are found to be smaller than those of Fan and Zhang (2000). However, in succeeding sections, the componentwise smoothing spline estimators are shown not as accurate as it is expected in the variances under both the finite sample and the infinite sample. This is mainly caused by the unexpected non-negative terms, which are functions of the moments of the covariates $\boldsymbol{X}$ and the parameter curves $\boldsymbol{\beta}(t)$, in the variances of the estimators.

Instead of using the componentwise estimation methods, we propose the simultaneous smoothing spline estimation methods based on the penalized least squares of Hoover *et al.* (1998) with adjustment for the variations of the covariates in the penalized terms, which avoid the penalized terms being influenced by the scales of the covariates. There are two features of the revised smoothing spline estimation methods: First, the proposed estimators are unlike the smoothing spline estimators of Hoover *et al.* (1998), which are smoothen by the random smoothing parameters and are more complicated in terms of deriving the properties of the estimators. Note that the asymptotic properties for the smoothing splines of Hoover *et al.* (1998) have not been developed. Second, when the smoothers within each estimator are set to be equal, it is shown that the mean squared errors of our simultaneous smoothing spline estimators are smaller than the corresponding componentwise ones.

The contents of this paper are organized as follows. In Section 2, we introduce the simultaneous smoothing spline estimation methods, and summarize the componentwise smoothing spline estimation methods. The asymptotic mean squared errors of the proposed estimators with or without equal smoothers for each estimator are developed in Section 3. For the sake of comparison, the asymptotic mean squared errors of the componentwise smoothing splines are also stated in this section. In Section 4, a Monte Carlo simulation is implemented to examine the finite sample properties of the simultaneous smoothing spline estimators. Applications of our estimation methods are also demonstrated in Section 5 through two empirical examples—a CD4 depletion study and an opioid detoxification study. Finally, the proofs of the main results are shown in the Appendix.

2. Estimation

Assume that the support of the design time points $\{t_{ij}\}$ is contained in a compact set $[a,b]$, and $\beta_l(t)$, $l=0,\ldots,k$, are twice differentiable. Also, let $\mathcal{H}_{[a,b]}$ be the set of compact supported functions such that

$$\mathcal{H}_{[a,b]} = \{g(\cdot) : g \text{ and } g^{(1)} \text{ are absolutely continuous, and } g^{(2)} \in L^2[a,b]\}.$$

The simultaneous smoothing spline estimators $\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda}) = (\widehat{\beta}_{0(s)}(t;\boldsymbol{\lambda}),\ldots,\widehat{\beta}_{k(s)}(t;\boldsymbol{\lambda}))^T$ of $\boldsymbol{\beta}(t)$ proposed here are obtained by minimizing the penalized least squares with adjustment for the variations of the covariates in the penalized terms

$$J_{1s}(\boldsymbol{\beta};\boldsymbol{\lambda}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} w_i \left(Y_{ij} - \sum_{l=0}^{k} X_{il}\beta_l(t_{ij})\right)^2 + \sum_{l=0}^{k} \lambda_l s_l^2 \int_a^b (\beta_l^{(2)}(t))^2 dt, \tag{2.1}$$

where $\boldsymbol{\lambda} = (\lambda_0,\ldots,\lambda_k)$ are non-negative smoothing parameters, $\boldsymbol{w} = (w_1,\ldots,w_n)$ are non-negative constant weights with $\sum_{i=1}^n m_i w_i = 1$, $s_l^2 = \sum_{i=1}^n (w_i m_i) X_{il}^2$, and $\beta_l^{(p)}(t)$ denotes the p-th derivative of $\beta_l(t)$ with respect to t. In practice, w_i's are usually assigned to be $1/N$ and $1/(nm_i)$ which provide equal weight to each single observation and each single subject, respectively. However, as mentioned in Remark 8 of Chiang *et al.* (2001), there may not have the explicit risk representations for the estimators with $w_i = 1/(nm_i)$ or more general weights. When the numbers of the repeated measurements are bounded, it was suggested by Lin and Carroll (2000) that $w_i = 1/N$ leads to asymptotically optimal kernel smoothers for the generalized estimating equations. For the sake of comparison with the existing estimators in the literature, the weights are assigned to be $1/N$ in the succeeding discussions. Setting the Gateaux derivatives of $J_{1s}(\boldsymbol{\beta};\boldsymbol{\lambda})$ to zero, $\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})$ uniquely minimize $J_{1s}(\boldsymbol{\beta};\boldsymbol{\lambda})$ if they satisfy the following normal equations

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\frac{X_{il}}{Ns_l^2}\left(Y_{ij} - \sum_{l_1=0}^{k}\widehat{\beta}_{l_1(s)}(t_{ij};\boldsymbol{\lambda})X_{il_1}\right)g_l(t_{ij})\right] \\ = \lambda_l \int_a^b \widehat{\beta}_{l(s)}^{(2)}(t;\boldsymbol{\lambda})g_l^{(2)}(t)dt,\end{aligned} \tag{2.2}$$

for $l=0,\ldots,k$, and all g_l's in a dense subset of $\mathcal{H}_{[a,b]}$. A similar argument as in Wahba (1990) shows that there is a symmetric function $S_{\lambda_l,X_l}(t,s)$, which belongs to $\mathcal{H}_{[a,b]}$ when either t or s is fixed, so that $\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})$ is given by

$$\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}\left(Y_{ij} - \sum_{l_1\neq l}\widehat{\beta}_{l_1(s)}(t_{ij};\boldsymbol{\lambda})X_{il_1}\right)S_{\lambda_l,X_l}(t,t_{ij}). \tag{2.3}$$

By substituting (2.3) into (2.2) and rearranging terms, we have the characterization of the S-spline function $S_{\lambda_l,X_l}(t,t_{ij})$,

$$\int_a^b S_{\lambda_l,X_l}(t,t_{ij})g_l(t)dF_{N,X_l}(t) + \lambda_l\int_a^b S_{\lambda_l,X_l}^{(2)}(t,t_{ij})g_l^{(2)}(t)dt = g_l(t_{ij}), \tag{2.4}$$

where $F_{N,X_l}(t) = \sum_{i=1}^n\sum_{j=1}^{m_i}(X_{il}^2/Ns_l^2)1_{[t_{ij}\le t]}$.

In (2.3), we can see that the estimator $\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})$ is influenced not only by the smoothing parameter λ_l but also by the other smoothers. To make $\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})$ being smoothen only by λ_l, the smoothing parameters $\boldsymbol{\lambda}$ are set equal to $\lambda_l\mathbf{1}$. Then, the by-product estimators, denoted by $\widehat{\beta}_{l(s)}(t;\lambda_l)$, can be expressed as

$$\widehat{\beta}_{l(s)}(t;\lambda_l)=\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}\left(Y_{ij}-\sum_{l_1\neq l}\widehat{\beta}_{l_1(s)}(t_{ij};\lambda_l)X_{il_1}\right)S_{\lambda_l,X_l}(t,t_{ij}). \tag{2.5}$$

For the smoothing spline estimation methods of Hoover *et al.* (1998), their estimators are obtained by minimizing

$$J_{2s}(\boldsymbol{\beta};\boldsymbol{\lambda})=\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(Y_{ij}-\sum_{l=0}^{k}X_{il}\beta_l(t_{ij})\right)^2+\sum_{l=0}^{k}\lambda_l\int_a^b(\beta_l^{(2)}(t))^2dt. \tag{2.6}$$

