Submitted to the Annals of Statistics

MULTIVARIATE VARYING COEFFICIENT MODEL AND ITS APPLICATION TO NEUROIMAGING DATA

BY HONGTU ZHU*, RUNZE LI† AND LINGLONG KONG

University of North Carolina and The Pennsylvania State University

Motivated by recent work studying massive imaging data in the neuroimaging literature, we propose multivariate varying coefficient models (MVCM) for modeling the relation between multiple functional responses and a set of covariates. We develop several statistical inference procedures for MVCM and systematically study their theoretical properties. We first establish the weak convergence of the local linear estimate of coefficient functions, as well as its asymptotic bias and variance, and then we derive asymptotic bias and mean integrated squared error of smoothed individual functions and their uniform convergence rate. We establish the uniform convergence rate of the estimated covariance function of the individual functions and its associated eigenvalue and eigenfunctions. We propose a global test for linear hypotheses of varying coefficient functions, and derive its asymptotic distribution under the null hypothesis. We also propose a simultaneous confidence band for each individual effect curve. We conduct Monte Carlo simulation to examine the finite-sample performance of the proposed procedures.

1. Introduction. With modern imaging techniques, massive imaging data can be observed over both time and space [37, 12, 31, 4, 14, 19]). Such imaging techniques include functional magnetic resonance imaging (fMRI), electroencephalography (EEG), diffusion tensor imaging (DTI), positron emission tomography (PET), and single photon emission-computed tomography (SPECT) among many other imaging techniques. See, for example, a recent review of multiple biomedical imaging techniques and their applications in cancer detection and prevention in Fass [12]. Among them, predominant functional imaging techniques including fMRI and EEG have been widely used in behavioral and cognitive neuroscience to understand

*The research of Zhu and Kong was supported by NSF grant BCS-08-26844 and NIH grants RR025747-01, P01CA142538-01, MH086633, EB005149-01 and AG033387.

†Li's research was supported by NSF grant DMS 0348869, NIH grants P50-DA10075 and R21-DA024260 and NNSF of China 11028103. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF or the NIH.

AMS 2000 subject classifications: Primary 62G05, 62G08; secondary 62G20

Keywords and phrases: Functional response, Global test statistic, Multivariate varying coefficient model, Simultaneous confidence band, Weak convergence

functional segregation and integration of different brain regions in a single subject and across different populations [14, 13, 23]. In DTI, multiple diffusion properties are measured along common major white matter fiber tracts across multiple subjects to characterize the structure and orientation of white matter structure in human brain *in vivo* [2, 3, 49].

A common feature of most imaging techniques is that massive functional data are observed at the same design points, such as time for functional images (e.g., PET and fMRI) or arc length for white matter fiber tract, for different individuals. As an illustration, we plot a diffusion property, called fractional anisotropy (FA), measured at 75 grid points along the right internal capsule tract (Fig. 1 (a)) from 40 randomly selected infants and the values of FA increase with gestational age at nearly all grid points (Figs. 1 (b) and (c)). Scientists are particularly interested in delineating the structure of the variability of these functional FA data and their association with a set of covariates of interest, such as age. Furthermore, as another illustration, we consider the BOLD fMRI signal, which is based on hemodynamic responses secondary to neural activity. We plot the estimated hemodynamic response functions (HRF) corresponding to two stimulus categories from 14 subjects at a selected voxel of a common template space (Fig. 1 (d)). Although the canonical form of the HRF is often used, when applying fMRI in a clinical population with possibly altered hemodynamic responses (Figs. 1 (e) and (f)), using the subjects own HRF in fMRI data analysis may be advantageous because HRF variability is greater across subjects than across brain regions within a subject [29, 1]. We are particularly interested in delineating the structure of the variability of the HRF and their association with a set of covariates of interest, such as diagnostic group and age [28]. A varying-coefficient model, which allows its regression coefficients to vary over some predictors of interest, is a powerful statistical tool for addressing these scientific questions. Since it was systematically introduced to statistical literature by Hastie and Tibshirani [18], many varying-coefficient models have been widely studied and developed for longitudinal, time series, and functional data [9, 41, 8, 11, 39, 20, 33, 22, 21, 44, 17]). However, most varying-coefficient models in the existing literature are developed for univariate response.

Let $\mathbf{y}_i(s) = (y_{i,1}(s), \ldots, y_{i,J}(s))^T$ be an J-dimensional functional response vector for subject i, $i = 1, \ldots, n$, and $\mathbf{x}_i$ be its associated $p \times 1$ vector of covariates of interest. Moreover, s varies in $[0, L_0]$ and denotes the design point, such as time for functional images (e.g., PET). A *multivariate varying coefficient model* is defined as

$$y_{i,j}(s) = \mathbf{x}_i^T B_j(s) + \eta_{i,j}(s) + \epsilon_{i,j}(s) \quad \text{for} \quad j = 1, \ldots, J, \tag{1.1}$$

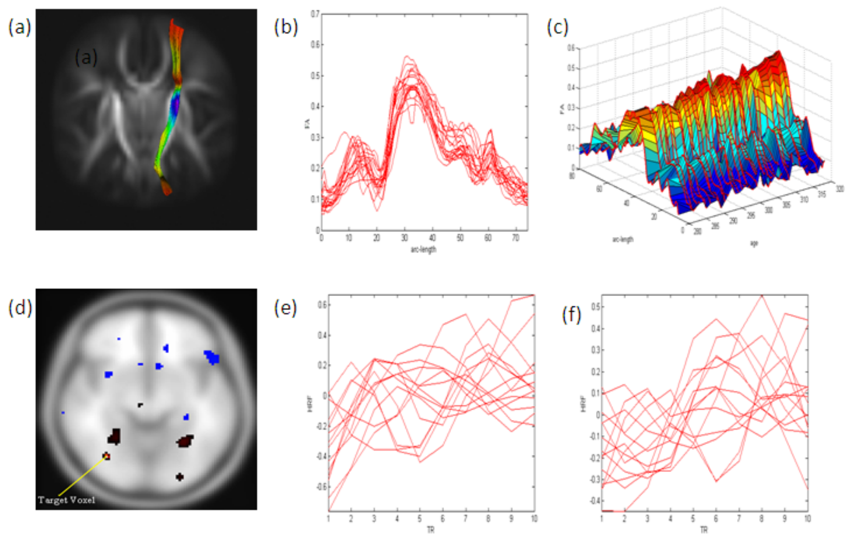

FIG 1. *Representative functional neuroimaging data: (a) the right internal capsule tract, (b) fractional anisotropy (FA) values measured at 75 grid points from 40 randomly selected infants, (c) 3D plot of FA functional curves from 40 randomly selected infants, (d) the Bonferroni corrected p values of the t statistics from a selected slice and a selected voxel, (e) and (f) the estimated hemodynamic response functions (HRF) corresponding to two stimulus categories from 14 subjects.*

where $B_j(s) = (b_{j1}(s), \ldots, b_{jp}(s))^T$ is a $p \times 1$ vector of functions of s, $\epsilon_{i,j}(s)$ are measurement errors, and $\eta_{i,j}(s)$ characterizes individual curve variations from $\mathbf{x}_i^T B_j(s)$. Moreover $\{\eta_{i,j}(s) : s \in [0, L_0]\}$ is assumed to be a stochastic process indexed by $s \in [0, L_0]$ and used to characterize the within-curve dependence. For image data, it is typical that the m functional responses $\mathbf{y}_i(s)$ are measured at the same location for all subjects and exhibit both the within-curve and between-curve dependence structure. Thus, for both ease of notation and the nature of imaging data, it is assumed throughout this paper that $\mathbf{y}_i(s)$ was measured at the same M location points $s_1 = 0 \leq s_2 \leq \ldots \leq s_M = L_0$ for all i.

Most varying coefficient models in the existing literature coincide model (1.1) with $J = 1$ and without the within-curve dependence. Statistical inferences for these varying coefficient models have been well studied. Particularly, Hoover et al. [20] and Wu et al. [42] were among the first to introduce the time-varying coefficient models for analysis of longitudinal data. Recently, Fan and Zhang [11] gave a comprehensive review of various statistical procedures proposed for many varying coefficient models. It is of particular interest in data analysis to construct simultaneous confidence bands (SCB) for any linear combination of $B_j(\cdot)$ instead of pointwise confidence intervals and to develop global test statistics for the general hypothesis testing problem on $B_j(\cdot)$. For varying coefficient models (i.e., $J = 1$) without the within-curve dependence, Fan and Zhang [10] constructed SCB using the limit theory for the maximum of the normalized deviation of the estimate from its expected value. It has been technically difficult to carry out statistical inferences including simultaneous confidence band and global test statistic on $B_j(t)$ in the presence of the within-curve dependence.

There have been several recent attempts to solve this problem in various settings. For time series data, which may be viewed as a case with $n = 1$ and $M \to \infty$, asymptotic SCB for coefficient functions in varying coefficient models can be built by using local kernel regression and a Gaussian approximation result for non-stationary time series [45]. For sparse irregular longitudinal data, Ma et al. [30] construct asymptotic SCB for the mean function of the functional regression model by using piecewise constant spline estimation and a strong approximation result. For functional data, Zhang and Chen [44] adopt the method of "smoothing first, then estimation" and propose a global test statistic for testing $B_j(\cdot)$, but their results cannot be used for constructing SCB for $B_j(\cdot)$.

In this paper, we propose an estimation procedure for the multivariate varying coefficient model (1.1) by using local linear regression techniques, and derive a simultaneous confidence band for the regression coefficient functions. We further develop a test for linear hypotheses of coefficient functions. The major aim of this paper is to investigate the theoretic properties of the proposed estimation procedure and test statistics. The theoretic development is challenging but of great interest for carrying out statistical inferences on $B_j(\cdot)$. The major contributions of this paper are summarized as follows.

1. We first establish the weak convergence of the local-linear estimator of $B_j(\cdot)$, denoted by $\hat{B}_j(\cdot)$, by using advanced empirical process methods [38, 26]. We further derive the bias and asymptotic variance of $\hat{B}_j(\cdot)$. These results provide insight into how the direct estimation procedure for $B_j(\cdot)$ using observations from all subjects outperforms the estimation procedure with the strategy of "smoothing first, then estimation."
2. After calculating $\hat{B}_j(\cdot)$, we reconstruct all individual functions $\eta_{i,k}(\cdot)$ and establish their uniform convergence rates.
3. We derive uniform convergence rates of the proposed estimate for the covariance matrix of $\eta(s)$ and its associated eigenvalue and eigenvector functions by using related results in Li and Hsing [27].

4. Using the weak convergence of the local linear estimator of $B_j(\cdot)$, we further establish the asymptotic distribution of a global test statistic for linear hypotheses of the regression coefficient functions, and construct an asymptotic SCB for each varying coefficient function.

The rest of this paper is organized as follows. In Section 2, we describe MVCM and its estimation procedure and establish the theoretical properties. In Section 3, we establish the asymptotic distribution of a global test statistic for linear hypotheses of the regression coefficient functions and construct an asymptotic SCB for each varying coefficient function. In Section 4, we present two sets of simulation studies with the known ground truth to examine the finite sample performance of the global test statistic and SCB for each individual varying coefficient function. Technical conditions and proofs are given in Section 5.

2. Estimation Procedures. Throughout this paper, we assume that $\boldsymbol{\epsilon}_i(s) = (\epsilon_{i,1}(s), \ldots, \epsilon_{i,J}(s))^T$ and $\boldsymbol{\eta}_i(s) = (\eta_{i,1}(s), \ldots, \eta_{i,J}(s))^T$ are mutually independent, and $\boldsymbol{\eta}_i(s)$ and $\boldsymbol{\epsilon}_i(s)$ are independent and identical copies of $\mathrm{SP}(\mathbf{0}, \Sigma_\eta)$ and $\mathrm{SP}(\mathbf{0}, \Sigma_\epsilon)$, respectively, where $\mathrm{SP}(\mu, \Sigma)$ denotes a stochastic process vector with mean function $\mu(t)$ and covariance function $\Sigma(s,t)$. Moreover, $\epsilon_i(s)$ and $\epsilon_i(t)$ are assumed to be independent for $s \neq t$ and $\Sigma_\epsilon(s,t)$ takes the form of $\Sigma_\epsilon(s,s)\mathbf{1}(s=t)$, where $\mathbf{1}(\cdot)$ is an indicator function. Therefore, the covariance structure of $\mathbf{y}_i(s)$, denoted by $\Sigma_y(s,t)$, is given by

$$\Sigma_y(s,t) = \mathrm{Cov}(\mathbf{y}_i(s), \mathbf{y}_i(t)) = \Sigma_\eta(s,t) + \Sigma_\epsilon(s,s)\mathbf{1}(s=t). \tag{2.1}$$

2.1. *Estimating varying coefficient functions.* We will employ local linear regression [7] to estimate the coefficient functions $B_j(s)$. Specifically, we apply the Taylor expansion for $B_j(s_m)$ at s as follows

$$B_j(s_m) = B_j(s) + \dot{B}_j(s)(s_m - s) = A_j(s)\mathbf{z}_{h_j}(s_m - s), \tag{2.2}$$

where $\mathbf{z}_{h_j}(s_m - s) = (1, (s_m - s)/h_j)^T$ and $A_j(s) = [B_j(s) \quad h_j\dot{B}_j(s)]$ is a $p \times 2$ matrix, in which $\dot{B}_j(s) = (\dot{b}_{j1}(s), \ldots, \dot{b}_{jp}(s))^T$ is a $p \times 1$ vector and $\dot{b}_{jl}(s) = db_{jl}(s)/ds$ for $l = 1, \ldots, p$. Let $K(\cdot)$ be a kernel function and $K_h(\cdot) = h^{-1}K(\cdot/h)$ be the rescaled kernel function with a bandwidth h. We estimate $A_j(s)$ by minimizing the following weighted least squares function:

$$\sum_{i=1}^{n}\sum_{m=1}^{M}[y_{i,j}(s_m) - \mathbf{x}_i^T A_j(s)\mathbf{z}_{h_j}(s_m - s)]^2 K_{h_j}(s_m - s). \tag{2.3}$$

Let us now introduce some matrix operators. Let $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any vector $\mathbf{a}$ and $C \otimes D$ be the Kronecker product of two matrices C and D. For an $M_1 \times M_2$ matrix $C = (c_{j,l})$, denote $\mathrm{vec}(C) = (c_{1,1}, \ldots, c_{1,M_2}, \ldots, c_{M_1,1}, \ldots, c_{M_1,M_2})^T$. Let $\hat{A}_j(s)$ be the minimizer of (2.3). Then

$$\mathrm{vec}(\hat{A}_j(s)) = \Sigma(h_j, s)^{-1} \sum_{i=1}^{n} \sum_{m=1}^{M} K_{h_j}(s_m - s)[\mathbf{x}_i \otimes \mathbf{z}_{h_j}(s_m - s)] y_{i,j}(s_m), \tag{2.4}$$

where $\Sigma(h_j, s) = \sum_{i=1}^{n} \sum_{m=1}^{M} K_{h_j}(s_m - s)[\mathbf{x}_i^{\otimes 2} \otimes \mathbf{z}_{h_j}(s_m - s)^{\otimes 2}]$. Thus, we have

$$\hat{B}_j(s) = (\hat{b}_{j1}(s), \ldots, \hat{b}_{jp}(s))^T = [I_p \otimes (1, 0)]\mathrm{vec}(\hat{A}_j(s)), \tag{2.5}$$

where I_p is a $p \times p$ identity matrix.

Define $u_r(K) = \int t^r K(t)dt$, $v_r(K) = \int t^r K^2(t)dt$, $\dot{f}(s) = df(s)/ds$, $\ddot{f}(s) = d^2 f(s)/ds^2$, $\dddot{f}(s) = d^3 f(s)/ds^3$, and $g^{(l_1,l_2)}(s,t) = \partial^{a+b} g(s,t)/\partial^a s \partial^b t$ for any smooth functions $f(s)$ and $g(s,t)$, where r, a and b are any nonnegative integers. Let $\mathbf{H} = \mathrm{diag}(h_1, \ldots, h_J)$, $\mathbf{B}(s) = [B_1(s), \ldots, B_J(s)]$, $\hat{\mathbf{B}}(s) = [\hat{B}_1(s), \ldots, \hat{B}_J(s)]$ and $\ddot{\mathbf{B}}(s) = [\ddot{B}_1(s), \ldots, \ddot{B}_J(s)]$, where $\ddot{B}_j(s) = (\ddot{b}_{j1}(s), \ldots, \ddot{b}_{jp}(s))^T$ is a $p \times 1$ vector. The following theorem establishes the weak convergence of $\{\hat{B}(s), s \in [0, L_0]\}$, which is essential for constructing global test statistics and SCB for $\mathbf{B}(\cdot)$. The assumptions and proofs of all theorems are given in Section 5.

Theorem 1. *Suppose that Assumptions (C1)-(C6) in Section 5 hold. The following results hold:*

(i) $\sqrt{n}\{vec(\hat{\mathbf{B}}(s) - \mathbf{B}(s) - 0.5\ddot{\mathbf{B}}(s)\mathbf{U}_2(K; s, \mathbf{H})\mathbf{H}^2[1 + o_p(1)]) : s \in [0, L_0]\}$ *converges weakly to a centered Gaussian process* $G(\cdot)$ *with covariance matrix* $\Sigma_\eta(s, s') \otimes \Omega_X^{-1}$, *where* $\Omega_X = E[\mathbf{x}^{\otimes 2}]$ *and* $\mathbf{U}_2(K; s, \mathbf{H})$ *is a* $J \times J$ *diagonal matrix, whose diagonal elements will be defined in Lemma 5 of Section 5.*

(ii) The asymptotic bias and conditional variance of $\hat{B}_j(s)$ *given* $\mathcal{S} = \{s_1, \ldots, s_M\}$ *for* $s \in (0, L_0)$ *are, respectively, given by* $0.5h_j^2 u_2(K)\ddot{B}_j(s)[1 + o_p(1)]$ *and*

$$\begin{aligned}
&n^{-1}e_n(s) + (nMh_j)^{-1}\pi(s)^{-1}v_0(K)[\Sigma_{\eta,jj}(s,s) + \Sigma_{\epsilon,jj}(s,s)]\Omega_X^{-1}[1 + O(h_j)] \\
&+n^{-1}\{\Sigma_{\eta,jj}(s,s) + h_j^2 u_2(K)[\Sigma_{\eta,jj}^{(2,0)}(s,s)\pi(s) + 2\Sigma_{\eta,jj}^{(1,0)}(s,s)\dot{\pi}(s) \\
&+\Sigma_{\eta,jj}(s,s)\ddot{\pi}(s)]\pi(s)^{-1} + O_p(M^{-1}) + o_p(h_j^2)\}\Omega_X^{-1},
\end{aligned}$$

where $e_n(s) = O_p((Mh_j)^{-1/2})$ *is a random matrix of* $\mathcal{S}$ *and* $\Sigma_\eta(s, s')$ *for* $s \neq s'$ *with* $E[e_n(s)] = 0$. *This will be defined in Section 5.*

Before discussing practical implementation issues of the proposed procedure, let us discuss the implications of Theorem 1.

