

## MULTIVARIATE VARYING COEFFICIENT MODEL FOR FUNCTIONAL RESPONSES

BY HONGTU ZHU<sup>1</sup>, RUNZE LI<sup>2</sup> AND LINGLONG KONG<sup>1</sup>

*University of North Carolina at Chapel Hill, Pennsylvania State University  
and University of Alberta*

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. We apply MVCM to investigate the development of white matter diffusivities along the genu tract of the corpus callosum in a clinical study of neurodevelopment.

**1. Introduction.** With modern imaging techniques, massive imaging data can be observed over both time and space [4, 17, 19, 25, 37, 41]. 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 [17]. Among them, predominant functional imaging techniques including fMRI and EEG have been widely used in behavioral and cognitive neuroscience to understand functional segregation and integration of different brain

---

Received March 2012; revised September 2012.

<sup>1</sup>Supported by NIH Grants RR025747-01, P01CA142538-01, MH086633, EB005149-01 and AG033387.

<sup>2</sup>Supported by NSF Grant DMS-03-48869, 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.

*MSC2010 subject classifications.* Primary 62G05, 62G08; secondary 62G20.

*Key words and phrases.* Functional response, global test statistic, multivariate varying coefficient model, simultaneous confidence band, weak convergence.

regions in a single subject and across different populations [18, 19, 29]. 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, 54].

A common feature of many imaging techniques is that massive functional data are observed/calculated at the same design points, such as time for functional images (e.g., PET and fMRI). As an illustration, we present two smoothed functional data as an illustration and a real imaging data in Section 6, that we encounter in neuroimaging studies. First, we plot two diffusion properties, called fractional anisotropy (FA) and mean diffusivity (MD), measured at 45 grid points along the genu tract of the corpus callosum [Figure 1(a) and (b)] from 40 randomly selected infants from a clinical study of neurodevelopment with more than 500 infants. Scientists are particularly interested in delineating the structure of the variability of these functional FA and MD data and their association with a set of covariates of interest, such as age. We will systematically investigate the development of FA and MD along the genu of the corpus callosum tract in Section 6. Second, 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 randomly selected

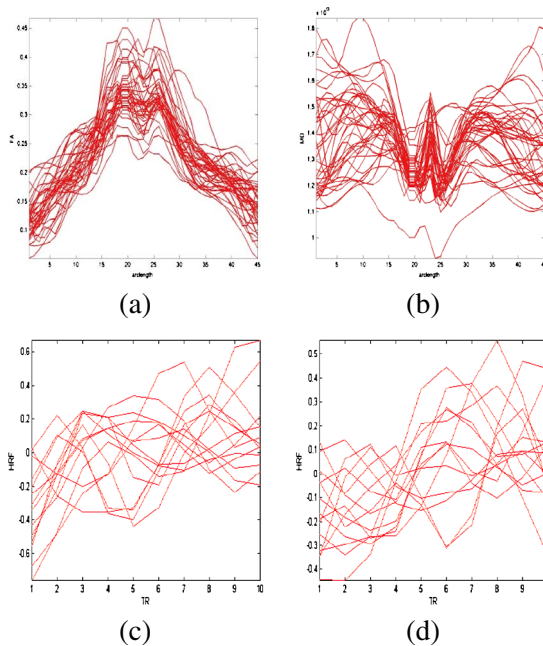


FIG. 1. *Representative functional neuroimaging data: (a) and (b) FA and MD along the genu tract of the corpus callosum from 40 randomly selected infants; and (c) and (d) the estimated hemodynamic response functions (HRF) corresponding to two stimulus categories from 14 subjects.*

subjects at a selected voxel of a common template space from a clinical study of Alzheimer's disease with more than 100 infants. Although the canonical form of the HRF is often used, when applying fMRI in a clinical population with possibly altered hemodynamic responses [Figure 1(c) and (d)], using the subject's own HRF in fMRI data analysis may be advantageous because HRF variability is greater across subjects than across brain regions within a subject [1, 33]. 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 [34].

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 [24], many varying-coefficient models have been widely studied and developed for longitudinal, time series and functional data [12, 13, 15, 23, 26–28, 38, 44, 47, 51]. However, most varying-coefficient models in the existing literature are developed for univariate response. Let  $\mathbf{y}_i(s) = (y_{i1}(s), \dots, y_{iJ}(s))^T$  be a  $J$ -dimensional functional response vector for subject  $i$ ,  $i = 1, \dots, n$ , and  $\mathbf{x}_i$  be its associated  $p \times 1$  vector of covariates of interest. Moreover,  $s$  varies in a compact subset of Euclidean space and denotes the design point, such as time for functional images and voxel for structural and functional images. For notational simplicity, we assume  $s \in [0, 1]$ , but our results can be easily extended to higher dimensions. A *multivariate varying coefficient model (MVCM)* is defined as

$$(1.1) \quad y_{ij}(s) = \mathbf{x}_i^T B_j(s) + \eta_{ij}(s) + \varepsilon_{ij}(s) \quad \text{for } j = 1, \dots, J,$$

where  $B_j(s) = (b_{j1}(s), \dots, b_{jp}(s))^T$  is a  $p \times 1$  vector of functions of  $s$ ,  $\varepsilon_{ij}(s)$  are measurement errors, and  $\eta_{ij}(s)$  characterizes individual curve variations from  $\mathbf{x}_i^T B_j(s)$ . Moreover,  $\{\eta_{ij}(s) : s \in [0, 1]\}$  is assumed to be a stochastic process indexed by  $s \in [0, 1]$  and used to characterize the within-curve dependence. For image data, it is typical that the  $J$  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 ease of notation, it is assumed throughout this paper that  $\mathbf{y}_i(s)$  was measured at the same  $M$  location points  $s_1 = 0 \leq s_2 \leq \dots \leq s_M = 1$  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 relatively well studied. Particularly, Hoover et al. [26] and Wu, Chiang and Hoover [48] were among the first to introduce the time-varying coefficient models for analysis of longitudinal data. Recently, Fan and Zhang [15] 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  $B_j$  instead of pointwise confidence intervals and to develop

global test statistics for the general hypothesis testing problem on  $B_j$ . For univariate varying coefficient models without the within-curve dependence, Fan and Zhang [14] constructed SCB using the limit theory for the maximum of the normalized deviation of the estimate from its expected value. Faraway [16], Chiou, Müller and Wang [8], and Cardot [5] proposed several varying coefficient models and their associated estimators for univariate functional response, but they did not give functional central limit theorem and simultaneous confidence band for their estimators. It has been technically difficult to carry out statistical inferences including simultaneous confidence band and global test statistic on  $B_j$  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  $M \rightarrow \infty$ , 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 [52]. For sparse irregular longitudinal data, Ma, Yang and Carroll [35] constructed 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, Degras [9] constructed asymptotic SCB for the mean function of the functional linear model without considering any covariate, while Zhang and Chen [51] adopted the method of “smoothing first, then estimation” and propose a global test statistic for testing  $B_j$ , but their results cannot be used for constructing SCB for  $B_j$ . Recently, Cardot et al. [6], Cardot and Josseland [7] built asymptotic SCB for Horvitz–Thompson estimators for the mean function, but their models and estimation methods differ significantly from ours.

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 theoretical properties of the proposed estimation procedure and test statistics. The theoretical development is challenging, but of great interest for carrying out statistical inferences on  $B_j$ . The major contributions of this paper are summarized as follows. We first establish the weak convergence of the local linear estimator of  $B_j$ , denoted by  $\hat{B}_j$ , by using advanced empirical process methods [31, 42]. We further derive the bias and asymptotic variance of  $\hat{B}_j$ . These results provide insight into how the direct estimation procedure for  $B_j$  using observations from all subjects outperforms the estimation procedure with the strategy of “smoothing first, then estimation.” After calculating  $\hat{B}_j$ , we reconstruct all individual functions  $\eta_{ij}$  and establish their uniform convergence rates. We derive uniform convergence rates of the proposed estimate for the covariance matrix of  $\eta_{ij}$  and its associated eigenvalue and eigenvector functions by using related results in Li and Hsing [32]. Using the weak convergence of the local linear estimator of  $B_j$ , 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 MVCMM and its estimation procedure. In Section 3, we propose a global test statistic for linear hypotheses of the regression coefficient functions and construct an asymptotic SCB for each coefficient function. In Section 4, we discuss the theoretical properties of estimation and inference procedures. Two sets of simulation studies are presented in Section 5 with the known ground truth to examine the finite sample performance of the global test statistic and SCB for each individual varying coefficient function. In Section 6, we use MVCMM to investigate the development of white matter diffusivities along the genu tract of the corpus callosum in a clinical study of neurodevelopment.

**2. Estimation procedure.** Throughout this paper, we assume that  $\boldsymbol{\varepsilon}_i(s) = (\varepsilon_{i1}(s), \dots, \varepsilon_{iJ}(s))^T$  and  $\boldsymbol{\eta}_i(s) = (\eta_{i1}(s), \dots, \eta_{iJ}(s))^T$  are mutually independent, and  $\boldsymbol{\eta}_i(s)$  and  $\boldsymbol{\varepsilon}_i(s)$  are independent and identical copies of  $SP(\mathbf{0}, \Sigma_\eta)$  and  $SP(\mathbf{0}, \Sigma_\varepsilon)$ , respectively, where  $SP(\mu, \Sigma)$  denotes a stochastic process vector with mean function  $\mu(t)$  and covariance function  $\Sigma(s, t)$ . Moreover,  $\boldsymbol{\varepsilon}_i(s)$  and  $\boldsymbol{\varepsilon}_i(t)$  are assumed to be independent for  $s \neq t$ , and  $\Sigma_\varepsilon(s, t)$  takes the form of  $S_\varepsilon(t)\mathbf{1}(s = t)$ , where  $S_\varepsilon(t) = (s_{\varepsilon, jj'}(t))$  is a  $J \times J$  matrix of functions of  $t$  and  $\mathbf{1}(\cdot)$  is an indicator function. Therefore, the covariance structure of  $\mathbf{y}_i(s)$ , denoted by  $\Sigma_y(s, t)$ , is given by

$$(2.1) \quad \Sigma_y(s, t) = \text{Cov}(\mathbf{y}_i(s), \mathbf{y}_i(t)) = \Sigma_\eta(s, t) + S_\varepsilon(t)\mathbf{1}(s = t).$$

2.1. *Estimating varying coefficient functions.* We employ local linear regression [11] to estimate the coefficient functions  $B_j$ . Specifically, we apply the Taylor expansion for  $B_j(s_m)$  at  $s$  as follows:

$$(2.2) \quad B_j(s_m) \approx B_j(s) + \dot{B}_j(s)(s_m - s) = A_j(s)\mathbf{z}_{h_{1j}}(s_m - s),$$

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

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

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

$C = (c_{jl})$ , denote  $\text{vec}(C) = (c_{11}, \dots, c_{1M_2}, \dots, c_{M_11}, \dots, c_{M_1M_2})^T$ . Let  $\hat{A}_j(s)$  be the minimizer of (2.3). Then

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

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

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

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

In practice, we may select the bandwidth  $h_{1j}$  by using leave-one-curve-out cross-validation. Specifically, for each  $j$ , we pool the data from all  $n$  subjects and select a bandwidth  $h_{1j}$ , denoted by  $\hat{h}_{1j}$ , by minimizing the cross-validation score given by

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

where  $\hat{B}_j(s, h_{1j})^{(-i)}$  is the local linear estimator of  $B_j(s)$  with the bandwidth  $h_{1j}$  based on data excluding all the observations from the  $i$ th subject.