The same reasoning as in the derivation of $\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})$ shows that the corresponding minimizers, say, $\widehat{\beta}^*_{l(s)}(t;\boldsymbol{\lambda}^*)$ of $J_{2s}(\boldsymbol{\beta};\boldsymbol{\lambda})$ are given by

$$\widehat{\beta}^*_{l(s)}(t;\boldsymbol{\lambda}^*)=\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}\left(Y_{ij}-\sum_{l_1\neq l}\widehat{\beta}^*_{l_1(s)}(t_{ij};\boldsymbol{\lambda}^*)X_{il_1}\right)S_{\lambda_l^*,X_l}(t,t_{ij}), \tag{2.7}$$

where $\boldsymbol{\lambda}^*=(\lambda_0^*,\ldots,\lambda_k^*)$ with $\lambda_l^*=\lambda_l/s_l^2$. As we can see, the random smoothing parameters $\boldsymbol{\lambda}^*$ will cause complexity in deriving the properties of $\widehat{\beta}^*_{l(s)}(t;\boldsymbol{\lambda}^*)$. However, the spline estimators $\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})$ can avoid this problem, and are more common in many practical applications. When each $\beta_l(t)$ has the function form $\beta_l(t)=\sum_{j=1}^{m}\delta_{jl}B_{jl}(t)$, where $B_{jl}(t)$'s are basis functions , the minimizers of (2.6) can be derived to be the estimators of Huang *et al.* (2002) with B-spline bases. In the following sections, we will focus only on the discussion of $\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})$.

To avoid intensive computation in the estimation of the coefficient curves $\boldsymbol{\beta}(t)$, Chiang *et al.* (2001) proposed the componentwise smoothing spline estimation methods for the varying coefficient model (1.2). Even though these methods are fast in computation, the variances of the estimators, as shown later, have a higher variability. Especially, computing considerations are no longer a major consideration in modern computer equipment. For the sake of comparison in succeeding sections, we summarize here their estimation methods.

Let $E_{\boldsymbol{XX}^T}=E[\boldsymbol{XX}^T]$ and assume that the inverse of $E_{\boldsymbol{XX}^T}$, denoted by $E^{-1}_{\boldsymbol{XX}^T}$, exists. By multiplying the both sides of (1.2) by $\boldsymbol{X}$ and taking expectations, $\boldsymbol{\beta}(t)$ can be expressed as

$$\boldsymbol{\beta}(t)=E[\boldsymbol{Z}(t)], \tag{2.8}$$

where $\boldsymbol{Z}(t)=(Z_0(t),\ldots,Z_k(t))^T$ with $Z_l(t)=\sum_{r=0}^{k}e_{lr}X_rY(t)$ and e_{lr} the (l,r)-th element of $E^{-1}_{\boldsymbol{XX}^T}$. Since $E_{\boldsymbol{XX}^T}$ is unknown and is invariant with respect to time t, it is naturally estimated by the sample mean

$$\widehat{E}_{\boldsymbol{XX}}=n^{-1}\sum_{i=1}^{n}(\boldsymbol{X}_i\boldsymbol{X}_i^T). \tag{2.9}$$

Assume further that the inverse of $\widehat{E}_{\boldsymbol{XX}}$, denoted by $\widehat{E}_{\boldsymbol{XX}}^{-1}$, exists. Substituting $E_{\boldsymbol{XX}^T}^{-1}$ with $\widehat{E}_{\boldsymbol{XX}}^{-1}$, the componentwise estimator, say, $\widehat{\beta}_{l(c)}(t;\lambda_l)$ of $\beta_l(t)$ is obtained by minimizing the following penalized least squares

$$J_c(\beta_l;\lambda_l) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\widehat{Z}_{ijl} - \beta_l(t_{ij}))^2 + \lambda_l \int_a^b (\beta_l^{(2)}(t))^2 dt, \tag{2.10}$$

where $\widehat{Z}_{ijl} = \sum_{r=0}^{k} \widehat{e}_{lr} X_{il} Y_{ij}$ is the estimated observed value of Z_{ijl} with $\widehat{e}_{lr}$ being the (l,r)-th element of $\widehat{E}_{\boldsymbol{XX}^T}^{-1}$. The minimizer $\widehat{\beta}_{l(c)}(t;\lambda_l)$ of $J_c(\beta_l;\lambda_l)$ can be expressed as

$$\widehat{\beta}_{l(c)}(t;\lambda_l) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\widehat{Z}_{ijl} S_{\lambda_l}(t, t_{ij}), \tag{2.11}$$

where $S_{\lambda_l}(t,s) = S_{\lambda_l, X_0}(t,s)$ is in $\mathcal{H}_{[a,b]}$.

As derived in that paper, the asymptotic variance of the estimator $\widehat{\beta}_{l(c)}(t;\lambda_l)$ contains the unexpected non-negative terms, which are functions of the moments of the covariates and the parameter curves. In Section 3, we will show that both $\widehat{\beta}_{l(s)}(t;\lambda_l)$ and $\widehat{\beta}_{l(c)}(t;\lambda_l)$ have the same asymptotic bias, but the asymptotic variance of $\widehat{\beta}_{l(s)}(t;\lambda_l)$ is smaller. Through a Monte Carlo simulation in Section 4 and two empirical examples in Section 5, it appears that the variance of $\widehat{\beta}_{l(s)}(t;\lambda_l)$ is smaller than that of $\widehat{\beta}_{l(c)}(t;\lambda_l)$. Except the evidence from the finite and the infinite sample properties of the estimators, the unexpected terms in the variances of $\widehat{\beta}_{l(c)}(t;\lambda_l)$ can also be explained by the following reasoning: From (1.2) and the definition of $\boldsymbol{Z}(t)$, $\boldsymbol{Z}(t)$ can be reexpressed as

$$\begin{aligned} \boldsymbol{Z}(t) &= E_{\boldsymbol{XX}^T}^{-1}\boldsymbol{X}Y(t) \\ &= E_{\boldsymbol{XX}^T}^{-1}(\boldsymbol{XX}^T)\boldsymbol{\beta}(t) + E_{\boldsymbol{XX}^T}^{-1}\boldsymbol{X}\varepsilon(t) \\ &= \boldsymbol{\beta}(t) + \varepsilon^*(t), \end{aligned} \tag{2.12}$$

where

$$\varepsilon^*(t) = E_{\boldsymbol{XX}^T}^{-1}(\boldsymbol{XX}^T - E_{\boldsymbol{XX}^T})\boldsymbol{\beta}(t) + E_{\boldsymbol{XX}^T}^{-1}\boldsymbol{X}\varepsilon(t).$$

It is observed that the new error process $\varepsilon^*(t)$ consists of two components: the variabilities of the covariates and the original stochastic error process $\varepsilon(t)$. Thus, the variances of $\widehat{\beta}_{l(c)}(t;\lambda_l)$'s are enlarged by the extra non-negative terms.

3. Asymptotic properties

The asymptotic mean squared errors of the simultaneous smoothing spline estimators $\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})$, $l = 0,\ldots,k$, will be derived in this section. Without loss of generality, we focus on the interval $[0,1]$. Extension to the general interval $[a,b]$ can be carried out by the affine transformation $u = (t-a)/(b-a)$ for $t \in [a,b]$. For the succeeding discussions, we make the following assumptions:

(A1) The time design points $\{t_{ij}\}$ are nonrandom and satisfy

$$D_N = \sup_{t\in[0,1]} |F_N(t) - F(t)| \to 0, \qquad \text{as} \quad n \to \infty,$$

for some distribution function F with strictly positive density f on $[0,1]$, where $F_N(t) = N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} 1_{[t_{ij}\le t]}$, and f is three times differentiable and uniformly continuous on $[0,1]$ with $f^{(\nu)}(0) = f^{(\nu)}(1) = 0$ for $\nu = 1, 2$.