1. The major challenge in proving Theorem 1 (i) is dealing with within-subject dependence. This is because the dependence between $\boldsymbol{\eta}(s)$ and $\boldsymbol{\eta}(s')$ in the newly proposed multivariate varying coefficient model does not converge to zero due to the within-curve dependence. It is worth noting that for any given s, the corresponding asymptotic normality of $\hat{\mathbf{B}}(s)$ may be established by using related techniques in Zhang and Chen [44]. However, the marginal asymptotic normality does not imply the weak convergence of $\hat{\mathbf{B}}(s)$ as a stochastic process in $[0, L_0]$, since we need to verify the asymptotic continuity of $\{\hat{\mathbf{B}}(s) : s \in [0, L_0]\}$ to establish its weak convergence. In addition, Zhang and Chen [44] considered "smoothing first, then estimation", which requires a stringent assumption such that $n = O(M^{4/5})$. Readers are referred to Condition A.4 and Theorem 4 in Zhang and Chen [44] for more details. In contrast, directly estimating $\mathbf{B}(s)$ using local kernel smoothing avoids such stringent assumption on the numbers of grid points and subjects.
2. When $\mathbf{x}_i$ only contains the intercept, model (1.1) reduces to the standard model for functional principal component analysis (FPCA). Although there is an extensive literature on establishing the theoretical results of FPCA and its extensions [27, 43, 16, 33, 32], the existing results are primarily marginal convergence and uniform convergence rate of mean function, and covariance function and associated eigenvalues and eigenfunctions. By using Theorem 1 (i), we can establish the weak convergence of the mean functional curve in the model for FPCA, which is a new result to the best of our knowledge. Moreover, it is interesting to extend model (1.1) by incorporating covariate into $\eta_{i,j}(s)$ [24] and investigate its associated statistical methods.
3. Theorem 1 (ii) only provides us the asymptotic bias and conditional variance of $\hat{B}_j(s)$ given $\mathcal{S}$ for the interior points of $(0, L_0)$. The asymptotic bias and conditional variance at the boundary points 0 and L_0 are given in Lemma 5 in Section 5. The asymptotic bias of $\hat{B}_j(s)$ is of the order h_j^2, as the one in nonparametric regression setting. Moreover, the asymptotic conditional variance of $\hat{B}_j(s)$ has a complicated form due to the within-curve dependence. The leading term in the asymptotic conditional variance is at the order of n^{-1}, which is slower than the standard nonparametric rate $(nMh_j)^{-1}$ with the assumption $h_j \to 0$ and $Mh_j \to \infty$. There is a new term at the order of $n^{-1}h_j^2$, which is also introduced by the within-curve dependence. The new term $e_n(s)$ is at the order of $(Mh_j)^{-1/2}$ and has zero mean. If we consider the asymptotic conditional variance of $\hat{B}_j(s)$, then $e_n(s)$ is dropped due to $E[e_n(s)] = 0$.

4. Choosing an optimal bandwidth h_j is not a trivial task for model (1.1). Generally, any bandwidth h_j satisfying the assumption $h_j \to 0$ and $Mh_j \to \infty$ can ensure the weak convergence of $\{\hat{\mathbf{B}}(s) : s \in [0, L_0]\}$. Ignoring the terms ne_n^{-1} and $n^{-1}h_j^2 O(1)$ leads to an optimal bandwidth for estimating $\mathbf{B}(s)$, $\hat{h}_j = O_p((nM)^{-1/5})$. In this case, $n^{-1}h_j^2$ and $(nM)^{-1}$ reduce to $O_p(n^{-7/5}M^{-2/5})$ and $(nM)^{-6/5}$, respectively, and their contributions depend on the relative size of n over M.

In practice, we may select the bandwidth h_j by using cross-validation. Specifically, for each j, we pool the data from all n subjects and select a bandwidth h_j, denoted by $\hat{h}_j$, by minimizing the cross-validation score given by

$$\mathrm{CV}_{j,1}(h_j) = (nM)^{-1} \sum_{i=1}^{n} \sum_{m=1}^{M} [y_{i,j}(s_m) - \mathbf{x}_i^T \hat{B}_j(s_m, h_j)^{(-i)}]^2, \tag{2.6}$$

where $\hat{B}_j(s, h_j)^{(-i)}$ is the local linear estimator of $B_j(s)$ with the bandwidth h_j based on data excluding all the observations from the i-th subject.

2.2. *Smoothing individual functions.* We also employ the local linear regression technique to estimate all individual functions $\eta_{i,j}(s)$. Specifically, we have the Taylor expansion for $\eta_{i,j}(s_m)$ at s:

$$\eta_{i,j}(s_m) = \mathbf{d}_{i,j}(s)^T \mathbf{z}_{h_j^{(2)}}(s_m - s), \tag{2.7}$$

where $\mathbf{d}_{i,j}(s) = (\eta_{i,j}(s), h_j^{(2)} \dot{\eta}_{i,j}(s))^T$ is a 2×1 vector. We develop an algorithm to estimate $\mathbf{d}_{i,j}(s)$ as follows. For each k and i, we estimate $\mathbf{d}_{i,j}(s)$ by minimizing the weighted least squares function

$$\sum_{m=1}^{M} [y_{i,j}(s_m) - \mathbf{x}_i^T \hat{B}_j(s_m) - \mathbf{d}_{i,j}(s)^T \mathbf{z}_{h_j^{(2)}}(s_m - s)]^2 K_{h_j^{(2)}}(s_m - s). \tag{2.8}$$

Thus,

$$\begin{aligned} \hat{\mathbf{d}}_{i,j}(s) &= [\sum_{m=1}^{M} K_{h_j^{(2)}}(s_m - s) \mathbf{z}_{h_j^{(2)}}(s_m - s)^{\otimes 2}]^{-1} \\ &\quad \times \sum_{m=1}^{M} K_{h_j^{(2)}}(s_m - s) \mathbf{z}_{h_j^{(2)}}(s_m - s)[y_{i,j}(s_m) - \mathbf{x}_i^T \hat{B}_j(s_m)]. \end{aligned}$$

Let $\mathbf{e}_{1,2} = (1, 0)^T$. Then, $\eta_{i,j}(s)$ can be estimated by

$$\hat{\eta}_{i,j}(s) = \mathbf{e}_{1,2}^T \hat{\mathbf{d}}_{i,j}(s) = \sum_{m=1}^{M} \tilde{K}^0_{h_j^{(2)}}(s_j - s, s)[y_{i,j}(s_m) - \mathbf{x}_i^T \hat{B}_j(s_m)], \tag{2.9}$$

where $\tilde{K}^0_{h_j^{(2)}}(\cdot,\cdot)$ are the empirical equivalent kernels [7]. Finally, let $S_{i,j}$ be the smoother matrix for the j-th measurement of the i-th subject and $R_{i,j} = (y_{i,j}(s_1) - \mathbf{x}_i^T \hat{B}_j(s_1), \ldots, y_{i,j}(s_M) - \mathbf{x}_i^T \hat{B}_j(s_M))^T$, we can obtain

$$\hat{\boldsymbol{\eta}}_{i,j} = (\hat{\eta}_{i,j}(s_1), \ldots, \hat{\eta}_{i,j}(s_M))^T = S_{i,j} R_{i,j}. \tag{2.10}$$

We next study the asymptotic bias and covariance of $\hat{\eta}_{i,j}(s)$ as follows. We distinguish between two cases. The first one is conditioning on the design points in $\mathcal{S}$, $\mathbf{X}$, and $\boldsymbol{\eta}$. The other is conditioning on the design points in $\mathcal{S}$ and $\mathbf{X}$. We define $K^*((s-t)/h) = \int K(u)K(u+(s-t)/h)du$.

Theorem 2. *Under assumptions (C1)-(C6) in Section 5, the following results hold for all $s \in (0, L_0)$.*

(a) Conditioning on $(\mathcal{S}, \mathbf{X}, \boldsymbol{\eta})$, we have

$$\begin{aligned}
& Bias[\hat{\eta}_{i,j}(s)|\mathcal{S}, \boldsymbol{\eta}, \mathbf{x}_i] \\
= \;& 0.5u_2(K)[\ddot{\eta}_{i,j}(s)h_j^{(2)2} + \mathbf{x}_i^T \ddot{B}_j(s_m)h_j^2][1+o_p(1)] + O_p(n^{-1/2}), \\
& Cov[\hat{\eta}_{i,j}(s), \hat{\eta}_{i,j}(t)|\mathcal{S}, \boldsymbol{\eta}, \mathbf{x}_i] \\
= \;& K^*((s-t)/h_j^{(2)})\Sigma_{\epsilon,jj}(s,s)\pi(t)^{-1}(Mh_j^{(2)})^{-1}[1+o_p(1)] \\
& \quad -\mathbf{x}_i^T \Omega_X^{-1}\mathbf{x}_i (nMh_j)^{-1}\pi(s)^{-1}\pi(t)^{-1}O_p(1).
\end{aligned}$$

(b) The asymptotic bias and covariance of $\hat{\eta}_{i,j}(s)$ conditioning on $\mathcal{S}$ and $\mathbf{X}$ are given by

$$\begin{aligned}
& Bias[\hat{\eta}_{i,j}(s)|\mathcal{S}, \mathbf{X}] = 0.5u_2(K)\mathbf{x}_i^T \ddot{B}_j(s_m)h_j^2[1+o_p(1)], \\
& Cov(\hat{\eta}_{i,j}(s) - \eta_{i,j}(s), \hat{\eta}_{i,j}(t) - \eta_{i,j}(t)|\mathcal{S}, \mathbf{X}) \\
= \;& [1+o_p(1)][0.25u_2(K)^2 h_j^{(2)4}\Sigma_{\eta,jj}^{(2,2)}(s,t) + K^*((s-t)/h_j^{(2)}) \\
& \quad \times \Sigma_{\epsilon,jj}(s,s)\pi(t)^{-1}(Mh_j^{(2)})^{-1} + n^{-1}\mathbf{x}_i^T \Omega_X^{-1}\mathbf{x}_i \Sigma_{\eta,jj}(s,t)].
\end{aligned}$$

(c) The mean integrated squared error (MISE) of all $\hat{\eta}_{i,j}(s)$ is given by

$$\begin{aligned}
& n^{-1}\sum_{i=1}^n \int_0^{L_0} E\{[\hat{\eta}_{i,j}(s) - \eta_{i,j}(s)]^2|\mathcal{S}\}\pi(s)ds \qquad (2.11) \\
= \;& [1+o_p(1)] \times \{v_0(K)(Mh_j^{(2)})^{-1}\int_0^{L_0}\Sigma_{\epsilon,jj}(s,s)\pi(s)ds \\
& +n^{-1}\int_0^{L_0}\Sigma_{\eta,jj}(s,s)\pi(s)ds \\
& +0.25u_2^2(K)\int_0^{L_0}[\ddot{B}_j(s)^T\Omega_X\ddot{B}_j(s)h_j^4 + \Sigma_{\eta,jj}^{(2,2)}(s,s)h_j^{(2)4}]\pi(s)ds\}.
\end{aligned}$$

(d) The optimal bandwidth for minimizing MISE (2.11) is given by

$$\hat{h}_j^{(2)} = \left(\frac{v_0(K) \int_0^{L_0} \Sigma_{\epsilon,jj}(s,s)\pi(s)ds}{u_2^2(K) \int_0^{L_0} \Sigma_{\eta,jj}^{(2,2)}(s,s)\pi(s)ds} \right)^{1/5} M^{-1/5}. \tag{2.12}$$

(e) The first order LPK reconstructions $\hat{\eta}_{i,j}(s)$ using $\hat{h}_j^{(2)}$ in (2.12) satisfy

$$\sup_{s\in[0,L_0]} |\hat{\eta}_{i,j}(s) - \eta_{i,j}(s)| = O_p(|\log(M)|^{1/2} M^{-2/5} + h_j^2 + n^{-1/2}), \tag{2.13}$$

for $i = 1, \cdots, n$.

REMARK. Theorem 2 characterizes the statistical properties of smoothing individual curves $\boldsymbol{\eta}_i(s)$ after first estimating $B_j(s)$. Conditioning on individual curves $\eta_{i,j}(s)$, Theorem 2 (a) shows that $\text{Bias}[\hat{\eta}_{i,j}(s)|\mathcal{S}, \mathbf{X}, \boldsymbol{\eta}]$ is associated with $0.5u_2(K)\mathbf{x}_i^T \ddot{B}_j(s_m)h_j^2$, which is the bias term of $\hat{B}_j(s)$ introduced in the estimation step, and $0.5u_2(K)\ddot{\eta}_{i,j}(s)h_j^{(2)2}$ is introduced in the smoothing individual functions step. Without conditioning on $\eta_{i,j}(s)$, Theorem 2 (b) shows that the bias of $\hat{\eta}_{i,j}(s)$ is mainly controlled by the bias in the estimation step. The MISE of $\hat{\eta}_{i,j}(s)$ in Theorem 2 (c) is the sum of $O_p(n^{-1} + h_j^4)$ introduced by the estimation of $B_j(s)$ and $O_p((Mh_j^{(2)})^{-1} + h_j^{(2)4})$ introduced by the reconstruction of $\eta_{i,j}(s)$. The optimal bandwidth for minimizing the MISE of $\hat{\eta}_{i,j}(s)$ is a standard bandwidth for LPK. If we use the optimal bandwidth in Theorem 2 (d), then the MISE of $\hat{\eta}_{i,j}(s)$ can achieve the order of $n^{-1} + h_j^4 + M^{-4/5}$.

For each j, we pool the data from all n subjects and select the optimal bandwidth $h_j^{(2)}$, denoted by $\hat{h}_j^{(2)}$, by minimizing the generalized cross-validation score given by

$$\text{GCV}_{j,2}(h_j^{(2)}) = \sum_{i=1}^{n} \frac{R_{i,j}^T (I_M - S_{i,j})^T (I_M - S_{i,j}) R_{i,j}}{[1 - M^{-1}\text{tr}(S_{i,j})]^2}. \tag{2.14}$$

Based on $\hat{h}_j^{(2)}$, we can use (2.9) to estimate $\eta_{i,j}(s)$ and $\boldsymbol{\eta}_i(s)$ for all i and j.

2.3. *Functional principal component analysis.* We consider a spectral decomposition of $\Sigma_\eta(s,t) = (\Sigma_{\eta,jj'}(s,t))$ and a representation of $\eta_{i,j}(s)$ for each j. After obtaining $\hat{\boldsymbol{\eta}}_i(s)$, we estimate $\Sigma_\eta(s,t)$ by using the empirical covariance of the estimated $\hat{\boldsymbol{\eta}}_i(s)$ as follows:

$$\hat{\Sigma}_\eta(s,t) = (n-p)^{-1} \sum_{i=1}^{n} \hat{\boldsymbol{\eta}}_i(s) \hat{\boldsymbol{\eta}}_i(t)^T.$$

For each j, a spectral decomposition of $\Sigma_{\eta,jj}(s,t)$ is assumed as follows:

$$\Sigma_{\eta,jj}(s,t) = \sum_{l=1}^{\infty} \lambda_{j,l}\psi_{j,l}(s)\psi_{j,l}(t), \tag{2.15}$$

where $\lambda_{j,1} \geq \lambda_{j,2} \geq \ldots \geq 0$ are ordered values of the eigenvalues of a linear operator determined by $\Sigma_{\eta,jj}$ with $\sum_{l=1}^{\infty} \lambda_{j,l} < \infty$ and the $\psi_{l,k}(t)$'s are the corresponding orthonormal eigenfunctions (or principal components) [27, 43, 16]. The eigenfunctions form an orthonormal basis on the space of square-integrable functions on $[0, L_0]$, and thus $\eta_{i,j}(s)$ admits the Karhunen-Loeve expansion as $\eta_{i,j}(s) = \sum_{l=1}^{\infty} \xi_{ij,l}\psi_{j,l}(t)$, where $\xi_{ij,l} = \int_0^{L_0} \eta_{i,j}(s)\psi_{j,l}(s)ds$ is referred to as the (j,l)-th functional principal component scores of the ith subject. The $\xi_{ij,l}$ are uncorrelated random variables with $E(\xi_{ij,l}) = 0$ and $E(\xi_{ij,l}^2) = \lambda_{j,l}$. Furthermore, for $j \neq j'$, we have

$$\Sigma_{\eta,jj'} = \sum_{l=1}^{\infty}\sum_{l'=1}^{\infty} E(\xi_{j,l}\xi_{j',l'})\psi_{j,l}(s)\psi_{j',l'}(t).$$

Following Rice and Silverman [34], we can calculate the spectral decomposition of $\hat{\Sigma}_{\eta,jj}(s,t)$ for each j as follows:

$$\hat{\Sigma}_{\eta,jj}(s,t) = \sum_{l} \hat{\lambda}_{j,l}\hat{\psi}_{j,l}(s)\hat{\psi}_{j,l}(t), \tag{2.16}$$

where $\hat{\lambda}_{j,1} \geq \hat{\lambda}_{j,2} \geq \ldots \geq 0$ are estimated eigenvalues and the $\hat{\psi}_{j,l}(t)$s' are the corresponding estimated principal components. Furthermore, the (j,l)th functional principal component scores can be computed using $\hat{\xi}_{ij,l} = \sum_{m=1}^{M} \hat{\eta}_{i,j}(s_m)\hat{\psi}_{j,l}(s_m)(s_m - s_{m-1})$. We further show the uniform convergence rate of $\hat{\Sigma}_{\eta}(s,t)$ and its associated eigenvalues and eigenfunctions. This result is useful for constructing the global and local test statistics for testing the covariate effects.

