Statistica Sinica **29** (2019), 2007-2033
doi:https://doi.org/10.5705/ss.202017.0505

# FMEM: FUNCTIONAL MIXED-EFFECTS MODELS FOR LONGITUDINAL FUNCTIONAL RESPONSES

Hongtu Zhu, Kehui Chen, Xinchao Luo, Ying Yuan and Jane-Ling Wang

*The University of Texas MD Anderson Cancer Center, University of Pittsburgh, Janssen R&D, LLC and University of California at Davis*

*Abstract:* The aim of this study is to conduct a systematic and theoretical analysis of estimations and inferences for a class of functional mixed-effects models (FMEM). FMEMs consist of fixed effects that characterize the association between longitudinal functional responses and covariates of interest and random effects that capture the spatial-temporal correlations of longitudinal functional responses. We propose local linear estimates of refined fixed-effect functions and establish their weak convergence, along with a simultaneous confidence band for each fixed-effect function. We propose a global test for the linear hypotheses of varying coefficient functions and derive the associated asymptotic distribution under the null hypothesis and the asymptotic power under the alternative hypothesis. We also establish the convergence rates of the estimated spatial-temporal covariance operators and their associated eigenvalues and eigenfunctions. We conduct extensive simulations and apply our method to a white-matter fiber data set from a national database for autism research to examine the finite-sample performance of the proposed estimation and inference procedures.

*Key words and phrases:* Functional response, global test statistic, mixed effects, spatial-temporal correlation, weak convergence.

## 1. Introduction

There is an increasing interest in analyses of massive functional data sets, many of which originate from brain imaging in large-scale longitudinal biomedical studies, such as the Alzeimer's Disease Neuroimaging Initiative (ADNI) (Evans and Group (2006); Mueller et al. (2005); Greven et al. (2010); Yuan et al. (2014); Zipunnikov et al. (2014)). In such studies, longitudinal functional data from $n$ different subjects are usually observed at, or are registered to, a large number of locations in a common space, denoted by $\mathcal{S}$, across multiple time points $\{t_{ij} : j = 1, \ldots, T_i; i = 1, \ldots, n\}$, where $T_i$ is the total number of time points for the $i$-th subject. Here, we use the term "functional data" for data that are measured

densely in $\mathcal{S}$, "spatial correlation" for correlations within the functional data, and "longitudinal data" and "temporal correlation" for data that are measured sparingly in $\{t_{ij} : j = 1, \ldots, T_i, i = 1, \ldots, n\}$.

The sheer size and complexity of the longitudinal functional data pose substantial challenges to most existing statistical methods for analyzing univariate or multivariate longitudinal data (Diggle et al. (2002); Fitzmaurice, Laird and Ware (2004)). These include: (i) the complexity of the temporal-spatial covariance structure, (ii) determining how to take advantage of the spatial-temporal smoothness, and (iii) theoretical justifications of the inference procedures. The first challenge is the introduction of random effects to characterize the spatial-temporal covariance structure of longitudinal functional responses. The second is the incorporation of spatial-temporal smoothness into both estimation and inference procedures to improve statistical efficiency (Ramsay and Silverman (2005)). The third is to systematically investigate the theoretical properties (e.g., consistency) of estimation and inference procedures for statistical models developed for longitudinal functional data.

Models for longitudinal functional data fall within a general functional mixed-effects modeling framework, which serves to characterize functional data with various levels of hierarchical structures (Guo (2002); Wu and Zhang (2002, 2006); Morris and Carroll (2006); Di et al. (2009); Greven et al. (2010); Zhou et al. (2010); Zhu, Brown and Morris (2011); Shi and Choi (2011); Cao, Yang and Todem (2012); Chen and Müller (2012); Horvath and Kokoszka (2012); Meyer et al. (2015); Reiss et al. (2014); Scheipl, Staicu and Greven (2015); Zipunnikov et al. (2014); Staicu, Lahiri and Carroll (2015); Cederbaum et al. (2016)). The term functional mixed-effects models (FMEMs) for correlated functional data was introduced by Guo (2002). Subsequently, Morris and Carroll (2006) developed general functional mixed-effects models with multiple levels of random-effect functions, as well as curve-to-curve deviations. Recently, a general framework for functional additive mixed models was introduced by Scheipl, Staicu and Greven (2015). Moreover, several FMEMs have been developed for longitudinal functional data (Greven et al. (2010); Yuan et al. (2014); Zipunnikov et al. (2014); Di, Crainiceanu and Jank (2014)). To the best of our knowledge, most studies on functional mixed-effects models focus on challenges (i) and (ii), described above. Here we focus on the third challenge, namely the theoretical challenges.

To address challenge (iii), we provide a comprehensive theoretical analysis for a class of FMEMs. Our FMEM consists of a measurement model at each grid point $s \in \mathcal{S}$ and a hierarchical factor model. The measurement model primarily

includes fixed effects to characterize the varying association between longitudinal functional responses and the covariates of interest. The hierarchical factor model primarily uses random effects to capture the medium-to-long-range spatial covariance and local covariance structure. Formally, we establish the weak convergence of the estimated varying association function, the uniform convergence rate of the spatial-temporal covariance estimator, the asymptotic distribution of a global test statistic for linear hypotheses of the regression coefficient functions, and an asymptotic simultaneous confidence band (SCB) for each varying fixed-effect function. The code and documentation for the FMEM, written in Matlab, are freely accessible from the "`http://www.nitrc.org/projects/fadtts`."

## 2. FMEM: Functional Mixed-Effects Model

### 2.1. Model setup

Suppose that we observe longitudinal functional data and clinical variables from $n$ independent subjects. Let $T_i$ be the total number of longitudinal measurements for the $i$-th subject, $i = 1, \ldots, n$, and let $t_{ij}$ be the $j$-th measurement time point for the $i$-th subject; thus, $j = 1, \ldots, T_i$. Throughout this paper, we focus on a fixed number of time points and sparse longitudinal data; that is, $\max_{i \le n} T_i < T_0 < \infty$. Let $s_m$ represent a specific grid point of the functional template space $\mathcal{S}$ for $m = 1, \ldots, M$. Specifically, for the $i$-th subject at time $t_{ij}$, we observe functional data, denoted by $y_{ij}(s_m) = y_i(t_{ij}, s_m)$ for $1 \le m \le M$, and a $p_x$-dimensional covariate vector $x_i$ of interest, denoted by $x_{ij} = x_i(t_{ij})$, at time $t_{ij}$. Here $x_i$ may include time-independent and time-dependent covariates, such as age, gender, and genetic markers. For ease of notation, we assume that $\mathcal{S} = [0, 1]$ and $0 = s_1 \le \cdots \le s_M = 1$. However, our results can easily be extended to higher dimensions when $\mathcal{S}$ is a compact subset of a Euclidean space.

We consider an FMEM consisting of a measurement model and a hierarchical factor model. This model aims to extend the conventional linear mixed-effects model to accommodate the additional spatial component. The measurement model associated with the FMEM characterizes the varying association between functional responses and their covariates at any $s \in \mathcal{S}$, as follows:

$$y_{ij}(s) = \mu(x_{ij}, \beta(s)) + z_{ij}^T b_i(s) + e_{ij}(s), \tag{2.1}$$

where $\mu(\cdot, \cdot)$ is a known function, $\beta(s) = (\beta_1(s), \ldots, \beta_{p_\beta}(s))^T$ is a $p_\beta \times 1$ vector of the fixed-effect functions of $s$, and $z_{ij} = z_i(t_{ij}) = (z_{ij1}, \ldots, z_{ijp_z})^T$ is a $p_z \times 1$ vector of the random-effect covariates associated with the random effects $b_i(s)$. Here,

$b_i(s) = (b_{i1}(s), \ldots, b_{ip_z}(s))^T$ is a vector of the random effects that characterize the spatial-temporal correlation structures across the functional domain space. In contrast, $e_{ij}(s)$ is a spatial random process delineated from $b_i(s)$, that is, after filtering out $z_{ij}^T b_i(s)$. Moreover, $e_{ij}(s)$ and $b_i(s)$ are independent. In many applications, $\mu(x_{ij}, \beta(s)) = x_{ij}^T \beta(s)$ is a linear function of $x_{ij}$, similar to the setting of the traditional linear mixed-effects model. Therefore, we focus on this special linear case in this paper. Extensions to nonlinear cases are discussed in Remark 1. Marginally, for a fixed $s$, model (2.1) with $\mu(x_{ij}, \beta(s)) = x_{ij}^T \beta(s)$ is a standard linear mixed-effects model. This motivates us to adopt the standard notation for linear mixed-effects models. Moreover, because $z_{ij}$ may include time-independent and time-dependent covariates, the inclusion of $z_{ij}^T b_i(s)$ allows us to capture a large portion of the variation in the spatial and temporal correlation structures.