(A2) The coefficient curve $\beta_l(t)$ is four times differentiable and satisfies the boundary conditions $\beta_l^{(\nu)}(0) = \beta_l^{(\nu)}(1) = 0$ for $\nu = 2, 3$. The fourth derivative $\beta_l^{(4)}(t)$ is Lipschitz continuous in the sense that $|\beta_l^{(4)}(s_1) - \beta_l^{(4)}(s_2)| \le c_{1l}|s_1 - s_2|^{c_{2l}}$ for all $s_1, s_2 \in [0,1]$ and some positive constants c_{1l} and c_{2l}.

(A3) The fourth moment $E[X_l^4]$ exists.

(A4) Define $D_{N,l} = \sup_{t\in[0,1]}|F_{N,X_l}(t) - F(t)|$. $\lambda_l = O(\lambda_0) \to 0$, $N\lambda_l^{1/4} \to \infty$, and $\lambda_l^{-5/4} D_{N,l} \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$.

(A5) Define $\sigma^2(t) = E[\varepsilon^2(t)]$ and $\rho(t) = \lim_{t'\to t} E[\varepsilon(t)\varepsilon(t')]$. Both $\sigma^2(t)$ and $\rho(t)$ are continuous at t.

Under the validity of assumption A3, it is straightforward to show by the law of large numbers that $\sup_{t\in[0,1]}|F_{N,X_l}(t) - F_N(t)|$ converges to zero almost surely. With assumption A1 and the property $D_{N,l} \le D_N + \sup_{t\in[0,1]}|F_{N,X_l}(t) - F_N(t)|$, we can show that $D_{N,l}$ in assumption A4 converges to zero almost surely. In general, $\sigma^2(t)$ need not be equal to $\rho(t)$. As discussed in Zeger and Diggle (1994), the strict inequality appears when $\varepsilon(t)$ consists of a stationary process of t and an independent measurement error. Since the spline function $S_{\lambda_l,X_l}(t,s)$ in (2.4) does not have an explicit expression, it may be approximated by the Green function $G_{\lambda_l}(t,s)$ of the 4th order differential equation

$$\lambda_l g_l^{(4)}(t) + f(t)g_l(t) = f(t)\beta_l(t), \qquad t \in [0,1], \tag{3.1}$$

with $g_l^{(\nu)}(0) = g_l^{(\nu)}(1)$ for $\nu = 2, 3$. By substituting the smoothing spline function $S_{\lambda_l,X_l}(t,s)$ with the Green function $G_{\lambda_l}(t,s)$ and using the exponential bound of $|S_{\lambda_l,X_l}(t,s) - G_{\lambda_l}(t,s)|$, the asymptotic properties of $\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})$ can be conveniently derived. It was also shown by Abramovich and Grinshtein (1999) and Chiang *et al.* (2001) that the Green function $G_{\lambda_l}(t,s)$ can be approximated by

$$\begin{aligned} H_{\lambda_l}(t,s) = {} & \frac{(\lambda_l/\gamma^4)^{-1/4}}{2}\Gamma^{(1)}(s)(f(s))^{-1}\sin\left(\frac{\pi}{4} + \frac{(\lambda_l/\gamma^4)^{-1/4}}{\sqrt{2}}|\Gamma(t) - \Gamma(s)|\right) \\ & \times \exp\left(-\frac{(\lambda_l/\gamma^4)^{-1/4}}{\sqrt{2}}|\Gamma(t) - \Gamma(s)|\right), \end{aligned} \tag{3.2}$$

where $\gamma = \int_0^1 (f(s))^{1/4}ds$ and $\Gamma(t) = \gamma^{-1}\int_0^t (f(s))^{1/4}ds$. Some important properties of the functions $S_{\lambda_l,X_l}(t,s)$, $G_{\lambda_l}(t,s)$ and $H_{\lambda_l}(t,s)$, which will be used in the main results, can be derived along the same lines as the proof in Lemma 3.1 of Chiang *et al.* (2001). Let $\boldsymbol{B}(\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})) = (B(\widehat{\beta}_{0(s)}(t;\boldsymbol{\lambda})),\ldots,B(\widehat{\beta}_{k(s)}(t;\boldsymbol{\lambda})))$ and $\boldsymbol{V}(\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})) = [(\mathrm{cov}(\widehat{\beta}_{l_1(s)}(t;\boldsymbol{\lambda}), \widehat{\beta}_{l_2(s)}(t;\boldsymbol{\lambda})))]$ be the bias and the variance of $\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})$. By the decomposition principle of the mean squared error, we can separately evaluate the bias and the variance of $\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})$.

THEOREM 3.1. *Suppose that assumptions* (A1)–(A5) *are satisfied and* $t \in (0,1)$. *Then, for sufficiently large* n, *the bias and the variance of* $\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})$ *are given by*

$$\boldsymbol{B}(\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})) = -(f(t))^{-1}[(E_{\boldsymbol{XX}^T}^{-1}\Lambda E_{\boldsymbol{XX}^T})\boldsymbol{\beta}^{(4)}(t)]^T(1 + o(1)), \tag{3.3}$$

and

$$V(\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})) = \frac{1}{2\sqrt{2}N}(f(t))^{-3/4}\sigma^2(t)(E^{-1}_{\boldsymbol{X}\boldsymbol{X}^T}\Lambda_{\boldsymbol{X}\boldsymbol{X}^T}E^{-1}_{\boldsymbol{X}\boldsymbol{X}^T})(1+o(1)) + \left(\sum_{i=1}^{n}\left(\frac{m_i}{N}\right)^2\rho(t,t)\right)E^{-1}_{\boldsymbol{X}\boldsymbol{X}^T}(1+o(1)), \tag{3.4}$$

where $\Lambda = \mathrm{diag}(\lambda_0,\ldots,\lambda_k)$, $\boldsymbol{\beta}^{(4)}(t) = (\beta_0^{(4)}(t),\ldots,\beta_k^{(4)}(t))^T$, *and* $\Lambda_{\boldsymbol{X}\boldsymbol{X}^T} = [(\lambda_{l_1l_2}E[X_{l_1}X_{l_2}])]$ *with* $\lambda_{l_1l_2} = (\lambda_{l_1}^{1/4}+\lambda_{l_2}^{1/4})^{-1}$.

PROOF. See Appendix. □

Following the arguments in Hoover *et al.* (1998), the variance of $\widehat{\boldsymbol{\beta}}_{l(s)}(t;\boldsymbol{\lambda})$ asymptotically converges to zero if and only if $\max_{1\le i\le n}(m_i/N)\to 0$. When the smoothing parameters $\boldsymbol{\lambda}$ in $\widehat{\boldsymbol{\beta}}_{l(s)}(t;\boldsymbol{\lambda})$ are set to be equal to $\lambda_l\mathbf{1}$, the asymptotic properties of $\widehat{\beta}_{l(s)}(t;\lambda_l)$ in (2.5) can be derived straightforward from Theorem 3.1.