Theorem 3. *(i) Under assumptions (C1)-(C7) in Section 5, it follows that*

$$\sup_{(s,t)\in[0,L_0]^2} |\hat{\Sigma}_{\eta}(s,t) - \Sigma_{\eta}(s,t)| = O_p(n^{-1/2} + (Mh_j^{(2)})^{-1} + h_j^2 + h_j^{(2)2} + (\log n/n)^{1/2}).$$

(ii) Under assumption (C1)-(C9) in Section 5, if the optimal bandwidths $h_j^{(m)}$ for $m = 1, 2$ are used to reconstruct $\hat{B}_j(s)$ and $\hat{\eta}_{i,j}(s)$ for all j, then for $l = 1, \ldots, K_j$, we have the following results:

(a) $\int_0^{L_0} [\hat{\psi}_{j,l}(s) - \psi_{j,l}(s)]^2 ds = O_p(n^{-1/2} + (Mh_j^{(2)})^{-1} + h_j^2 + h_j^{(2)2} + (\log n/n)^{1/2})$*;*

(b) $|\hat{\lambda}_{j,l} - \lambda_{j,l}| = O_p(n^{-1/2} + (Mh_j^{(2)})^{-1} + h_j^2 + h_j^{(2)2} + (\log n/n)^{1/2})$.

REMARK. Theorem 3 characterizes the uniform weak convergence rates of $\hat{\Sigma}_\eta(s,t)$, $\hat{\psi}_{j,l}(\cdot)$, and $\hat{\lambda}_{j,l}$ for all j. It can be regarded as an extension of Theorems 3.3-3.6 in Li and Hsing [27], which established the uniform strong convergence rates of these estimates with the sole presence of intercept and $J=1$ in model (1.1). Another difference is that Li and Hsing [27] employed all cross products $y_{ij}y_{ik}$ for $j \neq k$ and then used the local polynomial kernel to estimate $\Sigma_\eta(s,t)$. As discussed in Li and Hsing [27], their approach can release the assumption on the differentiability of the individual curves. In contrast, following Hall et al. [16] and Zhang and Chen [44], we directly fit a smooth curve to $\eta_{ij}(s)$ for each i and estimate $\Sigma_\eta(s,t)$ by the sample covariance functions. Our approach is computationally simple and can ensure that all $\hat{\Sigma}_{\eta,j}(s,t)$ are positive definite, whereas the approach in Li and Hsing [27] cannot. Since we cannot release the assumption on the differentiability of the individual curves in establishing the weak convergence of $\hat{B}_j(s)$, we do not employ Li and Hsing [27]'s approach.

We construct a nonparametric estimator of the covariance matrices $\Sigma_\epsilon(s,s)$ as follows. Let $\hat{\boldsymbol{\epsilon}}_i(s_m) = \mathbf{y}_i(s_m) - \hat{\mathbf{B}}(s_m)^T\mathbf{x}_i - \hat{\boldsymbol{\eta}}_i(s_m)$ be estimated residuals for $i=1,\ldots,n$ and $m=1,\ldots,M$. We consider the kernel estimate of $\Sigma_\epsilon(s,s)$ given by

$$\hat{\Sigma}_\epsilon(s,s) = (n-J)^{-1}\sum_{i=1}^{n}\sum_{m=1}^{M}\frac{K_{h^{(3)}}(s_m-s)[\hat{\boldsymbol{\epsilon}}_i(s_m)]^{\otimes 2}}{\sum_{m=1}^{M}K_{h^{(3)}}(s_m-s)}. \tag{2.17}$$

Let $\tilde{\Sigma}_\epsilon(s_m,s_m) = (n-m)^{-1}\sum_{i=1}^{n}[\hat{\boldsymbol{\epsilon}}_i(s_m)]^{\otimes 2}$ for $m=1,\ldots,M$. To select the optimal bandwidth $h^{(3)}$, denoted by $\hat{h}^{(3)}$, we minimize the cross-validation score given by

$$\mathrm{CV}(h^{(3)}) = (nM)^{-1}\sum_{i=1}^{n}\sum_{m=1}^{M}\mathrm{tr}\{[\hat{\boldsymbol{\epsilon}}_i(s_m)^{\otimes 2} - \hat{\Sigma}_\epsilon(s_m,s_m;h^{(3)})^{(-i)}]\tilde{\Sigma}_\epsilon(s_m,s_m)^{-1}\}^2,$$

where $\hat{\Sigma}_\epsilon(s,s;h^{(3)})^{(-i)}$ is the weighted least squares estimator of $\hat{\Sigma}_\epsilon(s,s)$ based on observed data with the observations from the $i-$th subject entirely excluded.

We obtain the uniform convergence of $\hat{\Sigma}_\epsilon(s,s)$. We do not include the uniform convergence rate of $\hat{\Sigma}_\epsilon(s,s)$ here because our focus is to derive the asymptotic distribution of a global test statistic, which does not involve $\Sigma_\epsilon(s,s)$.

Corollary 1. *Under assumptions (C1)-(C8) in Section 5, it follows that*

$$\sup_{s\in[0,L_0]}|\hat{\Sigma}_\epsilon(s,s) - \Sigma_\epsilon(s,s)| = o_p(1).$$

3. Test of Hypothesis and Simultaneous Confidence Bands. In this section, we study global tests for linear hypotheses of coefficient functions and SCB for each varying coefficient function. They are essential for statistical inference on the coefficient functions.

3.1. *Hypothesis test.* Consider the linear hypotheses of $\mathbf{B}(s)$ as follows:

$$H_0 : \mathbf{C}\text{vec}(\mathbf{B}(s)) = \mathbf{b}_0(s) \text{ for all } s \text{ vs. } H_1 : \mathbf{C}\text{vec}(\mathbf{B}(s)) \neq \mathbf{b}_0(s), \tag{3.1}$$

where $\mathbf{C}$ is a $r \times Jp$ matrix with rank r, and $\mathbf{b}_0(s)$ is a given $r \times 1$ vector of functions. Define a global test statistic S_n as

$$S_n = \int_0^{L_0} \mathbf{d}(s)^T [\mathbf{C}(\hat{\Sigma}_\eta(s,s) \otimes \hat{\Omega}_X^{-1})\mathbf{C}^T]^{-1}\mathbf{d}(s)ds, \tag{3.2}$$

where $\hat{\Omega}_X = \sum_{i=1}^n \mathbf{x}_i^{\otimes 2}$ and $\mathbf{d}(s) = \mathbf{C}\text{vec}(\hat{\mathbf{B}}(s) - \text{bias}(\hat{\mathbf{B}}(s))) - \mathbf{b}_0(s)$. We will derive the limiting distribution of S_n under H_0.

To calculate S_n, we need to estimate the bias of $\hat{B}_j(s)$ for all j. Based on (2.5), we have

$$\begin{aligned} \text{bias}(\hat{B}_j(s)) = [I_p \otimes (1,0)]\text{vec}(\Sigma(\hat{h}_j, s)^{-1} \sum_{i=1}^n \sum_{m=1}^M K_{\hat{h}_j}(s_m - s) \\ \times [\mathbf{x}_i \otimes \mathbf{z}_{\hat{h}_j}(s_m - s)]\mathbf{x}_i(s_m)^T [B_j(s_m) - \hat{A}_j(s)\mathbf{z}_{\hat{h}_j}(s_m - s)]). \end{aligned} \tag{3.3}$$

By using Taylor's expansion, we have

$$B_j(s_m) - \hat{A}_j(s)\mathbf{z}_{\hat{h}_j}(s_m - s) \approx 2^{-1}\ddot{B}_j(s)(s_m - s)^2 + 6^{-1}\dddot{B}_j(s)(s_m - s)^3,$$

where $\ddot{B}_j(s) = d^2B_j(s)/ds^2$ and $\dddot{B}_j(s) = d^3B_j(s)/ds^3$. Following the pre-asymptotic substitution method of Fan and Gijbels [7], we replace $B_j(s_m) - \hat{A}_j(s)\mathbf{z}_{\hat{h}_j}(s_m - s)$ by $2^{-1}\widehat{\ddot{B}}_j(s)(s_m - s)^2 + 6^{-1}\widehat{\dddot{B}}_j(s)(s_m - s)^2$, in which $\widehat{\ddot{B}}_j(s)$ and $\widehat{\dddot{B}}_j(s)$ are estimators obtained by using local cubic fit with a pilot bandwidth selected by $\text{CV}_{j,1}(h_j)$ in (2.6).

We formally characterize the asymptotic distribution of S_n as follows. Let $X_{\mathbf{C}}(\cdot)$ be a Gaussian process with zero mean and covariance structure $\Sigma_{\mathbf{C}}(s,t)$, which is the limit of $\hat{\Sigma}_{\mathbf{C}}(s,t)$ given by

$$[\mathbf{C}(\hat{\Sigma}_\eta(s,s) \otimes \hat{\Omega}_X^{-1})\mathbf{C}^T]^{-1}[\mathbf{C}(\hat{\Sigma}_\eta(s,t) \otimes \hat{\Omega}_X^{-1})\mathbf{C}^T][\mathbf{C}(\hat{\Sigma}_\eta(t,t) \otimes \hat{\Omega}_X^{-1})\mathbf{C}^T]^{-1}.$$

It follows from Theorem 1 that $\sqrt{n}[\mathbf{C}(\hat{\Sigma}_\eta(s,s) \otimes \hat{\Omega}_X^{-1})\mathbf{C}^T]^{-1}\mathbf{d}(s)$ converges weakly to $X_{\mathbf{C}}(s)$. Therefore, let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes; it follows from the continuous mapping theorem that as both n and M converge to infinity, we have

$$S_n \Rightarrow \int_0^{L_0} X_{\mathbf{C}}(s)^T X_{\mathbf{C}}(s) ds. \tag{3.4}$$

Theorem 4. *If assumptions (C1)-(C8) in Section 5 are true, then (3.4) is true.*

REMARK. Theorem 4 is similar to Theorem 7 of Zhang and Chen [44]. Both characterize the asymptotic distribution of S_n. In particular, Zhang and Chen [44] delineate the distribution of $\int_0^{L_0} X_{\mathbf{C}}(s)^T X_{\mathbf{C}}(s) ds$ as a χ^2-type mixture. All discussions associated Theorem 7 with Zhang and Chen [44] are valid here and therefore, we do not repeat them for the sake of space.

It is difficult to directly calculate the percentiles of $\int_0^{L_0} X_{\mathbf{C}}(s)^T X_{\mathbf{C}}(s) ds$. Thus, we propose using a wild bootstrap method to obtain critical values of S_n. The wild bootstrap consists of the following three steps.

Step 1, Fit model (1.1) under the null hypothesis H_0, which yields $\hat{B}^*(s_m)$, $\hat{\boldsymbol{\eta}}^*_{i,o}(s_m)$ and $\hat{\boldsymbol{\epsilon}}^*_{i,o}(s_m)$ for $i = 1, \ldots, n$ and $m = 1, \ldots, M$.

Step 2, Generate a random sample $\tau_i^{(g)}$ and $\tau_i(s_m)^{(g)}$ from a $N(0, 1)$ generator for $i = 1, \ldots, n$ and $m = 1, \ldots, M$ and then construct

$$\hat{\mathbf{y}}_i(s_m)^{(g)} = \hat{B}^*(s)^T \mathbf{x}_i + \tau_i^{(g)} \hat{\boldsymbol{\eta}}^*_{i,o}(s_m) + \tau_i(s_m)^{(g)} \hat{\boldsymbol{\epsilon}}^*_{i,o}(s_m).$$

Then, based on $\hat{\mathbf{y}}_i(s_m)^{(g)}$, we recalculate $\mathbf{H}^{(1)}$, $\hat{\mathbf{B}}(s)^{(g)}$, $\text{bias}(\hat{\mathbf{B}}(s)^{(g)})$, and $\mathbf{d}(s)^{(g)} = \mathbf{C}\text{vec}(\hat{\mathbf{B}}(s)^{(g)} - \text{bias}(\hat{\mathbf{B}}(s)^{(g)})) - \mathbf{b}_0(s)$. We also note that $\mathbf{C}\text{vec}(\hat{\mathbf{B}}(s)^{(g)}) \approx \mathbf{b}_0$ and $\mathbf{C}\text{vec}(\text{bias}(\hat{\mathbf{B}}(s)^{(g)})) \approx \mathbf{0}$. Thus, we can drop the term $\text{bias}(\hat{\mathbf{B}}(s)^{(g)})$ in $\mathbf{d}(s)^{(g)}$ for computational efficiency. Subsequently, we compute

$$S_n^{(g)} \;=\; n \int_0^{L_0} \mathbf{d}(s)^{(g)T} [\mathbf{C}(\hat{\Gamma}_\eta(s,s) \otimes \hat{\Omega}_X^{-1}) \mathbf{C}^T]^{-1} \mathbf{d}(s)^{(g)} ds.$$

Step 3, Repeat Step 2 G times to obtain $\{S_n^{(g)} : g = 1, \ldots, G\}$ and then calculate $p = G^{-1} \sum_{g=1}^{G} 1(S_n^{(g)} \geq S_n)$. If p is smaller than a pre-specified significance level α, say 0.05, then one rejects the null hypothesis H_0.

3.2. *Simultaneous confidence bands.* Construction of SCB for coefficient functions is of great interest in statistical inference for model (1.1). Theorem 1 allows us to construct SCB for coefficient functions $b_{jl}(s)$. For a given confidence level α, we construct SCB for each $b_{jl}(s)$ as follows:

$$P(\hat{b}_{jl}^{L,\alpha}(s) < b_{jl}(s) < \hat{b}_{jl}^{U,\alpha}(s) \ \text{ for all } \ s \in [0, L_0]) = 1 - \alpha, \tag{3.5}$$

where $\hat{b}_{jl}^{L,\alpha}(s)$ and $\hat{b}_{jl}^{U,\alpha}(s)$ are the lower and upper limits of SCB. Let $\mathbf{e}_{jl}$ be an $Mp \times 1$ vector such that $\mathbf{e}_{jl}\text{vec}(C) = c_{jl}$ for any $M \times p$ matrix $C = (c_{jl})$. It follows from Theorem 1 that

$$\sqrt{n}[\hat{b}_{jl}(\cdot) - b_{jl}(\cdot) - \text{bias}(\hat{b}_{jl}(\cdot))] \Rightarrow G_{jl}(\cdot), \tag{3.6}$$

where $G_{jl}(\cdot)$ is a centered Gaussian process indexed by $s \in [0, L_0]$. Therefore, by using the continuous mapping theorem, we have

$$\sup_{s\in[0,L_0]} |\sqrt{n}[\hat{b}_{jl}(s) - b_{jl}(s) - \text{bias}(\hat{b}_{jl}(s))]| \Rightarrow \sup_{s\in[0,L_0]} |G_{jl}(s)|. \tag{3.7}$$

We define $C_{jl}(\alpha)$ such that $P(\sup_{s\in[0,L_0]} |G_{jl}(s)| \leq C_{jl}(\alpha)) = 1 - \alpha$. Thus, based on (3.7), a $1 - \alpha$ simultaneous confidence band for $b_{jl}(s)$ is given as follows:

$$\left(\hat{b}_{jl}(s) - \text{bias}(\hat{b}_{jl}(s)) - \frac{C_{jl}(\alpha)}{\sqrt{n}}, \quad \hat{b}_{jl}(s) - \text{bias}(\hat{b}_{jl}(s)) + \frac{C_{jl}(\alpha)}{\sqrt{n}}\right). \tag{3.8}$$

The next issue is to determine $C_{jl}(\alpha)$.

Although there are several methods of determining $C_{jl}(\alpha)$ including random field theory [40, 35], we develop an efficient resampling method to approximate $C_{jl}(\alpha)$ as follows [48, 25]. Let $\mathbf{e}_l$ be a $p \times 1$ vector with the l-th element 1 and 0 otherwise, and $\hat{\mathbf{r}}_{i,j}(s_m) = y_{i,j}(s_m) - \mathbf{x}_i^T \hat{B}_j(s_m)$. For $g = 1, \ldots, G$, we independently simulate $\{\tau_i^{(g)} : i = 1, \ldots, n\}$ from $N(0, 1)$, and then we calculate a stochastic process $G_j(s)^{(g)}$, which is defined as follows:

$$\sqrt{n}[I_p \otimes (1,0)]\text{vec}(\Sigma(h_j, s)^{-1} \sum_{i=1}^{n} \tau_i^{(g)} \sum_{m=1}^{M} K_{h_j}(s_m - s)[\mathbf{x}_i \otimes \mathbf{z}_{h_j}(s_m - s)]\hat{\mathbf{r}}_{i,j}(s_m)).$$

It will be shown in Theorem 5 below that conditioning on the data, we have $\mathbf{e}_l G_j(\cdot)^{(g)} \Longrightarrow G_{jl}(\cdot)$. Finally, we calculate $\sup_{s\in[0,L_0]} |\mathbf{e}_l G_j(s)^{(g)}|$ for all g and use their $1 - \alpha$ empirical percentile to estimate $C_{jl}(\alpha)$.

We consider conditional convergence for bootstrapped stochastic processes. Since the arguments for establishing the wild bootstrap method for approximating the null distribution of S_n and the bootstrapped process $\{G_j(s)^{(g)} : s \in [0, L_0]\}$ are similar, we focus on the bootstrapped process $\{G_j(s)^{(g)} : s \in [0, L_0]\}$ as follows.