The spatial random process $e_{ij}$ in (2.1) is further decomposed into two parts,

$$e_{ij}(s) = e_{ij,G}(s) + e_{ij,L}(s), \tag{2.2}$$

where $e_{ij,G}(s)$ is a smooth stochastic process representing the global dependency that depicts the medium-to-long-range spatial dependence, $e_{ij,L}(s)$ is a measurement error representing local variability, and $e_{ij_1,G}(\cdot)$ and $e_{ij_2,L}(\cdot)$ are independent for any $j_1$ and $j_2$. Because $e_{ij,L}(s)$ are measurement errors, we assume that $e_{ij_1,L}(s)$ and $e_{ij_2,L}(s')$ are mutually independent whenever either $j_1 \neq j_2$ or $s \neq s'$. We also assume that, for any $j_1 \neq j_2$, $e_{ij_1,G}(\cdot)$ and $e_{ij_2,G}(\cdot)$ are mutually independent. This assumption is equivalent to assuming that the random effects $b_i(\cdot) = (b_{i1}(\cdot), \ldots, b_{ip_z}(\cdot))^T$ explain all the within-subject correlation along the longitudinal direction, which is a common assumption in linear mixed-effects models. However, it does not exclude correlations along the functional direction because $e_{ij,G}(s)$ and $e_{ij,G}(s')$ are not required to be independent for $s \neq s'$.

Moreover, $b_i(s)$, $e_{ij,L}(s)$, and $e_{ij,G}(s)$ are mutually independent and are independent and identical copies of $\mathrm{SP}(0, \Sigma_{e,L})$, $\mathrm{SP}(0, \Sigma_b)$, and $\mathrm{SP}(0, \Sigma_{e,G})$, respectively, where $\mathrm{SP}(\mu, \Sigma)$ denotes a stochastic process vector with mean function (or function vector) $\mu(s)$ and covariance function (or function matrix) $\Sigma(s, s')$. Moreover, $\Sigma_b(s, s')$ is a $p_z \times p_z$ matrix with $\Sigma_{bkk'}(s, s')$ as the $(k, k')$-th element. The covariance structure of $y_i(s) = (y_{i1}(s), \ldots, y_{iT_i}(s))^T$, denoted by $\Sigma_{y,i}(s, s')$, is

$$\Sigma_{y,ij_1j_2}(s, s') = z_{ij_1}^T \Sigma_b(s, s') z_{ij_2} + \Sigma_{e,G}(s, s') 1(j_1 = j_2) + \Sigma_{e,L}(s, s') 1(j_1 = j_2, s = s'),$$

where $1(\cdot)$ is an indicator function.

### 2.2. Estimation procedure

Our primary goal is to find efficient procedures for estimations and inferences for $\beta(\cdot)$. Inspired by novel ideas from the literature (Yao, Müller and Wang (2005); Greven et al. (2010); Zipunnikov et al. (2014)), we develop a procedure to estimate $\beta(\cdot)$, $\Sigma_{bkk'}(\cdot,\cdot)$, $\Sigma_{e,G}(\cdot,\cdot)$, $\Sigma_{e,L}(\cdot,\cdot)$, and the eigenvalue-eigenvector pairs of $\Sigma_{bkk'}(\cdot,\cdot)$ and $\Sigma_{e,G}(\cdot,\cdot)$. Compared with the estimation methods of Greven et al. (2010) and Zipunnikov et al. (2014), our method is an improvement over the ordinary least squares methods used to estimate $\beta(\cdot)$ by incorporating spatial and/or temporal smoothness in longitudinal functional data. Explicitly, we incorporate the within-subject correlations between $T_i$ longitudinal observations to obtain statistical efficiency, as stated in Theorem 1.

Henceforth, we focus on $\mu(x_{ij},\beta(s)) = x_{ij}^T\beta(s)$. However, the proposed estimation procedure can be extended to a nonlinear mean function $\mu(x_{ij},\beta(s))$, as discussed at the end of Section 2.2. There are four key steps in the estimation procedure:

Step (I): Calculate an initial estimator $\hat{\beta}(s)$ of $\beta(s)$ for each $s \in \mathcal{S}$.

Step (II): Calculate estimates of the covariance operators $\Sigma_{bkk'}(\cdot,\cdot)$ and $\Sigma_{e,G}(\cdot,\cdot)$ and their spectral decompositions, and obtain an estimate of $\Sigma_{e,L}(\cdot,\cdot)$.

Step (III): Use the estimated covariance operators obtained from Step (II) to improve the estimate in step (I) using a refined estimator of $\beta(s)$, denoted by $\tilde{\beta}(s)$.

Step (IV): Obtain individual random-effect functions $u_{ij,G}(s) = z_{ij}^T b_i(s) + e_{ij,G}(s)$.

**Step (I)**: We employ a local linear smoother (Fan and Gijbels (1996)) to obtain an initial estimator of $\beta(\cdot)$ without incorporating the spatial-temporal correlation. Specifically, we apply a Taylor expansion for $\beta$ at $s$,

$$\beta(s_m) \approx \beta(s) + \dot{\beta}(s)(s_m - s) = A(s)s_{h_1}(s_m - s), \tag{2.3}$$

where $s_{h_1}(s_m - s) = (1, (s_m - s)/h_1)^T$ and $A(s) = [\beta(s)\ h_1\dot{\beta}(s)]$ is a $p_x \times 2$ matrix. Here, $\dot{\beta}(s) = (\dot{\beta}_1(s),\ldots,\dot{\beta}_{p_x}(s))^T$ is a $p_x \times 1$ vector and $\dot{\beta}_l(s) = d\beta_l(s)/ds$ for $l = 1,\ldots,p_x$. Let $K(s)$ be a kernel function and $K_h(s) = h^{-1}K(s/h)$ be the rescaled kernel function with bandwidth $h$. We estimate $A(s)$ by minimizing the

following weighted least squares function:

$$\sum_{i=1}^{n}\sum_{j=1}^{T_i}\sum_{m=1}^{M}\{y_{ij}(s_m)-x_{ij}^T A(s)s_{h_1}(s_m-s)\}^2 K_{h_1}(s_m-s). \tag{2.4}$$

Let $a^{\otimes 2}=aa^T$ for any vector $a$ 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 $\mathrm{vec}(C)=(c_{11},\ldots,c_{M_1 1},\ldots,c_{1M_2},\ldots,c_{M_1M_2})^T$. Let $\hat{A}(s)$ be the minimizer of (2.4). Then,

$$\mathrm{vec}(\hat{A}(s))=\Sigma(s,h_1)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{T_i}\sum_{m=1}^{M}K_{h_1}(s_m-s)\{s_{h_1}(s_m-s)\otimes x_{ij}\}y_{ij}(s_m), \tag{2.5}$$

where $\Sigma(s,h_1)=\sum_{i=1}^{n}\sum_{j=1}^{T_i}\sum_{m=1}^{M}K_{h_1}(s_m-s)\{s_{h_1}(s_m-s)^{\otimes 2}\otimes x_{ij}^{\otimes 2}\}$. Thus, we have $\hat{\beta}(s)=(\hat{\beta}_1(s),\ldots,\hat{\beta}_{p_x}(s))^T=\{(1,0)\otimes I_{p_x}\}\mathrm{vec}(\hat{A}(s))$, where $I_{p_x}$ is a $p_x\times p_x$ identity matrix. In practice, we may select the bandwidth $h_1$ by using leave-one-curve-out cross-validation. Specifically, we pool the data from all $n$ subjects and select a bandwidth $h_1$ by minimizing the cross-validation score given by

$$\mathrm{CV}(h_1)=\left(\sum_{i=1}^{n}T_iM\right)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{T_i}\sum_{m=1}^{M}\{y_{ij}(s_m)-x_i^T\hat{\beta}(s_m,h_1)^{(-i)}\}^2, \tag{2.6}$$

where $\hat{\beta}(s,h_1)^{(-i)}$ is a local linear estimator of $\beta(s)$ with bandwidth $h_1$, based on all data excluding the observations for the $i$-th subject.

**Step (II)**: We use a two-step procedure to estimate $\Sigma_b(s,s')$ and $\Sigma_{e,G}(s,s')$. Let $\Sigma_e(s,s')$ be the covariance function of $e_{ij}(s)$.