LEMMA 3.1. *Suppose that assumptions* (A1)–(A5) *are satisfied and* $t\in(0,1)$. *Then, for sufficiently large* n, *the bias and the variance of* $\widehat{\beta}_{l(s)}(t;\lambda_l)$, $l=0,\ldots,k$ *are given by*

$$B(\widehat{\beta}_{l(s)}(t;\lambda_l)) = -(f(t))^{-1}\beta_l^{(4)}(t)\lambda_l(1+o(1)), \tag{3.5}$$

and

$$V(\widehat{\beta}_{l(s)}(t;\lambda_l)) = \frac{1}{4\sqrt{2}}(f(t))^{-3/4}(N\lambda_l^{1/4})^{-1}e_{ll}\sigma^2(t)(1+o(1)) + \sum_{i=1}^{n}\left(\frac{m_i}{N}\right)^2 e_{ll}\rho(t,t)(1+o(1)). \tag{3.6}$$

When the regularity conditions are satisfied, Chiang *et al.* (2001) derived that the asymptotic bias and the variance of $\widehat{\beta}_{l(c)}(t;\lambda_l)$, $l=0,\ldots,k$, are

$$B(\widehat{\beta}_{l(c)}(t;\lambda_l)) = -(f(t))^{-1}\beta_l^{(4)}(t)\lambda_l(1+o_p(1)) + O_p(n^{-1/2}), \tag{3.7}$$

and

$$V(\widehat{\beta}_{l(c)}(t;\lambda_l)) = \frac{1}{4\sqrt{2}}(f(t))^{-3/4}(N\lambda_l^{1/4})^{-1}(M_l(t)+e_{ll}\sigma^2(t))(1+o_p(1)) + \sum_{i=1}^{n}\left(\frac{m_i}{N}\right)^2(M_l(t)+e_{ll}\rho(t,t))(1+o_p(1)) + O_p(n^{-1}), \tag{3.8}$$

where $M_l(t) = \sum_{l_1}^{k}\sum_{l_2}^{k}(\beta_{l_1}(t)\beta_{l_2}(t)E[X_{l_1}X_{l_2}(\sum_{l_3=0}^{k}e_{ll_3}X_{l_3})^2]) - (\beta_l(t))^2$. It follows from Lemma 3.1 and (3.7)–(3.8) that both the dominating terms in the biases of $\widehat{\beta}_{l(s)}(t;\lambda_l)$ and $\widehat{\beta}_{l(c)}(t;\lambda_l)$ are same. However, the dominating term in the variance of $\widehat{\beta}_{l(s)}(t;\lambda_l)$ is smaller than that of $\widehat{\beta}_{l(c)}(t;\lambda_l)$ since $M_l(t)$ is nonnegative.

4. Monte Carlo simulation

Consider the varying coefficient model (1.2) with coefficient curves

$$\beta_0(t) = 3.5 + 6.5\sin\left(\frac{t\pi}{60}\right),$$

$$\beta_1(t) = -0.2 - 1.6\cos\left(\frac{(t-30)\pi}{60}\right),$$

$$\beta_2(t) = 0.25 - 0.0074\left(\frac{30-t}{10}\right)^3$$

and covariate vector $\boldsymbol{X} = (X_0, X_1, X_2)^T$, where X_1 and X_2 are independent Bernoulli and Gaussian random variables with joint density

$$f(x_1, x_2) = \frac{1}{8(2\pi)^{1/2}}\exp\left(-\frac{x_2^2}{32}\right)1_{\{0,1\}}(x_1)1_{(-\infty,\infty)}(x_2).$$

In this simulation, 400 subjects are scheduled to appear at equally spaced time points $0, 1, \ldots, 30$ with a 60% probability of missing for each of the 31 "appointments". The covariates of each subject are independently generated from the above distribution. Under the given time points $\{t_{ij}\}$, the errors $\varepsilon(t_{ij})$, which are independent of the covariates $\boldsymbol{X}_i$, are generated from the mean zero Gaussian process with covariance function

$$\operatorname{cov}(\varepsilon(t_{i_1j_1}), \varepsilon(t_{i_2j_2})) = \begin{cases} 0.0625\exp(-|t_{i_1j_1} - t_{i_2j_2}|), & \text{if} \quad i_1 = i_2, \\ 0, & \text{if} \quad i_1 \neq i_2. \end{cases}$$

Finally, the dependent variables Y_{ij} are automatically obtained by substituting $\boldsymbol{X}_i$, t_{ij} and $\varepsilon(t_{ij})$ into (1.3).

Based on the above design, the longitudinal data are repeatedly generated 500 times. In each set of simulated data, $\widehat{\beta}_{l(s)}(t;\lambda_l)$ and $\widehat{\beta}_{l(c)}(t;\lambda_l)$ are computed by (2.5) and (2.11) with appropriate smoothing parameters. As mentioned in Chiang *et al.* (2001), the "leave one subject out" cross-validation procedure of Rice and Silverman (1991) may sometimes select inadequate smoothing parameters. It is usually preferable to have a set of smoothing parameters which has the corresponding cross validation score close to the minimum and gives better estimators. For the purpose of comparison, the smoothing parameters $(\lambda_0, \lambda_1, \lambda_2) = (1, 1, 1)$ from their simulation are used to both estimators. Table 1 through Table 3 show the true curves, the 500 averages of the estimated curves and the standard errors of the 500 simulation estimates at nine selected time points. As shown in these tables, the variances of the componentwise smoothing spline estimators are enlarged by the values of $M_l(t)$s, and thus are larger than those of the simultaneous smoothing splines. The results are consistent with the asymptotic properties discussed in Section 3. Moreover, based on (3.5)–(3.6), the asymptotic bias and standard deviation of $(\widehat{\beta}_{0(s)}(t;1), \widehat{\beta}_{1(s)}(t;1), \widehat{\beta}_{2(s)}(t;1))$ are computed to be $(-0.0015, 0.0004, 0)$ and $(0.0193, 0.0273, 0.0034)$. Tables 1–3 show that the asymptotic variances are slight larger than the actual variances. It also impractical to directly estimate the unknown quantities in the moments of the estimators. Thus, in applications, the bootstrapping methods will be used to construct the confidence intervals of the discussed smoothing estimators.

Table 1. The real curve $\beta_0(t)$, the averages of 500 estimated curves $\widehat{\beta}_{0(c)}(t;1)$ and $\widehat{\beta}_{0(s)}(t;1)$, and the standard errors of 500 simulation estimates at nine time points.

Time	3.0	6.0	9.0	12.0	15.0	18.0	21.0	24.0	27.0
$\beta_0(t)$	4.517	5.509	6.451	7.321	8.096	8.759	9.292	9.682	9.920
$m(\widehat{\beta}_{0(c)}(t;1))$	4.513	5.499	6.448	7.323	8.103	8.766	9.291	9.660	9.901
$(s.d.)$	0.0955	0.1111	0.1270	0.1395	0.1560	0.1771	0.1866	0.1877	0.2010
$m(\widehat{\beta}_{0(s)}(t;1))$	4.525	5.509	6.449	7.317	8.091	8.747	9.270	9.652	9.913
$(s.d.)$	0.0137	0.0110	0.0104	0.0106	0.0101	0.0098	0.0098	0.0101	0.0120

Table 2. The real curve $\beta_1(t)$, the averages of 500 estimated curves $\widehat{\beta}_{1(c)}(t;1)$ and $\widehat{\beta}_{1(s)}(t;1)$, and the standard errors of 500 simulation estimates at nine time points.

Time	3.0	6.0	9.0	12.0	15.0	18.0	21.0	24.0	27.0
$\beta_1(t)$	−0.450	−0.694	−0.926	−1.140	−1.331	−1.494	−1.626	−1.722	−1.780
$m(\widehat{\beta}_{1(c)}(t;1))$	−0.438	−0.678	−0.921	−1.144	−1.344	−1.513	−1.639	−1.714	−1.769
$(s.d.)$	0.1761	0.2040	0.2328	0.2552	0.2831	0.3187	0.3354	0.3376	0.3638
$m(\widehat{\beta}_{1(s)}(t;1))$	−0.461	−0.696	−0.924	−1.140	−1.323	−1.477	−1.598	−1.694	−1.785
$(s.d.)$	0.0172	0.0154	0.0150	0.0152	0.0150	0.0141	0.0144	0.0153	0.0169

Table 3. The real curve $\beta_2(t)$, the averages of 500 estimated curves $\widehat{\beta}_{2(c)}(t;1)$ and $\widehat{\beta}_{2(s)}(t;1)$, and the standard errors of 500 simulation estimates at nine time points.