Theorem 5. *If assumptions (C1)-(C8) in Section 5 are true, then $G_j(s)^{(g)}(\cdot)$ converges weakly to $G_j(\cdot)$ conditioning on the data.*

Theorem 5 validates the bootstrapped process of $G_j(s)^{(g)}(\cdot)$. An interesting observation is that the bias correction for $\hat{B}_j(s)$ in constructing $G_j(s)^{(g)}(\cdot)$ is unnecessary. It leads to substantial computational saving.

4. Simulation studies. In this section, we present two simulation example to demonstrate the performance of the proposed procedures.

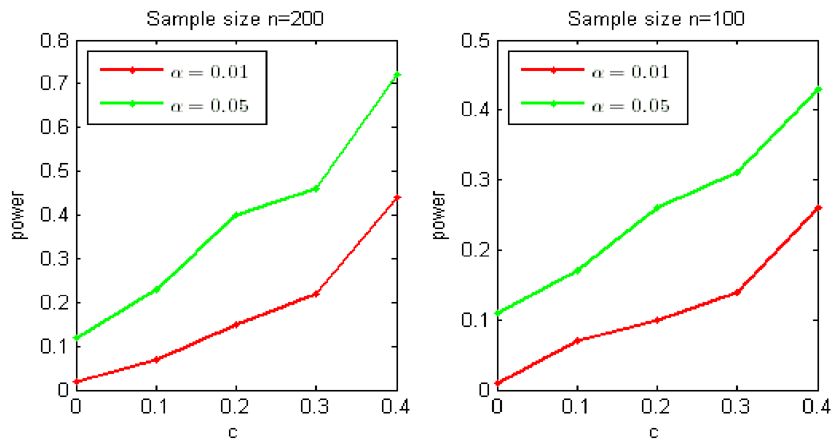

FIG 2. *Plot of Power Curves. Rejection rates of S_n based on the wild bootstrap method are calculated at five different values of c (0, 0.1, 0.2, 0.3, and 0.4) for two sample sizes of n (100 and 200) subjects at 5% (green) and 1% (red) significance levels.*

Example 1. This example is designed to evaluate the Type I error rate and power of the proposed global test S_n using Monte Carlo simulation. In this example, the data were generated from a bivariate MVCM as follows:

$$y_{i,j}(s_m) = \mathbf{x}_i^T B_j(s_m) + \eta_{i,j}(s_m) + \epsilon_{i,j}(s_m) \quad \text{for} \quad j = 1, 2, \tag{4.1}$$

where $s_m \sim U[0, 1]$, $\epsilon_{i,1} \sim N(0, \sigma_1^2)$, and $\epsilon_{i,2} \sim N(0, \sigma_2^2)$ for all $i = 1, \ldots, n$ and $m = 1, \ldots, M$. Let $\mathbf{x} = (1, x_1, x_2)$ and $\eta_j(s) = \xi_{j1}\psi_{j1}(s) + \xi_{j2}\psi_{j2}(s)$, where $x_1 \sim N(0, 1)$, $x_2 \sim N(0, 1)$, and $\mathrm{corr}(x_1, x_2) = 2^{-0.5}$, $\xi_{jl} \sim N(0, \lambda_{jl})$ for $j = 1, 2$ and $l = 1, 2$. Furthermore, s_m, (x_1, x_2), ξ_{11}, ξ_{12}, ξ_{21}, ξ_{22}, ϵ_1 and ϵ_2 are independent variables. The functional coefficients and eigenfunctions were chosen as

$$\begin{aligned}
&b_{11}(s) = s^2, \quad b_{12}(s) = (1-s)^2, \quad b_{13}(s) = 4s(1-s) - 0.4;\\
&\psi_{11}(s) = \sqrt{2}\sin(2\pi s), \quad \psi_{12}(s) = \sqrt{2}\cos(2\pi s);\\
&b_{21}(s) = 5(s-0.5)^2, \quad b_{22}(s) = s^{0.5}, \quad b_{23}(s) = 4s(1-s) - 0.4;\\
&\psi_{21}(s) = \sqrt{2}\cos(2\pi s), \quad \psi_{22}(s) = \sqrt{2}\sin(2\pi s);
\end{aligned}$$

and variance parameters $(\lambda_{11}, \lambda_{12}, \sigma_1^2, \lambda_{21}, \lambda_{22}, \sigma_2^2) = (1.2, 0.6, .2, 1, 0.5, 0.1)$. Then, except for $(b_{13}(s),\, b_{23}(s))$ for all s, we fixed all other parameters at the values specified above, whereas we assumed $(b_{13}(s), b_{23}(s)) = c(4s(1-s) - 0.4, 4s(1-s) - 0.4)$, where c is a scalar specified below.

We want to test the hypotheses $H_0 : b_{13}(s) = b_{23}(s) = 0$ for all s against $H_1 : b_{13}(s) \neq 0 \ \text{ or } \ b_{23}(s) \neq 0$ for at least one s. We set $c = 0$ to assess the Type I error rates for S_n, and set $c = 0.1, 0.2, 0.3$, and 0.4 to examine the power of S_n. We set $M = 50$, $n = 200$ and 100. For each simulation, the significance levels were set at $\alpha = 0.05$ and 0.01, and 100 replications were used to estimate the rejection rates.

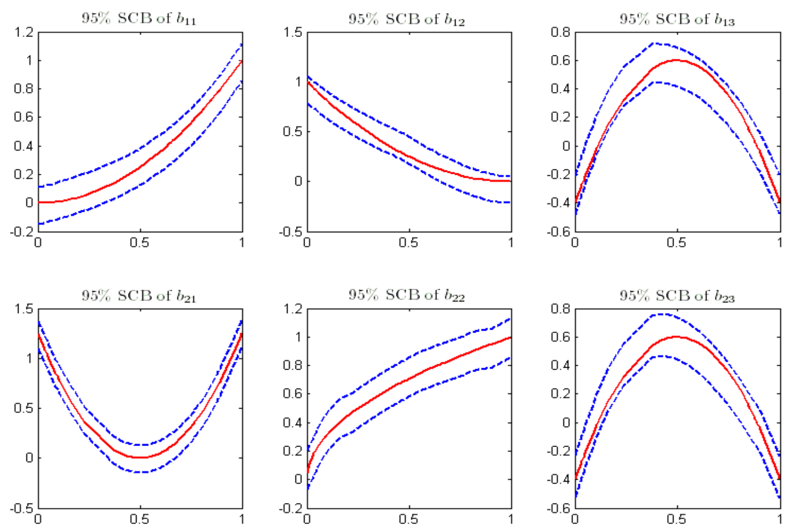

FIG 3. *Typical simultaneous confidence bands with $n = 500$ and $M = 50$. The red solid curves are the true coefficient functions, and the blue dashed curves are the confidence bands.*

Fig. 2 depicts the power curves. It can be seen from Fig. 2 that the rejection rates for S_n based on the resampling method are accurate for moderate sample sizes, such as ($n = 100$, or 200) at both significance levels ($\alpha = 0.01$ or 0.05). As expected, the power increases with the sample size.

Example 2. This example is used to evaluate the coverage probabilities of SCB of the functional coefficients $\mathbf{B}(s)$ based on the resampling method and the accuracy of the estimators of the eigenvalues and eigenfunctions of Σ_η and the estimators of Σ_ϵ. The data were generated from model (4.1) under

TABLE 1
Empirical coverage probabilities of $1-\alpha$ SCB for all components of $B_1(\cdot)$ and $B_2(\cdot)$ based on 200 simulated data sets.

M	b_{11}	b_{12}	b_{13}	b_{21}	b_{22}	b_{23}
	$\alpha = 0.05$					
25	0.915	0.930	0.945	0.920	0.915	0.945
50	0.925	0.940	0.945	0.930	0.925	0.950
75	0.945	0.950	0.955	0.945	0.945	0.955
	$\alpha = 0.01$					
25	0.985	0.965	0.985	0.985	0.990	0.980
50	0.995	0.980	0.985	0.985	0.995	0.985
75	0.990	0.985	0.990	0.995	0.990	0.990

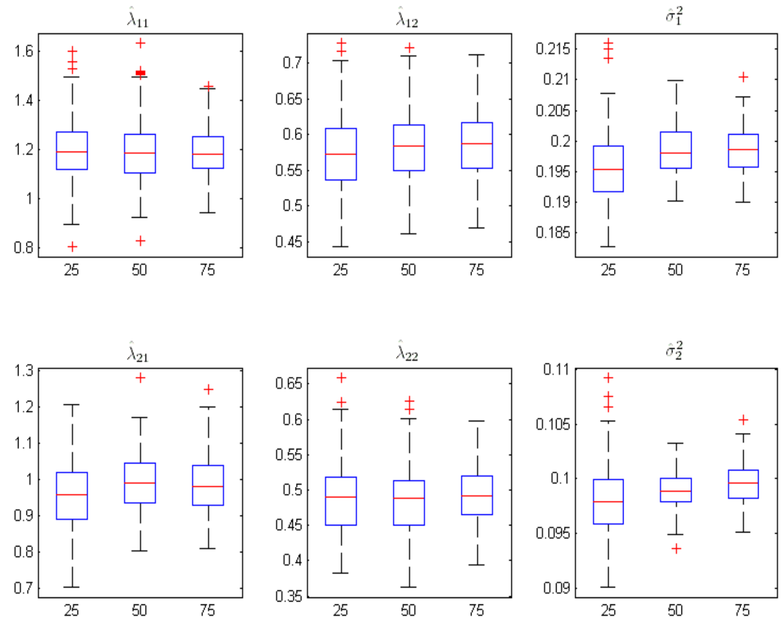

FIG 4. *Boxplot for the eigenvalues $\hat\lambda_{11}$, $\hat\lambda_{12}$, $\hat\lambda_{21}$, and $\hat\lambda_{22}$ and the variances $\hat\sigma_1^2$ and $\hat\sigma_2^2$, when $M = 25$, 50 and 75.*

the same parameter values. Recall that $\Sigma_\epsilon = \text{diag}(\sigma_1^2, \sigma_2^2)$. We set $n = 500$ and $M = 25$, 50, and 75 and generated 200 datasets for each combination.

Based on the generated data, we calculated SCB for each component of $B_1(s)$ and $B_2(s)$. Table 1 summarizes the empirical coverage probabilities based on 200 simulations for $\alpha = 0.01$ and $\alpha = 0.05$. The coverage

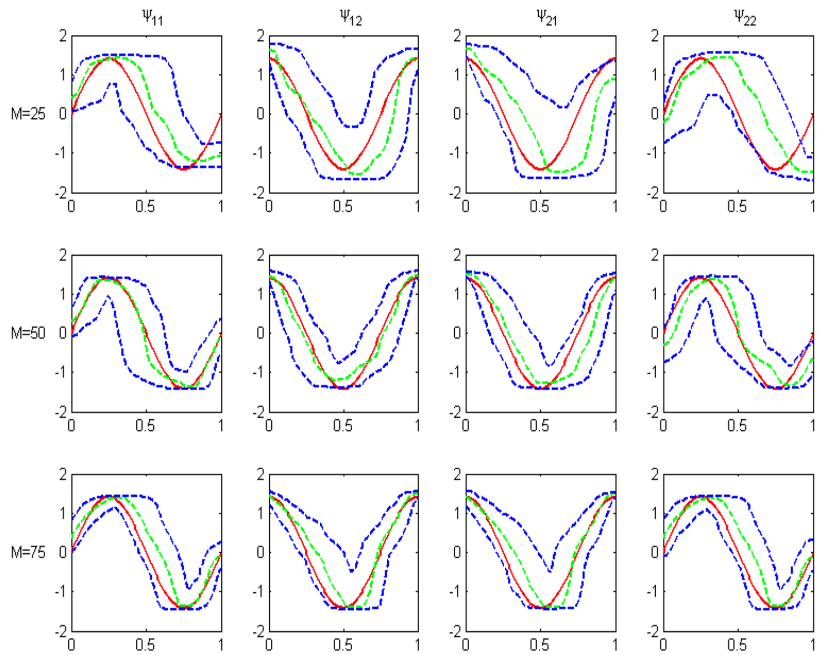

FIG 5. *Plot of eigenfunctions and their pointwise confidence intervals. The red solid line is the true eigenfunction, the middle green dashed line is the pointwise mean of estimated eigenfunctions and other two blue dashed lines are the pointwise 1% and 99% percentiles of the estimated eigenfunctions in* 200 *runs.*

probabilities improve with the number of grid points M. When $M = 75$, the differences between the coverage probabilities and the claimed confidence levels are fairly acceptable. The Monte Carol errors are of size $\sqrt{0.95 \times 0.05/200} \approx 0.015$ for $\alpha = 0.05$. Fig. 3 depicts typical simultaneous confidence bands, where $n = 500$ and $M = 50$.

We estimated the eigenvalues λ_{11}, λ_{12}, λ_{21}, and λ_{22} and the variances σ_1^2 and σ_2^2 for each simulated data set for $M = 25$, 50 and 75. The accuracy of estimators improves with M. The performance of the estimators for $M = 50$ is almost as good as their performance for $M = 75$. Fig. 4 shows the empirical distributions of $\hat{\lambda}_{jl}$ and $\hat{\sigma}_j^2$ for $j = 1, 2$ and $l = 1, 2$. The estimated eigenvalues and variances should be compared with the true ones, which are $(1.2, 0.6, 1, 0.5, 0.2, 0.1)$. When M is large, the estimated eigenvalues and variances are very close to their true values. We summarized the estimated results on $\psi_{jl}(s)$ for $j = 1, 2$ and $l = 1, 2$ in Fig. 5, in which we plotted the mean and the pointwise 1^{st} and 99^{th} percentiles of the estimated eigenfunctions with the true eigenfunctions. The performance of the estimated eigenfunctions improves with M increasing as expected.

5. Technical Conditions and Proofs.

5.1. *Assumptions.* Throughout the paper, the following assumptions are needed to facilitate the technical details, although they may not be the weakest conditions. We need to introduce some notation. Let $N(\mu, \Sigma)$ be a normal random vector with mean μ and covariance Σ. We define the fourth moments of $\eta_{i,j}(s)$ to be $\gamma_{jj'll'}(s_1, t_1, s_2, t_2) = E[\eta_{i,j}(s_1)\eta_{i,j'}(t_1)\eta_{i,l}(s_2)\eta_{i,l'}(t_2)]$ for any j, j', l, and l'. Moreover, we do not distinguish the differentiation and continuation at the boundary points from those in the interior of $[0, L_0]$. For instance, a continuous function at the boundary of $[0, L_0]$ means that this function is left continuous at 0 and right continuous at L_0.

Assumption C1. $\boldsymbol{\epsilon}_i(s)$ and $\boldsymbol{\eta}_i(s)$ are identical and independent copies of $\mathrm{SP}(0, \Sigma_\epsilon)$ and $\mathrm{SP}(0, \Sigma_\eta)$, respectively, and $\boldsymbol{\epsilon}_i(s)$ and $\boldsymbol{\epsilon}_i(t)$ are independent for $s \neq t$. Moreover, with probability one, the sample path of $\eta_{i,j}(s)$ has continuous second-order derivatives on $[0, L_0]$ and $E[\sup_{s\in[0,L_0]} ||\boldsymbol{\eta}(s)||_2^{r_1}] < \infty$ and $E\{\sup_{s\in[0,L_0]}[||\dot{\boldsymbol{\eta}}(s)||_2 + ||\ddot{\boldsymbol{\eta}}(s)||_2]^{r_2}\} < \infty$ for some $r_1, r_2 \in (2, \infty)$, where $|| \cdot ||_2$ is the Euclidean norm.

Assumption C2. All components of $\mathbf{B}(s)$ and $\Sigma_\epsilon(s, s)$ have continuous second derivatives on $[0, L_0]$. The fourth moments of $\epsilon_i(s)$ are continuous on $[0, L_0]$. All components of $\Sigma_\eta(s, t)$ have continuous second-order partial derivatives with respect to $(s, t) \in [0, L_0]^2$. Moreover, $\Sigma_\epsilon(s, s)$ and $\Sigma_\eta(s, s)$ are positive for all $s \in [0, L_0]$.

Assumption C3. The grid points $\mathcal{S} = \{s_m, m = 1, \ldots, M\}$ are independently and identically distributed with density function $\pi(s)$, which has the bounded support $[0, L_0]$. For some constants π_L and $\pi_U \in (0, \infty)$ and any $s \in [0, L_0]$, $\pi_L \leq \pi(s) \leq \pi_U$ and $\pi(s)$ has continuous second-order derivative.

Assumption C4. The kernel function $K(t)$ is a symmetric density function with a compact support $[-1, 1]$, and is Lipschitz continuous. Moreover, $0 < \inf_{h>0,s\in[0,L_0]} \lambda_{\min}(\Omega_1(h, s)) \leq \sup_{h>0,s\in[0,L_0]} \lambda_{\max}(\Omega_1(h, s)) < \infty$ where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively, denote the smallest and largest eigenvalues of matrix A, and $\Omega_1(h, s)$ is defined as

$$\Omega_1(h, s) = \int_0^{L_0} \begin{pmatrix} 1 & h^{-1}(u - s) \\ h^{-1}(u - s) & h^{-2}(u - s)^2 \end{pmatrix} K_h(u - s)\pi(u)du.$$

Assumption C5. The covariate vectors $\mathbf{x}_i$s are independently and identically distributed with $E\mathbf{x}_i = \mu_x$ and $E[||\mathbf{x}_i||_2^4] < \infty$. Assume that $E[\mathbf{x}_i^{\otimes 2}] = \Omega_X$ is invertible.

Assumption C6. Both n and M converge to ∞, $\max_j h_j = o(1)$, $Mh_j \to \infty$, and
$\max_j h_j^{-1} |\log h_j|^{1-2/q_1} \leq M^{1-2/q_1}$ for $j = 1, \ldots, J$, where $q_1 \in (2, 4)$.