2.2. *Smoothing individual functions.* By assuming certain smoothness conditions on  $\eta_{ij}(s)$ , we also employ the local linear regression technique to estimate all individual functions  $\eta_{ij}(s)$  [11, 38, 43, 45, 49, 51]. Specifically, we have the Taylor expansion for  $\eta_{ij}(s_m)$  at  $s$ ,

$$(2.7) \quad \eta_{ij}(s_m) \approx \mathbf{d}_{ij}(s)^T \mathbf{z}_{h_{2j}}(s_m - s),$$

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

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

Then,  $\eta_{ij}(s)$  can be estimated by

$$(2.9) \quad \begin{aligned} \hat{\eta}_{ij}(s) &= (1, 0) \hat{\mathbf{d}}_{ij}(s) \\ &= \sum_{m=1}^M \tilde{K}_{h_{2j}}(s_m - s) [y_{ij}(s_m) - \mathbf{x}_i^T \hat{B}_j(s_m)], \end{aligned}$$

where  $\tilde{K}_{h_{2j}}(s)$  are the empirical equivalent kernels and  $\hat{\mathbf{d}}_{ij}(s)$  is given by

$$\hat{\mathbf{d}}_{ij}(s) = \left[ \sum_{m=1}^M K_{h_{2j}}(s_m - s) \mathbf{z}_{h_{2j}}(s_m - s)^{\otimes 2} \right]^{-1} \times \sum_{m=1}^M K_{h_{2j}}(s_m - s) \mathbf{z}_{h_{2j}}(s_m - s) [y_{ij}(s_m) - \mathbf{x}_i^T \hat{\mathbf{B}}_j(s_m)].$$

Finally, let  $S_{ij}$  be the smoother matrix for the  $j$ th measurement of the  $i$ th subject [11], we can obtain

$$(2.10) \quad \hat{\boldsymbol{\eta}}_{ij} = (\hat{\eta}_{ij}(s_1), \dots, \hat{\eta}_{ij}(s_M))^T = S_{ij} R_{ij},$$

where  $R_{ij} = (y_{ij}(s_1) - \mathbf{x}_i^T \hat{\mathbf{B}}_j(s_1), \dots, y_{ij}(s_M) - \mathbf{x}_i^T \hat{\mathbf{B}}_j(s_M))^T$ .

A simple and efficient way to obtain  $h_{2j}$  is to use generalized cross-validation method. For each  $j$ , we pool the data from all  $n$  subjects and select the optimal bandwidth  $h_{2j}$ , denoted by  $\hat{h}_{2j}$ , by minimizing the generalized cross-validation score given by

$$(2.11) \quad \text{GCV}(h_{2j}) = \sum_{i=1}^n \frac{R_{ij}^T (I_M - S_{ij})^T (I_M - S_{ij}) R_{ij}}{[1 - M^{-1} \text{tr}(S_{ij})]^2}.$$

Based on  $\hat{h}_{2j}$ , we can use (2.9) to estimate  $\eta_{ij}(s)$  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 its approximation. According to Mercer’s theorem [36], if  $\Sigma_{\eta}(s, t)$  is continuous on  $[0, 1] \times [0, 1]$ , then  $\Sigma_{\eta, jj}(s, t)$  admits a spectral decomposition. Specifically, we have

$$(2.12) \quad \Sigma_{\eta, jj}(s, t) = \sum_{l=1}^{\infty} \lambda_{jl} \psi_{jl}(s) \psi_{jl}(t)$$

for  $j = 1, \dots, J$ , where  $\lambda_{j1} \geq \lambda_{j2} \geq \dots \geq 0$  are ordered values of the eigenvalues of a linear operator determined by  $\Sigma_{\eta, jj}$  with  $\sum_{l=1}^{\infty} \lambda_{jl} < \infty$  and the  $\psi_{jl}(t)$ ’s are the corresponding orthonormal eigenfunctions (or principal components) [22, 32, 50]. The eigenfunctions form an orthonormal system on the space of square-integrable functions on  $[0, 1]$ , and  $\eta_{ij}(t)$  admits the Karhunen–Loeve expansion as  $\eta_{ij}(t) = \sum_{l=1}^{\infty} \xi_{ijl} \psi_{jl}(t)$ , where  $\xi_{ijl} = \int_0^1 \eta_{ij}(s) \psi_{jl}(s) ds$  is referred to as the  $(jl)$ th functional principal component scores of the  $i$ th subject. For each fixed  $(i, j)$ , the  $\xi_{ijl}$ ’s are uncorrelated random variables with  $E(\xi_{ijl}) = 0$  and  $E(\xi_{ijl}^2) = \lambda_{jl}$ . Furthermore, for  $j \neq j'$ , we have

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

After obtaining  $\hat{\boldsymbol{\eta}}_i(s) = (\hat{\eta}_{i1}(s), \dots, \hat{\eta}_{iJ}(s))^T$ , we estimate  $\Sigma_\eta(s, t)$  by using the empirical covariance of the estimated  $\hat{\boldsymbol{\eta}}_i(s)$  as follows:

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

Following Rice and Silverman [39], we can calculate the spectral decomposition of  $\hat{\Sigma}_{\eta, jj}(s, t)$  for each  $j$  as follows:

$$(2.13) \quad \hat{\Sigma}_{\eta, jj}(s, t) = \sum_l \hat{\lambda}_{jl} \hat{\psi}_{jl}(s) \hat{\psi}_{jl}(t),$$

where  $\hat{\lambda}_{j1} \geq \hat{\lambda}_{j2} \geq \dots \geq 0$  are estimated eigenvalues and the  $\hat{\psi}_{jl}(t)$ 's are the corresponding estimated principal components. Furthermore, the  $(j, l)$ th functional principal component scores can be computed using  $\hat{\xi}_{ijl} = \sum_{m=1}^M \hat{\eta}_{ij}(s_m) \times \hat{\psi}_{jl}(s_m)(s_m - s_{m-1})$  for  $i = 1, \dots, n$ . 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.

**3. Inference procedure.** 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:

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

where  $\mathbf{B}(s) = [B_1(s), \dots, B_J(s)]$ ,  $\mathbf{C}$  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

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

where  $\hat{\Omega}_X = \sum_{i=1}^n \mathbf{x}_i^{\otimes 2}$  and  $\mathbf{d}(s) = \mathbf{C} \text{vec}(\hat{\mathbf{B}}(s) - \text{bias}(\hat{\mathbf{B}}(s))) - \mathbf{b}_0(s)$ .

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

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



By using Taylor’s expansion, we have

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

where  $\ddot{B}_j(s) = d^2B_j(s)/ds^2$  and  $\ddot{\ddot{B}}_j(s) = d^3B_j(s)/ds^3$ . Following the pre-asymptotic substitution method of Fan and Gijbels [11], we replace  $B_j(s_m) - \hat{A}_j(s)\mathbf{z}_{h_{1j}}(s_m - s)$  by  $2^{-1}\hat{\ddot{B}}_j(s)(s_m - s)^2 + 6^{-1}\hat{\ddot{\ddot{B}}}_j(s)(s_m - s)^3$ , in which  $\hat{\ddot{B}}_j(s)$  and  $\hat{\ddot{\ddot{B}}}_j(s)$  are estimators obtained by using local cubic fit with a pilot bandwidth selected in (2.6).

It will be shown below that the asymptotic distribution of  $S_n$  is quite complicated, and it is difficult to directly approximate the percentiles of  $S_n$  under the null hypothesis. Instead, 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  $\hat{B}^*(s_m)$ ,  $\hat{\eta}_{i,0}^*(s_m)$  and  $\hat{\epsilon}_{i,0}^*(s_m)$  for  $i = 1, \dots, n$  and  $m = 1, \dots, M$ .

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

$$\hat{\mathbf{y}}_i(s_m)^{(g)} = \hat{B}^*(s)^T \mathbf{x}_i + \tau_i^{(g)} \hat{\eta}_{i,0}^*(s_m) + \tau_i(s_m)^{(g)} \hat{\epsilon}_{i,0}^*(s_m).$$

Then, based on  $\hat{\mathbf{y}}_i(s_m)^{(g)}$ , we recalculate  $\hat{\mathbf{B}}(s)^{(g)}$ ,  $\text{bias}(\hat{\mathbf{B}}(s)^{(g)})$  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

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

STEP 3. Repeat Step 2  $G$  times to obtain  $\{S_n^{(g)} : g = 1, \dots, G\}$ , 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  $\alpha$ , 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). For a given confidence level  $\alpha$ , we construct SCB for each  $b_{jl}(s)$  as follows:

$$(3.4) \quad P(\hat{b}_{jl}^{L,\alpha}(s) < b_{jl}(s) < \hat{b}_{jl}^{U,\alpha}(s) \text{ for all } s \in [0, 1]) = 1 - \alpha,$$

where  $\hat{b}_{jl}^{L,\alpha}(s)$  and  $\hat{b}_{jl}^{U,\alpha}(s)$  are the lower and upper limits of SCB. Specifically, it will be shown below that a  $1 - \alpha$  simultaneous confidence band for  $b_{jl}(s)$  is given as follows:

$$(3.5) \quad \left( \hat{b}_{jl}(s) - \text{bias}(\hat{b}_{jl}(s)) - \frac{C_{jl}(\alpha)}{\sqrt{n}}, \hat{b}_{jl}(s) - \text{bias}(\hat{b}_{jl}(s)) + \frac{C_{jl}(\alpha)}{\sqrt{n}} \right),$$

where  $C_{jl}(\alpha)$  is a scalar. Since the calculation of  $\hat{b}_{jl}(s)$  and  $\text{bias}(\hat{b}_{jl}(s))$  has been discussed in (2.5) and (3.3), the next issue is to determine  $C_{jl}(\alpha)$ .

Although there are several methods of determining  $C_{jl}(\alpha)$  including random field theory [40, 46], we develop an efficient resampling method to approximate  $C_{jl}(\alpha)$  as follows [30, 55]:

- We calculate  $\hat{r}_{ij}(s_m) = y_{ij}(s_m) - \mathbf{x}_i^T \hat{B}_j(s_m)$  for all  $i, j$ , and  $m$ .
- For  $g = 1, \dots, G$ , we independently simulate  $\{\tau_i^{(g)} : i = 1, \dots, n\}$  from  $N(0, 1)$  and calculate a stochastic process  $G_j(s)^{(g)}$  given by

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

- We calculate  $\sup_{s \in [0, 1]} |\mathbf{e}_l G_j(s)^{(g)}|$  for all  $g$ , where  $\mathbf{e}_l$  be a  $p \times 1$  vector with the  $l$ th element 1 and 0 otherwise, and use their  $1 - \alpha$  empirical percentile to estimate  $C_{jl}(\alpha)$ .

**4. Asymptotic properties.** In this section, we systematically examine the asymptotic properties of  $\hat{\mathbf{B}}(s)$ ,  $\hat{\eta}_{ij}(s)$ ,  $\hat{\Sigma}_\eta(s, t)$  and  $S_n$  developed in Sections 2 and 3. Let us first define some notation. Let  $u_r(K) = \int t^r K(t) dt$  and  $v_r(K) = \int t^r K^2(t) dt$ , where  $r$  is any integer. For any smooth functions  $f(s)$  and  $g(s, t)$ , define  $\dot{f}(s) = df(s)/ds$ ,  $\ddot{f}(s) = d^2 f(s)/ds^2$ ,  $\dddot{f}(s) = d^3 f(s)/ds^3$  and  $g^{(a,b)}(s, t) = \partial^{a+b} g(s, t) / \partial^a s \partial^b t$ , where  $a$  and  $b$  are any nonnegative integers. Let  $\mathbf{H} = \text{diag}(h_{11}, \dots, h_{1J})$ ,  $\mathbf{B}(s) = [B_1(s), \dots, B_J(s)]$ ,  $\hat{\mathbf{B}}(s) = [\hat{B}_1(s), \dots, \hat{B}_J(s)]$  and  $\ddot{\mathbf{B}}(s) = [\ddot{B}_1(s), \dots, \ddot{B}_J(s)]$ , where  $\ddot{B}_j(s) = (\ddot{b}_{j1}(s), \dots, \ddot{b}_{jp}(s))^T$  is a  $p \times 1$  vector. Let  $\mathcal{S} = \{s_1, \dots, s_M\}$ .