Time	3.0	6.0	9.0	12.0	15.0	18.0	21.0	24.0	27.0
$\beta_2(t)$	0.104	0.148	0.181	0.207	0.225	0.237	0.245	0.248	0.250
$m(\widehat{\beta}_{2(c)}(t;1))$	0.103	0.146	0.181	0.207	0.225	0.235	0.243	0.248	0.251
$(s.d.)$	0.02178	0.0266	0.0298	0.0314	0.0329	0.0366	0.0410	0.0421	0.0474
$m(\widehat{\beta}_{2(s)}(t;1))$	0.103	0.146	0.181	0.207	0.225	0.237	0.245	0.248	0.250
$(s.d.)$	0.0021	0.0021	0.0021	0.0020	0.0020	0.0020	0.0019	0.0020	0.0021

5. Application

In this section, the proposed simultaneous smoothing spline estimation methods and the componentwise ones are applied to two empirical examples. These longitudinal data sets arise from a CD4 depletion study and an opioid detoxification study.

5.1 *A CD4 depletion study*

The first data set is from the Multicenter AIDS Cohort Study (MACS), which includes 283 homosexual men who were infected by HIV-1 virus. Measurements taken include CD4 percentage, the cigarette smoking status, pre-HIV infection CD4 percentage, and age at HIV infection. Individuals were repeatedly measured at scheduled semi-annual visits between 1984 and 1991. During the study period, many individuals missed some of their scheduled visits. Thus, the numbers of repeated measurements may differ among subjects. Details of the design and the methods of this study are described in Kaslow *et al.* (1987).

In this study, the objective is to evaluate the effects of cigarette smoking, pre-

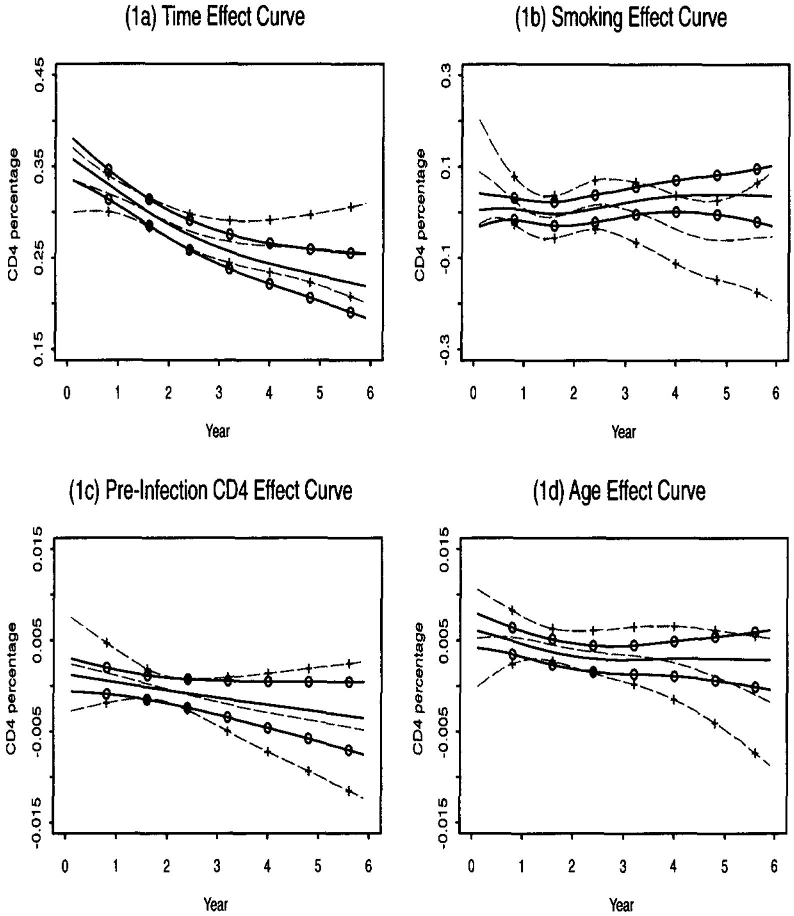

Fig. 1. The simultaneous smoothing splines $\widehat{\beta}_{l(s)}(t;\lambda_l)$ (solid curve) and the componentwise smoothing splines $\widehat{\beta}_{l(c)}(t;\lambda_l)$ (dashed curve) with the corresponding 95% bootstrap confidence intervals labeled o and $+$.

HIV infection CD4 percentage, and age at HIV infection on the mean post-HIV CD4 percentage at any given time since the infection among seroconverters. Based on model (1.2), the simultaneous smoothing spline estimation methods and the componentwise smoothing spline estimation methods are used to estimate the effects of the concerned covariates. Estimators $\widehat{\beta}_{l(s)}(t;\lambda_l)$ and $\widehat{\beta}_{l(c)}(t;\lambda_l)$ are separately computed from (2.5) and (2.11) with the smoothing parameters $(\lambda_0,\lambda_1,\lambda_2,\lambda_3)=(0.1,0.01,1,0.1)$, which have the corresponding cross-validation score close to the minimum.

Figures (1a)–(1d) show the estimated curves and their 95% pointwise bootstrap confidence intervals. From these graphs, we can see that $\widehat{\beta}_{l(c)}(t;\lambda_l)$ and $\widehat{\beta}_{l(s)}(t;\lambda_l)$ have similar physical explanations. However, the confidence bands of $\widehat{\beta}_{l(c)}(t;\lambda_l)$ are wider than $\widehat{\beta}_{l(s)}(t;\lambda_l)$. As mentioned in Section 3 and Section 4, the simultaneous smoothing spline estimation methods are more reliable. From Fig. (1a), the mean CD4 percentage for the non-smoking group with average pre-infection CD4 percentage and average age at HIV infection appears depleting rather quickly at the beginning of HIV infection, but

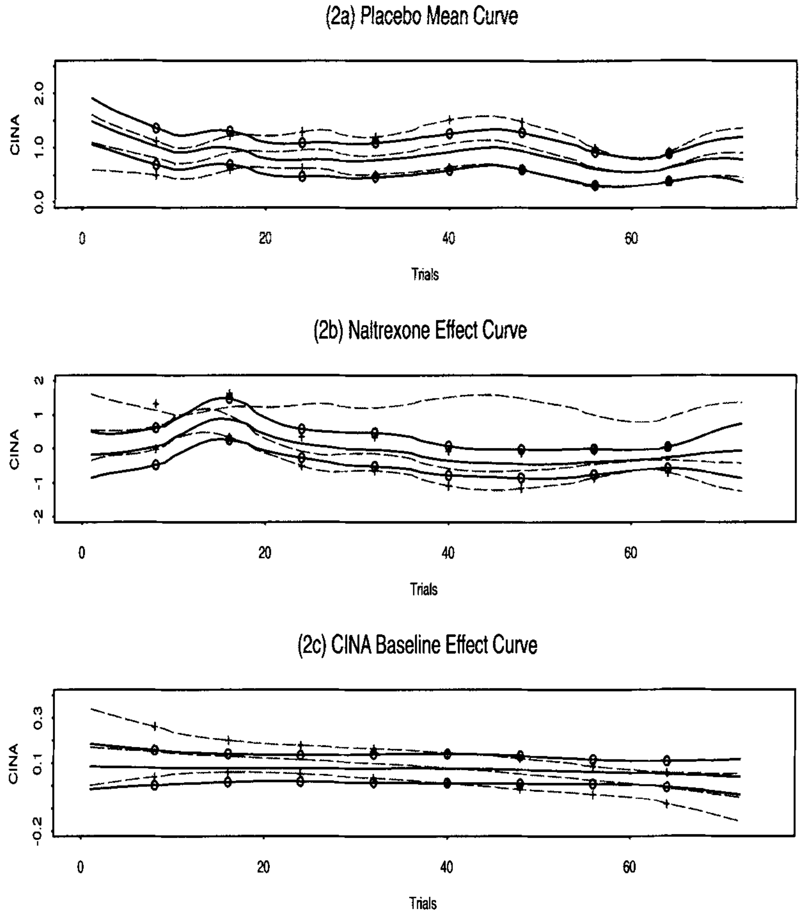

Fig. 2. The simultaneous smoothing splines $\widehat{\beta}_{l(s)}(t;\lambda_l)$ (solid curve) and the componentwise smoothing splines $\widehat{\beta}_{l(c)}(t;\lambda_l)$ (dashed curve) with the corresponding 95% bootstrap confidence intervals labeled o and $+$.

the rate of depletion seems to be slowing down for the later period of the study after infection. No significant effects are detected in Figs. (1b) and (1d) for cigarette smoking and age at HIV infection. However, it appears in Fig. (1c) that the pre-HIV infection CD4 percentage associated positively with higher CD4 percentage after the infection.