Assumption C7. $E[|\epsilon_{i,j}(s_m)|^{q_2}] < \infty$ for some $q_2 \in (4, \infty)$ and all j; $\max_j h_j^{(2)} = o(1)$, $Mh_j^{(2)} \to \infty$, and $\max_j (h_j^{(2)})^{-4} (\log n/n)^{1-2/q_2} = o(1)$ for $j = 1, \ldots, J$.

Assumption C8. $E[|\epsilon_{i,j}(s_m)|^{q_3}] < \infty$ for some $q_3 \in (4, \infty)$ and all j; $h^{(3)} = o(1)$, $Mh^{(3)} \to \infty$, and $(h^{(3)})^{-2} (\log n/n)^{1-2/q_3} = o(1)$.

Assumption C9. There is a positive integer $E_j < \infty$ such that $\lambda_{j,1} > \ldots > \lambda_{j,E_j} > 0$ and $\lambda_{j,E_j+1} = \ldots = 0$ for $j = 1, \cdots, J$.

REMARK. Assumption C1 requires sample-path differentiability, which is a sufficient condition for establishing the weak convergence of local linear estimator for $\mathbf{B}(s)$ [16, 44]. We are unable to relax this assumption. In Assumption C2, we assume that $\Sigma_\eta(s, s') > 0$ as $|s - s'|$ is small. The bounded support restriction on $K(\cdot)$ in Assumption C4 is not essential and can be removed if we put a restriction on the tail of $K(\cdot)$. Assumptions C6-C8 on bandwidths are similar to the conditions used in [27, 6].

5.2. *Proofs.* Detailed proofs are given in the supplementary material of this paper [46]. The proof of Theorem 1 will be based on the following lemmas. Throughout the proofs, C_ks stand for a generic constant, and it may vary from line to line. Let

$$T_{\epsilon,j}(h_j, s) = \sum_{i=1}^{n} \sum_{m=1}^{M} K_{h_j}(s_m - s)[\mathbf{x}_i \otimes \mathbf{z}_{h_j}(s_m - s)] \epsilon_{i,j}(s_m),$$

whose order is given in the following lemma.

Lemma 1. Under Assumptions C1-C6, we have that for each j,

$$\sup_{s \in [0, L_0]} n^{-1/2} h_j |T_{\epsilon,j}(h_j, s)| = O_p(\sqrt{Mh_j |\log h_j|}) = o_p(Mh_j). \tag{5.1}$$

Proof. Let $F_n(s_m) = n^{-1/2} \sum_{i=1}^{n} \mathbf{x}_i \epsilon_{i,j}(s_m)$. Then it follows by the definition of $T_{\epsilon,j}(h_j, s)$ that

$$n^{-1/2} h_j T_{\epsilon,j}(h_j, s) = h_j \sum_{m=1}^{M} K_{h_j}(s_m - s) F_n(s_m) \otimes \mathbf{z}_{h_j}(s_m - s).$$

Let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $\tilde{T}_{\epsilon,j}(h_j, s) = \{T'_{\epsilon,j}(h_j, s) - E[T'_{\epsilon,j}(h_j, s)|\mathbf{X}, \mathcal{S}]\}$, where

$$T'_{\epsilon,j}(h_j, s) = \sqrt{n} \sum_{m=1}^{M} K_{h_j}(s_m - s) F_n(s_m) \mathbf{1}(||F_n(s_m)||_2 \le \gamma_M) \otimes \mathbf{z}_{h_j}(s_m - s),$$

in which γ_M is a positive number to be specified below. The proof of Lemma 1 consists of three steps. In Step 1, we show that

$$\sup_{s \in [0, L_0]} n^{-1/2} h_j ||T_{\epsilon,j}(h_j, s) - \tilde{T}_{\epsilon,j}(h_j, s)||_2 = o_p(\sqrt{M h_j |\log h_j|}). \tag{5.2}$$

In Step 2, we define an equally-spaced grid $\tilde{\mathbf{S}} = \{\tilde{s}_l = l h_j : l = 0, \ldots, L_0 h_j^{-1}\}$ and then show that

$$\max_l h_j ||n^{-1/2} \tilde{T}_{\epsilon,j}(h_j, \tilde{s}_l)||_2 = O_p(\sqrt{M h_j |\log h_j|}). \tag{5.3}$$

In Step 3, we show that

$$\max_l \sup_{s \in [\tilde{s}_{l-1}, \tilde{s}_l]} n^{-1/2} h_j ||\tilde{T}_{\epsilon,j}(h_j, \tilde{s}_{l-1}) - \tilde{T}_{\epsilon,j}(h_j, s)||_2 = O_p(\sqrt{M h_j |\log h_j|}). \tag{5.4}$$

It is easy to see that the proof of Lemma 1 is completed by combing (5.2)-(5.4).

We first show (5.2). It follows from Assumptions C1 and C4 and $s_m, s \in [0, L_0]$ that

$$n^{-1/2} h_j ||T_{\epsilon,j}(h_j, s) - \tilde{T}_{\epsilon,j}(h_j, s)||_2 \le C_1 \sum_{m=1}^{M} ||F_n(s_m)||_2 \mathbf{1}(||F_n(s_m)||_2 \ge \gamma_M)$$
$$+ C_1 \sum_{m=1}^{M} E[||F_n(s_m)||_2 \mathbf{1}(||F_n(s_m)||_2 \ge \gamma_M)|\mathbf{X}, \mathcal{S}],$$

for a positive constant C_1. Let $\gamma_M = \delta(M/|\log h_j|)^{1/q_1}$, where δ is a positive scalar. It follows from Assumption C6 that $(|\log h_j|/M)^{1-2/q_1} \le h_j \to 0$ and $1 - 2/q_1 > 0$, which yields that $|\log h_j|/M \to 0$ and $\gamma_M \to \infty$. As $\gamma_M \to \infty$, we can show that

$$\max_m E[||F_n(s_m)||_2^{q_1} \mathbf{1}(||F_n(s_m)||_2 \ge \gamma_M)|\mathbf{X}, \mathcal{S}] = o(1). \tag{5.5}$$

For notational simplicity, we only consider the case $p = 1$, (i.e., $\mathbf{x}_i$ is scalar). For any $c > 0$ with $q_1 + c < 4$, equation (5.5) is followed from Assumptions

C1, C5 and C6 and the partial sum moment inequality [5] as follows:

$$\begin{aligned}
& \max_m E[||F_n(s_m)||_2^{q_1}\mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X},\mathcal{S}] \\
\leq\ & \max_m E[|n^{-1/2}\sum_{i=1}^n \mathbf{x}_i\epsilon_{ik}(s_m)|^{q_1+c}|\mathbf{X},\mathcal{S}]/\gamma_M^c \\
\leq\ & \max_m n^{-(q_1+c)/2}C(q_1)n^{(q_1+c)/2-1}\sum_{i=1}^n |\mathbf{x}_i|^{q_1+c}E[|\epsilon_{ik}(s_m)|^{q_1+c}|]/\gamma_M^c = o(1),
\end{aligned}$$

where $C(q_1)$ is a universal constant independent of n. It follows from Assumption C6 and (5.5) that

$$\begin{aligned}
(5.6)\qquad & \sum_{m=1}^M E[||F_n(s_m)||_2\mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X},\mathcal{S}] \\
\leq\ & o(1)M^{1/q_1}|\log h_j|^{1-1/q_1} \leq o(\sqrt{Mh_j|\log h_j|}).
\end{aligned}$$

Furthermore, we show in Zhu et al. [46] that

$$\begin{aligned}
(5.7)\qquad & \mathrm{Var}(\sum_{m=1}^M ||F_n(s_m)||_2\mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X},\mathcal{S}) \\
\leq\ & \max_m E[||F_n(s_m)||_2^{q_1}\mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X},\mathcal{S}]Mh_j = o(Mh_j|\log h_j|).
\end{aligned}$$

Therefore, combing equations (5.6) and (5.7), we have

$$\sum_{m=1}^M ||F_n(s_m)||_2\mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M) = o_p(\sqrt{Mh_j|\log h_j|}),$$

which yields (5.2).

We next prove (5.3). It follows from Assumption C4 that

$$\begin{aligned}
& h_j||K_{h_j}(s_m-s)\{F_n(s_m)\mathbf{1}(||F_n(s_m)||_2 \leq \gamma_M) - \\
& \quad E[F_n(s_m)\mathbf{1}(||F_n(s_m)||_2 \leq \gamma_M)|\mathbf{X},\mathcal{S}]\} \otimes \mathbf{z}_{h_j}(s_m-s)||_2 \\
\leq\ & C_2(M/|\log h_j|)^{1/q_1} \leq C_2\sqrt{Mh_j/|\log h_j|},
\end{aligned}$$

where $C_2 = 4\delta\sup_{t\in[-1,1]}|K(t)|$. Furthermore, let $E_{\mathcal{S}}$ denote the expectation to s_m. As shown in Zhu et al. [46], we have

$$\mathrm{Var}(\sum_{m=1}^M h_jK_{h_j}(s_m-s)F_n(s_m)\mathbf{1}(||F_n(s_m)||_2 \leq \gamma_M)\otimes\mathbf{z}_{h_j}(s_m-s)|\mathbf{X}) = O_p(Mh_j).$$

Therefore, by applying Bernstein's inequality to each component of $h_j n^{-1/2}\tilde{T}_{\epsilon,j}(h_j, \tilde{s}_l)$ [38], we can prove (5.3). For instance, let $\mathbf{e}_1$ be a $\dim(\tilde{T}_{\epsilon,j}(h_j, \tilde{s}_l)) \times 1$ vector with the first element 1 and zero otherwise; we have

$$\begin{aligned} (5.8) \qquad & P(\max_l |\mathbf{e}_1 h_j n^{-1/2}\tilde{T}_{\epsilon,j}(h_j, \tilde{s}_l)| > t|\mathbf{X}) \\ \leq \quad & C_3(L_0 h_j^{-1} + 1)E[\exp(-\frac{1}{2}\frac{t^2}{v(\mathbf{X}) + tC_3\sqrt{Mh_j/|\log h_j|}/3})|\mathbf{X}], \end{aligned}$$

where $C_3 = O(1)$, t is a positive scalar, and $v(\mathbf{X}) \geq \mathrm{Var}(\mathbf{e}_1 h_j n^{-1/2}\tilde{T}_{\epsilon,j}(h_j, \tilde{s}_l)|\mathbf{X})$ for all l. By setting $t = C_4\sqrt{Mh_j|\log h_j|}$ for large $C_4 > 0$, we can show that the right hand side of (5.8) is of order $h_j^{C_5}$, where C_5 is a positive scalar. Thus, for sufficiently large $C_4 > 0$, we have

$$P(\max_l |\mathbf{e}_1 h_j n^{-1/2}\tilde{T}_{\epsilon,j}(h_j, \tilde{s}_l)| > C_4\sqrt{Mh_j|\log h_j|}) \to 0 \quad \text{as} \quad h_j \to 0.$$

In Step 3, we focus on the first component of $\mathbf{z}_{h_j}(s_m - s)$. We first consider the following function class:

$$\mathcal{E}_l = \{w_l(S; s) = h_j[K_{h_j}(S-\tilde{s}_l) - K_{h_j}(S-s)]F_n(S)\mathbf{1}(||F_n(S)||_2 \leq \gamma_M) : s \in [\tilde{s}_{l-1}, \tilde{s}_l]\}.$$

It follows from Assumption C4 and γ_M that $\mathcal{E}_l$ is a pointwise measurable class of functions and $\sup_{s\in[0,L_0]} |w_l(S; s)| \leq C_6\gamma_M \leq C_7\sqrt{Mh_j/|\log h_j|}$. Let $||\phi||_D = \sup_{z\in D} |\phi(z)|$ for any real valued function ϕ defined on a set D and $\tau_1, \ldots, \tau_M$ be a sequence of independent Rademacher random variables independent of observed data. It follows from an inequality of Talagrand [36, 6] that conditioning on $\mathbf{X}$, we have for suitable finite constants $A_1, A_2 > 0$

$$\begin{aligned} & P\{||\sum_{m=1}^M [w_l(s_m; s) - E[w_l(s_m; s)|\mathbf{X}]||_{\mathcal{E}_l} \geq A_1(E[||\sum_{m=1}^M \tau_j w_l(s_m; s)||_{\mathcal{E}_l}|\mathbf{X}] + t)|\mathbf{X}\} \\ & \leq 2[\exp(-A_2 t^2/(MV_{\mathcal{E}_l}(\mathbf{X}))) + \exp(-A_2 t/(C_7\sqrt{Mh_j/|\log h_j|}))], \end{aligned}$$

where $V_{\mathcal{E}_l}(\mathbf{X}) = \sup_{s\in[\tilde{s}_{l-1}, \tilde{s}_l]} \mathrm{Var}(w_l(S; s)|\mathbf{X})$. It can be shown that

$$\begin{aligned} V_{\mathcal{E}_l}(\mathbf{X}) \leq & \sup_{s\in[\tilde{s}_{l-1}, \tilde{s}_l]} E_S\{h_j^2[K_{h_j}(S-\tilde{s}_l) - K_{h_j}(S-s)]^2 E[F_n(S)^2|\mathbf{X}]\} \\ \leq \quad & C_8 h_j n^{-1}\sum_{i=1}^n \mathbf{x}_i^{\otimes 2}, \end{aligned}$$

where C_8 is a positive scalar. By setting $t = C_9\sqrt{Mh_j|\log h_j|}$ for a large $C_9 > 0$, we can show that $A_2t^2/(MV_{\mathcal{E}_l}(\mathbf{X})) = C_{10}|\log h_j|$ and $A_2t/(C_7\sqrt{Mh_j/|\log h_j|}) = C_{11}|\log h_j|$. Moreover, it follows from Assumption C4 that $\mathcal{E}_l$ is a pointwise measurable Vapnik and Cervonenkis (VC) class [38]. By using Proposition A.1 of [6], we can show that $\max_l E[||\sum_{m=1}^M \tau_j w_l(s_m;s)||_{\mathcal{E}_l}|\mathbf{X}] \leq O(\sqrt{Mh_j|\log h_j|})$. This yields (5.4).

Lemma 2. Under Assumption C1-C6, we have that for any $r \geq 0$ and j,

$$(5.9)\quad \sup_{s\in[0,L_0]}\left|\int K_{h_j}(u-s)\frac{(u-s)^r}{h_j^r}d[\Pi_M(u)-\Pi(u)]\right| = O_p(M^{-1/2}h_j^{-1}),$$

$$(5.10)\quad \sup_{s\in[0,L_0]}\left|\int K_{h_j}(u-s)\frac{(u-s)^r}{h_j^r}\epsilon_{i,j}(u)d\Pi_M(u)\right| = O_p(\sqrt{\frac{|\log(h_j)|}{Mh_j}}),$$

where $\Pi_M(\cdot)$ is the sampling distribution function based on $\mathcal{S} = \{s_1,\ldots,s_M\}$, and $\Pi(\cdot)$ is the distribution function of s_m.

Proof. Equation (5.9) follows from the integration by parts, while (5.10) can be proved by using similar arguments of Lemma 1.

Define

$$\Delta_j(s;\boldsymbol{\eta}_i,h_j) = M^{-1}\sum_{m=1}^M K_{h_j}(s_m-s)\mathbf{z}_{h_j}(s_m-s)\eta_{i,j}(s_m) - \int K_{h_j}(u-s)\mathbf{z}_{h_j}(u-s)\eta_{i,j}(u)\pi(u)du.$$

Lemma 3. Suppose that Assumptions (C1)-(C6) hold. Then we have

$$(5.11)\quad \sup_{s\in[0,L_0]}|n^{-1/2}\sum_{i=1}^n \mathbf{x}_i\otimes\Delta_j(s;\boldsymbol{\eta}_i,h_j)| = O_p(h_j^{-1/2}M^{-1/2}+h_j^{-1}M^{-1/2}).$$

Proof. Let $\tau_1,\ldots,\tau_n$ be a sequence of independent Rademacher random variables independent of observed data. It follows from the symmetrization inequality [38] that

$$E||n^{-1/2}\sum_{i=1}^n \mathbf{x}_i\otimes\Delta_j(s;\boldsymbol{\eta}_i,h_j)||_{[0,L_0]} \leq 2E\{E_\tau[||n^{-1/2}\sum_{i=1}^n \tau_i\mathbf{x}_i\otimes\Delta_j(s;\boldsymbol{\eta}_i,h_j)||_{[0,L_0]}]\}.$$

Thus, it is sufficient to show that $E[||n^{-1/2}\sum_{i=1}^n \tau_i\mathbf{x}_i\otimes\Delta_j(s;\boldsymbol{\eta}_i,h_j)||_{[0,L_0]}] = o(1)$. We consider the following function class:

$$\mathcal{E}_\eta = \{f(\mathbf{x},\boldsymbol{\eta};s) = \mathbf{x}\otimes\Delta_j(s;\boldsymbol{\eta},h_j) : s\in[0,L_0]\}.$$

Let $N(\epsilon, \mathcal{E}_\eta, d_Q)$ be the minimal number of d_Q-balls with radius ϵ needed to cover $\mathcal{E}_\eta$. We define

$$D_n^2 = \sup_{s\in[0,L_0]} |n^{-1}\sum_{i=1}^n \mathrm{tr}\{[\mathbf{x}_i \otimes \Delta_j(s;\boldsymbol{\eta}_i, h_j)]^{\otimes 2}\}|, \tag{5.12}$$

$$d_{2;x,\eta}(s_1,s_2)^2 = \sum_{i=1}^n \mathrm{tr}\{\mathbf{x}_i^{\otimes 2}\}||\Delta_j(s_1;\boldsymbol{\eta}_i,h_j) - \Delta_j(s_2;\boldsymbol{\eta}_i,h_j)||_2^2.$$

It follows from Corollary 2.2.8 of van der Vaar and Wellner [38] that for any fixed $\{(\mathbf{x}_1,\boldsymbol{\eta}_1),\ldots,(\mathbf{x}_n,\boldsymbol{\eta}_n)\}$, we have

$$\begin{aligned} &E_\tau||n^{-1/2}\sum_{i=1}^n \tau_i\mathbf{x}_i\otimes\Delta_j(s;\boldsymbol{\eta}_i,h_j)||_{[0,L_0]} \\ \leq\ & E_\tau||n^{-1/2}\sum_{i=1}^n \tau_i\mathbf{x}_i\otimes\Delta_j(s_0;\boldsymbol{\eta}_i,h_j)||_2 + C_1\int_0^{D_n}\sqrt{\log N(\epsilon,\mathcal{E}_\eta,d_{2;x,\eta})}d\epsilon, \end{aligned} \tag{5.13}$$

where s_0 is any point in $[0, L_0]$ and C_1 is a positive universal constant.