(S1) First, we use the least squares method to estimate $\Sigma_b(s_m,s_{m'})$ and $\Sigma_e(s_m,s_{m'})$ for $m,m'=1,\ldots,M$. Let $\hat{u}_{ij}(s)=y_{ij}(s)-x_{ij}^T\hat{\beta}(s)$. We estimate $\Sigma_b(s_m,s_{m'})$ and $\Sigma_e(s_m,s_{m'})$ by minimizing the following least squares function:

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{j_1\neq j_2}\{\hat{u}_{ij_1}(s_m)\hat{u}_{ij_2}(s_{m'})-z_{ij_1}^T\Sigma_b(s_m,s_{m'})z_{ij_2}\}^2\\&+\sum_{i=1}^{n}\sum_{j=1}^{T_i}\{\hat{u}_{ij}(s_m)\hat{u}_{ij}(s_{m'})-z_{ij}^T\Sigma_b(s_m,s_{m'})z_{ij}-\Sigma_e(s_m,s_{m'})\}^2,\end{aligned} \tag{2.7}$$

where $\sum_{j_1\neq j_2}$ denotes the sum over all $j_1,j_2=1,\ldots,T_i$, such that $j_1\neq j_2$. The least squares method in (2.7) has been considered in the literature (Di et al. (2009); Greven et al. (2010); Cederbaum et al. (2016)), where previous authors used penalized splines smoothing instead of a local linear regression. Let $\hat{\Sigma}_b^{LS}(s_m,s_{m'})$ and $\hat{\Sigma}_e^{LS}(s_m,s_{m'})$ be the minimizers of (2.7). Then, we

have $\mathrm{vec}(\hat{\Sigma}_b^{LS}(s_m, s_{m'})) = G\{u(s_m, s_{m'}) - \hat{\Sigma}_e^{LS}(s_m, s_{m'})g\}$ and $\hat{\Sigma}_e^{LS}(s_m, s_{m'}) = (1 - a_2 g^T g)^{-1}\{v(s_m, s_{m'}) - a_2 g^T G u(s_m, s_{m'})\}$, where $a_2 = (\sum_{i=1}^n T_i)^{-1}$, $g = \sum_{i=1}^n \sum_{j=1}^{T_i} z_{ij} \otimes z_{ij}$, $G = \{\sum_{i=1}^n \sum_{j_1,j_2}^{T_i} (z_{ij_1} \otimes z_{ij_2})^{\otimes 2}\}^{-1}$,

$$v(s_m, s_{m'}) = a_2 \sum_{i=1}^n \sum_{j=1}^{T_i} \hat{u}_{ij}(s_m)\hat{u}_{ij}(s_{m'}), \text{and}$$

$$u(s_m, s_{m'}) = \sum_{i=1}^n \sum_{j_1,j_2=1}^{T_i} \hat{u}_{ij_1}(s_m)\hat{u}_{ij_2}(s_{m'})(z_{ij_1} \otimes z_{ij_2}).$$

(S2) Next, for each $(k, k')$, with $1 \le k, k' \le p_z$, we apply a local constant smoother to $\hat{\Sigma}_{bkk'}^{LS}(s_m, s_{m'})$ for $s_m, s_{m'} \in \mathcal{S} \times \mathcal{S}$ and $m, m' = 1, \ldots, M$. This provides the final estimate for $\Sigma_b(s, s')$. Similarly, we can obtain an estimate of $\Sigma_{e,G}(s, s')$ using a local constant smoother, where the diagonal elements of $\hat{\Sigma}_e^{LS}(s_m, s_{m'})$ (i.e., $\hat{\Sigma}_e^{LS}(s_m, s_m), m = 1, \ldots M$) are excluded from the estimation of $\Sigma_{e,G}(s, s')$.

Specifically, we estimate $\Sigma_{bkk'}(s, s')$ and $\Sigma_{e,G}(s, s')$ by minimizing the following weighted least squares functions:

$$\begin{aligned}
&\min_{\Sigma_{bkk'}(s,s')} \sum_{m,m'=1}^{M} \{\hat{\Sigma}_{bkk'}^{LS}(s_m, s_{m'}) - \Sigma_{bkk'}(s, s')\}^2 K_{h_2}(s_m - s) K_{h_2}(s_{m'} - s'), \\
&\min_{\Sigma_{e,G}(s,s')} \sum_{m \neq m'} \{\hat{\Sigma}_e^{LS}(s_m, s_{m'}) - \Sigma_{e,G}(s, s')\}^2 K_{h_3}(s_m - s) K_{h_3}(s_{m'} - s').
\end{aligned} \tag{2.8}$$

The bandwidths $h_2$ and $h_3$ are selected using the leave-one-curve-out cross-validation method.

Finally, we perform the spectral decomposition of $\hat{\Sigma}_{bkk'}(s, s')$ and $\hat{\Sigma}_{e,G}(s, s')$, and then calculate $\hat{\Sigma}_{e,L}(s_m, s_m)$ using

$$\hat{\Sigma}_{e,L}(s_m, s_m) = \{\hat{\Sigma}_e^{LS}(s_m, s_m) - \hat{\Sigma}_{e,G}(s_m, s_m)\} 1(\hat{\Sigma}_e^{LS}(s_m, s_m) - \hat{\Sigma}_{e,G}(s_m, s_m) > 0).$$

**Step (III)**: We incorporate the estimated covariance function to improve the local linear regression estimate of $\beta(\cdot)$. Similar, but different ideas have been used to iteratively improve the mean estimation (Cederbaum et al. (2016); Di, Crainiceanu and Jank (2014)). Letting $\Sigma_{y_i,G}(s, s')$ be the covariance function of $u_{i,G}(s) = (u_{i1,G}(s), \ldots, u_{iT_i,G}(s))^T$, we obtain its estimator $\hat{\Sigma}_{y_i,G}(s, s')$ based on $\hat{\Sigma}_b(s, s')$ and $\hat{\Sigma}_{e,G}(s, s')$ from step (II). Let $X_i = (x_{i1} \ldots x_{iT_i})$ be a $p_x \times T_i$ matrix.

We estimate $A(s)$ by minimizing the following weighted least squares function:

$$\sum_{i=1}^{n}\sum_{m=1}^{M}[\{y_i(s_m)-X_i^TA(s)s_{h_\beta}(s_m-s)\}^T\hat{\Sigma}_{y_i,G}(s_m,s_m)^{-1/2}]^{\otimes 2}K_{h_\beta}(s_m-s), \quad (2.9)$$

where $h_\beta$ is a bandwidth.

Let $\tilde{A}(s)$ be the minimizer of (2.9). Then, we have

$$\begin{aligned}&\mathrm{vec}(\tilde{A}(s))\\ &=\tilde{\Sigma}(s,h_\beta)^{-1}\sum_{i=1}^{n}\sum_{m=1}^{M}K_{h_\beta}(s_m-s)\{s_{h_\beta}(s_m-s)\otimes X_i\}\{\hat{\Sigma}_{y_i,G}(s_m,s_m)\}^{-1}y_i(s_m),\end{aligned}$$

where $\tilde{\Sigma}(s,h_\beta)=\sum_{i=1}^{n}\sum_{m=1}^{M}K_{h_\beta}(s_m-s)[\{s_{h_\beta}(s_m-s)\otimes X_i\}\hat{\Sigma}_{y_i,G}(s_m,s_m)^{-1/2}]^{\otimes 2}$. We have

$$\tilde{\beta}(s)=(\tilde{\beta}_1(s),\ldots,\tilde{\beta}_{p_x}(s))^T=\{(1,0)\otimes I_{p_x}\}\mathrm{vec}(\tilde{A}(s)). \quad (2.10)$$

To select the bandwidth $h_\beta$, we pool the data from all $n$ subjects and select the bandwidth $h_\beta$ that minimizes $\mathrm{CV}(h_\beta)=(nM)^{-1}\sum_{i=1}^{n}\sum_{m=1}^{M}[\{y_i(s_m)-X_i^T\tilde{\beta}(s_m,h_\beta)^{(-i)}\}^T\hat{\Sigma}_{y_i,G}(s_m,s_m)^{-1/2}]^{\otimes 2}$. Here, $\tilde{\beta}(s,h_\beta)^{(-i)}$ is the local linear estimator of $\beta(s)$ with the bandwidth $h_\beta$, based on all data excluding the observations for the $i$-th subject.

**Step (IV)**: We use the local linear regression method to smooth $\{\tilde{u}_{ij}(s_m)=y_{ij}(s_m)-x_{ij}^T\tilde{\beta}(s_m)\}_{m=1}^{M}$, and then obtain an estimate of $u_{ij,G}(s)=z_{ij}^Tb_i(s)+e_{ij,G}(s)$ for each $i$ and $j$. Because the local linear regression is a standard method (Fan and Gijbels (1996); Wand and Jones (1995)), we omit the detailed steps for the approximation of $u_{ij,G}(s)$. Furthermore, to recover the subject-specific random effect $b_i(s)$, we can use the best linear unbiased predictors. These are commonly employed in linear mixed-effects models to estimate $b_i(s)$ at each point $s$ and then smooth over $s$.