5.2 *An opioid detoxification study*

The second data set is from the National Institute on Drug Abuse (NIDA) opioid detoxification study, which includes 60 opioid dependent (DSM-IV) heroin users seeking detoxification treatment. In the study design, 32 patients are randomly assigned to the naltrexone-buprenorphine group and 28 to the placebo-buprenorphine group. Measurements were taken at 9 scheduled times per day by a trained nurse for a total 72 (8×9) measurements. During an 8 day inpatient clinical trial, each patient was subjected to the observer-rated opioid withdrawal scale (OOW) measurement, a scale to rate opioid withdrawal symptoms. Since some patients randomly missed some scheduled measurements or quit the treatment altogether, the number of measurements may be different for each patient. Details of this design and its medical implications can be found in Umbricht-Schneiter *et al.* (1999).

The objective here is to detect the effects of treatment status and the centered baseline OOW score on the OOW scores over the trials. Similar to the process of analysis in Subsection 5.1, two estimation methods are used to detect the effects of the interesting covariates. Here, estimators $\widehat{\beta}_{l(c)}(t;\lambda_l)$ and $\widehat{\beta}_{l(s)}(t;\lambda_l)$ are computed with the smoothing parameters $(\lambda_0,\lambda_1,\lambda_2)=(0.001,0.001,0.1)$.

Figures (2a)–(2c) show the estimated curves and their 95% pointwise bootstrap confidence intervals. From these graphs, two estimation methods provide similar explanations. Also, the confidence bands for some of the componentwise smoothing spline estimators are close to the corresponding simultaneous ones. This can be explained by the small effect of $M_l(t)$ to the variance of $\widehat{\beta}_{l(c)}(t;\lambda_l)$ in (3.8) for some l. It is shown that the placebo mean stays very close to a constant throughout the trials, while the naltrexone treatment is generally associated with lower OOW scores roughly after the later half of the trial. The peak at the beginning of the trial for the naltrexone treatment is mainly caused by the patient's initial negative reaction to the treatment. As expected, the baseline OOW score has a positive association with the OOW scores.

Acknowledgements

The author's research was partially supported by the National Science Council grants 92-2118-M-002-005 and 93-2118-m-002-001 (Taiwan). I am grateful to Dr. A. Umbricht-Schneiter and the reviewers for valuable comments and suggestions.

Appendix A

Before conducting the proof of Theorem 3.1, a technical lemma is stated first.

LEMMA A.1. *Suppose assumptions* A1 *and* A4 *are satisfied, and let function* $M(t)$ *be continuous for all* $t\in[0,1]$. *Then, for sufficiently small* λ_l,

$$\int_0^1 G_{\lambda_{l_1}}(t,s)G_{\lambda_{l_2}}(t,s)M(s)f(s)ds=\frac{1}{2\sqrt{2}}(f(t))^{-3/4}M(t)\lambda_{l_1l_2}(1+o(1)), \tag{A.1}$$

and

$$\int_0^1 G_{\lambda_l}(t,s)M(s)f(s)ds=M(t)(1+o(1)) \tag{A.2}$$

for all $t\in[\tau,1-\tau]$ *with some* $\tau>0$.

PROOF. It is easy to see that the quantity $\int_0^1 G_{\lambda_{l_1}}(t,s)G_{\lambda_{l_2}}(t,s)M(s)f(s)ds$ can be expressed as

$$\begin{aligned}\int_0^1 & G_{\lambda_{l_1}}(t,s)G_{\lambda_{l_2}}(t,s)M(s)f(s)ds \\ &=\int_0^1 (G_{\lambda_{l_1}}(t,s)-H_{\lambda_{l_1}}(t,s))G_{\lambda_{l_2}}(t,s)M(s)f(s)ds \\ &\quad+\int_0^1 H_{\lambda_{l_1}}(t,s)(G_{\lambda_{l_2}}(t,s)\end{aligned} \tag{A.3}$$

$$- H_{\lambda_{l_2}}(t,s))M(s)f(s)ds + \int_0^1 H_{\lambda_{l_1}}(t,s)H_{\lambda_{l_2}}(t,s)M(s)f(s)ds.$$

From (3.2), Lemma 3.1 of Chiang *et al.* (2001), and the properties of double exponential distributions, there exists a positive constant c_1 so that, as $\lambda_{l_1} \to 0$, $\lambda_{l_2} \to 0$, and $\lambda_{l_1} = O(\lambda_{l_2})$,

$$\begin{aligned}(\mathrm{A.4})\quad & \left|\int_0^1 (G_{\lambda_{l_1}}(t,s) - H_{\lambda_{l_1}}(t,s))G_{\lambda_{l_2}}(t,s)M(s)f(s)ds\right| \\ &\le \int_0^1 |G_{\lambda_{l_1}}(t,s) - H_{\lambda_{l_1}}(t,s)||G_{\lambda_{l_2}}(t,s)||M(s)|f(s)ds \\ &\le \int_0^1 (\kappa_1)^2\lambda_{l_2}^{-1/4}\exp(-(\alpha_1\lambda_{l_1}^{-1/4} + \alpha_2\lambda_{l_2}^{-1/4})|t-s|)|M(s)|f(s)ds \\ &= c_1|M(t)|f(t)(1+o(1)).\end{aligned}$$

Similarly,

$$(\mathrm{A.5})\quad \left|\int_0^1 H_{\lambda_{l_1}}(t,s)(G_{\lambda_{l_2}}(t,s) - H_{\lambda_{l_2}}(t,s))M(s)f(s)ds\right| \le c_2|M(t)|f(t)(1+o(1))$$

for a positive constant c_2. Let $u = \Gamma(t)$ and $v = \Gamma(s)$. Again, using the properties of double exponential distributions, it can be shown that

$$\begin{aligned}(\mathrm{A.6})\quad & \int_0^1 H_{\lambda_{l_1}}(t,s)H_{\lambda_{l_2}}(t,s)M(s)f(s)ds \\ &= \frac{(\gamma^2(\lambda_{l_1}\lambda_{l_2})^{-1/4})}{4}\int_0^1 \sin\left(\frac{\pi}{4} + \gamma\frac{\lambda_{l1}|u-v|}{\sqrt{2}}\right)\sin\left(\frac{\pi}{4} + \gamma\frac{\lambda_{l2}|u-v|}{\sqrt{2}}\right) \\ &\quad\cdot \exp\left(-\gamma\frac{\lambda^*_{l_1l_2}|u-v|}{\sqrt{2}}\right)M(\Gamma^{-1}(v))\gamma^{-1}(f(\Gamma^{-1}(v)))^{-3/4}dv \\ &= \frac{1}{2\sqrt{2}}(f(t))^{-3/4}M(t)\lambda_{l_1l_2}(1+o(1)),\end{aligned}$$

where $\lambda^*_{l_1l_2} = \lambda_{l_1}^{-1/4} + \lambda_{l_2}^{-1/4}$. By substituting (A.4), (A.5) and (A.6) into (A.3), (A.1) is then obtained. Similar arguments can be used to show that (A.2) also holds.