4.1. *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  $\mu$  and covariance  $\Sigma$ . Let  $\Omega_1(h, s) = \int (1, h^{-1}(u - s))^{\otimes 2} K_h(u - s) \pi(u) du$ . Moreover, we do not distinguish the differentiation and continuation at the boundary points from those in the interior of  $[0, 1]$ . For instance, a continuous function at the boundary of  $[0, 1]$  means that this function is left continuous at 0 and right continuous at 1.

ASSUMPTION (C1). For all  $j = 1, \dots, J$ ,  $\sup_{s_m} E[|\varepsilon_{ij}(s_m)|^q] < \infty$  for some  $q > 4$  and all grid points  $s_m$ .

ASSUMPTION (C2). Each component of  $\{\eta(s) : s \in [0, 1]\}$ ,  $\{\eta(s)\eta(t)^T : (s, t) \in [0, 1]^2\}$  and  $\{\mathbf{x}\eta^T(s) : s \in [0, 1]\}$  are Donsker classes.

ASSUMPTION (C3). The covariate vectors  $\mathbf{x}_i$ 's are independently and identically distributed with  $E\mathbf{x}_i = \mu_x$  and  $\|\mathbf{x}_i\|_\infty < \infty$ . Assume that  $E[\mathbf{x}_i^{\otimes 2}] = \Omega_X$  is positive definite.

ASSUMPTION (C4). The grid points  $\mathcal{S} = \{s_m, m = 1, \dots, M\}$  are randomly generated from a density function  $\pi(s)$ . Moreover,  $\pi(s) > 0$  for all  $s \in [0, 1]$  and  $\pi(s)$  has continuous second-order derivative with the bounded support  $[0, 1]$ .

ASSUMPTION (C4b). The grid points  $\mathcal{S} = \{s_m, m = 1, \dots, M\}$  are prefixed according to  $\pi(s)$  such that  $\int_0^{s_m} \pi(s) ds = m/M$  for  $M \geq m \geq 1$ . Moreover,  $\pi(s) > 0$  for all  $s \in [0, 1]$  and  $\pi(s)$  has continuous second-order derivative with the bounded support  $[0, 1]$ .

ASSUMPTION (C5). The kernel function  $K(t)$  is a symmetric density function with a compact support  $[-1, 1]$ , and is Lipschitz continuous. Moreover,  $0 < \inf_{h \in (0, h_0], s \in [0, 1]} \det(\Omega_1(h, s))$ , where  $h_0 > 0$  is a small scalar and  $\det(\Omega_1(h, s))$  denotes the determinant of  $\Omega_1(h, s)$ .

ASSUMPTION (C6). All components of  $\mathbf{B}(s)$  have continuous second derivatives on  $[0, 1]$ .

ASSUMPTION (C7). Both  $n$  and  $M$  converge to  $\infty$ ,  $\max_j h_{1j} = o(1)$ ,  $Mh_{1j} \rightarrow \infty$  and  $\max_j h_{1j}^{-1} |\log h_{1j}|^{1-2/q_1} \leq M^{1-2/q_1}$  for  $j = 1, \dots, J$ , where  $q_1 \in (2, 4)$ .

ASSUMPTION (C7b). Both  $n$  and  $M$  converge to  $\infty$ ,  $\max_j h_{1j} = o(1)$ ,  $Mh_{1j} \rightarrow \infty$  and  $\log(M) = o(Mh_{1j})$ . There exists a sequence of  $\gamma_n > 0$  such that  $\gamma_n \rightarrow \infty$ ,  $\max_j n^{1/2} \gamma_n^{1-q} h_{1j}^{-1} = o(1)$  and  $n^{-1/2} \gamma_n \log(M) = o(1)$ .

ASSUMPTION (C8). For all  $j$ ,  $\max_j (h_{2j})^{-4} (\log n/n)^{1-2/q_2} = o(1)$  for  $q_2 \in (2, \infty)$ ,  $\max_j h_{2j} = o(1)$ , and  $Mh_{2j} \rightarrow \infty$  for  $j = 1, \dots, J$ .

ASSUMPTION (C9). The sample path of  $\eta_{ij}(s)$  has continuous second-order derivative on  $[0, 1]$  and  $E[\sup_{s \in [0, 1]} \|\boldsymbol{\eta}(s)\|_2^{r_1}] < \infty$  and  $E\{\sup_{s \in [0, 1]} [\|\dot{\boldsymbol{\eta}}(s)\|_2 + \|\ddot{\boldsymbol{\eta}}(s)\|_2]^{r_2}\} < \infty$  for some  $r_1, r_2 \in (2, \infty)$ , where  $\|\cdot\|_2$  is the Euclidean norm.

ASSUMPTION (C9b).  $E[\sup_{s \in [0, 1]} \|\boldsymbol{\eta}(s)\|_2^{r_1}] < \infty$  for some  $r_1 \in (2, \infty)$  and all components of  $\Sigma_\eta(s, t)$  have continuous second-order partial derivatives with respect to  $(s, t) \in [0, 1]^2$  and  $\inf_{s \in [0, 1]} \Sigma_\eta(s, s) > 0$ .

ASSUMPTION (C10). There is a positive fixed integer  $E_j < \infty$  such that  $\lambda_{j,1} > \dots > \lambda_{j,E_j} > \lambda_{j,E_j+1} \geq \dots \geq 0$  for  $j = 1, \dots, J$ .

REMARK. Assumption (C1) requires the uniform bound on the high-order moment of  $\varepsilon_{ij}(s_m)$  for all grid points  $s_m$ . Assumption (C2) avoids smoothness conditions on the sample path  $\eta(s)$ , which are commonly assumed in the literature [9, 22, 51]. Assumption (C3) is a relatively weak condition on the covariate vector, and the boundedness of  $\|\mathbf{x}_i\|_2$  is not essential. Assumption (C4) is a weak condition on the random grid points. In many neuroimaging applications,  $M$  is often much larger than  $n$  and for such large  $M$ , a regular grid of voxels is fairly well approximated by voxels generated by a uniform distribution in a compact subset of Euclidean space. For notational simplicity, we only state the theoretical results for the random grid points throughout the paper. Assumption (C4b) is a weak condition on the fixed grid points. We will prove several key results for the fixed grid point case in Lemma 8 of the supplemental article [53]. The bounded support restriction on  $K(\cdot)$  in Assumption (C5) is not essential and can be removed if we put a restriction on the tail of  $K(\cdot)$ . Assumption (C6) is the standard smoothness condition on  $\mathbf{B}(s)$  in the literature [12, 13, 15, 23, 26–28, 38, 44, 47, 51]. Assumptions (C7) and (C8) on bandwidths are similar to the conditions used in [10, 32]. Assumption (C7b) is a weak condition on  $n$ ,  $M$ ,  $h_{1j}$  and  $\gamma_n$  for the fixed grid point case. For instance, if we set  $\gamma_n = n^{1/2} \log(M)^{-1-c_0}$  for a positive scalar  $c_0 > 0$ , then we have  $n^{1/2} \gamma_n^{1-q} h_{1j}^{-1} = n^{1-q/2} \log(M)^{(1+c_0)(q-1)} h_{1j}^{-1} = o(1)$  and  $n^{-1/2} \gamma_n \log(M) = \log(M)^{-c_0} = o(1)$ . As shown in Theorem 1 below, if  $h_{1j} = O((nM)^{-1/5})$  and  $\gamma_n = n^{1/2} \log(M)^{-1-c_0}$ ,  $n^{1/2} \gamma_n^{1-q} h_{1j}^{-1}$  reduces to  $n^{6/5-q/2} \log(M)^{(1+c_0)(q-1)} M^{1/5}$ . For relatively large  $q$  in Assumption (C1),  $n^{6/5-q/2} \log(M)^{(1+c_0)(q-1)} M^{1/5}$  can converge to zero. Assumptions (C9) and (C3) are sufficient conditions of Assumption (C2). Assumption (C9b) on the sample path is the same as Condition C6 used in [32]. Particularly, if we use the method for estimating  $\Sigma_\eta(s, s')$  considered in Li and Hsing [32], then the differentiability of  $\eta(s)$  in Assumption (C9) can be dropped. Assumption (C10) on simple multiplicity of the first  $E_j$  eigenvalues is only needed to investigate the asymptotic properties of eigenfunctions.

4.2. *Asymptotic properties of  $\hat{\mathbf{B}}(s)$ .* The following theorem establishes the weak convergence of  $\{\hat{\mathbf{B}}(s), s \in [0, 1]\}$ , which is essential for constructing global test statistics and SCB for  $\mathbf{B}(s)$ .

THEOREM 1. *Suppose that Assumptions (C1)–(C7) hold. The following results hold:*

- (i)  $\sqrt{n}\{\text{vec}(\hat{\mathbf{B}}(s) - \mathbf{B}(s) - 0.5\ddot{\mathbf{B}}(s)\mathbf{U}_2(K; s, \mathbf{H})\mathbf{H}^2[1 + o_p(1)]) : s \in [0, 1]\}$  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 in the Appendix.

(ii) *The asymptotic bias and conditional variance of  $\hat{B}_j(s)$  given  $\mathcal{S}$  for  $s \in (0, 1)$  are given by  $0.5h_{1j}^2 u_2(K) \ddot{B}_j(s)[1 + o_p(1)]$  and  $n^{-1} \Sigma_{\eta, jj}(s, s) \Omega_X^{-1} [1 + o_p(1)]$ , respectively.*

REMARKS. (1) The major challenge in proving Theorem 1(i) is dealing with within-subject dependence. This is because the dependence between  $\eta(s)$  and  $\eta(s')$  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 [51]. However, the marginal asymptotic normality does not imply the weak convergence of  $\hat{\mathbf{B}}(s)$  as a stochastic process in  $[0, 1]$ , since we need to verify the asymptotic continuity of  $\{\hat{\mathbf{B}}(s) : s \in [0, 1]\}$  to establish its weak convergence. In addition, Zhang and Chen [51] 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 [51] 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) 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  $(0, 1)$ . The asymptotic bias and conditional variance at the boundary points 0 and 1 are given in Lemma 5. The asymptotic bias of  $\hat{B}_j(s)$  is of the order  $h_{1j}^2$ , 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 of order  $n^{-1}$ , which is slower than the standard nonparametric rate  $(nMh_{1j})^{-1}$  with the assumption  $h_{1j} \rightarrow 0$  and  $Mh_{1j} \rightarrow \infty$ .

(3) Choosing an optimal bandwidth  $h_{1j}$  is not a trivial task for model (1.1). Generally, any bandwidth  $h_{1j}$  satisfying the assumptions  $h_{1j} \rightarrow 0$  and  $Mh_{1j} \rightarrow \infty$  can ensure the weak convergence of  $\{\hat{\mathbf{B}}(s) : s \in [0, 1]\}$ . Based on the asymptotic bias and conditional variance of  $\hat{\mathbf{B}}(s)$ , we can calculate an optimal bandwidth for estimating  $\mathbf{B}(s)$ ,  $h_{1j} = O_p((nM)^{-1/5})$ . In this case,  $n^{-1}h_{1j}^2$  and  $(nM)^{-1}h_{1j}$  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  $M$ .