**Remark 1.** To extend the estimation procedure to nonlinear mean functions $\mu(x_{ij},\beta(s))$, such as exponential functions or power functions, we need to modify steps (I) and (III) by applying a Taylor expansion for $\mu(x_{ij},\beta(s_m))$ at $s$, as follows:

$$\mu(x_{ij},\beta(s_m))\approx\mu(x_{ij},\beta(s))+\dot{\mu}(x_{ij},\beta(s))\dot{\beta}(s)(s_m-s)=\mu_{ij}(s)s_{h_1}(s_m-s),$$

where $\dot{\mu}(x_{ij},\beta(s))=\partial\mu(x_{ij},\beta(s))/\partial\beta(s)$ and $\mu_{ij}(s)=(\mu(x_{ij},\ \beta(s)),\ \dot{\mu}(x_{ij},\ \beta(s))\dot{\beta}(s)h_1)$. Then, we estimate $A(s)$ by minimizing the following nonlinear weighted least squares function: $L_n(A(s))=\sum_{i=1}^{n}\sum_{j=1}^{T_i}\sum_{m=1}^{M}\{y_{ij}(s_m)-\mu_{ij}(s)\ s_{h_1}(s_m-s)\}^2K_{h_1}(s_m-s)$. In this general case, $\hat{A}(s)$ does not have an explicit form, but it can be estimated using optimization algorithms, such as the Gaus-

sian Newton algorithm or the Levenberg–Marquardt algorithm (Seber and Wild (1989)). Similarly to $L_n(A(s))$, we can modify (2.9) in step (III).

### 2.3. Computational complexity

The computational complexity of our estimation procedure is extremely important for high-dimensional neuroimaging data, which usually contain a large number of locations, especially when they correspond to the voxel locations of an image. For instance, $M$ can have a magnitude of tens of thousands. For the linear mean function, the computational complexity of our estimation procedure in Section 2.2 is $O(nh_1T_0M^2 + nT_0(R_0M)^2 + nT_0h_sM^2)$. If we use leave-one-out cross-validation, then the computational effort increases by a factor of $n$.

We first discuss steps (I) and (III). In step (I), we need to calculate the local linear estimator of $\beta(s_m)$ at each grid point $s_m$ across $\mathcal{S}_0 = \{s_m, m = 1, \ldots, M\}$. The computational complexity of step (I) is almost the same as that of the standard pointwise linear regression analysis. An alternative is to fit a linear mixed-effect model at each grid point $s_m$ using the maximum likelihood. However, this step is not necessary because it only applies to an initial estimate, which is then improved in step (III).

For step (III), we only need to calculate the weighted least squares estimators $\tilde{\beta}(s_m)$ in (2.10) across $s_m \in \mathcal{S}_0$, which is computationally straightforward. The computational complexity is $O(nT_0h_1M)$ for each $s_m$; thus, the overall complexity is $O(nT_0h_1M^2)$.

To improve the computational efficiency, we standardize all covariates and then use a single tuning parameter $h_1$ to smooth all coefficient functions $\beta_j(s)$. This strategy works best for coefficient functions that exhibit similar degrees of smoothness. Thus, it may be necessary to use different tuning parameters for different coefficient functions (Fan and Zhang (2008)) when the functions have different levels of smoothness.

Next, we discuss the computational complexity of step (II). First, estimating $\hat{u}_{ij}(s)$ is computationally fast for all possible $(i, j)$. Second, we do not need to calculate $\Sigma_b(s, s')$ and $\Sigma_{e,G}(s, s')$ for all possible $(s, s')$. As discussed in step (III), we only need the estimates of $\Sigma_b(s_m, s_m)$ and $\Sigma_{e,G}(s_m, s_m)$ for all $s_m \in \mathcal{S}_0$. Therefore, in step (S2), we focus on solving $\Sigma_b(s_m, s_m)$ and $\Sigma_{e,G}(s_m, s_m)$, with all $(s_m, s_{m'})$ in $\{(s_m, s_{m'}) \in \mathcal{S}_0 \times \mathcal{S}_0 : |s_m - s_{m'}| \leq R_0\}$, where $R_0$ is a positive scalar. In this case, step (II) is computationally feasible even for large $M$ when $R_0$ is relatively small. The computational complexity is at most $O(nT_0(R_0M)^2)$ for $(s_m, s_{m'}) \in \mathcal{S}_0 \times \mathcal{S}_0$.

A major computational hurdle is the calculation of $\Sigma_b(s, s')$ and $\Sigma_{e,G}(s, s')$ for all possible $(s, s')$. If $M$ is relatively large, it can be computationally challenging to estimate $\Sigma_b(s_m, s'_m)$ and $\Sigma_{e,G}(s_m, s'_m)$ across all possible $(s_m, s'_m) \in \mathcal{S}_0 \times \mathcal{S}_0$. We take two different approaches. The first estimates $\Sigma_b(s_m, s'_m)$ and $\Sigma_{e,G}(s_m, s'_m)$ for a small subset of $S_0 \times S_0$. Specifically, we bin the data to reduce the number of grid points to a much smaller number $M_0 << M$. Then, we estimate $\Sigma_b(s, s')$ and $\Sigma_{e,G}(s, s')$ on those $M_0$ points and interpolate the results elsewhere. The second approach applies the approaches proposed by Zipunnikov et al. (2014) and Xiao et al. (2016) to estimate $\Sigma_b(s, s')$ and $\Sigma_{e,G}(s, s')$. These methods include a fast implementation of the sandwich smoother for covariance smoothing, and a two-step procedure where we first obtain the singular value decomposition of the data matrix and then smooth the eigenvectors.

Note that with regard to the computational complexity of step (IV), similarly to step (II), smoothing $u_{ij,P}(s)$ for all possible $(i, j)$ is computationally light. The overall computational complexity is approximately $O(nT_0h_sM^2)$, where $h_s$ is the bandwidth of the local linear method.

**Remark 2.** We discuss two possible extensions of (2.2). The first is to extend the estimation procedure from $\mathcal{S} = [0, 1]$ to a $D$-dimensional compact subset of a Euclidean space. For this, we only need to modify steps (I) and (III) by changing $\dot{\beta}_l(s)$ and $s_m - s$ into $D \times 1$ vectors. The second extension is to assume that $e_{ij_1,G}(s)$ and $e_{ij_2,G}(s)$, for $j_1 \neq j_2$, are dependent and have a separable covariance structure, $\text{cov}(e_{ij_1,G}(s), e_{ij_2,G}(s)) = \Sigma_{e,G}(s, s')\rho(t_{ij_1}, t_{ij_2}; \theta)$. Here, $\rho(t_{ij_1}, t_{ij_2}; \theta)$ is usually a prespecified correlation function of unknown parameter $\theta$, such as the exponential correlation model with $\rho(t_{ij_1}, t_{ij_2}; \theta) = \exp(-\theta|t_{ij_1} - t_{ij_2}|)$ (Diggle et al. (2002); Fitzmaurice, Laird and Ware (2004)). However, we found empirically that using of the correlation function significantly increases the computational complexity, but does not yield much of an efficiency gain when estimating $\beta(\cdot)$.

## 3. Theoretical Results

We systematically investigate the asymptotic properties of all estimators proposed in Section 2.2 and investigate several inference procedures based on the asymptotic properties. For any smooth function $f(s)$, we use the notation $\dot{f}(s) = df(s)/ds$ and $\ddot{f}(s) = d^2f(s)/ds^2$. We use $u_q = \int K(v)v^q dv$ and $v_q = \int K^q(v)dv$, for $q = 1$ and 2, and $||\cdot||_2$ for the Euclidean norm.

### 3.1. Assumptions

Throughout the paper, the following assumptions are used to facilitate the technical details. Some of the assumptions might be weakened, but the current version simplifies the proof.

(A.1) The grid points in $\mathcal{S}_0 = \{s_m, m = 1, \ldots, M\}$ are independently and identically distributed with a density function $f(s)$, which has a continuous second-order derivative and bounded support $\mathcal{S}$. Moreover, for some $f_l > 0$ and $f_u < \infty$, $f_l < f(s) < f_u$ for all $s \in \mathcal{S}$.

