行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　　※
※　　變異係數模式之多項式平滑迴歸法(2/2)　　※
※　　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：☑個別型計畫　□整合型計畫
計畫編號：NSC90－2118－M－002－013
執行期間：90 年 8 月 1 日至 91 年 7 月 31 日

計畫主持人：江金倉
共同主持人：
計畫參與人員：

本成果報告包括以下應繳交之附件：
- □赴國外出差或研習心得報告一份
- □赴大陸地區出差或研習心得報告一份
- ☑出席國際學術會議心得報告及發表之論文各一份
- □國際合作研究計畫國外研究報告書一份

執行單位：台灣大學數學系

中　華　民　國　91　年　10　月 28 日

行政院國家科學委員會專題研究計畫成果報告

變異係數模式之多項式平滑迴歸法(2/2)

(Simultaneous Smoothing Splines for Varying Coefficient Models)

計畫編號：NSC 90-2118-M-002-013
執行期限：90 年 8 月 1 日至 91 年 7 月 31 日

主持人：江金倉　　執行機構及單位名稱：台灣大學數學系
共同主持人：　　　執行機構及單位名稱
計畫參與人員：　　執行機構及單位名稱

一、中文摘要

提出多項式平滑回歸估計方法，當中之反應變數為時間函數，而解釋變數為一常值，在此估計方法中，吾等考慮將其解釋變數之變異消除，此目的乃在於避免估計式受隨機平滑參數之影響及理論推導之複雜性。對此研究主題，吾已推導出估計式之大樣本理論並借助實例及模擬發現好的有限樣本之性質。

關鍵詞：

Abstract

Keywords:

二、緣由與目的

三、結果與討論

四、計畫成果自評

五、參考文獻

(以上討論於所附之研究成果)

Simultaneous Smoothing Splines for Varying Coefficient Models

Chin-Tsang Chiang
Department of Mathematics
National Taiwan University
Taipei 106, Taiwan, R.O.C.

Abstract

In this paper, simultaneous smoothing spline estimation methods are proposed to estimate the coefficient curves on the varying coefficient model with repeatedly measured (longitudinal) dependent variable and time invariant covariates. The estimators are 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 properties of the estimators. We also develop the asymptotic properties of the simultaneous smoothing splines. When the smoothing parameters within each estimator are set to be equal, the simultaneous smoothing spline estimators are shown to have smaller variances than the componentwise smoothing spline estimators. Through a Monte Carlo simulation and two empirical examples, the simultaneous smoothing splines are also found to be more accurate in the variances.

1 Introduction

In biomedical and epidemiology 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

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

subjects. For the ith subject, let m_i, t_{ij}, Y_{ij}, and $\mathbf{X}_i^T = (X_{i0}, \cdots, X_{ik})$ with $X_{i0} = 1$, respectively denote the number of the repeated measurements, the time of the jth 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 $\mathbf{X}^T(t) = (X_0(t), \cdots, X_k(t))$ with $X_0(t) = 1$, Hoover, Rice, Wu and Yang (1998) considered a more flexible varying coefficient model of Hastie and Tibshirani (1993)

$$Y(t) = \mathbf{X}^T(t)\boldsymbol{\beta}(t) + \varepsilon(t), \tag{1}$$

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 $\mathbf{X}(t)$. They also proposed a class of smoothing methods to estimate the coefficient curves. Based on model (1), Hoover, *et al.* (1998), and Wu, Chiang and Hoover (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, Yu and Chiang (2000) also proposed a two-step smoothing method to avoid large biases appearing on the estimates.

In this paper, we focus on the covariates which are invariant with respect to the time points. Under this data setting, model (1) will be reduced to

$$Y(t) = \mathbf{X}^T\boldsymbol{\beta}(t) + \varepsilon(t), \tag{2}$$

where $\mathbf{X}^T = (X_0, \cdots, X_k)$ with $X_0 = 1$. Substituting $Y(t)$, $\mathbf{X}$ and t with observations Y_{ij}, $\mathbf{X}_i$ and t_{ij}, model (2) becomes

$$Y_{ij} = \mathbf{X}_i^T\boldsymbol{\beta}(t_{ij}) + \varepsilon(t_{ij}), \qquad i = 1, \cdots n; j = 1, \cdots m_i. \tag{3}$$

Based on (3), Wu and Chiang (2000), and Chiang, Rice and Wu (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 $\mathbf{X}$ 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 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 our simultaneous smoothing spline estimation methods: First, our 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. 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 smoothing splines.