4.3. *Asymptotic properties of  $\hat{\eta}_{ij}(s)$ .* We next study the asymptotic bias and covariance of  $\hat{\eta}_{ij}(s)$  as follows. We distinguish between two cases. The first one is conditioning on the design points in  $\mathcal{S}$ ,  $\mathbf{X}$ , and  $\eta$ . 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) and (C3)–(C8), the following results hold for all  $s \in (0, L)$ :*

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

$$\begin{aligned} & \text{Bias}[\hat{\eta}_{ij}(s)|\mathcal{S}, \boldsymbol{\eta}, \mathbf{x}_i] \\ &= 0.5u_2(K)[\hat{\eta}_{ij}(s)h_{2j}^2 + \mathbf{x}_i^T \ddot{B}_j(s_m)h_{1j}^2][1 + o_p(1)] + O_p(n^{-1/2}), \\ & \text{Cov}[\hat{\eta}_{ij}(s), \hat{\eta}_{ij}(t)|\mathcal{S}, \boldsymbol{\eta}, \mathbf{x}_i] \\ &= \mathbf{K}^*((s - t)/h_{2j})\pi(t)^{-1}(Mh_{2j})^{-1}O_p(1) - \mathbf{x}_i^T \Omega_X^{-1} \mathbf{x}_i (nMh_{1j})^{-1}O_p(1). \end{aligned}$$

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

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

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

$$\begin{aligned} & n^{-1} \sum_{i=1}^n \int_0^1 E\{[\hat{\eta}_{ij}(s) - \eta_{ij}(s)]^2|\mathcal{S}\}\pi(s) ds \\ (4.1) \quad &= [1 + o_p(1)] \\ &\quad \times \left\{ O((Mh_{2j})^{-1}) + n^{-1} \int_0^1 \Sigma_{\eta,jj}(s, s)\pi(s) ds \right. \\ &\quad \left. + 0.25u_2^2(K) \int_0^1 [\ddot{B}_j(s)^T \Omega_X \ddot{B}_j(s)h_{1j}^4 + \Sigma_{\eta,jj}^{(2,2)}(s, s)h_{2j}^4]\pi(s) ds \right\}. \end{aligned}$$

(d) *The optimal bandwidth for minimizing MISE (4.1) is given by*

$$(4.2) \quad \hat{h}_{2j} = O(M^{-1/5}).$$

(e) *The first order LPK reconstructions  $\hat{\eta}_{ij}(s)$  using  $\hat{h}_{2j}$  in (4.2) satisfy*

$$(4.3) \quad \sup_{s \in [0,1]} |\hat{\eta}_{ij}(s) - \eta_{ij}(s)| = O_p(|\log(M)|^{1/2}M^{-2/5} + h_{1j}^2 + n^{-1/2})$$

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

REMARK. Theorem 2 characterizes the statistical properties of smoothing individual curves  $\eta_{ij}(s)$  after first estimating  $B_j(s)$ . Conditioning on individual curves  $\eta_{ij}(s)$ , Theorem 2(a) shows that  $\text{Bias}[\hat{\eta}_{ij}(s)|\mathcal{S}, \mathbf{X}, \boldsymbol{\eta}]$  is associated with

$0.5u_2(K)\mathbf{x}_i^T \ddot{B}_j(s_m)h_{1j}^2$ , which is the bias term of  $\hat{B}_j(s)$  introduced in the estimation step, and  $0.5u_2(K)\dot{\eta}_{ij}(s)h_{2j}^2$  is introduced in the smoothing individual functions step. Without conditioning on  $\eta_{ij}(s)$ , Theorem 2(b) shows that the bias of  $\hat{\eta}_{ij}(s)$  is mainly controlled by the bias in the estimation step. The MISE of  $\hat{\eta}_{ij}(s)$  in Theorem 2(c) is the sum of  $O_p(n^{-1} + h_{1j}^4)$  introduced by the estimation of  $B_j(s)$  and  $O_p((Mh_{2j})^{-1} + h_{2j}^4)$  introduced by the reconstruction of  $\eta_{ij}(s)$ . The optimal bandwidth for minimizing the MISE of  $\hat{\eta}_{ij}(s)$  is a standard bandwidth for LPK. If we use the optimal bandwidth in Theorem 2(d), then the MISE of  $\hat{\eta}_{ij}(s)$  can achieve the order of  $n^{-1} + h_{1j}^4 + M^{-4/5}$ .

4.4. *Asymptotic properties of  $\hat{\Sigma}_\eta(s, t)$ .* In this section, we study the asymptotic properties of  $\hat{\Sigma}_\eta(s, t)$  and its spectrum decomposition.

THEOREM 3. (i) *Under Assumptions (C1) and (C3)–(C9), it follows that*

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

(ii) *Under Assumptions (C1) and (C3)–(C10), if the optimal bandwidths  $h_{mj}$  for  $m = 1, 2$  are used to reconstruct  $\hat{B}_j(s)$  and  $\hat{\eta}_{ij}(s)$  for all  $j$ , then for  $l = 1, \dots, E_j$ , we have the following results:*

- (a)  $\int_0^1 [\hat{\psi}_{jl}(s) - \psi_{jl}(s)]^2 ds = O_p((Mh_{2j})^{-1} + h_{1j}^2 + h_{2j}^2 + (\log n/n)^{1/2});$
- (b)  $|\hat{\lambda}_{jl} - \lambda_{jl}| = O_p((Mh_{2j})^{-1} + h_{1j}^2 + h_{2j}^2 + (\log n/n)^{1/2}).$

REMARK. Theorem 3 characterizes the uniform weak convergence rates of  $\hat{\Sigma}_\eta(s, t)$ ,  $\hat{\psi}_{jl}$  and  $\hat{\lambda}_{jl}$  for all  $j$ . It can be regarded as an extension of Theorems 3.3–3.6 in Li and Hsing [32], 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 [32] 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 [32], their approach can relax the assumption on the differentiability of the individual curves. In contrast, following Hall, Müller and Wang [22] and Zhang and Chen [51], 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  $\hat{\Sigma}_{\eta, jj}(s, t)$  are positive semi-definite, whereas the approach in Li and Hsing [32] cannot. This is extremely important for high-dimensional neuroimaging data, which usually contains a large number of locations (called voxels) on a two-dimensional (2D) surface or in a 3D volume. For instance, the number of  $M$  can number in the tens of thousands to millions, and thus it can be numerically infeasible to directly operate on  $\hat{\Sigma}_\eta(s, s')$ .

We use  $\tilde{\Sigma}_\eta(s, s')$  to denote the local linear estimator of  $\Sigma_\eta(s, s')$  proposed in Li and Hsing [32]. Following the arguments in Li and Hsing [32], we can easily obtain the following result.

COROLLARY 1. *Under Assumptions (C1)–(C8) and (C9b), it follows that*

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

4.5. *Asymptotic properties of the inference procedures.* In this section, we discuss the asymptotic properties of the global statistic  $S_n$  and the critical values of SCB. Theorem 1 allows us to construct SCB for coefficient functions  $b_{jl}(s)$ . It follows from Theorem 1 that

$$(4.4) \quad \sqrt{n}[\hat{b}_{jl}(s) - b_{jl}(s) - \text{Bias}(\hat{b}_{jl}(s))] \Rightarrow G_{jl}(s),$$

where  $\Rightarrow$  denotes weak convergence of a sequence of stochastic processes, and  $G_{jl}(s)$  is a centered Gaussian process indexed by  $s \in [0, 1]$ . Therefore, let  $X_C(s)$  be a centered Gaussian process, and we have

$$(4.5) \quad \begin{aligned} & [\mathbf{C}(\hat{\Sigma}_\eta(s,s) \otimes \hat{\Omega}_X^{-1})\mathbf{C}^T]^{-1/2} \mathbf{d}(s) \Rightarrow X_C(s), \\ & \sup_{s \in [0,1]} |\sqrt{n}[\hat{b}_{jl}(s) - b_{jl}(s) - \text{Bias}(\hat{b}_{jl}(s))]| \Rightarrow \sup_{s \in [0,1]} |G_{jl}(s)|. \end{aligned}$$

We define  $C_{jl}(\alpha)$  such that  $P(\sup_{s \in [0,1]} |G_{jl}(s)| \leq C_{jl}(\alpha)) = 1 - \alpha$ . Thus, the confidence band given in (3.5) is a  $1 - \alpha$  simultaneous confidence band for  $b_{jl}(s)$ .

THEOREM 4. *If Assumptions (C1)–(C9) are true, then we have*

$$(4.6) \quad S_n \Rightarrow \int_0^1 X_C(s)^T X_C(s) ds.$$

REMARK. Theorem 4 is similar to Theorem 7 of Zhang and Chen [51]. Both characterize the asymptotic distribution of  $S_n$ . In particular, Zhang and Chen [51] delineate the distribution of  $\int_0^1 X_C(s)^T X_C(s) ds$  as a  $\chi^2$ -type mixture. All discussions associated with Theorem 7 of Zhang and Chen [51] are valid here, and therefore, we do not repeat them for the sake of space.

We consider conditional convergence for bootstrapped stochastic processes. We focus on the bootstrapped process  $\{G_j(s)^{(g)} : s \in [0, 1]\}$  as the arguments for establishing the wild bootstrap method for approximating the null distribution of  $S_n$  and the bootstrapped process  $\{G_j(s)^{(g)} : s \in [0, 1]\}$  are similar.

THEOREM 5. *If Assumptions (C1)–(C9) are true, then  $G_j(s)^{(g)}(s)$  converges weakly to  $G_j(s)$  conditioning on the data, where  $G_j(s)$  is a centered Gaussian process indexed by  $s \in [0, 1]$ .*

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

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



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 MVCMM as follows:

$$(5.1) \quad y_{ij}(s_m) = \mathbf{x}_i^T B_j(s_m) + \eta_{ij}(s_m) + \varepsilon_{ij}(s_m) \quad \text{for } j = 1, 2,$$

where  $s_m \sim U[0, 1]$ ,  $(\varepsilon_{i1}(s_m), \varepsilon_{i2}(s_m))^T \sim N((0, 0)^T, S_\varepsilon(s_m) = \text{diag}(\sigma_1^2, \sigma_2^2))$  and  $\mathbf{x}_i = (1, x_{i1}, x_{i2})$  for all  $i = 1, \dots, n$  and  $m = 1, \dots, M$ . Moreover,  $(x_{i1}, x_{i2})^T \sim N((0, 0)^T, \text{diag}(1 - 2^{-0.5}, 1 - 2^{-0.5}) + 2^{-0.5}(1, 1)^{\otimes 2})$  and  $\eta_{ij}(s) = \xi_{ij1}\psi_{j1}(s) + \xi_{ij2}\psi_{j2}(s)$ , where  $\xi_{ijl} \sim N(0, \lambda_{jl})$  for  $j = 1, 2$  and  $l = 1, 2$ . Furthermore,  $s_m, (x_{i1}, x_{i2}), \xi_{i11}, \xi_{i12}, \xi_{i21}, \xi_{i22}, \varepsilon_{i1}(s_m)$ , and  $\varepsilon_{i2}(s_m)$  are independent random variables. We set  $(\lambda_{11}, \lambda_{12}, \sigma_1^2, \lambda_{21}, \lambda_{22}, \sigma_2^2) = (1.2, 0.6, 0.2, 1, 0.5, 0.1)$  and the functional coefficients and eigenfunctions as follows:

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

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  $H_1: b_{13}(s) \neq 0$  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.

Figure 2 depicts the power curves. It can be seen from Figure 2 that the rejection rates for  $S_n$  based on the wild bootstrap method are accurate for moderate sample

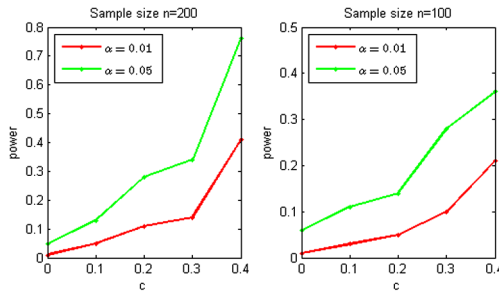


FIG. 2. Plot of power curves. Rejection rates of  $S_n$  based on the wild bootstrap method are calculated at five different values of  $c$  (0, 0.1, 0.2, 0.3, and 0.4) for two sample sizes of  $n$  (100 and 200) subjects at 5% (green) and 1% (red) significance levels.