(A.1b) The grid points $\mathcal{S}_0 = \{s_m, m = 1, \ldots, M\}$ are prefixed according to a design density function $f(s)$ such that $\int_0^{s_m} f(s)ds = m/M$ for $m \geq 1$. Here, $f(s)$ has a continuous second-order derivative and bounded support $[0, 1]$, and $f_l < f(s) < f_u$ for all $s \in [0, 1]$, for some positive $f_l > 0$ and $f_u < \infty$.

(A.2) The covariate vectors $x_{ij} = (x_{ij1}, \ldots, x_{ijp_x})^T$ and $z_{ij} = z_i(t_{ij}) = (z_{ij1}, \ldots, z_{ijp_z})^T$ may or may not be time-dependent. Nevertheless, we use the notation $x_{ijl} = x_{il}(t_{ij})$ for $1 \leq l \leq p_x$, and $z_{ijl} = z_{il}(t_{ij})$ for $1 \leq l \leq p_z$. We assume that $\sup_{t\in\mathcal{T}} |x_{il}(t)|$ and $\sup_{t\in\mathcal{T}} |z_{il}(t)|$ are almost surely bounded, where $\mathcal{T}$ is a finite time domain.

(A.3) The kernel function $K(t)$ is a symmetric density function with compact support $[-1, 1]$, and is Lipschitz continuous.

(A.4) All components of $\beta(s)$ have continuous second-order derivatives on $\mathcal{S}$.

(A.5) With probability one, the sample paths of $e_{ij,G}(\cdot)$ and $b_i(\cdot)$ are Lipschitz continuous.

(A.6) $\max_i T_i < T_0$, $n, M \to \infty$, $h \to 0$, $Mh \to \infty$ and $n^a h \to \infty$ for some $a > 0$, where $T_0$ is a fixed constant, and $h$ can be $h_1$, $h_\beta$, $h_2$, or $h_3$.

(A.7) $E\{\sup_{s\in[0,1]} |e_{ij,G}(s)|^{2q}\} + E\{\sup_{s\in S_0} |e_{ij,L}(s)|^{2q}\} < \infty$, for some $q > 2$.

(A.8) $E\{\sup_{s\in[0,1]} \|b_i(s)\|_2^{2q}\} < \infty$, for some $q > 2$.

(A.9) $E\{X_i \Sigma_{y_i,G}(s, s)^{-1} \Sigma_{y_i,G}(s, s') \Sigma_{y_i,G}(s', s')^{-1} X_i^T\}$ exists for any $(s, s')$.

(A.10) There is a positive fixed integer $E < \infty$ such that the eigenvalues of $\Sigma_{e,G}$ satisfy $\lambda_1^e > \cdots > \lambda_E^e > \lambda$, for some constant $\lambda > 0$. There is an analogous case for the eigenvalues of $\Sigma_b$.

**Remark 3.** Our theoretical results hold for both random and fixed designs. Assumptions (A.1) is a standard condition on random design points $s$, whereas (A.1b) applies to fixed designs. Assumption (A.2) is a condition on the boundedness of the covariate vectors. The bounded support restriction on $K(\cdot)$ in assumption (A.3) is not essential and can be removed if we include restrictions on the tail of $K(\cdot)$. Assumptions (A.4)–(A.5) are smoothness conditions on the

coefficient functions, random functions, and their covariances. The smoothness condition in assumption (A.5) can be relaxed with substantial additional effort (Zhu, Li and Kong (2012)). Assumption (A.6) is a weak condition on $n$, $M$, and $h$, where $h_1$ is the bandwidth used in Step (I) for the initial estimate of $\beta$. Assumptions (A.7) and (A.8) require uniform bounds on certain high-order moments of the random functions, which are standard assumptions in the literature (Zhu, Li and Kong (2012); Li and Hsing (2010)). Assumption (A.10) on the simple multiplicity of the first $E$ eigenvalues is only needed to investigate the asymptotic properties of the eigenfunctions. It is also a standard assumption in the literature.

### 3.2. Asymptotics of estimation procedure

We state the following theorems, for which detailed proofs can be found in the Supplementary Material. The first theorem tackles the theoretical properties of $\{\tilde{\beta}(s) : s \in \mathcal{S}\}$ obtained from step (III).

**Theorem 1.** *Under* (A.1) (or (A.1b)) *and* (A.2)–(A.9)*, we have the following results:*
*(i) The asymptotic bias and covariance of* $\tilde{\beta}(s)$ *for* $s \in (0,1)$ *are*

$$Bias(\tilde{\beta}(s)|\mathcal{S}) = \frac{1}{2}\ddot{\beta}(s)h_\beta^2 u_2\{1+o(1)\}, \tag{3.1}$$

$$var(\tilde{\beta}(s)|\mathcal{S}) = n^{-1}\left\{n^{-1}\sum_{i=1}^{n} E(X_i\{\Sigma_{y_i,G}(s,s)\}^{-1}X_i^T)\right\}^{-1}\{1+o(1)\}.$$

*(ii) If* $\log M = o(Mh_\beta)$ *and there exists* $\gamma_n \to \infty$*, with* $n^{1/2}\gamma_n^{1-q} = o(1)$ *and* $n^{-1/2}\gamma_n \log M = o(1)$ *for some* $q > 2$ *that satisfies (A.7), then as* $n \to \infty$*,* $\sqrt{n}\{\tilde{\beta}(s) - E(\tilde{\beta}(s)|\mathcal{S})\}$ *converges weakly to a centered Gaussian process* $G(\cdot) \sim \mathcal{G}(0,R)$*, where* $R(s,s') = \{Q^*(s,s)\}^{-1}Q^*(s,s')\{Q^*(s',s')\}^{-1}$*, with* $Q^*(s,s') = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E(X_i\{\Sigma_{y_i,G}(s,s)\}^{-1}\Sigma_{y_i,G}(s,s')\{\Sigma_{y_i,G}(s',s')\}^{-1}X_i^T)$.

Theorem 1 (i) provides a theoretical justification for steps (I)–(III) for the refined estimator $\tilde{\beta}(s)$. It has several important implications. First, the estimator $\hat{\beta}(s)$ obtained in step (I) has the asymptotic covariance

$$n^{-1}\left\{n^{-1}\sum_{i=1}^{n} E(X_iX_i^T)\right\}^{-1} n^{-1}\sum_{i=1}^{n} E(X_i\Sigma_{y_i,G}(s,s)X_i^T)\left\{n^{-1}\sum_{i=1}^{n} E(X_iX_i^T)\right\}^{-1}$$

(details can be found in the proof of Theorem 1), which is larger than that of $\tilde{\beta}(s)$. The improvement by the refined estimator $\tilde{\beta}(s)$ is the result of incorporating within-subject correlations between $T_i$ longitudinal observations, and can lead

to a substantial efficiency gain when estimating $\{\beta(s) : s \in \mathcal{S}\}$. Second, if we use the maximum likelihood (or the restricted maximum likelihood) estimators for each observation at $s_m$, the asymptotic covariance, given by $\{\sum_{i=1}^{n} E(X_i\{\Sigma_{y_i}(s_m, s_m)\}^{-1}X_i^T)\}^{-1}$, is larger than that of $\tilde{\beta}(s_m)$. The improvement achieved by $\tilde{\beta}(s_m)$ is the result of incorporating the smoothness in the functional data. Therefore, we can construct more efficient estimators of $\beta(s)$ by simultaneously accounting for the smoothness in the functional data and the within-subject covariance, because these functions are measured repeatedly and longitudinally. Moreover, the asymptotic bias of $\tilde{\beta}(s)$ is of order $h_\beta^2$, which is similar to that of a nonparametric regression for independent responses. In contrast, the asymptotic variance of $\tilde{\beta}(s)$ is of order $n^{-1}$.

Note that the efficiency gain discussed above does not conflict with the results of Lin and Carroll (2001), who show that the most efficient estimator of the nonparametric function using kernel smoothing is achieved by ignoring the dependence structure among the functional observations. In our setting, this means that kernel smoothing in the direction of $s$ should be implemented as in Step (I), that is, by ignoring the dependence structure among functional observations. However, in the FMEM setting of longitudinal functional data, it is possible to improve the $\beta$ estimate as we did in Step (III) by incorporating the covariance structure $\Sigma_{y_i,G}(s, s)$. The analogy here is with the standard linear mixed-effects model with longitudinal data only (i.e., no functional components), because the FMEM is an extension of the linear mixed-effects model. Clearly, in a linear mixed-effects model, we need to perform a weighted least squares procedure to improve the efficiency of the $\beta$ estimator. This was done in Step (III) to refine the $\beta$ estimator using a weighted least squares estimator with weights from $\Sigma_{yi,G}(s, s)$. Note that we could implement Step (III) only after we obtaining a covariance estimate in Step (II), which relies on an initial unweighted least squares estimator of $\beta$ in Step (I). This explains why three steps are necessary for the estimation of $\beta$.