We investigate the two terms on the right hand side of (5.13). First, we note that

$$[E_\tau||n^{-1/2}\sum_{i=1}^n \tau_i\mathbf{x}_i\Delta_j(s_0;\boldsymbol{\eta}_i,h_j)||_2]^2 \leq n^{-1}\sum_{i=1}^n \mathbf{x}_i^T\mathbf{x}_i||\Delta_j(s_0;\boldsymbol{\eta}_i,h_j)||_2^2. \tag{5.14}$$

With some calculations, we show in [46] that

$$E[||\Delta_j(s_0;\boldsymbol{\eta}_i,h_j)||_2^2] = O(M^{-1}+M^{-1}h_j^{-1}).$$

Thus, we have $[E_\tau||n^{-1/2}\sum_{i=1}^n \tau_i\mathbf{x}_i\Delta_j(s_0;\eta_i,h_j)||_2]^2 = O_p(M^{-1}+M^{-1}h_j^{-1})$.

Second, with some calculations, we have

$$\Delta_j(s;\boldsymbol{\eta}_i,h_j) = \int K_{h_j}(u-s)\mathbf{z}_{h_j}(u-s)\eta_{i,j}(u)d[\Pi_M(u)-\Pi(u)].$$

It follows from an integration by parts that $||\Delta_j(s;\boldsymbol{\eta}_i,h_j)||_2$ is bounded from above by

$$C_2h_j^{-1}\sup_{s\in[0,L_0]}|\Pi_M(s)-\Pi(s)|\sup_{s\in[0,L_0]}[|\eta_{i,j}(s)|+|\dot{\eta}_{i,j}(s)|], \tag{5.15}$$

where C_2 is a positive constant. Therefore, we have

$$D_n \leq C_2h_j^{-1}\sup_{s\in[0,L_0]}|\Pi_M(s)-\Pi(s)|\sqrt{n^{-1}\sum_{i=1}^n\{\mathbf{x}_i^T\mathbf{x}_i\sup_{s\in[0,L_0]}[|\eta_{i,j}(s)|+|\dot{\eta}_{i,j}(s)|]^2\}}.$$

Let $H_{h_j}(u-s) = K_{h_j}(u-s)\mathbf{z}_{h_j}(u-s)$. It follows from the integration by parts that there is a positive scalar C_3 such that

$$\begin{aligned} &\Delta_M(s_1;\boldsymbol{\eta}_i,h_j) - \Delta_M(s_2;\boldsymbol{\eta}_i,h_j) \\ \leq\ & |s_1-s_2|C_3h_j^{-1}\sup_{s\in[0,L_0]}|\Pi_M(s)-\Pi(s)|\sup_{s\in[0,L_0]}[|\eta_{i,j}(s)|+|\dot{\eta}_{i,j}(s)|]. \end{aligned} \tag{5.16}$$

For any given $s_1, s_2 \in [0, L_0]$, it follows (5.16) and $\sup_{s\in[0,L_0]}|\Pi_M(s) - \Pi(s)| = O_p(M^{-1/2})$ that

$$d_{2;x,y}(s_1,s_2) \leq C_4|s_1-s_2|h_j^{-1}\sup_{s\in[0,L_0]}|\Pi_M(s)-\Pi(s)|,$$

where $\tilde{c}$ is positive scalar. Therefore, let $A_n = h_j^{-1}\sup_{s\in[0,L_0]}|\Pi_M(s)-\Pi(s)|$; we have
$N(\epsilon A_n, \mathcal{E}_\eta, d_{2;x,\eta}) = O(\epsilon^{-1})$ and

$$\int_0^{D_n}\sqrt{\log N(\epsilon,\mathcal{E}_\eta,d_{2;x,\eta})}d\epsilon \leq \int_0^1\sqrt{\log N(tA_n,\mathcal{E}_\eta,d_{2;x,\eta})}dtA_n = O_p(h_j^{-1}M^{-1/2}),$$

in which we have used a change of variable $t = \epsilon/A_n$. This completes the proof of Lemma 3.
Lemma 4 . If Assumptions C1-C6 hold, then for any $s \in (0, L_0)$, we have

$$\begin{aligned} &E[\hat{B}_j(s)|\mathcal{S}] - B_j(s) \\ =\ & 0.5h_j^2u_2(K)\ddot{B}_j(s)[1+O_p(n^{-1/2}+h_j+(Mh_j)^{-1/2})] \\ =\ & 0.5h_j^2u_2(K)\ddot{B}_j(s)[1+o_p(1)], \end{aligned} \tag{5.17}$$

and

$$\begin{aligned} &Var[\hat{B}_j(s)|\mathcal{S}] = (nMh_j)^{-1}\pi(s)^{-1}v_0(K)[\Sigma_{\eta,jj}(s,s)+\Sigma_{\epsilon,jj}(s,s)]\Omega_X^{-1}[1+O_p(h_j)] \\ &+n^{-1}\{\Sigma_{\eta,jj}(s,s)+h_j^2u_2(K)[\Sigma_{\eta,jj}^{(2,0)}(s,s)\pi(s)+2\Sigma_{\eta,jj}^{(1,0)}(s,s)\dot{\pi}(s)+\Sigma_{\eta,jj}(s,s)\ddot{\pi}(s)]\pi(s)^{-1} \\ &+e_n(s)+o_p(h_j^2)\}\Omega_X^{-1}, \end{aligned}$$

where

$$e_n(s) = 2\pi(s)^{-2}\sum_{m=1}^M[P_{\eta,11}(s_m)-\theta_{11}(s)]/M + \tilde{E}_{n,11}(s) = O_p((Mh_j)^{-1/2}).$$

Thus, $e_n(s) = O_p((Mh_j)^{-1/2})$ with $E[e_n(s)] = 0$.
The proof of Lemma is given in the supplementary material Zhu et al. [46].

Lemma 5. If Assumptions C1-C6 hold, then for $s = 0$ or L_0, we have

$$E[\hat{B}_j(s)|\mathcal{S}] - B_j(s) = 0.5h_j^2 r_u(K; s, h_j)\ddot{B}_j(s)[1 + o_p(1)], \quad Var[\hat{B}_j(s)|\mathcal{S}] = n^{-1}\Sigma_{\eta,jj}(s, s)\Omega_X^{-1}[1 + o_p(1)]; \tag{5.18}$$

where $r_u(K; s, h) = [u_2(K; s, h)^2 - u_1(K; s, h)u_3(K; s, h)]/[u_0(K; s, h)u_2(K; s, h) - u_1(K; s, h)^2]$, *in which* $u_r(K; s, h) = \int_0^{L_0} h^{-r}(u - s)^r K_h(u - s)du$ *for* $r \geq 0$. The proof of Lemma is given in the supplementary material Zhu et al. [46].

We define some notation as follows:

$$T_{B,k}(h_j, s) = \sum_{i=1}^{n}\sum_{m=1}^{M} K_{h_j}(s_m - s)[\mathbf{x}_i \otimes \mathbf{z}_{h_j}(s_m - s)]\mathbf{x}_i^T B_j(s_m),$$
$$T_{\eta,k}(h_j, s) = \sum_{i=1}^{n}\sum_{m=1}^{M} K_{h_j}(s_m - s)[\mathbf{x}_i \otimes \mathbf{z}_{h_j}(s_m - s)]\eta_{i,j}(s_m),$$

$$\Omega_{2,k}(h_j, s) = h_j\int_0^{L_0}\Sigma_{\epsilon,jj}(u, u)\begin{pmatrix} 1 & h_j^{-1}(u - s) \\ h_j^{-1}(u - s) & h_j^{-2}(u - s)^2\end{pmatrix}[K_{h_j}(u - s)]^2\pi(u)du.$$

Proof of Theorem 1. Define

$$\mathbf{U}_2(K; s, \mathbf{H}) = \mathrm{diag}(r_u(K; s, h_1), \ldots, r_u(K; s, h_J)),$$
$$X_n(s) = \sqrt{n}\{\mathrm{vec}(\hat{\mathbf{B}}(s) - \mathbf{B}(s) - 0.5\ddot{\mathbf{B}}(s)\mathbf{U}_2(K; s, \mathbf{H})\mathbf{H}^2) + o_p(||\mathbf{H}^2||_2)\},$$
$$X_{n,j}(s) = \sqrt{n}\{\hat{B}_j(s) - B_j(s) - 0.5r_u(K; s, h_j)h_j^2\ddot{B}_j(s) + o_p(h_j^2)\}.$$

The proof of Theorem 1 (i) consists of two steps. The first step is to show the finite convergence of $\{X_n(s) : s \in [0, L_0]\}$. The second step is to check the asymptotic continuity of $X_n(s)$ as s varies in $[0, L_0]$. Moreover, Theorem 1 (ii) is a consequence of Theorem 1 (i) and Lemma 4.

In the first step, we only show that at a single point s, $X_n(s)$ converges weakly to $N(0, \Sigma_\eta(s, s)\otimes\Omega_X^{-1})$. The finite convergence can be directly verified by generalizing the asymptotic distribution of $X_n(s)$ at one point to any finite number of points using the Cramer-Wold theorem [5].

We show that $X_{n,j}(s)$ converges weakly to $N(0, \Sigma_{\eta,jj}(s, s)\Omega_X^{-1})$ as $n \to \infty$. It follows from Lemma 4 and the standard central limit theorem that

$$\sqrt{n}[\mathbf{I}_p \otimes (1, 0)]\Sigma(h_j, s)^{-1}[T_{\epsilon,j}(h_j, s) + T_{\eta,k}(h_j, s)] \to^L N(\mathbf{0}, \Sigma_{\eta,jj}(s, s)\Omega_X^{-1}), \tag{5.19}$$

where $\to^L$ denotes convergence in distribution. This yields that $X_{n,k}(s)$ converges weakly to $N(0, \Sigma_{\eta,jj}(s, s)\Omega_X^{-1})$. We can use the Cramer-Wald theorem to show that as $n \to \infty$,

$$X_n(s) \to^L N(\mathbf{0}, \Sigma_\eta(s, s) \otimes \Omega_X^{-1}) \quad \text{for each} \quad s \in [0, L_0].$$

In the second step, we will show the stochastic continuity of $X_n(s)$, which is equivalent to the stochastic continuity of $X_{n,j}(s)$ for $j = 1, \ldots, J$. It follows from Lemmas 1 and 2 that

$$(nM)^{-1}\Sigma(h_j, s) = L_0^{-1}\Omega_X \otimes \Omega_1(h_j, s) + O_p(|\log(h_j)|(Mh_j)^{-1}),$$
$$n^{-1/2}\sum_{i=1}^n \mathbf{x}_i \otimes \{M^{-1}\sum_{m=1}^M K_{h_j}(s_m - s)\mathbf{z}_{h_j}(s)\epsilon_{i,j}(s_m)\} = O_p(|\log(h_j)|(Mh_j)^{-1})$$

hold uniformly for all $s \in [0, L_0]$. It follows from the change of variable that

$$\begin{aligned}
&\Omega_X^{-1}\mathbf{x}_i \otimes \Omega_1(h_j, s)^{-1}\int_0^{L_0} H_{h_j}(u - s)\eta_{i,j}(u)\pi(u)du \\
&= \Omega_X^{-1}\mathbf{x}_i \otimes \Omega_1(h_j, s)^{-1}\eta_{i,j}(s)\pi(s)\int_{\max(-sh_j^{-1},-1)}^{\min((L_0-s)h_j^{-1},1)} K(u)(1, u)^T du \\
&+\Omega_X^{-1}\mathbf{x}_i \otimes \Omega_1(h_j, s)^{-1}\int_{\max(-sh_j^{-1},-1)}^{\min((L_0-s)h_j^{-1},1)} K(u)(1, u)^T[\eta_{i,j}(s + hu)\pi(s + hu) - \eta_{i,j}(s)\pi(s)]du.
\end{aligned}$$

It follows from Assumptions C1 and C3 that

(5.20)
$$\begin{aligned}
&|| \int_{\max(-sh_j^{-1},-1)}^{\min((L_0-s)h_j^{-1},1)} K(u)(1, u)^T[\eta_{i,j}(s + hu)\pi(s + hu) - \eta_{i,j}(s)\pi(s)]du||_2 \\
&\leq h \sup_{s\in[0,L_0]} |\dot{\eta}_{i,j}(s)\pi(s) + \eta_{i,j}(s)\dot{\pi}(s)| \times || \int_{\max(-sh_j^{-1},-1)}^{\min((L_0-s)h_j^{-1},1)} K(u)(|u|, u^2)du||_2 \\
&\leq h \sup_{s\in[0,L_0]} |\dot{\eta}_{i,j}(s)\pi(s) + \eta_{i,j}(s)\dot{\pi}(s)| \times || \int_{-1}^{1} K(u)(|u|, u^2)du||_2.
\end{aligned}$$

Thus, we define

(5.21)
$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^n \Omega_X^{-1}\mathbf{x}_i \otimes [\Omega_1(h_j, s)^{-1}M^{-1}\sum_{m=1}^M H_{h_j}(s_m - s)\eta_{i,j}(s_m)] \\
= \;& n^{-1/2}\sum_{i=1}^n \Omega_X^{-1}\mathbf{x}_i \otimes \Omega_1(h_j, s)^{-1}\{\Delta_j(s; \boldsymbol{\eta}_i, h_j) \\
&+ \int_{\max(-sh_j^{-1},-1)}^{\min((L_0-s)h_j^{-1},1)} K(u)(1, u)^T[\eta_{i,j}(s + hu)\pi(s + hu) - \eta_{i,j}(s)\pi(s)]du \\
&+\eta_{i,j}(s)\pi(s)\int_{\max(-sh_j^{-1},-1)}^{\min((L_0-s)h_j^{-1},1)} K(u)(1, u)^T du\}.
\end{aligned}$$

It follows from Lemma 3 and Assumption C6 that the first term on the right hand side of (5.21) converges to zero uniformly. It follows from Assumptions C1 and C6 and (5.20) that the second term on the right hand side of (5.21) also converges to zero uniformly. Define

$$\hat{X}_{n,j}(s) = n^{-1/2}\sum_{i=1}^{n}\Omega_X^{-1}\mathbf{x}_i\otimes(1,0)\Omega_1(h_j,s)^{-1}\eta_{i,j}(s)\pi(s)\int_{\max(-sh_j^{-1},-1)}^{\min((L_0-s)h_j^{-1},1)}K(u)(1,u)^T du.$$

Thus, we need to prove the asymptotic tightness of $\hat{X}_{n,j}(s)$.

It follows from the change of variable that all elements of $\Omega_1(h,s)$ can be written as

$$\int_0^{L_0} h^{-r}(u-s)^r K_h(u-s)\pi(u)du = \pi(s)u_r(K;s,h) + O(h) \tag{5.22}$$

for $r = 0, 1$, and 2.

The asymptotic tightness of $\hat{X}_{n,j}(s)$ can be proved using the empirical process techniques [38]. It follows from (5.22) that

$$\begin{aligned}&(1,0)\Omega_1(h_j,s)^{-1}\pi(s)\int_{\max(-sh_j^{-1},-1)}^{\min((L_0-s)h_j^{-1},1)}K(u)(1,u)^T du\\ =&\frac{u_2(K;s,h_j)u_0(K;s,h_j)-u_1(K;s,h_j)^2+o(h_j)}{u_2(K;s,h_j)u_0(K;s,h_j)-u_1(K;s,h_j)^2+o(h_j)} = 1+o(h_j).\end{aligned}$$

Thus, $\hat{X}_{n,j}(s)$ can be simplified as

$$\hat{X}_{n,j}(s) = [1+o(h_j)]n^{-1/2}\sum_{i=1}^{n}\eta_{i,j}(s)\Omega_X^{-1}\mathbf{x}_i.$$

We consider a function class $\mathcal{E}_\eta = \{f(s;\mathbf{x},\eta_{\cdot,j}) = \Omega_X^{-1}\mathbf{x}\eta_{\cdot,j}(s) : s \in [0, L_0]\}$ with $F(\mathbf{x},\eta_{\cdot,j}) = ||\mathbf{x}||_2\lambda_{\max}(\Omega_X^{-1})||\eta_{\cdot,j}||_{[0,L_0]}$ as an envelope function. Due to the differentiability of $\eta_{\cdot,j}(s)$ in Assumption C1, it follows from Theorem 2.7.11 of [38] that $\mathcal{E}_\eta$ is a $P-$Donsker class. Therefore, the proof of Theorem 1 is completed.