The contents of this paper are organized as follows. In Section 2, we propose 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 smoothing parameters 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 $\beta_l(t), l = 0, \cdots, k$, are twice differentiable. Also, let $\mathcal{H}_{[a,b]}$ be the set of compact supported functions such that

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

The simultaneous smoothing splines $\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda}) = (\widehat{\beta}_{0(s)}(t;\boldsymbol{\lambda}), \cdots, \widehat{\beta}_{k(s)}(t;\boldsymbol{\lambda}))^T$ 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

$$J_{1s}(\boldsymbol{\beta};\boldsymbol{\lambda}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} w_i \left[Y_{ij} - \left(\sum_{l=0}^{k} X_{il}\beta_l(t_{ij}) \right) \right]^2 + \sum_{l=0}^{k} \lambda_l s_l^2 \int_a^b \left(\beta_l^{(2)}(t) \right)^2 dt, \tag{4}$$

where $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_k)$ are non-negative smoothing parameters, $\mathbf{w} = (w_1, \cdots, w_n)$ are non-negative constant weights with $\sum_{i=1}^{n} m_i w_i = 1$, $s_l = \sum_{i=1}^{n}(w_i m_i)X_{il}^2$, and $\beta_l^{(p)}(t)$ denotes the *pth* 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

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\frac{X_{il}}{Ns_l^2}\left(Y_{ij}-\sum_{l_1=0}^{k}\widehat{\beta}_{l_1(s)}(t_{ij};\boldsymbol{\lambda})X_{il_1}\right)g_l(t_{ij})\right] = \lambda_l\int_a^b \widehat{\beta}_{l(s)}^{(2)}(t;\boldsymbol{\lambda})g_l^{(2)}(t)dt, \tag{5}$$

for $l = 0,\cdots,k$, and all g_l's in a dense subset of $\mathcal{H}_{[a,b]}$. A similar argument as in Wahba (1975) shows that there is a symmetric function $S_{\lambda_l,X_l}(t,s)$ in $\mathcal{H}_{[a,b]}$ so that $\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})$ is given by

$$\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}(Y_{ij}-\sum_{l_1\neq l}\widehat{\beta}_{l_1(s)}(t_{ij};\boldsymbol{\lambda})X_{il_1})S_{\lambda_l,X_l}(t,t_{ij}). \tag{6}$$

By substituting (6) into (5) and rearranging the terms, we have the characterization of the spline function $S_{\lambda_l,X_l}(t,t_{ij})$,

$$\int_a^b S_{\lambda_l,X_l}(t,t_{ij})g_l(t)dF_{N,X_l}(t) + \lambda_l\int_a^b S_{\lambda_l,X_l}^{(2)}(t,t_{ij})g_l^{(2)}(t)dt = g_l(t_{ij}), \tag{7}$$

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

In (6), 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

$$\widehat{\beta}_{l(s)}(t;\lambda_l) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{X_{il}}{Ns_l^2}(Y_{ij}-\sum_{l_1\neq l}\widehat{\beta}_{l_1(s)}(t_{ij};\lambda_l)X_{il_1})S_{\lambda_l,X_l}(t,t_{ij}). \tag{8}$$

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

$$J_{2s}(\boldsymbol{\beta}; \boldsymbol{\lambda}) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(Y_{ij} - \sum_{l=0}^{k} X_{il} \beta_l(t_{ij}) \right)^2 + \sum_{l=0}^{k} \lambda_l \int_a^b \left(\beta_l^{(2)}(t) \right)^2 dt. \qquad (9)$$