Proof of (3.3). From (1.3) and (2.3), we can derive the equations

$$\begin{aligned}(\mathrm{A.7})\quad & \widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda}) - \sum_{i=1}^n\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}\left(\beta_l(t_{ij})X_{il} - \sum_{l_1\ne l}B_{l_1(s)}(t_{ij};\boldsymbol{\lambda})X_{il_1}\right)S_{\lambda_l,X_l}(t,t_{ij}) \\ &= \sum_{i=1}^n\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}\varepsilon(t_{ij})S_{\lambda_l,X_l}(t,t_{ij}), \qquad l = 0,\dots,k,\end{aligned}$$

where $B_{l_1(s)}(t_{ij};\boldsymbol{\lambda}) = (\widehat{\beta}_{l_1(s)}(t_{ij};\boldsymbol{\lambda}) - \beta_{l_1}(t_{ij}))$. By the law of large numbers and assumptions A1, A3, it can be shown that

$$(\mathrm{A.8})\qquad \sum_{i=1}^n\sum_{j=1}^{m_i}\frac{X_{il}X_{il_1}}{Ns_l^2}1_{[t_{ij}\le t]} \overset{\text{a.s.}}{\to} \left(\frac{E[X_lX_{l_1}]}{E[X_l^2]}\right)F(t).$$

Using the properties in Lemma 3.1 of Chiang *et al.* (2001), (A.8), assumption A4, and taking expectation of the left hand side of (A.7),

$$\text{(A.9)}\quad E\left[\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda}) - \sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}\left(\beta_l(t_{ij})X_{il} - \sum_{l_1\neq l} B_{l_1(s)}(t_{ij};\boldsymbol{\lambda})X_{il_1}\right)S_{\lambda_l,X_l}(t,t_{ij})\right]$$
$$= B_I + B_{II},$$

where

$$B_I = E[\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})] - \int_0^1 \beta_l(s)G_{\lambda_l}(t,s)f(s)ds(1+o(1)),$$

and

$$B_{II} = \sum_{l_1\neq l}\frac{E[X_lX_{l_1}]}{E[X_l^2]}\int_0^1 (E[\widehat{\beta}_{l_1(s)}(s;\boldsymbol{\lambda})] - \beta_{l_1}(s))G_{\lambda_l}(t,s)f(s)ds(1+o(1)).$$

Let $g_l(t) = \int_0^1 G_{\lambda_l}(t,s)\beta_l(s)f(s)ds$. It follows by the definition of the Green function in (3.1) and Lemma 6.1 of Nychka (1995) that

$$\text{(A.10)}\qquad g_l(t) - \beta_l(t) = \frac{-\lambda_l}{f(t)}g_l^{(4)}(t)(1+o(1)) = \frac{-\lambda_l}{f(t)}\beta_l^{(4)}(t)(1+o(1)).$$

Thus, from (A.10) and Lemma A.1, we can get

$$\text{(A.11)}\qquad B_I = E[\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})] - \left(\beta_l(t) - \frac{\lambda_l}{f(t)}\beta_l^{(4)}(t) + o(\lambda_l)\right)(1+o(1))$$
$$= \left(E[\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})] - \beta_l(t) + \frac{\lambda_l}{f(t)}\beta_l^{(4)}(t)\right)(1+o(1))$$

and

$$\text{(A.12)}\qquad B_{II} = \sum_{l_1\neq l}\frac{E[X_lX_{l_1}]}{E[X_l^2]}\left(E[\widehat{\beta}_{l_1(s)}(t;\boldsymbol{\lambda})] - \left(\beta_{l_1}(t) - \frac{\lambda_l}{f(t)}\beta_{l_1}^{(4)}(t) + o(\lambda_l)\right)\right)$$
$$\cdot(1+o(1))$$
$$= \sum_{l_1\neq l}\frac{E[X_lX_{l_1}]}{E[X_l^2]}\left(E[\widehat{\beta}_{l_1(s)}(t;\boldsymbol{\lambda})] - \beta_{l_1}(t) + \frac{\lambda_l}{f(t)}\beta_{l_1}^{(4)}(t)\right)(1+o(1))$$
$$= \sum_{l_1\neq l}\frac{E[X_lX_{l_1}]}{E[X_l^2]}\left(B(\widehat{\beta}_{l_1(s)}(t;\boldsymbol{\lambda})) + \frac{\lambda_l}{f(t)}\beta_{l_1}^{(4)}(t)\right)(1+o(1)).$$

By substituting (A.11) and (A.12) into (A.9),

$$\text{(A.13)}\qquad E\left[\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda}) - \sum_{\{i,j\}}\frac{X_{il}}{Ns_l^2}\left(\beta_l(t_{ij})X_{il} - \sum_{l_1\neq l} B_{l_1(s)}(t_{ij};\boldsymbol{\lambda})X_{il_1}\right)S_{\lambda_l,X_l}(t,t_{ij})\right]$$
$$= \sum_{l_1=0}^{k}\frac{E[X_lX_{l_1}]}{E[X_l^2]}\left(B(\widehat{\beta}_{l_1(s)}(t;\boldsymbol{\lambda})) + \frac{\lambda_l}{f(t)}\beta_{l_1}^{(4)}(t)\right)(1+o(1)).$$

Since $\varepsilon_i(t)$ is a mean zero stochastic process and is independent of the covariates $\boldsymbol{X}$, it implies that

$$E\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}\varepsilon(t_{ij})S_{\lambda_l,X_l}(t,t_{ij})\right]=0. \tag{A.14}$$

Finally, from (A.13) and (A.14), the bias of $\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})$ in (3.3) is obtained.

PROOF OF (3.4). Following (A.7), we can get the following equations

$$\begin{aligned}\prod_{\{l=l_1,l_2\}}&\left(\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})\right.\\ &\left.-\sum_{\{i,j\}}\frac{X_{il}}{Ns_l^2}\left(\beta_l(t_{ij})X_{il}-\sum_{l_3\neq l}B_{l_3(s)}(t_{ij};\boldsymbol{\lambda})X_{il_3}\right)S_{\lambda_l,X_l}(t,t_{ij})\right)\\ =\prod_{\{l=l_1,l_2\}}&\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}\varepsilon(t_{ij})S_{\lambda_l}(t,t_{ij})\right),\qquad \forall l_1,l_2\in\{0,\ldots,k\}.\end{aligned} \tag{A.15}$$

From Lemma 3.1 of Chiang *et al.* (2001), (A.8), (A.10), assumptions A1–A4, and Lemma A.1, the expectation of the left hand side of (A.15) is derived as