Proof of Theorem 2. Let $\tilde{K}_{M,h}(s) = \tilde{K}_M(s/h)/h$, where $\tilde{K}_M(s)$ is the empirical equivalent kernels for the first-order local polynomial kernel [7]. Thus, we have

$$\begin{aligned}&\hat{\eta}_{i,j}(s)-\eta_{i,j}(s)\\ =&\sum_{m=1}^{M}\tilde{K}_{M,h_j^{(2)}}(s_m-s)\mathbf{x}_i^T[B_j(s_m)-\hat{B}_j(s_m)]\\ &+\sum_{m=1}^{M}\tilde{K}_{M,h_j^{(2)}}(s_m-s)[\eta_{i,j}(s_m)+\epsilon_{i,j}(s_m)-\eta_{i,j}(s)].\end{aligned} \tag{5.23}$$

It follows from a Taylor's expansion that

$$\sum_{m=1}^{M} \tilde{K}_{M,h_j^{(2)}}(s_m - s)[\eta_{i,j}(s_m) - \eta_{i,j}(s)] = 0.5u_2(K)\ddot{\eta}_{i,j}(s)h_j^{(2)2}[1 + o_p(1)],$$

and

$$\begin{aligned}
& \sum_{m=1}^{M} \tilde{K}_{M,h_j^{(2)}}(s_m - s)\mathbf{x}_i^T\{B_j(s_m) - E[\hat{B}_j(s_m)|\mathcal{S}, \boldsymbol{\eta}, \mathbf{X}]\} \\
= \ & [0.5h_j^2 u_2(K)\mathbf{x}_i^T\ddot{B}_j(s) + O_p(n^{-1/2})] \\
& \times[1 + O_p(h_j + h_j^{(2)} + n^{-1/2}) + O_p((Mh_j^{(2)})^{-1/2} + (Mh_j)^{-1/2})],
\end{aligned}$$

which leads to Bias$[\hat{\eta}_{i,j}(s)|\mathcal{S}, \boldsymbol{\eta}, \mathbf{X}]$. See Zhu et al. [46] for detailed arguments.

Furthermore, it can be shown that

$$\begin{aligned}
& \hat{\eta}_{i,j}(s) - E[\hat{\eta}_{i,j}(s)|\mathcal{S}, \boldsymbol{\eta}, \mathbf{X}] \\
= \ & \sum_{m=1}^{M} \tilde{K}_{M,h_j^{(2)}}(s_m - s)\{\epsilon_{i,j}(s_m) - \mathbf{x}_i^T[\mathbf{I}_p \otimes (1,0)]\Sigma(h_j, s_m)^{-1}T_{\epsilon,j}(h_j, s_m)\} \\
= \ & \sum_{m=1}^{M} \tilde{K}_{M,h_j^{(2)}}(s_m - s)\{\epsilon_{i,j}(s_m) - \mathbf{x}_i^T[\mathbf{I}_p \otimes (1,0)]\Sigma(h_j, s_m)^{-1}\sum_{i'=1}^{n}\mathbf{x}_{i'} \\
& \otimes \sum_{m'=1}^{M} H_{h_j}(s_{m'} - s_m)\epsilon_{i',j}(s_{m'})\}.
\end{aligned}$$

With tedious calculations, we have

$$\begin{aligned}
& \mathrm{Cov}(\hat{\eta}_{i,j}(s) - \eta_{i,j}(s), \hat{\eta}_{i,j}(t) - \eta_{i,j}(t)|\mathcal{S}, \boldsymbol{\eta}, \mathbf{X}) \\
= \ & K^*((s-t)/h_j^{(2)})\Sigma_{\epsilon,jj}(s,s)\pi(t)^{-1}(Mh_j^{(2)})^{-1}[1 + o_p(1)] - \mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i(nMh_j)^{-1}\pi(s)^{-1}\pi(t)^{-1} \\
& \times\{\pi(t)\Sigma_{\epsilon,jj}(t,t)\int K(u)K(v)K([t - s + h_j^{(2)}(u-v)]/h_j)dudv \\
& +\pi(s)\Sigma_{\epsilon,jj}(s,s)\int K(u)K(v)K([s - t + h_j^{(2)}(u-v)]/h_j)dudv \\
& -\pi(s)\Sigma_{\epsilon,jj}(s,s)\int K(u)K(v)K(w)K(w + [s - t + h_j^{(2)}(u-v)]/h_j)dudvdw\}[1 + o_p(1)].
\end{aligned}$$

Furthermore, for $i = 1, \ldots, n$, after dropping some higher order terms, we have

$$
\begin{aligned}
& E\{[\hat{\eta}_{i,j}(s)-\eta_{i,j}(s)]^2|\mathcal{S},\boldsymbol{\eta},\mathbf{X}\} \\
= & \{E[\hat{\eta}_{i,j}(s)-\eta_{i,j}(s)|\mathcal{S},\boldsymbol{\eta},\mathbf{X}]\}^2+\mathrm{Var}[\hat{\eta}_{i,j}(s)-\eta_{i,j}(s)|\mathcal{S},\boldsymbol{\eta},\mathbf{X}] \\
= & [0.5h_j^2u_2(K)\mathbf{x}_i^T\ddot{B}_j(s_m)+0.5u_2(K)\ddot{\eta}_{i,j}(s)h_j^{(2)2}+\mathbf{x}_i^T\Omega_X^{-1}n^{-1}\sum_{i'=1}^n\mathbf{x}_{i'}\eta_{i',j}(s_m)]^2[1+o_p(1)] \\
& +v_0(K)\Sigma_{\epsilon,jj}(s,s)\pi(s)^{-1}(Mh_j^{(2)})^{-1}[1+o_p(1)]-\mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i(nMh_j)^{-1}\pi(s)^{-1}\Sigma_{\epsilon,jj}(s,s) \\
& \times\{2\int K(u)K(v)K(h_j^{(2)}(u-v)/h_j)dudv \\
& -\int K(u)K(v)K(w)K(w+h_j^{(2)}(u-v)/h_j)dudvdw\}[1+o_p(1)].
\end{aligned}
$$

This completes the proof of Theorem 2 (a).

It follows from (5.17) that

$$
\begin{aligned}
& \sum_{m=1}^M\tilde{K}_{M,h_j^{(2)}}(s_m-s)\mathbf{x}_i^T\{B_j(s_m)-E[\hat{B}_j(s_m)|\mathcal{S},\mathbf{X}]\} \\
= & \sum_{m=1}^M\tilde{K}_{M,h_j^{(2)}}(s_m-s)0.5h_j^2u_2(K)\mathbf{x}_i^T\ddot{B}_j(s_m)[1+O_p(h_j+n^{-1/2}+(Mh_j)^{-1/2})].
\end{aligned}
$$

Furthermore, it can be shown that

$$
\begin{aligned}
& \hat{\eta}_{i,j}(s)-\eta_{i,j}(s)-E[\hat{\eta}_{i,j}(s)|\mathcal{S},\mathbf{X}] \\
= & \sum_{m=1}^M\tilde{K}_{M,h_j^{(2)}}(s_m-s)[\eta_{i,j}(s_m)+\epsilon_{i,j}(s_m)-\eta_{i,j}(s)] \\
& -\sum_{m=1}^M\tilde{K}_{M,h_j^{(2)}}(s_m-s)\{\mathbf{x}_i^T[\mathbf{I}_p\otimes(1,0)]\Sigma(h_j,s_m)^{-1}[T_{\eta,j}(h_j,s_m)+T_{\epsilon,j}(h_j,s_m)]\}.
\end{aligned}
$$

With tedious calculations, we have

$$
\begin{aligned}
& \mathrm{Cov}(\hat{\eta}_{i,j}(s)-\eta_{i,j}(s),\hat{\eta}_{i,j}(t)-\eta_{i,j}(t)|\mathcal{S},\mathbf{X}) \\
= & K^*((s-t)/h_j^{(2)})\Sigma_{\epsilon,jj}(s,s)\pi(t)^{-1}(Mh_j^{(2)})^{-1}[1+o_p(1)] \\
& -\mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i(nMh_j)^{-1}\pi(s)^{-1}\pi(t)^{-1}O_p(1) \\
& +[1+o_p(1)]\{0.25u_2(K)^2h_j^{(2)4}\Sigma_{\eta,jj}^{(2,2)}(s,t)+n^{-1}\mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i\Sigma_{\eta,jj}(s,t) \\
& -0.5n^{-1}u_2(K)h_j^{(2)2}\mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i[\Sigma_{\eta,jj}^{(2,0)}(s,t)\pi(s)^{-1}+\Sigma_{\eta,jj}^{(0,2)}(s,t)\pi(t)^{-1}]\}.
\end{aligned}
$$

It follows from (5.17) that

$$
\begin{aligned}
& E\{[\hat{\eta}_{i,j}(s)-\eta_{i,j}(s)]^2|\mathcal{S},\mathbf{X}] \\
= \; & \{E[\hat{\eta}_{i,j}(s)-\eta_{i,j}(s)|\mathcal{S},\mathbf{X}]\}^2+\mathrm{Var}[\hat{\eta}_{i,j}(s)-\eta_{i,j}(s)|\mathcal{S},\mathbf{X}] \\
= \; & \{0.25h_j^4u_2(K)^2[\mathbf{x}_i^T\ddot{B}_j(s)]^2+0.25u_2(K)^2h_j^{(2)4}\Sigma_{\eta,jj}^{(2,2)}(s,t) \\
& +n^{-1}\mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i\Sigma_{\eta,jj}(s,t)+v_0(K)\Sigma_{\epsilon,jj}(s,s)\pi(s)^{-1}(Mh_j^{(2)})^{-1}\}[1+o_p(1)],
\end{aligned}
$$

which leads to Theorem 2 (b). Furthermore, by noting that $E\{[\hat{\eta}_{i,j}(s)-\eta_{i,j}(s)]^2|\mathcal{S}]=E(E\{[\hat{\eta}_{i,j}(s)-\eta_{i,j}(s)]^2|\mathcal{S},\mathbf{X}]|\mathcal{S})$, we can easily get Theorem 2 (c) and (d).

We define

$$
\begin{aligned}
\bar{\epsilon}_{i,j}(s) &= \sum_{m=1}^{M}\tilde{K}_{M,h_j^{(2)}}(s_m-s)\epsilon_{i,j}(s_m), \\
\Delta\eta_{i,j}(s) &= \sum_{m=1}^{M}\tilde{K}_{M,h_j^{(2)}}(s_m-s)[\eta_{i,j}(s_m)-\eta_{i,j}(s)], \\
\Delta B_j(s) &= \sum_{m=1}^{M}\tilde{K}_{M,h_j^{(2)}}(s_m-s)[B_j(s_m)-\hat{B}_j(s_m)], \\
\Delta_{i,j}(s) &= \bar{\epsilon}_{i,j}(s)+\Delta\eta_{i,j}(s)+\mathbf{x}_i^T\Delta B_j(s).
\end{aligned}
$$

Recall from (5.23) that

$$
\hat{\eta}_{i,j}(s)-\eta_{i,j}(s)=\Delta_{i,j}(s)=\bar{\epsilon}_{i,j}(s)+\Delta\eta_{i,j}(s)+\mathbf{x}_i^T\Delta B_j(s). \tag{5.24}
$$

It follows from Lemma 2 and a Taylor's expansion that

$$
\sup_{s\in[0,L_0]}|\bar{\epsilon}_{i,j}(s)|=O_p\Big(\sqrt{\frac{|\log(h_j^{(2)})|}{Mh_j^{(2)}}}\Big) \text{ and } \sup_{s\in[0,L_0]}|\Delta\eta_{i,j}(s)|=O_p(1)\sup_{s\in[0,L_0]}|\ddot{\eta}_{i,j}(s)|h_j^{(2)2}.
$$

Since $\sqrt{n}\{\hat{B}_j(\cdot)-B_j(\cdot)-0.5u_2(K)^2h_j^2\ddot{B}_j(\cdot)[1+o_p(1)]\}$ weakly converges to a Gaussian process in $\ell^\infty([0,L_0])$ as $n\to\infty$, $\sqrt{n}\{\hat{B}_j(\cdot)-B_j(\cdot)-0.5u_2(K)^2h_j^2\ddot{B}_j(\cdot)[1+o_p(1)]\}$ is asymptotically tight. Thus, we have

$$
\begin{aligned}
& \Delta B_{i,j}(s)=-\sum_{m=1}^{M}\tilde{K}_{M,h_j^{(2)}}(s_j-s)0.5u_2(K)^2h_j^2\ddot{B}_j(s_m)[1+o_p(1)] \\
& +\sum_{m=1}^{M}\tilde{K}_{M,h_j^{(2)}}(s_j-s)[0.5u_2(K)^2h_j^2\ddot{B}_j(s_m)[1+o_p(1)]+B_j(s_m)-\hat{B}_j(s_m)], \\
& \sup_{s\in[0,L_0]}||\Delta B_j(s)||=O_p(n^{-1/2})+O_p(h_j^2).
\end{aligned}
$$

Combining these results, we have

$$\sup_{s\in[0,L_0]} |\hat{\eta}_{i,j}(s)-\eta_{i,j}(s)| = O_p(|\log(h_j^{(2)})|^{1/2}(Mh_j^{(2)})^{-1/2}+h_j^{(2)2}+h_j^2+n^{-1/2}).$$

Lemma 6. Under Assumptions (C1), (C3), (C4) and (C7), we have

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^n \bar{\epsilon}_{i,j}(s)\eta_{i,j}(t)| = O_p(n^{-1/2}(\log n)^{1/2}). \tag{5.25}$$

The proof of Lemma 6 is the supplementary material Zhu et al. [46].

Lemma 7. Under Assumptions (C1), (C3), (C4) and (C7), we have

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^n \bar{\epsilon}_{i,j}(s)\bar{\epsilon}_{i,j}(t)| = O((Mh_j^{(2)})^{-1}+(\log n/n)^{1/2}) = o_p(1). \tag{5.26}$$

The proof of Lemma 6 is the supplementary material Zhu et al. [46].

Proof of Theorem 3. Recall that $\hat{\eta}_{i,j}(s)=\eta_{i,j}(s)+\Delta_{i,j}(s)$, we have

$$\begin{aligned} & n^{-1}\sum_{i=1}^n \hat{\eta}_{i,j}(s)\hat{\eta}_{i,j}(t) \\ = \ & n^{-1}\sum_{i=1}^n \Delta_{i,j}(s)\Delta_{i,j}(t)+n^{-1}\sum_{i=1}^n \eta_{i,j}(s)\Delta_{i,j}(t) \\ & +n^{-1}\sum_{i=1}^n \Delta_{i,j}(s)\eta_{i,j}(t)+n^{-1}\sum_{i=1}^n \eta_{i,j}(s)\eta_{i,j}(t). \end{aligned} \tag{5.27}$$

This proof consists of two steps. The first step is to show that the first three terms on the right hand side of (5.27) converge to zero uniformly for all $(s,t)\in[0,L_0]^2$ in probability. The second step is to show the uniform convergence of $n^{-1}\sum_{i=1}^n \eta_{i,j}(s)\eta_{i,j}(t)$ to $\Sigma_\eta(s,t)$ over $(s,t)\in[0,L_0]^2$ in probability.

We first show that

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^n \Delta_{i,j}(s)\eta_{i,j}(t)| = O_p(n^{-1/2}+h_j^2+h_j^{(2)2}+(\log n/n)^{1/2}). \tag{5.28}$$

Since

$$\begin{aligned} & \sum_{i=1}^n \Delta_{i,j}(s)\eta_{i,j}(t) \\ \leq \ & n^{-1}\{|\sum_{i=1}^n \bar{\epsilon}_{i,j}(s)\eta_{i,j}(t)|+|\sum_{i=1}^n \Delta\eta_{i,j}(s)\eta_{i,j}(t)|+|\sum_{i=1}^n \mathbf{x}_i^T\Delta B_j(s)\eta_{i,j}(t)|\}, \end{aligned} \tag{5.29}$$

it is sufficient to focus on the three terms on the right-hand side of (5.29). Since

$$|\mathbf{x}_i^T \Delta B_j(s)\eta_{i,j}(t)| \le ||\mathbf{x}_i||_2 \sup_{s\in[0,L_0]} ||\Delta B_k(s)||_2 \sup_{t\in[0,L_0]} |\eta_{i,j}(t)|,$$

we have

$$n^{-1}|\sum_{i=1}^n \mathbf{x}_i^T \Delta B_j(s)\eta_{i,j}(t)| \le \sup_{s\in[0,L_0]} ||\Delta B_k(s)||_2 n^{-1} \sum_{i=1}^n ||\mathbf{x}_i||_2 |\eta_{i,j}(t)| = O_p(n^{-1/2}+h_j^2).$$

Similarly, we have

$$n^{-1}|\sum_{i=1}^n \Delta\eta_{i,j}(s)\eta_{i,j}(t)| \le n^{-1}\sum_{i=1}^n \sup_{s,t\in[0,L_0]} |\Delta\eta_{i,j}(s)\eta_{i,j}(t)| = O_p(h_j^{(2)2}) = o_p(1).$$

It follows from Lemma 6 that $\sup_{(s,t)} n^{-1}\{|\sum_{i=1}^n \bar{\epsilon}_{i,j}(s)\eta_{i,j}(t)| = O((\log n/n)^{1/2})$. Similarly, we can show that $\sup_{(s,t)} n^{-1}|\sum_{i=1}^n \Delta_{i,j}(t)\eta_{i,j}(s)| = O_p(n^{-1/2} + h_j^2 + h_j^{(2)2} + (\log n/n)^{1/2})$.

We can show that

$$\sup_{(s,t)} |n^{-1}\sum_{i=1}^n [\eta_{i,j}(s)\eta_{i,j}(t) - \Sigma_{\eta,jj}(s,t)]| = O_p(n^{-1/2}). \tag{5.30}$$

Note that

$$\begin{aligned} & |\eta_{i,j}(s_1)\eta_{i,j}(t_1) - \eta_{i,j}(s_2)\eta_{i,j}(t_2)| \\ \le\ & 2(|s_1-s_2|+|t_1-t_2|) \sup_{s\in[0,L_0]} |\dot{\eta}_{i,j}(s)| \sup_{s\in[0,L_0]} |\eta_{i,j}(s)| \end{aligned}$$

holds for any (s_1,t_1) and (s_2,t_2), the functional class $\{\eta_j(u)\eta_j(v) : (u,v) \in [0,L_0]^2\}$ is a Vapnik and Cervonenkis (VC) class [38, 26]. Thus, it yields that (5.30) is true.