Theorem 1 (ii) establishes the weak convergence of the centered estimator $\tilde{\beta}(s) - E(\tilde{\beta}(s))$, which is essential to the statistical inference for $\beta(s)$ in Section 3.3. Let $h = n^\alpha$, $M = n^\beta$, and $\gamma_n = n^\gamma$. Anything that satisfies $\alpha < 0$, $\alpha + \beta > 0$, and $-1/(2(1-q)) < \gamma < 1/2$ will satisfy the assumptions, where $q > 2$ is a constant that satisfies the moment condition given in (A.7).

The second theorem provides a theoretical analysis of the estimators of $\Sigma_{e,G}(s, s')$ obtained from step (II). Similar results can be obtained for $\Sigma_{b,kk'}(s, s')$, $1 \le k, k' \le p_z$, and are provided in the online Supplementary Material.

**Theorem 2.** *Under* (A.1) (or (A.1b)), (A.2)–(A.8), *and* (A.10), *if* $h_1 = O((\log n/n)^{1/4})$ *and* $h_3 = O(\log n/n)^{1/4}$, *then we have the following results:*
*(i)* $\sup_{s,s'} |\hat{\Sigma}_{e,G}(s,s') - \Sigma_{e,G}(s,s')| = O_p((\log n/n)^{1/2})$;
*(ii) For* $1 \le l \le E$, $\{\int_0^1 |\hat{\psi}_l^e(s) - \psi_l^e(s)|^2 ds\}^{1/2} = O_p((\log n/n)^{1/2})$;
*(iii) For* $1 \le l \le E$, $|\hat{\lambda}_l^e - \lambda_l^e| = O_p((\log n/n)^{1/2})$.

Theorem 2 characterizes the uniform convergence rates of $\hat{\Sigma}_{e,G}(s,s')$ and the associated eigenvalues and eigenfunctions. It can be regarded as an extension of Theorems 3.3–3.6 of Li and Hsing (2010), who established the strong uniform convergence rates of these estimates under a simpler model.

### 3.3. Asymptotics of inference procedure

In this subsection, we derive the asymptotic theory of a global test for testing linear hypotheses of $\beta(\cdot)$ and the theory for the SCB for each component of $\beta(\cdot)$. These are key tools for statistical inferences for the coefficient functions.

We first consider linear hypotheses for $\beta(s)$,

$$H_0 : C\beta(s) = \beta_0(s) \ \text{ for all } s \ \text{ vs. } \ H_1 : C\beta(s) \neq \beta_0(s) \ \text{ for some } \ s, \tag{3.2}$$

where $C$ is a $q \times p_x$ matrix with rank $q$, and $\beta_0(s)$ is a given $q \times 1$ vector of functions. We define a global test statistic $S_n$ as

$$S_n = \int_0^1 d(s)^T \left[ C \left\{ \sum_{i=1}^n X_i \hat{\Sigma}_{y_i,G}(s,s)^{-1} X_i^T \right\}^{-1} C^T \right]^{-1} d(s) ds, \tag{3.3}$$

where $d(s) = C\tilde{\beta}(s) - \text{bias}(C\tilde{\beta}(s)) - \beta_0(s)$. For simplicity and computational efficiency, we do not consider estimating the bias of $C\tilde{\beta}(s)$ because our simulation results show that it is negligible. It follows from Theorem 1 that under $H_0$, we have

$$\left[ C \left\{ \sum_{i=1}^n X_i \hat{\Sigma}_{y_i,G}(s,s)^{-1} X_i^T \right\}^{-1} C^T \right]^{-1/2} d(s) \Rightarrow G_C(s),$$

where $\Rightarrow$ denotes weak convergence and $G_C(\cdot)$ is a centered Gaussian process with covariance function $\{CQ^*(s,s)C^T\}^{-1/2} R(s,s') \{CQ^*(s',s')C^T\}^{-1/2}$. Thus, we can derive the asymptotic distribution of $S_n$ under the null hypothesis and its asymptotic power under local alternative hypotheses.

**Theorem 3.** *Under assumptions* (A.1)–(A.9), *if* $\log M = o(Mh_\beta)$ *and there exists* $\gamma_n \to \infty$, *with* $n^{1/2}\gamma_n^{1-q} = o(1)$ *and* $n^{-1/2}\gamma_n \log M = o(1)$ *for some* $q > 2$ *that satisfies* (A.7), *we have the following results:*
*(i)* $S_n \Rightarrow \int_0^1 G_C(s)^T G_C(s) ds$ *under the null hypothesis* $H_0$,

*(ii)* $P(S_n \geq S_{n,\alpha}|H_{1n}) \overset{n\to\infty}{\longrightarrow} 1$ *for a sequence of local alternatives* $H_{1n}$ : $C\beta(s)-\beta_0(s) = n^{-\tau/2}d(s)$, *where* $\tau$ *is any scalar in* $[0,1)$, $S_{n,\alpha}$ *is the upper* $100\alpha$ *percentile of* $S_n$ *under* $H_0$, *and* $0 < \int_{\mathcal{S}} ||d(s)||^2 ds < \infty$.

Theorem 3 can be regarded as a generalization of Theorem 7 of Zhang and Chen (2007) and Theorem 2 of Zhang (2011). The test statistic $S_n$ has a weighted $\chi^2$-type asymptotic distribution under $H_0$. Zhang and Chen (2007) (after Theorem 7) discuss the estimation of the null distribution of $S_n$ using a $\chi^2$-approximation and bootstrapping, which also applies to the case we consider here. It is easy to see that part (ii) still holds when the critical value $S_{n,\alpha}$ is replaced by some estimated critical value.

Next, we construct SCBs for the coefficient functions, which can then be used for statistical inferences for FMEM. For a given confidence level $\alpha$, we construct a SCB for each $\beta_l(s)$, $1 \leq l \leq p_x$, as follows:

$$P(\hat{\beta}_l^{L,\alpha}(s) < \beta_l(s) < \hat{\beta}_l^{U,\alpha}(s) \ \text{ for all } \ s \in [0,1]) = 1-\alpha, \tag{3.4}$$

where $\hat{\beta}_l^{L,\alpha}(s)$ and $\hat{\beta}_l^{U,\alpha}(s)$ are the lower and upper limits, respectively, of the SCB. Specifically, a $1-\alpha$ SCB for $\beta_l(s)$ is:

$$\left(\hat{\beta}_l(s) - \text{bias}(\hat{\beta}_l(s)) - \frac{C_l(\alpha)}{\sqrt{n}}, \quad \hat{\beta}_l(s) - \text{bias}(\hat{\beta}_l(s)) + \frac{C_l(\alpha)}{\sqrt{n}}\right), \tag{3.5}$$

where $C_l(\alpha)$ is the critical value of $\sup_{s\in\mathcal{S}} |G(s)|$ associated with $\hat{\beta}_l(s)$ in Theorem 1.

To carry out the inference procedure developed above, we approximate both $C_l(\alpha)$ and $S_{n,\alpha}$. Because 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, we use a wild bootstrap method to approximate the critical values of $S_n$. The wild bootstrap idea has been used by Zhu, Li and Kong (2012); details are presented in the Appendix. Let $G^{(q)}(\cdot)$ be the bootstrapped samples for $q = 1,\ldots,Q$, where $Q$ is the total number of wild bootstrap samples. The following theorem lays a foundation for the wild bootstrap method used to construct a SCB of $\beta(s)$ and to approximate the null distribution of $S_n$.

**Theorem 4.** *Under assumptions* (A.1)–(A.9) *and given data, the bootstrapped process* $G^{(q)}(s)$ *converges in distribution to* $\mathcal{G}(0,R)$, *which is defined in part (ii) of Theorem* 1, *as* $n \to \infty$.

## 4. Simulation Studies

In this section, we present four sets of simulations to examine the finite-

sample performance of the proposed estimation and inference procedures. In the first two simulations, we consider two competing methods: wavelet-based functional mixed models (WFMM) (Morris and Carroll (2006)) and functional additive mixed models (FAMM) (Scheipl, Staicu and Greven (2015)). All computations for these numerical examples were carried out on a computer with Windows 7, a 3.60 GHz quard-core Intel Core i7 CPU, and 16 GB DDR3 1,066 MHz memory. The computation time for FMEMs can be reduced further by using other computer languages, such as C++.