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

$$\widehat{\beta}^*_{l(s)}(t; \boldsymbol{\lambda}^*) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{X_{il}}{N s_l^2} \left(Y_{ij} - \sum_{l_1 \neq l} \widehat{\beta}^*_{l_1(s)}(t_{ij}; \boldsymbol{\lambda}^*) X_{il_1} \right) S_{\lambda_l^*, X_l}(t, t_{ij}), \qquad (10)$$

where $\boldsymbol{\lambda}^* = (\lambda_0^*, \cdots, \lambda_k^*)$ with $\lambda_l^* = \lambda_l / s_l^2$. As we can see, our procedures work like the usual smoothing methods of Hoover, *et al.* (1998). However, the random smoothing parameters $\boldsymbol{\lambda}^*$ will cause complexity in deriving the properties of $\widehat{\beta}^*_{l(s)}(t; \boldsymbol{\lambda}^*)$. By contrast, the simultaneous smoothing spline estimators $\widehat{\beta}_{l(s)}(t; \boldsymbol{\lambda})$ can avoid this problem, and are more common in many practical applications. In the following sections, we will focus only on this 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 (2). Even though these methods are fast in computation, the variances of the estimators, as shown later, are not accurate enough. 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.

Assume that the inverse of $E[\mathbf{X}\mathbf{X}^T]$, denoted by $E^{-1}_{\mathbf{X}\mathbf{X}^T}$, exists. By multiplying the both sides of (2) by $\mathbf{X}$ and taking expectations, $\boldsymbol{\beta}(t)$ can be expressed as

$$\boldsymbol{\beta}(t) = E[\mathbf{Z}(t)], \qquad (11)$$

where $\mathbf{Z}(t) = (Z_0(t), \cdots, Z_k(t))^T$ with $Z_l(t) = \sum_{r=0}^{k} e_{lr} X_r Y(t)$ and e_{lr} the $(l, r)th$ element of $E^{-1}_{\mathbf{X}\mathbf{X}^T}$. Since $E[\mathbf{X}\mathbf{X}^T]$ is unknown and is invariant with respect to time t, it is naturally estimated by the sample mean

$$\widehat{E[\mathbf{X}\mathbf{X}^T]} = n^{-1} \sum_{i=1}^{n} (\mathbf{X}_i \mathbf{X}_i^T). \tag{12}$$

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

$$J_c(\beta_l; \lambda_l) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(\widehat{Z}_{ijl} - \beta_l(t_{ij}) \right)^2 + \lambda_l \int_a^b \left(\beta_l^{(2)}(t) \right)^2 dt, \tag{13}$$

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 $\widehat{E}^{-1}_{\mathbf{X}\mathbf{X}}$. The minimizer $\widehat{\beta}_{l(c)}(t; \lambda_l)$ of $J_c(\beta_l; \lambda_l)$ can be expressed as

$$\widehat{\beta}_{l(c)}(t; \lambda_l) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \widehat{Z}_{ijl} S_{\lambda_l}(t, t_{ij}), \tag{14}$$

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 $\widehat{\beta}_{l(s)}(t; \lambda_l)$ is more accurate in variance than $\widehat{\beta}_{l(c)}(t; \lambda_l)$. 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 (2) and the definition of $\mathbf{Z}(t)$, $\mathbf{Z}(t)$ can be reexpressed as

$$\begin{aligned} \mathbf{Z}(t) &= E^{-1}_{\mathbf{X}\mathbf{X}^T}\mathbf{X}Y(t) \\ &= E^{-1}_{\mathbf{X}\mathbf{X}^T}(\mathbf{X}\mathbf{X}^T)\boldsymbol{\beta}(t) + E^{-1}_{\mathbf{X}\mathbf{X}^T}\mathbf{X}\varepsilon(t) \\ &= \boldsymbol{\beta}(t) + \varepsilon^*(t), \end{aligned} \tag{15}$$

where

$$\varepsilon^*(t) = E^{-1}_{\mathbf{X}\mathbf{X}^T}(\mathbf{X}\mathbf{X}^T - E[\mathbf{X}\mathbf{X}^T])\boldsymbol{\beta}(t) + E^{-1}_{\mathbf{X}\mathbf{X}^T}\mathbf{X}\varepsilon(t).$$

One can observe 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 $\widehat{\beta}_{l(s)}(t;\boldsymbol{\lambda})$, $l = 0, \cdots, k$, 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

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

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 $\lambda_l^{-5/4} D_{N,l} \overset{a.s.}{\to} 0$ 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 (7) 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

$$\lambda_l g_l^{(4)}(t) + f(t) g_l(t) = f(t)\beta_l(t), \qquad t \in [0,1], \tag{16}$$