Finally, we can show that

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^n \Delta_{i,j}(s)\Delta_{i,j}(t)| = O_p((Mh_j^{(2)})^{-1} + (\log n/n)^{1/2} + h_j^4 + h_j^{(2)4}). \tag{5.31}$$

It follows from the Cauchy-Schwartz inequality that

$$\begin{aligned} & |\sum_{i=1}^n \Delta_{i,j}(s)\Delta_{i,j}(t)| \\ \le\ & C_1 \sup_{(s,t)}[|\sum_{i=1}^n \bar{\epsilon}_{i,j}(s)\bar{\epsilon}_{i,j}(t)| + |\sum_{i=1}^n \Delta\eta_{i,j}(s)\Delta\eta_{i,j}(t)| + |\sum_{i=1}^n \mathbf{x}_i^T \Delta B_j(s)\Delta B_j(t)\mathbf{x}_i|], \end{aligned}$$

for a positive constant C_1.

It follows from Lemma 7 that

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n}\bar{\epsilon}_{i,j}(s)\bar{\epsilon}_{i,j}(t)| = O_p((Mh_j^{(2)})^{-1} + (\log n/n)^{1/2}).$$

Since

$$\sup_{s\in[0,L_0]}|\Delta\eta_{i,j}(s)| = C_2 \sup_{s\in[0,L_0]}|\ddot{\eta}_{i,j}(s)|h_j^{(2)2},$$

we have $\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n}\Delta\eta_{i,j}(s)\Delta\eta_{i,j}(t)| = O(h_j^{(2)4})$. Furthermore, since $\sup_{s\in[0,L_0]}||\Delta\mathbf{B}(s)|| = O_p(n^{-1/2}+h_j^2)$, we have

$$|\sum_{i=1}^{n}\mathbf{x}_i^T\Delta B_j(s)\Delta B_j(t)\mathbf{x}_i| = O_p(n^{-1}+h_j^4).$$

Note that the arguments for (5.28)-(5.31) hold for $\hat{\Sigma}_{\eta,jj'}(\cdot,\cdot)$ for any $j\neq j'$. Thus, combining (5.28)-(5.31) leads to Theorem 3 (i).

To prove Theorem 3 (ii), we follow the same arguments in Lemma 6 of Li and Hsing [27]. For completion, we highlight several key steps below. We define

$$(\Delta\psi_{j,j})(s) = \int_0^{L_0}[\hat{\Sigma}_{\eta,jj}(s,t)-\Sigma_{\eta,jj}(s,t)]\psi_{j,j}(t)dt. \tag{5.32}$$

Following Hall and Hosseini-Nasab [15] and the Cauchy-Schwarz inequality, we have

$$\{\int_0^{L_0}[\hat{\psi}_{j,j}(s)-\psi_{j,j}(s)]^2ds\}^{1/2} \leq C_2 \sup_{(s,t)\in[0,L_0]^2}|\hat{\Sigma}_{\eta,jj}(s,t)-\Sigma_{\eta,jj}(s,t)|$$

for some constant C_2. This yields Theorem 3 (ii.a). See Zhu et al. [46] for more details.

Using (4.9) in Hall et al. [16], we have

$$\begin{aligned}
& |\hat{\lambda}_{j,j}-\lambda_{j,j}| \\
\leq\ & |\int_0^{L_0}\int_0^{L_0}[\hat{\Sigma}_{\eta,jj}-\Sigma_{\eta,jj}](s,t)\psi_{j,j}(s)\psi_{j,j}(t)dsdt + O(\int_0^{L_0}(\Delta\psi_{j,j})(s)^2ds) \\
\leq\ & C_3 \sup_{(s,t)\in[0,L_0]^2}|\hat{\Sigma}_{\eta,jj}(s,t)-\Sigma_{\eta,jj}(s,t)|
\end{aligned}$$

for some constant C_3. This yields Theorem 3 (ii.b). This complete the proof.

Since the proof of Corollary 1 is similar to the proof of Theorem 3 (i) and Theorem 4 follows from the continuous mapping theorem, we omit them for the sake of space.

We need to introduce some notation to establish the weak convergence of a sequence of stochastic processes indexed by $s \in [0, L_0]$ [38]. The uniform metric is used here to define the weak convergence. Let $\ell^\infty([0, L_0])$ be the space of all uniformly bounded, real functions on $[0, L_0]$, and endow $\ell^\infty([0, L_0])$ with the uniform metric. We consider $\mathcal{B}L_1(\ell^\infty([0, L_0])$ to be the space of real-valued functions on $\ell^\infty([0, L_0])$ with Lipschitz norm bounded by 1; that is, for any $k(\cdot) \in \mathcal{B}L_1(\ell^\infty([0, L_0]))$, $\sup_{x(s)\in\ell^\infty([0,L_0])} |k(x)| \le 1$ and $|k(x) - k(y)| \le ||x-y||_{[0,L_0]}$. As $n \to \infty$, a stochastic process $G_j(\cdot)$ weakly converges to $X(\cdot)$ on $\ell^\infty([0, L_0])$ if and only if $\sup_{k\in\mathcal{B}L_1(\ell^\infty([0,L_0]))} |Ek(G_j) - Ek(X)| \to 0$.

Proof of Theorem 5. We define $\mathbf{r}_{i,j}(s) = y_{i,j}(s) - \mathbf{x}_i^T B_j(s)$ and

(5.33)

$$\tilde{G}_j(s)^{(g)} = \sqrt{n}[I_p\otimes(1,0)]\mathrm{vec}(\Sigma(h_j, s)^{-1}\sum_{i=1}^n \tau_i^{(g)} \sum_{m=1}^M \mathbf{x}_i\otimes H_{h_j}(s_m-s)\mathbf{r}_{i,j}(s_m)).$$

Following the arguments in Kosorok [25] and Zhu and Zhang [47], we will prove Theorem 5 in three steps. In Step 1, we will prove the unconditional weak convergence of $\tilde{G}_j(s)^{(g)}$. In Step 2, we will prove the weak convergence of $\tilde{G}_j(s)^{(g)}$ conditional on the data. In Step 3, we will prove the weak convergence of $G_j(s)^{(g)}$ conditional on the data by showing that $\tilde{G}_j(s)^{(g)}$ and $G_j(s)^{(g)}$ are asymptotically equivalent as $n \to \infty$.

In Step 1, we note that $\mathbf{r}_{i,j}(s_m) = \eta_{i,j}(s_m) + \epsilon_{i,j}(s_m)$ and

$$\tilde{G}_j(s)^{(g)} = \sqrt{n}[I_p\otimes(1,0)]\mathrm{vec}(\Sigma(h_j, s)^{-1}\sum_{i=1}^n \tau_i^{(g)}\mathbf{x}_i\otimes\sum_{m=1}^M H_{h_j}(s_m-s)[\eta_{i,j}(s_m)+\epsilon_{i,j}(s_m)].$$

Therefore, by treating $\tau_i^{(g)}\mathbf{x}_i$ as the new 'covariate' vector, we can apply the same arguments in the proof of Theorem 1 to prove that $\tilde{G}_j^{(g)}$ converges to G_j in distribution; that is, $\tilde{G}_j^{(g)}$ is asymptotically measurable.

In Step 2, we define

$$\begin{aligned} S_j(s,t) &= n^{-1}n_G^{-2}\sum_{i=1}^n \mathbf{x}_i^{\otimes 2}\otimes\sum_{m,m'=1}^M H_{h_j}(s_m-s)H_{h_j}(s_{m'}-t)\mathbf{r}_{i,j}(s_m)\mathbf{r}_{i,j}(s_{m'}), \\ S_{j,\eta\eta}(s,t) &= n^{-1}n_G^{-2}\sum_{i=1}^n \mathbf{x}_i^{\otimes 2}\otimes\sum_{m,m'=1}^M H_{h_j}(s_m-s)H_{h_j}(s_{m'}-t)\eta_{i,j}(s_m)\eta_{i,j}(s_{m'}), \end{aligned}$$

$$\begin{aligned} S_{j,\eta\epsilon}(s,t) &= n^{-1}n_G^{-2}\sum_{i=1}^{n}\mathbf{x}_i^{\otimes 2}\otimes\sum_{m,m'=1}^{M}H_{h_j}(s_m-s)H_{h_j}(s_{m'}-t)\eta_{i,j}(s_m)\epsilon_{i,j}(s_{m'}),\\ S_{j,\epsilon\epsilon}(s,t) &= n^{-1}n_G^{-2}\sum_{i=1}^{n}\mathbf{x}_i^{\otimes 2}\otimes\sum_{m,m'=1}^{M}H_{h_j}(s_m-s)H_{h_j}(s_{m'}-t)\epsilon_{i,j}(s_m)\epsilon_{i,j}(s_{m'}). \end{aligned}$$

Thus, conditioning on the data, $\tilde{G}_j(s)^{(g)}$ is a normal random vector with zero mean and covariance given by $(nM)^{-2}[I_p\otimes(1,0)]\Sigma(h_j,s)^{-1}S_j(s,t)\Sigma(h_j,s)^{-1}[I_p\otimes(1,0)^T]$. It is easy to see that

$$S_j(s,t)=S_{j,\eta\eta}(s,t)+S_{j,\eta\epsilon}(s,t)+S_{j,\eta\epsilon}(t,s)+S_{j,\epsilon\epsilon}(s,t). \tag{5.34}$$

Following the arguments of Lemmas 6 and 7, we can show that $S_{j,\eta\epsilon}(s,t)+S_{j,\eta\epsilon}(t,s)+S_{j,\epsilon\epsilon}(s,t)=o(1)$. Furthermore, it can be shown that $E[S_{j,\eta\eta}(s,t)]=\Omega_X\otimes\text{diag}(1,0)\Sigma_{\eta,jj}(s,t)+O(h_j)$ and $\text{Cov}[S_{j,\eta\eta}(s,t)]=O(n^{-1})$. Therefore, $\text{Cov}_\tau[\tilde{G}_j(s)^{(g)},\tilde{G}_j(t)^{(g)}]$ converges to $\Sigma_{\eta,jj}(s,t)\Omega_X^{-1}$ in probability, where the expectation is taken with respect to $\tau_i^{(g)}$ conditioning on the data. We can obtain the marginal convergence of $\tilde{G}_j(s)^{(g)}$ in the conditional central limit theorem by using the Cramer-Wald method.

For each $\delta>0$, let $\tilde{\mathbf{S}}_\delta=\{l\delta: l=0,\ldots,L_0\delta^{-1}\}$ be an equally $\delta-$spaced grid and $[0,L_0]_\delta(s)$ assign to each $s\in[0,L_0]$ a closest element of $\tilde{\mathbf{S}}_\delta$. The finite convergence results yield

$$\sup_{k(\cdot)\in\mathcal{B}L_1(\ell^\infty([0,L_0])}|E_\tau k(\tilde{G}_j^{(g)}([0,L_0]_\delta))-Ek(G_j([0,L_0]_\delta))|\to 0$$

in probability, as $n\to\infty$. Due to the continuity of $G_j(s)$, we have $G_j([0,L_0]_\delta(s))\to G_j(s)$ almost surely as $\delta\to 0$; that is $\lim_{\delta\to 0}\sup_{k(\cdot)\in\mathcal{B}L_1(\ell^\infty([0,L_0])}|E_\tau k(\tilde{G}_j([0,L_0]_\delta)-E_\tau k(G_j([0,L_0]))|=0$. Finally, we have

$$\begin{aligned} &\sup_{k(\cdot)\in\mathcal{B}L_1(\ell^\infty([0,L_0])}|E_\tau k(\tilde{G}_j^{(g)}([0,L_0]_\delta(\cdot)))-E_\tau k(\tilde{G}_j^{(g)}([0,L_0](\cdot)))|\\ \leq\ & E_\tau(\sup_{|s-s'|_2\leq\delta}|\tilde{G}_j^{(g)}(s)-\tilde{G}_j^{(g)}(s')|). \end{aligned}$$

Thus, the expectation on the left side of the above equation is smaller than $E(\sup_{|s-s'|_2\leq\delta}|\tilde{G}_j^{(g)}(s)-\tilde{G}_j^{(g)}(s')|)$, which was established by the unconditional weak convergence of $\tilde{G}_j^{(g)}(\cdot)$ in Step 1. This finishes the proof of Step 2.

In Step 3, following the arguments in Theorem 3 of Kosorok [25], we only need to prove that $\Delta_{n,B} = o_p(1)$, where

$$\Delta_{n,B} = \sup_{s\in[0,L_0]} n^{-1}\sum_{i=1}^{n} \mathrm{tr}\{\mathbf{x}_i^{\otimes 2}\otimes\{M^{-1}\sum_{m=1}^{M} H_{h_j}(s_m - s)\mathbf{x}_i^T[\hat{B}_j(s_m) - B_j(s_m)]\}^{\otimes 2}\}.$$

It follows from the proof of Theorem 3 that $\Delta_{n,B} = O_p(n^{-1} + \hat{h}_j^4)$, which converges to zero in probability. This finishes the proof of Theorem 5.

References.

[1] Aguirre, G. K., Zarahn, E., and D'Esposito, M. (1998). The variability of human, bold hemodynamic responses. *NeuroImage*, 8:360–369.

[2] Basser, P. J., Mattiello, J., and LeBihan, D. (1994a). Estimation of the effective self- diffusion tensor from the nmr spin echo. *Journal of Magnetic Resonance Ser. B*, 103:247–254.

[3] Basser, P. J., Mattiello, J., and LeBihan, D. (1994b). Mr diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267.

[4] Buzsaki, G. (2006). *Rhythms of The Brain*. Oxford University Press.

[5] DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer Verlag, New York.

[6] Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *Journal of Theoretical Probability*, 13:1–37.

[7] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

[8] Fan, J., Yao, Q., and Cai, Z. (2003). Adaptive varying-coefficient linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65(1):57–80.

[9] Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.*, 27(5):1491–1518.

[10] Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.*, 27(4):715–731.

[11] Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Stat. Interface*, 1(1):179–195.

[12] Fass, L. (2008). Imaging and cancer: a review. *Molecular Oncology*, 2:115–152.

[13] Friston, K. J. (2007). *Statistical Parametric Mapping: the Analysis of Functional Brain Images*. Academic Press, London.

[14] Friston, K. J. (2009). Modalities, modes, and models in functional neuroimaging. *Science*, 326:399–403.

[15] Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society B*, 68:109–126.

[16] Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.*, 34(3):1493–1517.

[17] Hall, P., Müller, H.-G., and Yao, F. (2008). Modelling sparse generalized longitudinal observations with latent Gaussian processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 70(4):703–723.

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

[19] Heywood, I., Cornelius, S., and Carver, S. (2006). *An Introduction to Geographical Information Systems.* Prentice Hall., 3rd edition.

[20] 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(4):809–822.

[21] Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128.

[22] Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica*, 14(3):763–788.

[23] Huettel, S. A., Song, A. W., and McCarthy, G. (2004). *Functional Magnetic Resonance Imaging.* Sinauer Associates, Inc, London.

[24] Jiang, C. and Wang, J. (2010). Covariate-adjusted functional principal components analysis for longitudinal data. *The Annals of Statistics*, 38(2):1194–1226.

[25] Kosorok, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Multivariate Anal.*, 84:299–318.

[26] Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer: New York.

[27] Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38:3321–3351.

[28] Lindquist, M. (2008). The statistical analysis of fmri data. *Statistical Science*, 23:439–464.

[29] Lindquist, M., Loh, J. M., Atlas, L., and Wager, T. (2008). Modeling the hemodynamic response function in fmri: Efficiency, bias and mis-modeling. *NeuroImage*, 45:S187–S198.

[30] Ma, S., Yang, L., and Carroll, R. J. (2011). A simultaneous confidence band for sparse longitudinal regression. *Statistica Sinica*, 21.

[31] Niedermeyer, E. and da Silva, F. L. (2004). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields.* Lippincot Williams & Wilkins.

[32] Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis.* Springer Series in Statistics. Springer-Verlag, New York. Methods and case studies.

[33] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* Springer-Verlag, New York.

[34] Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B*, 53(1):233–243.

[35] Sun, J. and Loader, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing. *The Annals of Statistics*, 22:1328–1345.

[36] Talagrand, M. (1994). Sharper bounds for gaussian and empirical processes,. *Ann. Probab*, 22:28–76.

[37] Towle, V., Bolaños, J., Suarez, D., Tan, K., Grzeszczuk, R., Levin, D. N., Cakmur, R., Frank, S. A., and Spire, J. P. (1993). The spatial location of eeg electrodes: locating the best-fitting sphere relative to cortical anatomy. *Electroencephalogr Clin Neurophysiol*, 86:1–6.

[38] van der Vaar, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag Inc.

[39] Wang, L., Li, H., and Huang, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *J. Amer. Statist. Assoc.*, 103(484):1556–1569.

[40] Worsley, K. J., Taylor, J. E., Tomaiuolo, F., and Lerch, J. (2004). Unified univariate and multivariate random field theory. *NeuroImage*, 23:189–195.

[41] Wu, C. O. and Chiang, C.-T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statist. Sinica*, 10(2):433–456.

[42] 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. *J. Amer. Statist. Assoc.*, 93:1388–1402.

[43] Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(1):3–25.

[44] Zhang, J. and Chen, J. (2007). Statistical inference for functional data. *The Annals of Statistics*, 35:1052–1079.

[45] Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. R. Statist. Soc. B*, 72:513–531.

[46] Zhu, H., Li, R., and Kong, L. (2011). Supplementary material for "multivariate varying coefficient model and its application to neuroimaging data".

[47] Zhu, H. and Zhang, H. P. (2006). Generalized score test of homogeneity for mixed effects models. *Annals of Statistics*, 34:1545–1569.

[48] Zhu, H. T., Ibrahim, J. G., Tang, N., Rowe, D., Hao, X., Bansal, R., and Peterson, B. S. (2007a). A statistical analysis of brain morphology using wild bootstrapping. *IEEE Trans Med Imaging*, 26:954–966.

[49] Zhu, H. T., Zhang, H. P., Ibrahim, J. G., and Peterson, B. G. (2007b). Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data (with discussion). *Journal of the American Statistical Association*, 102:1085–1102.

Hongtu Zhu and Linglong Kong
Departments of Biostatistics
and Biomedical Research Imaging Center
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599, USA
E-mail: hzhu@bios.unc.edu
linglong.kong2008@gmail.com

Runze Li
Department of Statistics
The Pennsylvania State University
University Park, PA 16802
E-mail: rli@stat.psu.edu