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

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 wild bootstrap method. The data were generated from model (5.1) under the same parameter values. 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 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 Carlo errors are of size  $\sqrt{0.95 \times 0.05/200} \approx 0.015$  for  $\alpha = 0.05$ . Figure 3 depicts typical simultaneous confidence bands, where  $n = 500$  and  $M = 50$ . Additional simulation results are given in the supplemental article [53].

**6. Real data analysis.** The data set consists of 128 healthy infants (75 males and 53 females) from the neonatal project on early brain development. The gestational ages of these infants range from 262 to 433 days, and their mean gestational age is 298 days with standard deviation 17.6 days. The DTIs and T1-weighted images were acquired for each subject. For the DTIs, the imaging parameters were as follows: the six noncollinear directions at the  $b$ -value of  $1000 \text{ s/mm}^2$  with a reference scan ( $b = 0$ ), the isotropic voxel resolution = 2 mm, and the in-plane field of view = 256 mm in both directions. A total of five repetitions were acquired to improve the signal-to-noise ratio of the DTIs.

The DTI data were processed by two key steps including a weighted least squares estimation method [2, 54] to construct the diffusion tensors and a DTI atlas building pipeline [20, 56] to register DTIs from multiple subjects to create

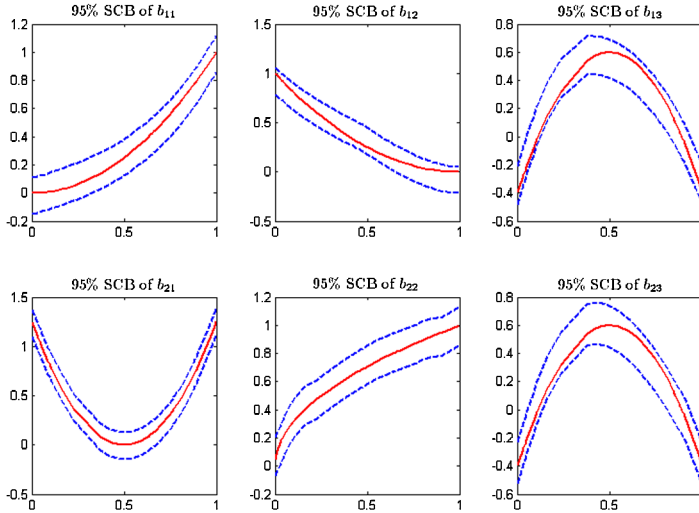


FIG. 3. Typical simultaneous confidence bands with  $n = 500$  and  $M = 50$ . The red solid curves are the true coefficient functions, and the blue dashed curves are the confidence bands.

a study specific unbiased DTI atlas, to track fiber tracts in the atlas space and to propagate them back into each subject's native space by using registration information. Subsequently, diffusion tensors (DTs) and their scalar diffusion properties were calculated at each location along each individual fiber tract by using DTs in neighboring voxels close to the fiber tract. Figure 1(a) displays the fiber bundle of the genu of the corpus callosum (GCC), which is an area of white matter in the brain. The GCC is the anterior end of the corpus callosum, and is bent downward and backward in front of the septum pellucidum; diminishing rapidly in thickness, it is prolonged backward under the name of the rostrum, which is connected below with the lamina terminalis. It was found that neonatal microstructural development of GCC positively correlates with age and callosal thickness.

The two aims of this analysis are to compare diffusion properties including FA and MD along the GCC between the male and female groups and to delineate the development of fiber diffusion properties across time, which is addressed by including the gestational age at MRI scanning as a covariate. FA and MD, respectively, measure the inhomogeneous extent of local barriers to water diffusion and the averaged magnitude of local water diffusion. We fit model (1.1) to the FA and MD values from all 128 subjects, in which  $\mathbf{x}_i = (1, G, \text{Age})^T$ , where  $G$  represents gender. We then applied the estimation and inference procedures to estimate  $\mathbf{B}(s)$  and calculate  $S_n$  for each hypothesis test. We approximated the  $p$ -value of  $S_n$  using the wild bootstrap method with  $G = 1000$  replications. Finally, we constructed the 95% simultaneous confidence bands for the functional coefficients of  $B_j(s)$  for  $j = 1, 2$ .

Figure 4 presents the estimated coefficient functions corresponding to 1, G and Age associated with FA and MD (blue solid lines in all panels of Figure 4). The intercept functions [panels (a) and (d) in Figure 4] describe the overall trend of FA and MD. The gender coefficients for FA and MD in Figure 4(b) and (e) are negative at most of the grid points, which may indicate that compared with female infants, male infants have relatively smaller magnitudes of local water diffusivity along the genu of the corpus callosum. The gestational age coefficients for FA [panel (c) of Figure 4] are positive at most grid points, indicating that FA measures increase with age in both male and female infants, whereas those corresponding to MD [panel (f) of Figure 4] are negative at most grid points. This may indicate a negative correlation between the magnitudes of local water diffusivity and gestational age along the genu of the corpus callosum.

We statistically tested the effects of gender and gestational age on FA and MD along the GCC tract. To test the gender effect, we computed the global test statistic  $S_n = 144.63$  and its associated  $p$ -value ( $p = 0.078$ ), indicating a weakly significant gender effect, which agrees with the findings in panels (b) and (e) of Figure 4. A moderately significant age effect was found with  $S_n = 929.69$  ( $p$ -value  $< 0.001$ ). This agrees with the findings in panel (f) of Figure 4, indicating that MD along the GCC tract changes moderately with gestational age. Furthermore, for FA and MD, we constructed the 95% simultaneous confidence bands of the varying-coefficients for  $G_i$  and  $age_i$  (Figure 4).

Figure 5 presents the first 10 eigenvalues and 3 eigenfunctions of  $\hat{\Sigma}_{\eta,jj}(s, t)$  for  $j = 1, 2$ . The relative eigenvalues of  $\hat{\Sigma}_{\eta,jj}$  defined as the ratios of the eigenvalues of  $\hat{\Sigma}_{\eta,jj}(s, t)$  over their sum have similar distributional patterns [panel (a) of Figure 5]. We observe that the first three eigenvalues account for more than 90% of the total and the others quickly vanish to zero. The eigenfunctions of FA corresponding to the largest three eigenvalues [Figure 5(b)] are different from those of MD [Figure 5(c)].

In the supplement article [53], we further illustrate the proposed methodology by an empirical analysis of another real data set.

APPENDIX

We introduce some notation. We define

$$\begin{aligned}
 T_{B,j}(h, s) &= \sum_{i=1}^n \sum_{m=1}^M K_h(s_m - s) [\mathbf{x}_i \otimes \mathbf{z}_h(s_m - s)] \mathbf{x}_i^T B_j(s_m), \\
 T_{\eta,j}(h, s) &= \sum_{i=1}^n \sum_{m=1}^M K_h(s_m - s) [\mathbf{x}_i \otimes \mathbf{z}_h(s_m - s)] \eta_{ij}(s_m), \\
 (A.1) \quad T_{\varepsilon,j}(h, s) &= \sum_{i=1}^n \sum_{m=1}^M K_h(s_m - s) [\mathbf{x}_i \otimes \mathbf{z}_h(s_m - s)] \varepsilon_{ij}(s_m),
 \end{aligned}$$

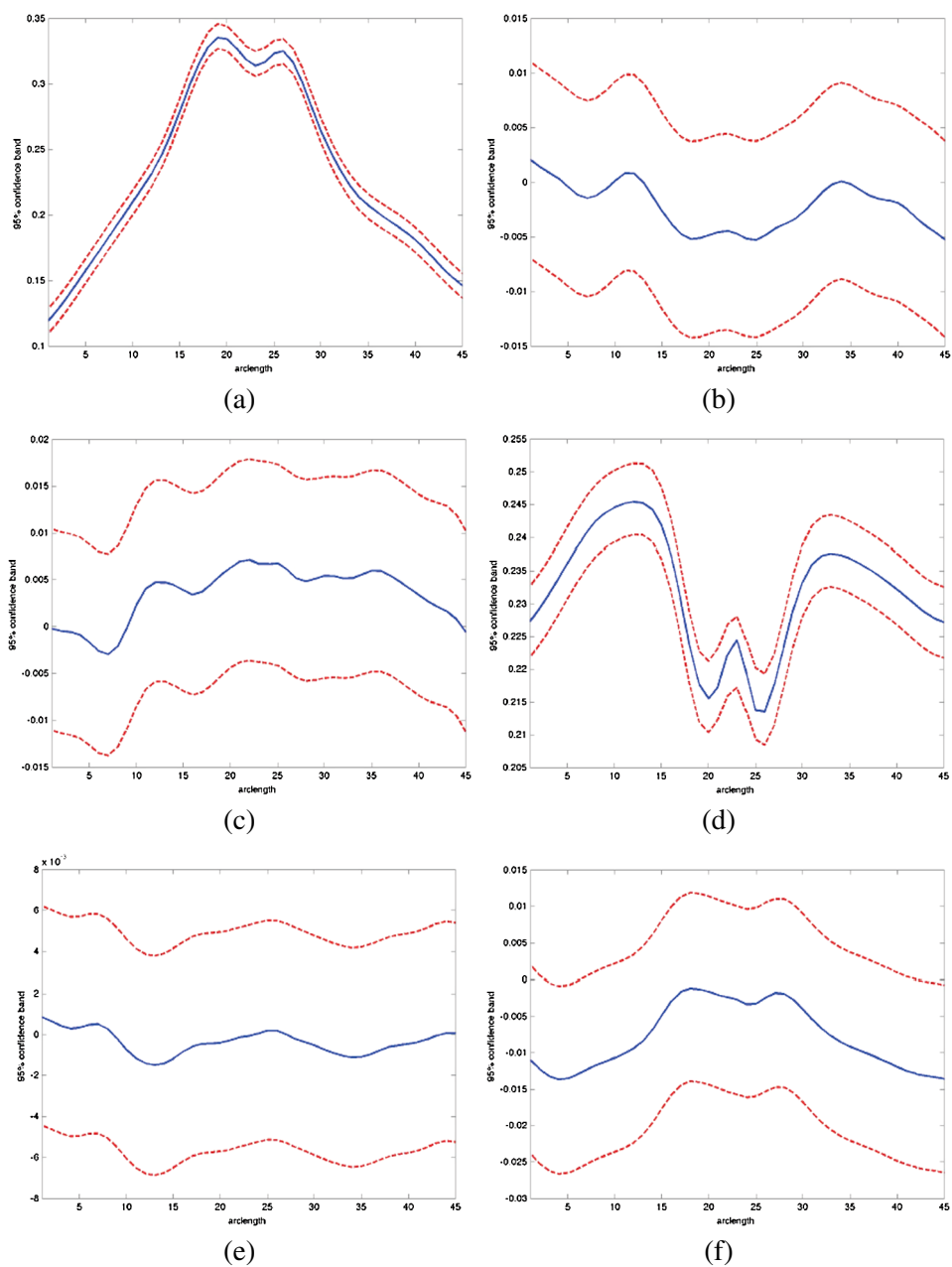


FIG. 4. Plot of estimated effects of intercept [(a), (d)], gender [(b), (e)], and age [(c), (f)] and their 95% confidence bands. The first three panels [(a), (b), (c)] are for FA and the last three panels [(d), (e) and (f)] are for MD. The blue solid curves are the estimated coefficient functions, and the red dashed curves are the confidence bands.