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

$$
\begin{aligned}
H_{\lambda_l}(t,s) &= \frac{(\lambda_l/\gamma^4)^{-1/4}}{2} \Gamma^{(1)}(s)(f(s))^{-1} \sin\left(\frac{\pi}{4} + \frac{(\lambda_l/\gamma^4)^{-1/4}}{\sqrt{2}} |\Gamma(t) - \Gamma(s)|\right) \\
&\quad \times \exp\left(-\frac{(\lambda_l/\gamma^4)^{-1/4}}{\sqrt{2}} |\Gamma(t) - \Gamma(s)|\right), \qquad (17)
\end{aligned}
$$

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 $S_{\lambda_l}(t,s)$, $G_{\lambda_l}(t,s)$ and $H_{\lambda_l}(t,s)$, which will be used in the main results, are stated in the following Lemma.

Lemma 1. Under assumptions A1 and A4, there are positive constants α_1, α_2, κ_1 and κ_2 so that

$$
|G_{\lambda_l}(t,s) - H_{\lambda_l}(t,s)| \leq \kappa_1 \exp\left(-\alpha_1 \lambda_l^{-1/4} |t-s|\right), \qquad (18)
$$

$$
\left|\frac{\partial^\nu}{\partial t^\nu} G_{\lambda_l}(t,s)\right| \leq \kappa_1 \lambda_l^{-(\nu+1)/4} \exp\left(-\alpha_2 \lambda_l^{-1/4} |t-s|\right), \quad 0 \leq \nu \leq 3, \qquad (19)
$$

$$
|S_{\lambda_l, X_l}(t,s) - G_{\lambda_l}(t,s)| \leq \kappa_2 \lambda_l^{-1/2} D_{N,l} \exp\left(-\alpha_1 \lambda_l^{-1/4} |t-s|\right) \ a.s., \qquad (20)
$$

and

$$
\left|\frac{\partial^\nu}{\partial t^\nu} S_{\lambda_l, X_l}(t,s)\right| \leq \kappa_2 \lambda_l^{-(\nu+1)/4} D_{N,l} \exp\left(-\alpha_2 \lambda_l^{-1/4} |t-s|\right) a.s. \qquad (21)
$$

hold uniformly for $t, s \in [0,1]$ and $0 \leq \nu \leq 3$.

Proof. The proof can be derived along the same lines as the proof in Lemma 3.1 of Chiang, *et al.* (2001). □

Let $\mathbf{B}(\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda}))$ and $\mathbf{V}(\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda}))$ 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 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

$$\mathbf{B}(\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})) = -(f(t))^{-1}\left[(E^{-1}_{\mathbf{XX}^T}\Lambda E[\mathbf{XX}^T])\boldsymbol{\beta}^{(4)}(t)\right]^T(1+o(1)), \tag{22}$$

and

$$\mathbf{V}(\widehat{\boldsymbol{\beta}}_{(s)}(t;\boldsymbol{\lambda})) = \frac{1}{2\sqrt{2}N}(f(t))^{-3/4}\sigma^2(t)(E^{-1}_{\mathbf{XX}^T}\Lambda_{\mathbf{XX}^T}E^{-1}_{\mathbf{XX}^T})\,(1+o(1))$$

$$+ \quad (\sum_{i=1}^{n}(\frac{m_i}{N})^2\rho(t,t))E^{-1}_{\mathbf{XX}^T}\,(1+o(1)), \tag{23}$$

where $\Lambda = diag(\lambda_0,\cdots,\lambda_k)$, $\boldsymbol{\beta}^{(4)}(t) = (\beta_0^{(4)}(t),\cdots,\beta_k^{(4)}(t))^T$, and $\Lambda_{\mathbf{XX}^T} = [(\lambda_{l_1l_2}E[X_{l_1}X_{l_2}]$ $)]$ 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 (8) can be derived straightforward from Theorem 1.

Lemma 2. Suppose that assumptions (A1)-(A5) are satisfied and $t \in (0,1)$. Then, for sufficiently large n, the bias and the variance of $\widehat{\beta}_{l(s)}(t;\lambda_l)$, $l = 0,\cdots,k$ are given by

$$B(\widehat{\beta}_{l(s)}(t;\lambda_l)) = -(f(t))^{-1}\beta_l^{(4)}(t)\lambda_l(1+o(1)), \tag{24}$$