$$\begin{aligned}E&\left[\prod_{\{l=l_1,l_2\}}\left(\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})\right.\right.\\ &\left.\left.-\sum_{\{i,j\}}\frac{X_{il}}{Ns_l^2}\left(\beta_l(t_{ij})X_{il}-\sum_{l_3\neq l}B_{l_3(s)}(t_{ij};\boldsymbol{\lambda})X_{il_3}\right)S_{\lambda_l,X_l}(t,t_{ij})\right)\right]\\ &=E\left[\prod_{\{l=l_1,l_2\}}\left(\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})-\int_0^1\beta_l(s)G_{\lambda_l}(t,s)f(s)ds(1+o_p(1))\right.\right.\\ &\qquad\left.\left.+\sum_{l_3\neq l}\frac{E[X_lX_{l_3}]}{E[X_l^2]}\int_0^1 B_{l_3(s)}(s;\boldsymbol{\lambda})G_{\lambda_l}(t,s)f(s)ds(1+o_p(1))\right)\right]\\ &=E\left[\prod_{\{l=l_1,l_2\}}\left(\sum_{l_3=0}^{k}\frac{E[X_lX_{l_3}]}{E[X_l^2]}\left(B_{l_3(s)}(t;\boldsymbol{\lambda})+\frac{\lambda_l}{f(t)}\beta_{l_3}^{(4)}(t)\right)(1+o_p(1))\right)\right]\\ &=E\left[\prod_{\{l=l_1,l_2\}}\left(\sum_{l_3=0}^{k}\frac{E[X_lX_{l_3}]}{E[X_l^2]}(\widehat{\beta}_{l_3(s)}(t;\boldsymbol{\lambda})-E[\widehat{\beta}_{l_3(s)}(t;\boldsymbol{\lambda})])(1+o_p(1))\right)\right]\\ &=\sum_{\{l_3,l_4\}}^{k}\left(\frac{E[X_{l_1}X_{l_3}]E[X_{l_2}X_{l_4}]}{E[X_{l_1}^2]E[X_{l_2}^2]}\right)\mathrm{Cov}(\widehat{\beta}_{l_3(s)}(t;\boldsymbol{\lambda}),\widehat{\beta}_{l_4(s)}(t;\boldsymbol{\lambda}))(1+o(1)).\end{aligned} \tag{A.16}$$

Taking expectation of the right hand side of (A.15),

$$E\left[\prod_{\{l=l_1,l_2\}}\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}\varepsilon(t_{ij})S_{\lambda_l,X_l}(t,t_{ij})\right)\right]=V_I+V_{II}+V_{III}, \tag{A.17}$$

where

$$V_I=E\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\frac{X_{il_1}X_{il_2}}{N^2s_{l_1}^2s_{l_2}^2}\right)\varepsilon^2(t_{ij})S_{\lambda_{l_1},X_{l_1}}(t,t_{ij})S_{\lambda_{l_2},X_{l_2}}(t,t_{ij})\right],$$

$$V_{II}=E\left[\sum_{i=1}^{n}\sum_{\{j_1\neq j_2\}}\left(\frac{X_{il_1}X_{il_2}}{N^2s_{l_1}^2s_{l_2}^2}\right)\varepsilon(t_{ij_1})\varepsilon(t_{ij_2})S_{\lambda_{l_1},X_{l_1}}(t,t_{ij_1})S_{\lambda_{l_2},X_{l_2}}(t,t_{ij_2})\right],$$

and

$$V_{III}=E\left[\sum_{\{i_1\neq i_2\}}\sum_{\{j_1,j_2\}}\left(\frac{X_{i_1l_1}X_{i_2l_2}}{N^2s_{l_1}^2s_{l_2}^2}\right)\varepsilon(t_{i_1j_1})\varepsilon(t_{i_2j_2})S_{\lambda_{l_1},X_{l_1}}(t,t_{i_1j_1})S_{\lambda_{l_2},X_{l_2}}(t,t_{i_2j_2})\right].$$

From assumption A4, Lemma 3.1 of Chiang *et al.* (2001), and Lemma A.1, it can be derived that

$$\begin{aligned}V_I&=\frac{E[X_{l_1}X_{l_2}]}{NE[X_{l_1}^2]E[X_{l_2}^2]}\int_0^1\sigma^2(t)G_{\lambda_{l_1}}(t,s)G_{\lambda_{l_2}}(t,s)f(s)ds(1+o(1))\\&=\frac{1}{2\sqrt{2}N}(f(t))^{-3/4}\left(\frac{\lambda_{l_1l_2}E[X_{l_1}X_{l_2}]}{E[X_{l_1}^2]E[X_{l_2}^2]}\right)\sigma^2(t)(1+o(1)),\end{aligned} \tag{A.18}$$

and

$$\begin{aligned}V_{II}&=\left(\sum_{i=1}^{n}\left(\frac{m_i}{N}\right)^2-1/N\right)\left(\frac{E[X_{l_1}X_{l_2}]}{E[X_{l_1}^2]E[X_{l_2}^2]}\right)\\&\quad\cdot\int_0^1\int_0^1\rho(s_1,s_2)G_{\lambda_{l_1}}(t,s_1)G_{\lambda_{l_2}}(t,s_2)f(s_1)f(s_2)ds_1ds_2(1+o(1))\\&=\left(\sum_{i=1}^{n}\frac{m_i}{N}\right)^2\left(\frac{E[X_{l_1}X_{l_2}]}{E[X_{l_1}^2]E[X_{l_2}^2]}\right)\rho(t,t)(1+o(1)).\end{aligned} \tag{A.19}$$

Since $\varepsilon_i(t)$ is a mean zero stochastic process and is independent of the covariates $\boldsymbol{X}$, it implies that $V_{III}=0$. Substituting (A.18), (A.19), and $V_{III}=0$ into (A.17), we can get

$$\begin{aligned}&E\left[\prod_{\{l=l_1,l_2\}}\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}\varepsilon_i(t_{ij})S_{\lambda_l,X_l}(t,t_{ij})\right)\right]\\&=\left(\frac{E[X_{l_1}X_{l_2}]}{E[X_{l_1}^2]E[X_{l_2}^2]}\right)\\&\quad\cdot\left(\frac{\lambda_{l_1l_2}}{2\sqrt{2}N}(f(t))^{-3/4}\sigma^2(t)+\left(\sum_{i=1}^{n}(m_i/N)^2\right)\rho(t,t)\right)(1+o(1)).\end{aligned} \tag{A.20}$$

From (A.16) and (A.20), the variance of $\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})$ in (3.4) is then obtained.

REFERENCES

Abramovich, F. and Grinshtein, V. (1999). Derivation of equivalent kernel for general spline smoothing: A systematic approach, *Bernoulli*, **5**, 359–379.

Chiang, C. T., Rice, J. A. and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variable, *Journal of the American Statistical Association*, **96**, 605–619.

Fan, J. Q. and Zhang, J. T. (2000). Functional linear models for longitudinal data, *Journal of the Royal Statistical Society, Series B*, **62**, 303–322.

Hastie, T. J. and Tibshirani, R. J. (1993). Varying coefficient model, *Journal of the Royal Statistical Society, Series B*, **55**, 757–796.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika*, **85**, 809–822.

Huang, J. Z., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis repeated measurements, *Biometrika*, **89**, 111–128.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F. and Rinaldo, C. R. (1987). The multicenter AIDS cohort study: Rationale, organization and selected characteristics of the participants, *American Journal of Epidemiology*, **126**, 310–318.

Lin, X. and Carroll, R. J. (2000). Semiparametric regression for the mean and rate functions of recurrent events, *Journal of the Royal Statistical Society, Series B*, **62**, 711–730.

Nychka, D. (1995). Splines as local smoothers, *The Annals of Statistics*, **23**, 1175–1197.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves, *Journal of the Royal Statistical Society, Series B*, **53**, 233–243.

Umbricht-Schneiter, A., Montoya, I. D., Hoover D. R., Demuch, K. L., Chiang, C.-T. and Preston, K. L. (1999). Naltrexone shortened opioid detoxification with buprenorphine, *Drug and Alcohol Dependence*, **56**, 181–190.

Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia, Pennsylvania.

Wu, C. O. and Chiang, C. T. (2000). Kernel estimation and its applications of a varying coefficient model with longitudinal data, *Statistica Sinica*, **10**, 433–456.

Wu, C. O., Chiang, C.-T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying coefficient model with longitudinal data, *Journal of the American Statistical Association*, **93**, 1388–1402.

Wu, C. O., Yu, K. F. and Chiang, C.-T. (2000). A two-step smoothing method for varying coefficient models with repeated measurements, *Annals of the Institute of Statistical Mathematics*, **52**, 519–543.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters, *Biometrics*, **50**, 689–699.