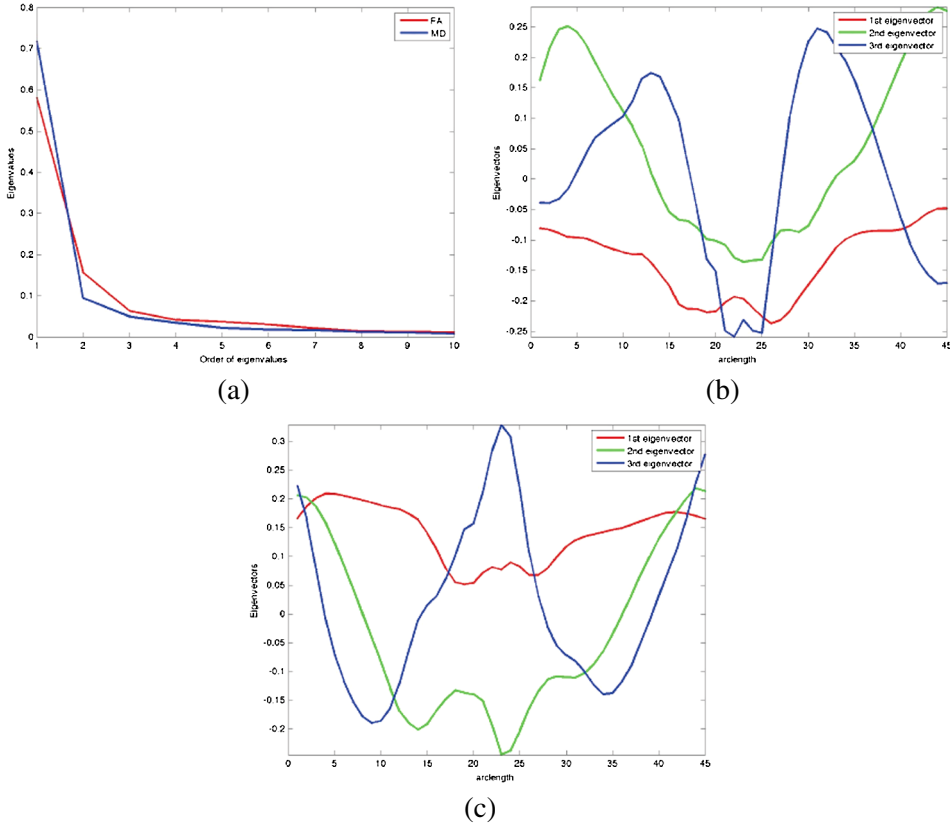


FIG. 5. Plot of the first 10 eigenvalues (a) and the first 3 eigenfunctions for FA (b) and MD (c).

$$r_u(K; s, h) = \frac{u_2(K; s, h)^2 - u_1(K; s, h)u_3(K; s, h)}{u_0(K; s, h)u_2(K; s, h) - u_1(K; s, h)^2},$$

$$H_h(s_m - s) = K_h(s_m - s)z_h(s_m - s),$$

$$\Delta_j(s; \eta_i, h_{1j}) = M^{-1} \sum_{m=1}^M H_{h_{1j}}(s_m - s)\eta_{ij}(s_m)$$

$$- \int_0^1 H_{h_{1j}}(u - s)\eta_{ij}(u)\pi(u) du,$$

where  $u_r(K; s, h) = \int_0^1 h^{-r} (u - s)^r K_h(u - s) du$  for  $r \geq 0$ . Throughout the proofs,  $C_k$ 's stand for a generic constant, and it may vary from line to line.

The proofs of Theorems 1–5 rely on the following lemmas whose proofs are given in the supplemental article [53].

LEMMA 1. Under Assumptions (C1), (C3)–(C5) and (C7), we have that for each  $j$ ,

$$(A.2) \quad \sup_{s \in [0,1]} n^{-1/2} h_{1j} |T_{\varepsilon,j}(h_{1j}, s)| = O_p(\sqrt{M h_{1j} |\log h_{1j}|}) = o_p(M h_{1j}).$$

LEMMA 2. Under Assumptions (C1), (C4), (C5) and (C7), we have that for any  $r \geq 0$  and  $j$ ,

$$\begin{aligned} \sup_{s \in [0,1]} \left| \int K_{h_{1j}}(u-s) \frac{(u-s)^r}{h_{1j}^r} d[\Pi_M(u) - \Pi(u)] \right| &= O_p((M h_{1j})^{-1/2}), \\ \sup_{s \in [0,1]} \left| \int K_{h_{1j}}(u-s) \frac{(u-s)^r}{h_{1j}^r} \varepsilon_{ij}(u) d\Pi_M(u) \right| &= O_p((M h_{1j})^{-1/2} \sqrt{|\log h_{1j}|}), \end{aligned}$$

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

LEMMA 3. Under Assumptions (C2)–(C5), we have

$$(A.3) \quad \sup_{s \in [0,1]} \left| n^{-1/2} \sum_{i=1}^n \mathbf{x}_i \otimes \Delta_j(s; \boldsymbol{\eta}_i, h_{1j}) \right| = o_p(1).$$

LEMMA 4. If Assumptions (C1) and (C3)–(C6) hold, then we have

$$(A.4) \quad \begin{aligned} E[\hat{B}_j(s)|\mathcal{S}] - B_j(s) &= 0.5 h_{1j}^2 u_2(K) \ddot{B}_j(s) [1 + o_p(1)], \\ \text{Var}[\hat{B}_j(s)|\mathcal{S}] &= n^{-1} \Sigma_{\eta,jj}(s, s) \Omega_X^{-1} [1 + o_p(1)], \end{aligned}$$

where  $e_n(s) = O_p((M h_{1j})^{-1/2})$  with  $E[e_n(s)] = 0$ .

LEMMA 5. If Assumptions (C1) and (C3)–(C6) hold, then for  $s = 0$  or  $1$ , we have

$$(A.5) \quad \begin{aligned} E[\hat{B}_j(s)|\mathcal{S}] - B_j(s) &= 0.5 h_{1j}^2 r_u(K; s, h_{1j}) \ddot{B}_j(s) [1 + o_p(1)], \\ \text{Var}[\hat{B}_j(s)|\mathcal{S}] &= n^{-1} \Sigma_{\eta,jj}(s, s) \Omega_X^{-1} [1 + o_p(1)]. \end{aligned}$$

LEMMA 6. Under Assumptions (C1)–(C9), we have

$$\begin{aligned} \sup_{(s,t)} n^{-1} \left| \sum_{i=1}^n \bar{\varepsilon}_{ij}(s) \eta_{ij}(t) \right| &= O_p(n^{-1/2} (\log n)^{1/2}), \\ \sup_{(s,t)} n^{-1} \left| \sum_{i=1}^n \bar{\varepsilon}_{ij}(s) \Delta \eta_{ij}(t) \right| &= O_p(n^{-1/2} (\log n)^{1/2}), \\ \sup_s n^{-1} \left| \sum_{i=1}^n \bar{\varepsilon}_{ij}(s) \mathbf{x}_i \right| &= O_p(n^{-1/2} (\log n)^{1/2}), \end{aligned}$$

$$\sup_s n^{-1} \left| \sum_{i=1}^n \Delta \eta_{ij}(s) \mathbf{x}_i \right| = O_p(n^{-1/2}(\log n)^{1/2}).$$

LEMMA 7. Under Assumptions (C1)–(C9), we have

$$\sup_{(s,t)} n^{-1} \left| \sum_{i=1}^n \bar{\varepsilon}_{ij}(s) \bar{\varepsilon}_{ij}(t) \right| = O((Mh_{2j})^{-1} + (\log n/n)^{1/2}) = o_p(1).$$

We present only the key steps in the proof of Theorem 1 below.

PROOF OF THEOREM 1. Define

$$\begin{aligned} \mathbf{U}_2(K; s, \mathbf{H}) &= \text{diag}(r_u(K; s, h_{11}), \dots, r_u(K; s, h_{1J})), \\ X_n(s) &= \sqrt{n} \{ \hat{\mathbf{B}}(s) - E[\hat{\mathbf{B}}(s)|\mathcal{S}] \}, \\ X_{n,j}(s) &= \sqrt{n} \{ \hat{B}_j(s) - E[\hat{B}_j(s)|\mathcal{S}] \}. \end{aligned}$$

According to the definition of  $\text{vec}(\hat{A}_j(s))$ , it is easy to see that

$$(A.6) \quad \text{vec}(\hat{A}_j(s)) = \Sigma(s, h_{1j})^{-1} [T_{B,j}(h_{1j}, s) + T_{\varepsilon,j}(h_{1j}, s) + T_{\eta,j}(h_{1j}, s)],$$

$$(A.7) \quad X_{n,j}(s) = \sqrt{n} [\mathbf{I}_p \otimes (1, 0)] \Sigma(s, h_{1j})^{-1} [T_{\varepsilon,j}(h_{1j}, s) + T_{\eta,j}(h_{1j}, s)].$$

The proof of Theorem 1(i) consists of two parts:

- Part 1 shows that  $\sqrt{n} \Sigma(s, h_{1j})^{-1} T_{\varepsilon,j}(h_{1j}, s) = o_p(1)$  holds uniformly for all  $s \in [0, 1]$  and  $j = 1, \dots, J$ .
- Part 2 shows that  $\sqrt{n} \Sigma(s, h_{1j})^{-1} T_{\eta,j}(h_{1j}, s)$  converges weakly to a Gaussian process  $G(\cdot)$  with mean zero and covariance matrix  $\Sigma_{\eta,jj}(s, s') \Omega_X^{-1}$  for each  $j$ .

In part 1, we show that

$$(A.8) \quad \sqrt{n} [\mathbf{I}_p \otimes (1, 0)] \Sigma(s, h_{1j})^{-1} T_{\varepsilon,j}(h_{1j}, s) = o_p(1).$$

It follows from Lemma 1 that

$$n^{-1/2} \sum_{i=1}^n \mathbf{x}_i \otimes \left\{ M^{-1} \sum_{m=1}^M K_{h_{1j}}(s_m - s) \mathbf{z}_{h_{1j}}(s) \varepsilon_{i,j}(s_m) \right\} = o_p(1)$$

hold uniformly for all  $s \in [0, 1]$ . It follows from Lemma 2 that

$$(A.9) \quad (nM)^{-1} \Sigma(s, h_{1j}) = \Omega_X \otimes \Omega_1(h_{1j}, s) + o_p(1)$$

hold uniformly for all  $s \in [0, 1]$ . Based on these results, we can finish the proof of (A.8).

In part 2, we show the weak convergence of  $\sqrt{n} [\mathbf{I}_p \otimes (1, 0)] \Sigma(s, h_{1j})^{-1} \times T_{\eta,j}(h_{1j}, s)$  for  $j = 1, \dots, J$ . Part 2 consists of two steps. In Step 1, it follows from the standard central limit theorem that for each  $s \in [0, 1]$ ,

$$(A.10) \quad \sqrt{n} [\mathbf{I}_p \otimes (1, 0)] \Sigma(s, h_{1j})^{-1} T_{\eta,j}(h_{1j}, s) \rightarrow^L N(\mathbf{0}, \Sigma_{\eta,jj}(s, s) \Omega_X^{-1}),$$



where  $\rightarrow^L$  denotes convergence in distribution.