All simulated data sets were generated from the following model:

$$\begin{aligned} y_{ij}(s) &= x_{ij}^T\beta(s) + z_{ij}^T b_i(s) + e_{ij,G}(s) + e_{ij,L}(s), \\ b_i(s) &= \sum_{k=1}^{2} b_{ik}\psi_k^b(s), \qquad e_{ij,G}(s) = \sum_{k=1}^{2} e_{ijk}\psi_k^e(s), \end{aligned} \tag{4.1}$$

where $x_{ij} = (1, x_{ij,1}, x_{ij,2})^T$, $z_{ij} = (1, x_{ij,2})$, $b_{ik} \sim N(0, \lambda_k^b)$, $e_{ijk} \sim N(0, \lambda_k^e)$, and $e_{ij,L}(s) \sim N(0, \Sigma_{e,L})$, for $i = 1, \ldots, n$. Each subject was observed up to three times in this sample, among which 5%, 30%, and 65% have only one, two, and all three observations, respectively. We set $s_m = (m - 0.5)/M$. The first covariate $x_{ij,1}$ was simulated from $N(0, 1)$ and fixed across time for subject $i$. The second covariate $x_{ij,2}$ was assumed to vary over time, where the increments $x_{ij,2} - x_{i(j-1),2}$ were independently sampled from a uniform distribution on $[0, 1]$. Both covariates were standardized to have a zero mean and unit variance. Moreover, we set $\lambda_k^b = \lambda_k^e = 2^{1-k}$ for $k = 1, 2$, and $\Sigma_{e,L} = 0.01$. The functional coefficients and eigenfunctions were selected as

$$\begin{aligned} &\beta_1(s) = s^2, \quad \beta_2(s) = (1-s)^2, \quad \beta_3(s) = 4s(1-s) - 0.4, \\ &\psi_1^b(s)^T = (\psi_{11}^b(s), \psi_{12}^b(s)) = (\sin(2\pi s), \cos(2\pi s)), \quad \psi_1^e(s) = \sqrt{3}(2s-1), \\ &\psi_2^b(s)^T = (\psi_{21}^b(s), \psi_{22}^b(s)) = \left(\frac{1}{\sqrt{2}}, \sin(2\pi s)\right), \quad \psi_2^e(s) = \sqrt{5}(6s^2 - 6s + 1). \end{aligned}$$

We fitted an FMEM, WFMM, and FAMM to each simulated data set and calculated all unknown quantities. The average computation times per simulated data set with $n = 100$ and $M = 40$ for the FMEM, WFMM, and FAMM are, respectively, 19.6 seconds, 2.32 seconds, and 1.15 hours.

*Simulation 1.* The first simulation evaluates the performance of the estimates for $\beta_j(\cdot)$. We set $n = 100$ and $M = 40$ and 60, and then simulated 1,000 data sets from model (4.1), as described above. Table 1 summarizes the mean integrated absolute error (MIAE) and mean integrated squared error (MISE) of all estimated coefficient functions, based on 1,000 simulations. The results in Table 1 indicate

Table 1. Simulation 1. MIAE$\times 10^{-2}$ and MISE $\times 10^{-2}$ and their standard deviations $\times 10^{-2}$. MIAE denotes the mean integrated absolute error and MISE denotes the mean integrated square error. Standard deviations are shown in parentheses. For each case, 100 simulated data sets were used.

| Method | | MIAE$\times 10^{-2}$ | | | MISE$\times 10^{-2}$ | | |
|---|---|---|---|---|---|---|---|
| | $M$ | $\beta_1(\cdot)$ | $\beta_2(\cdot)$ | $\beta_3(\cdot)$ | $\beta_1(\cdot)$ | $\beta_2(\cdot)$ | $\beta_3(\cdot)$ |
| WFMM | 40 | 1.63 (0.73) | 1.67 (0.77) | 1.88 (0.78) | 0.04 (0.04) | 0.05 (0.04) | 0.06 (0.04) |
| | 60 | 1.37 (0.61) | 1.39 (0.63) | 1.55 (0.64) | 0.03 (0.03) | 0.03 (0.03) | 0.04 (0.03) |
| FAMM | 40 | 3.36 (2.11) | 2.84 (1.88) | 4.26 (3.27) | 0.23 (0.56) | 0.16 (0.35) | 0.38 (0.77) |
| | 60 | 3.03 (1.93) | 2.51 (1.58) | 3.95 (3.29) | 0.18 (0.36) | 0.13 (0.21) | 0.34 (0.95) |
| FMEM | 40 | 1.57 (0.72) | 1.44 (0.65) | 1.69 (0.70) | 0.04 (0.03) | 0.03 (0.03) | 0.05 (0.03) |
| | 60 | 1.29 (0.60) | 1.23 (0.55) | 1.37 (0.53) | 0.03 (0.03) | 0.03 (0.01) | 0.03 (0.03) |

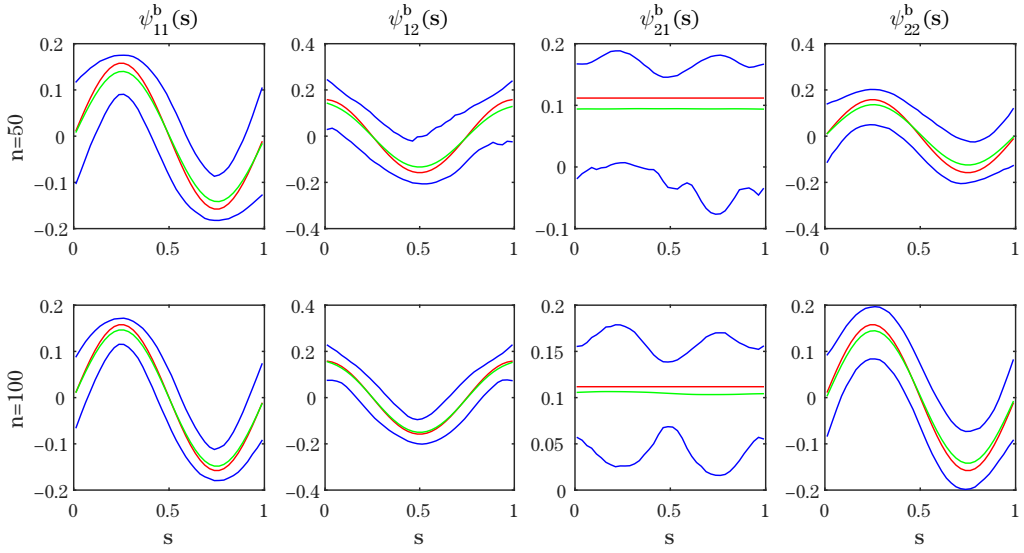


Figure 1. Simulations 2: the estimates of the first two eigenfunctions $\psi^b_{l,k}(\cdot)$, for $l, k = 1, 2$, and their pointwise confidence intervals. The red solid, green dashed, and blue solid, curves are, respectively, the true eigenfunctions, the pointwise means, and the pointwise $5th$ and $95th$ percentiles of the estimated eigenfunctions based on 1,000 replications.

satisfactory performance of our estimators because all MIAE and MISE values are quite small. As expected, the errors all decrease as the number of grid points increases. Moreover, the FMEM outperforms the WFMM and FAMM in terms of both the MIAE and the MISE. However, this comparison may be unfair to the WFMM, because it is designed for spiky data, not intrinsically smooth functional data.

*Simulation 2.* The second simulation evaluates the accuracy of the estima-

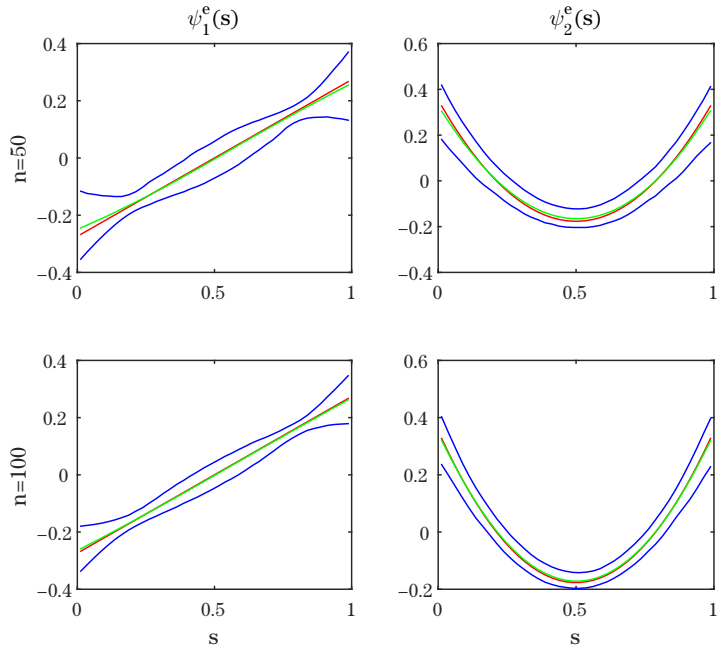


Figure 2. Simulations 2: the estimates of the first two eigenfunctions $\psi_k^e, k = 1, 2$, and their pointwise confidence intervals. The red solid, green dashed, and blue solid, curves are, respectively, the true eigenfunctions, the pointwise means and the pointwise $5th$ and $95th$ percentiles of the estimated eigenfunctions based on 1,000 replications.