Step 2 shows the asymptotic tightness of  $\sqrt{n}[\mathbf{I}_p \otimes (1, 0)]\Sigma(s, h_{1j})^{-1} \times T_{\eta,j}(h_{1j}, s)$ . By using (A.9) and (A.1),  $\sqrt{n}\Sigma(s, h_{1j})^{-1}T_{\eta,j}(h_{1j}, s)[1 + o_p(1)]$  can be approximated by the sum of three terms (I), (II) and (III) as follows:

$$\begin{aligned}
 \text{(I)} &= n^{-1/2} \sum_{i=1}^n \Omega_X^{-1} \mathbf{x}_i \otimes \Omega_1(h_{1j}, s)^{-1} \Delta_j(s; \boldsymbol{\eta}_i, h_{1j}), \\
 \text{(II)} &= n^{-1/2} \sum_{i=1}^n \Omega_X^{-1} \mathbf{x}_i \otimes \Omega_1(h_{1j}, s)^{-1} \eta_{ij}(s) \\
 &\quad \times \int_{\max(-sh_{1j}^{-1}, -1)}^{\min((1-s)h_{1j}^{-1}, 1)} K(u)(1, u)^T \pi(s + h_{1j}u) du, \\
 \text{(A.11)} \quad \text{(III)} &= n^{-1/2} \sum_{i=1}^n \Omega_X^{-1} \mathbf{x}_i \otimes \Omega_1(h_{1j}, s)^{-1} \\
 &\quad \times \int_{\max(-sh_{1j}^{-1}, -1)}^{\min((1-s)h_{1j}^{-1}, 1)} K(u) \begin{pmatrix} 1 \\ u \end{pmatrix} [\eta_{ij}(s + h_{1j}u) - \eta_{ij}(s)] \\
 &\quad \times \pi(s + h_{1j}u) du.
 \end{aligned}$$

We investigate the three terms on the right-hand side of (A.11) as follows. It follows from Lemma 3 that the first term on the right-hand side of (A.11) converges to zero uniformly. We prove the asymptotic tightness of (II) as follows. Define

$$\begin{aligned}
 \hat{X}_{n,j}(s) &= n^{-1/2} \sum_{i=1}^n \Omega_X^{-1} \mathbf{x}_i \otimes (1, 0)\Omega_1(h_{1j}, s)^{-1} \eta_{ij}(s) \\
 &\quad \times \int_{\max(-sh_{1j}^{-1}, -1)}^{\min((1-s)h_{1j}^{-1}, 1)} K(u)(1, u)^T \pi(s + h_{1j}u) du.
 \end{aligned}$$

Thus, we only need to prove the asymptotic tightness of  $\hat{X}_{n,j}(s)$ . The asymptotic tightness of  $\hat{X}_{n,j}(s)$  can be proved using the empirical process techniques [42]. It follows that

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

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

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

We consider a function class  $\mathcal{E}_\eta = \{f(s; \mathbf{x}, \eta_{\cdot,j}) = \Omega_X^{-1} \mathbf{x} \eta_{\cdot,j}(s) : s \in [0, 1]\}$ . Due to Assumption (C2),  $\mathcal{E}_\eta$  is a  $P$ -Donsker class.

Finally, we consider the third term (III) on the right-hand side of (A.11). It is easy to see that (III) can be written as

$$\begin{aligned} &\Omega_X^{-1} \otimes \Omega_1(h_{1j}, s)^{-1} \\ &\times \int_{\max(-sh_{1j}^{-1}, -1)}^{\min((1-s)h_{1j}^{-1}, 1)} K(u) \left[ n^{-1/2} \sum_{i=1}^n \mathbf{x}_i \{ \eta_{ij}(s + h_{1j}u) - \eta_{ij}(s) \} \right] \otimes \begin{pmatrix} 1 \\ u \end{pmatrix} \\ &\times \pi(s + h_{1j}u) du. \end{aligned}$$

Using the same argument of proving the second term (II), we can show the asymptotic tightness of  $n^{-1/2} \sum_{i=1}^n \mathbf{x}_i \eta_{ij}(s)$ . Therefore, for any  $h_{1j} \rightarrow 0$ ,

$$(A.12) \quad \sup_{s \in [0, 1], |u| \leq 1} \left| n^{-1/2} \sum_{i=1}^n \mathbf{x}_i \{ \eta_{ij}(s + h_{1j}u) - \eta_{ij}(s) \} \right| = o_p(1).$$

It follows from Assumptions (C5) and (C7) and (A.12) that (III) converges to zero uniformly. Therefore, we can finish the proof of Theorem 1(i). Since Theorem 1(ii) is a direct consequence of Theorem 1(i) and Lemma 4, we finish the proof of Theorem 1.  $\square$

**PROOF OF THEOREM 2.** Proofs of parts (a)–(d) are completed by some straightforward calculations. Detailed derivation is given in the supplemental document. Here we prove part (e) only. 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 [11]. Thus, we have

$$(A.13) \quad \begin{aligned} \hat{\eta}_{ij}(s) - \eta_{ij}(s) &= \sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_m - s) \mathbf{x}_i^T [B_j(s_m) - \hat{B}_j(s_m)] \\ &+ \sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_m - s) [\eta_{ij}(s_m) + \varepsilon_{ij}(s_m) - \eta_{ij}(s)]. \end{aligned}$$

We define

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

It follows from (A.13) that

$$(A.14) \quad \hat{\eta}_{ij}(s) - \eta_{ij}(s) = \Delta_{ij}(s) = \bar{\varepsilon}_{ij}(s) + \Delta\eta_{ij}(s) + \mathbf{x}_i^T \Delta B_j(s).$$

It follows from Lemma 2 and a Taylor expansion that

$$\sup_{s \in [0,1]} |\bar{\varepsilon}_{ij}(s)| = O_p\left(\sqrt{\frac{|\log(h_{2j})|}{Mh_{2j}}}\right)$$

and

$$\sup_{s \in [0,1]} |\Delta\eta_{ij}(s)| = O_p(1) \sup_{s \in [0,1]} |\ddot{\eta}_{ij}(s)| h_{1j}^{(2)2}.$$

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

$$\begin{aligned} \Delta B_{ij}(s) = & - \sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_j - s) 0.5u_2(K)^2 h_{1j}^2 \ddot{B}_j(s_m)[1 + o_p(1)] \\ & + \sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_j - s) \{0.5u_2(K)^2 h_{1j}^2 \ddot{B}_j(s_m)[1 + o_p(1)] \\ & \qquad \qquad \qquad + B_j(s_m) - \hat{B}_j(s_m)\}, \end{aligned}$$

$$\sup_{s \in [0,1]} \|\Delta B_j(s)\| = O_p(n^{-1/2}) + O_p(h_{1j}^2).$$

Combining these results, we have

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

This completes the proof of part (e).  $\square$

PROOF OF THEOREM 3. Recall that  $\hat{\eta}_{ij}(s) = \eta_{ij}(s) + \Delta_{i,j}(s)$ , we have

$$(A.15) \quad \begin{aligned} n^{-1} \sum_{i=1}^n \hat{\eta}_{ij}(s) \hat{\eta}_{ij}(t) = & n^{-1} \sum_{i=1}^n \Delta_{ij}(s) \Delta_{ij}(t) + n^{-1} \sum_{i=1}^n \eta_{ij}(s) \Delta_{ij}(t) \\ & + n^{-1} \sum_{i=1}^n \Delta_{ij}(s) \eta_{ij}(t) + n^{-1} \sum_{i=1}^n \eta_{ij}(s) \eta_{ij}(t). \end{aligned}$$

This proof consists of two steps. The first step is to show that the first three terms on the right-hand side of (A.15) converge to zero uniformly for all  $(s, t) \in [0, 1]^2$  in probability. The second step is to show the uniform convergence of  $n^{-1} \sum_{i=1}^n \eta_{ij}(s) \eta_{ij}(t)$  to  $\Sigma_\eta(s, t)$  over  $(s, t) \in [0, 1]^2$  in probability.

We first show that

$$(A.16) \quad \sup_{(s,t)} n^{-1} \left| \sum_{i=1}^n \Delta_{ij}(s) \eta_{ij}(t) \right| = O_p(n^{-1/2} + h_{1j}^2 + h_{2j}^2 + (\log n/n)^{1/2}).$$

Since

$$(A.17) \quad \begin{aligned} & \sum_{i=1}^n \Delta_{ij}(s) \eta_{ij}(t) \\ & \leq n^{-1} \left\{ \left| \sum_{i=1}^n \bar{\varepsilon}_{ij}(s) \eta_{ij}(t) \right| + \left| \sum_{i=1}^n \Delta \eta_{ij}(s) \eta_{ij}(t) \right| \right. \\ & \quad \left. + \left| \sum_{i=1}^n \mathbf{x}_i^T \Delta B_j(s) \eta_{ij}(t) \right| \right\}, \end{aligned}$$

it is sufficient to focus on the three terms on the right-hand side of (A.17). Since

$$|\mathbf{x}_i^T \Delta B_j(s) \eta_{ij}(t)| \leq \|\mathbf{x}_i\|_2 \sup_{s \in [0,1]} \|\Delta B_k(s)\|_2 \sup_{t \in [0,1]} |\eta_{ij}(t)|,$$

we have

$$\begin{aligned} n^{-1} \left| \sum_{i=1}^n \mathbf{x}_i^T \Delta B_j(s) \eta_{ij}(t) \right| & \leq \sup_{s \in [0,1]} \|\Delta B_k(s)\|_2 n^{-1} \sum_{i=1}^n \|\mathbf{x}_i\|_2 |\eta_{ij}(t)| \\ & = O_p(n^{-1/2} + h_{1j}^2). \end{aligned}$$

Similarly, we have

$$n^{-1} \left| \sum_{i=1}^n \Delta \eta_{ij}(s) \eta_{ij}(t) \right| \leq n^{-1} \sum_{i=1}^n \sup_{s,t \in [0,1]} |\Delta \eta_{ij}(s) \eta_{ij}(t)| = O_p(h_{1j}^{(2)2}) = o_p(1).$$

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

We can show that

$$(A.18) \quad \sup_{(s,t)} n^{-1} \sum_{i=1}^n [\eta_{ij}(s) \eta_{ij}(t) - \Sigma_{\eta,jj}(s,t)] = O_p(n^{-1/2}).$$

Note that

$$\begin{aligned} & |\eta_{ij}(s_1) \eta_{ij}(t_1) - \eta_{ij}(s_2) \eta_{ij}(t_2)| \\ & \leq 2(|s_1 - s_2| + |t_1 - t_2|) \sup_{s \in [0,1]} |\dot{\eta}_{ij}(s)| \sup_{s \in [0,1]} |\eta_{ij}(s)| \end{aligned}$$

holds for any  $(s_1, t_1)$  and  $(s_2, t_2)$ , the functional class  $\{\eta_j(u) \eta_j(v) : (u, v) \in [0, 1]^2\}$  is a Vapnik and Cervonenkis (VC) class [31, 42]. Thus, it yields that (A.18) is true.

Finally, we can show that

$$(A.19) \quad \sup_{(s,t)} n^{-1} \left| \sum_{i=1}^n \Delta_{ij}(s) \Delta_{ij}(t) \right| = O_p((Mh_{2j})^{-1} + (\log n/n)^{1/2} + h_j^4 + h_{1j}^{(2)4}).$$

With some calculations, for a positive constant  $C_1$ , we have

$$\begin{aligned} & \left| \sum_{i=1}^n \Delta_{ij}(s) \Delta_{ij}(t) \right| \\ & \leq C_1 \sup_{(s,t)} \left[ \left| \sum_{i=1}^n \bar{\varepsilon}_{ij}(s) \bar{\varepsilon}_{ij}(t) \right| + \left| \sum_{i=1}^n \bar{\varepsilon}_{ij}(s) \Delta \eta_{ij}(t) \right| \right. \\ & \quad + \left| \sum_{i=1}^n \Delta \eta_{ij}(t) \mathbf{x}_i^T \Delta B_j(s) \right| + \left| \sum_{i=1}^n \bar{\varepsilon}_{ij}(s) \mathbf{x}_i^T \Delta B_j(t) \right| \\ & \quad \left. + \left| \sum_{i=1}^n \Delta \eta_{ij}(s) \Delta \eta_{ij}(t) \right| + \left| \sum_{i=1}^n \mathbf{x}_i^T \Delta B_j(s) \Delta B_j(t) \mathbf{x}_i \right| \right]. \end{aligned}$$

It follows from Lemma 7 that

$$\begin{aligned} & \sup_{(s,t)} n^{-1} \left| \sum_{i=1}^n \bar{\varepsilon}_{ij}(s) \bar{\varepsilon}_{ij}(t) \right| = O_p((Mh_{2j})^{-1} + (\log n/n)^{1/2}), \\ & \sup_{(s,t)} n^{-1} \left[ \left| \sum_{i=1}^n \bar{\varepsilon}_{ij}(s) \Delta \eta_{ij}(t) \right| + \left| \sum_{i=1}^n \Delta \eta_{ij}(t) \mathbf{x}_i^T \Delta B_j(s) \right| + \left| \sum_{i=1}^n \bar{\varepsilon}_{ij}(s) \mathbf{x}_i^T \Delta B_j(t) \right| \right] \\ & = O_p((\log n/n)^{1/2}). \end{aligned}$$

Since  $\sup_{s \in [0,1]} |\Delta \eta_{ij}(s)| = C_2 \sup_{s \in [0,1]} |\dot{\eta}_{ij}(s)| h_{2j}^2$ , we have

$$\sup_{(s,t)} n^{-1} \left| \sum_{i=1}^n \Delta \eta_{ij}(s) \Delta \eta_{ij}(t) \right| = O(h_{1j}^{(2)4}).$$

Furthermore, since  $\sup_{s \in [0,1]} \|\Delta \mathbf{B}(s)\| = O_p(n^{-1/2} + h_j^2)$ , we have

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

Note that the arguments for (A.16)–(A.19) hold for  $\hat{\Sigma}_{\eta, jj'}(\cdot, \cdot)$  for any  $j \neq j'$ . Thus, combining (A.16)–(A.19) leads to Theorem 3(i).

To prove Theorem 3(ii), we follow the same arguments in Lemma 6 of Li and Hsing [32]. For completion, we highlight several key steps below. We define

$$(A.20) \quad (\Delta\psi_{j,j})(s) = \int_0^1 [\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)]\psi_{j,j}(t) dt.$$

Following Hall and Hosseini-Nasab [21] and the Cauchy–Schwarz inequality, we have

$$\begin{aligned} & \left\{ \int_0^1 [\hat{\psi}_{j,j}(s) - \psi_{j,j}(s)]^2 ds \right\}^{1/2} \\ & \leq C_2 \left\{ \left[ \int_0^1 (\Delta\psi_{j,j})(s)^2 ds \right]^{1/2} + \int_0^1 \int_0^1 [\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)]^2 ds dt \right\} \\ & \leq C_2 \left\{ \int_0^1 \int_0^1 [\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)]^2 ds dt \right\}^{1/2} \left\{ \int_0^1 [\psi_{j,j}(t)]^2 dt \right\}^{1/2} \\ & \quad + \int_0^1 \int_0^1 [\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)]^2 ds dt \\ & \leq C_3 \sup_{(s,t) \in [0,1]^2} |\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)|, \end{aligned}$$

which yields Theorem 3(ii)(a).

Using (4.9) in Hall, Müller and Wang [22], we have

$$\begin{aligned} & |\hat{\lambda}_{j,j} - \lambda_{j,j}| \\ & \leq \left| \int_0^1 \int_0^1 [\hat{\Sigma}_{\eta,jj} - \Sigma_{\eta,jj}](s,t)\psi_{j,j}(s)\psi_{j,j}(t) ds dt \right. \\ & \quad \left. + O\left( \int_0^1 (\Delta\psi_{j,j})(s)^2 ds \right) \right| \\ & \leq C_4 \sup_{(s,t) \in [0,1]^2} |\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)|, \end{aligned}$$

which yields Theorem 3(ii)(b). This completes the proof.  $\square$

**PROOF OF THEOREM 5.** The proof of Theorem 5 is given in the supplement article [53].  $\square$

**Acknowledgments.** The authors are grateful to the Editor Peter Bühlmann, the Associate Editor, and three anonymous referees for valuable suggestions, which have greatly helped to improve our presentation.

SUPPLEMENTARY MATERIAL

**Supplement to “Multivariate varying coefficient model for functional responses”** (DOI: [10.1214/12-AOS1045SUPP](https://doi.org/10.1214/12-AOS1045SUPP); .pdf). This supplemental material includes the proofs of all theorems and lemmas.

## 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. (1994). 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. (1994). MR diffusion tensor spectroscopy and imaging. *Biophys. J.* **66** 259–267.
- [4] BUZSÁKI, G. (2006). *Rhythms of the Brain*. Oxford Univ. Press, Oxford. [MR2270828](#)
- [5] CARDOT, H. (2007). Conditional functional principal components analysis. *Scand. J. Stat.* **34** 317–335. [MR2346642](#)
- [6] CARDOT, H., CHAOUCH, M., GOGA, C. and LABRUÈRE, C. (2010). Properties of design-based functional principal components analysis. *J. Statist. Plann. Inference* **140** 75–91. [MR2568123](#)
- [7] CARDOT, H. and JOSSEAND, E. (2011). Horvitz–Thompson estimators for functional data: Asymptotic confidence bands and optimal allocation for stratified sampling. *Biometrika* **98** 107–118. [MR2804213](#)
- [8] CHIOU, J.-M., MÜLLER, H.-G. and WANG, J.-L. (2004). Functional response models. *Statist. Sinica* **14** 675–693. [MR2087968](#)
- [9] DEGRAS, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* **21** 1735–1765. [MR2895997](#)
- [10] EINMAHL, U. and MASON, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.* **13** 1–37. [MR1744994](#)
- [11] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. Chapman & Hall, London. [MR1383587](#)
- [12] FAN, J., YAO, Q. and CAI, Z. (2003). Adaptive varying-coefficient linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 57–80. [MR1959093](#)
- [13] FAN, J. and ZHANG, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518. [MR1742497](#)
- [14] FAN, J. and ZHANG, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Stat.* **27** 715–731. [MR1804172](#)
- [15] FAN, J. and ZHANG, W. (2008). Statistical methods with varying coefficient models. *Stat. Interface* **1** 179–195. [MR2425354](#)
- [16] FARAWAY, J. J. (1997). Regression analysis for a functional response. *Technometrics* **39** 254–261. [MR1462586](#)
- [17] FASS, L. (2008). Imaging and cancer: A review. *Mol. Oncol.* **2** 115–152.
- [18] FRISTON, K. J. (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, London.
- [19] FRISTON, K. J. (2009). Modalities, modes, and models in functional neuroimaging. *Science* **326** 399–403.
- [20] GOODLETT, C. B., FLETCHER, P. T., GILMORE, J. H. and GERIG, G. (2009). Group analysis of DTI fiber tract statistics with application to neurodevelopment. *NeuroImage* **45** S133–S142.
- [21] HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 109–126. [MR2212577](#)
- [22] HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. [MR2278365](#)
- [23] 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** 703–723. [MR2523900](#)

- [24] HASTIE, T. and TIBSHIRANI, R. (1993). Varying-coefficient models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **55** 757–796. [MR1229881](#)
- [25] HEYWOOD, I., CORNELIUS, S. and CARVER, S. (2006). *An Introduction to Geographical Information Systems.*, 3rd ed. Prentice Hall, New York.
- [26] HOOVER, D. R., RICE, J. A., WU, C. O. and YANG, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822. [MR1666699](#)
- [27] 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** 111–128. [MR1888349](#)
- [28] 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** 763–788. [MR2087972](#)
- [29] HUETTEL, S. A., SONG, A. W. and MCCARTHY, G. (2004). *Functional Magnetic Resonance Imaging*. Sinauer, London.
- [30] KOSOROK, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Multivariate Anal.* **84** 299–318. [MR1965224](#)
- [31] KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York. [MR2724368](#)
- [32] LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38** 3321–3351. [MR2766854](#)
- [33] 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.
- [34] LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. [MR2530545](#)
- [35] MA, S., YANG, L. and CARROLL, R. J. (2012). A simultaneous confidence band for sparse longitudinal regression. *Statist. Sinica* **22** 95–122. [MR2933169](#)
- [36] MERCER, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **209** 415–446.
- [37] NIEDERMEYER, E. and DA SILVA, F. L. (2004). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Williams & Wilkins, Baltimore.
- [38] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. [MR2168993](#)
- [39] RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **53** 233–243. [MR1094283](#)
- [40] SUN, J. and LOADER, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.* **22** 1328–1345. [MR1311978](#)
- [41] TOWLE, V. L., 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.
- [42] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. [MR1385671](#)
- [43] WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing. Monographs on Statistics and Applied Probability* **60**. Chapman & Hall, London. [MR1319818](#)



- [44] 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** 1556–1569. [MR2504204](#)
- [45] WELSH, A. H. and YEE, T. W. (2006). Local regression for vector responses. *J. Statist. Plann. Inference* **136** 3007–3031. [MR2281238](#)
- [46] WORSLEY, K. J., TAYLOR, J. E., TOMAIUOLO, F. and LERCH, J. (2004). Unified univariate and multivariate random field theory. *NeuroImage* **23** 189–195.
- [47] WU, C. O. and CHIANG, C.-T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statist. Sinica* **10** 433–456. [MR1769751](#)
- [48] 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. [MR1666635](#)
- [49] WU, H. and ZHANG, J.-T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. Wiley, Hoboken, NJ. [MR2216899](#)
- [50] 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** 3–25. [MR2212572](#)
- [51] ZHANG, J.-T. and CHEN, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35** 1052–1079. [MR2341698](#)
- [52] ZHOU, Z. and WU, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 513–531. [MR2758526](#)
- [53] ZHU, H., LI, R. and KONG, L. (2012). Supplement to “Multivariate varying coefficient model for functional responses.” DOI:[10.1214/12-AOS1045SUPP](#).
- [54] ZHU, H., ZHANG, H., IBRAHIM, J. G. and PETERSON, B. S. (2007). Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data. *J. Amer. Statist. Assoc.* **102** 1085–1102. [MR2412530](#)
- [55] ZHU, H. T., IBRAHIM, J. G., TANG, N., ROWE, D. B., HAO, X., BANSAL, R. and PETERSON, B. S. (2007). A statistical analysis of brain morphology using wild bootstrapping. *IEEE Trans. Med. Imaging* **26** 954–966.
- [56] ZHU, H. T., STYNER, M., TANG, N. S., LIU, Z. X., LIN, W. L. and GILMORE, J. H. (2010). FRATS: Functional regression analysis of DTI tract statistics. *IEEE Trans. Med. Imaging* **29** 1039–1049.

H. ZHU  
DEPARTMENTS OF BIOSTATISTICS  
BIOMEDICAL RESEARCH IMAGING CENTER  
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL  
CHAPEL HILL, NORTH CAROLINA 27599  
USA  
E-MAIL: [hzhu@bios.unc.edu](mailto:hzhu@bios.unc.edu)

R. LI  
DEPARTMENT OF STATISTICS  
PENNSYLVANIA STATE UNIVERSITY  
UNIVERSITY PARK, PENNSYLVANIA 16802  
USA  
E-MAIL: [rli@stat.psu.edu](mailto:rli@stat.psu.edu)

L. KONG  
DEPARTMENT OF MATHEMATICAL  
AND STATISTICAL SCIENCES  
UNIVERSITY OF ALBERTA  
CAB 632  
EDMONTON, ALBERTA, T6G 2G1  
CANADA  
E-MAIL: [lkong@ualberta.ca](mailto:lkong@ualberta.ca)