tors of the eigenvalues and eigenfunctions of the covariance functions $\Sigma_b(.,.)$, $\Sigma_{e,G}(.,.)$, and $\Sigma_{e,L}$. We used the same parameter values as those in Simulation 1. We set $c = 0.1$ and $n = 50$ and 100, and generated 1,000 data sets for each combination. The accuracy of all of the estimators improves with the sample size. The estimated eigenfunctions are plotted in Figures 1 and 2, in which the mean and pointwise $5th$ and $95th$ percentiles of the estimated functions are plotted with the true eigenfunctions. Figure 3 shows the boxplots for the estimates of the eigenvalues and $\sigma^2$, which are quite close to their true values.

*Simulation 3.* The third simulation evaluates the type-I error rate and the power of the global test statistics, $S_n$. We are interested in testing $H_0 : \beta_3(s) = 0$, for all $s$, against $H_1 : \beta_3(s) \neq 0$, for some $s$. All parameters in the FMEM were specified as above, except that $\beta_3(s)$ was set as $4cs(1 - s) - 0.4c$. Here we first set $c = 0$ to assess the type-I error rate of $S_n$, and then set $c = 0.04, 0.06, 0.08$, and 0.1 to examine the power of $S_n$ at different effect sizes. Furthermore, we set $n = 50$ and 100 and used 1,000 replications to estimate the rejection rate of

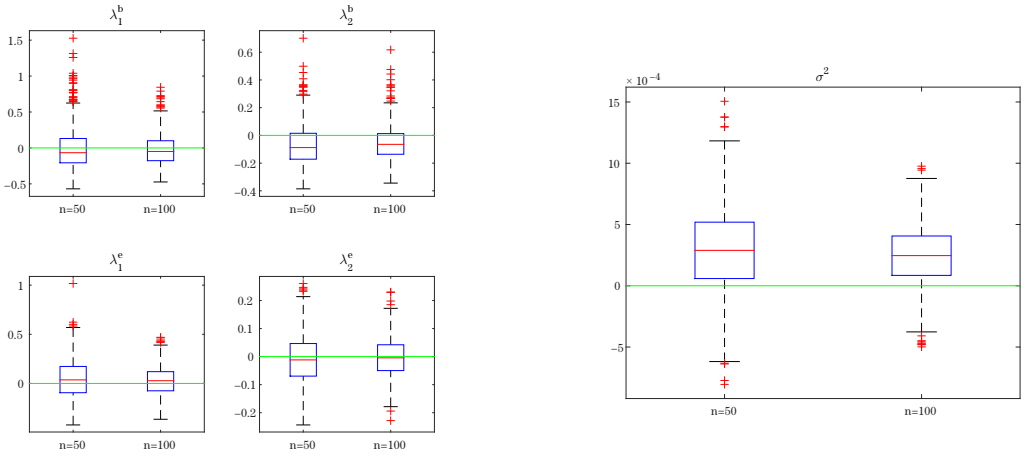


Figure 3. Simulation 2: boxplots of the differences between the estimated eigenvalues $\hat{\lambda}_k^b$ and $\hat{\lambda}_k^e$, for $k = 1, 2$, and their true values (left panel), and those between the estimated $\sigma^2$ and its true values (right panel) based on 1,000 replications.

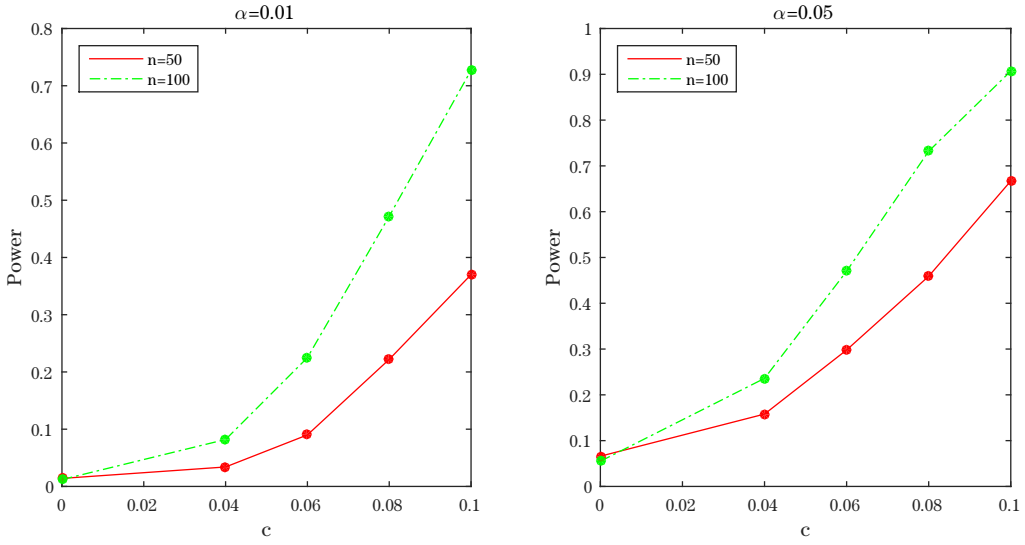


Figure 4. Simulation 3: Power curves as functions of $c$. The rejection rates of $S_n$ using the wild bootstrap method are calculated at five different values of the effect size $c$ ($c = 0, 0.04, 0.06, 0.08$ and $0.1$) for two sample sizes ($n = 50$ and $100$) at the 0.01 (a) and 0.05 (b) significance levels based on 1,000 replications.

$S_n$. The $p$-value of $S_n$ was approximated using the wild bootstrap method with $Q = 500$ bootstrap samples.

Fig. 4 presents the rejection rates of $S_n$ across all effect sizes at the two

Table 2. Autism spectrum disorder data analysis: demographic information for participants.

| Visit | Number of subjects | Age: mean(std) (years) | Age: range (years) |
|---|---|---|---|
| 1 | 58 | 10.53 (5.96) | [0, 18] |
| 2 | 148 | 12.25 (4.62) | [0, 21] |
| 3 | 160 | 12.29 (5.14) | [1, 22] |
| 4 | 19 | 1.84 (1.42) | [1, 6] |
| 5 | 7 | 1.57 (0.79) | [1, 3] |
| 6 | 10 | 2.70 (0.67) | [2, 4] |
| 7 | 6 | 3.17 (0.75) | [2, 4] |
| 8 | 5 | 3.40 (1.14) | [2, 5] |
| 9 | 3 | 3.67 (1.15) | [3, 5] |
| Gender | Male/Female | | 126/127 |

significance levels $\alpha = 0.05$ and 0.01. The type-I error rates are well maintained at the two significance levels for $n = 100$. Specifically, at $\alpha = 0.05$ (or 0.01), the type I error rates of $S_n$ are 0.066 (or 0.014) for $n = 50$ and 0.055 (or 0.012) for $n = 100$. As expected, the statistical power for rejecting the null hypothesis increases with the sample size, the effect size $c$, and the significance level.

## 5. Data Analysis

The data set was taken from the national database for autism research (NDAR) (`http://ndar.nih.gov/`), an NIH-funded research data repository that aims to accelerate progress in autism spectrum disorder (ASD) research through data sharing, data harmonization, and the reporting of research results. A total of 416 MRI scans are selected for 253 normal children (126 males and 127 females), following the standard protocol. Table 3 contains demographic information and the distribution of scan availability.

The diffusion tensor imaging (DTI) data were processed by two key steps, including a weighted least squares estimation method (Basser, Mattiello and LeBihan (1994)) to construct the diffusion tensors, and a pipeline for tract-based spatial statistics (TBSS) (Smith et al. (2006)) to register DTIs from multiple subjects to create a mean image and a mean skeleton. Specifically, maps of fractional anisotropy (FA) were computed for all subjects from the DTI after Eddy current correction and automatic brain extraction using the FMRIB software library. FA maps were then fed into the TBSS tool, which is also part of the FSL. In the TBSS analysis, the FA data for all subjects were aligned into a common space by a nonlinear registration method. Then, the the mean FA images were

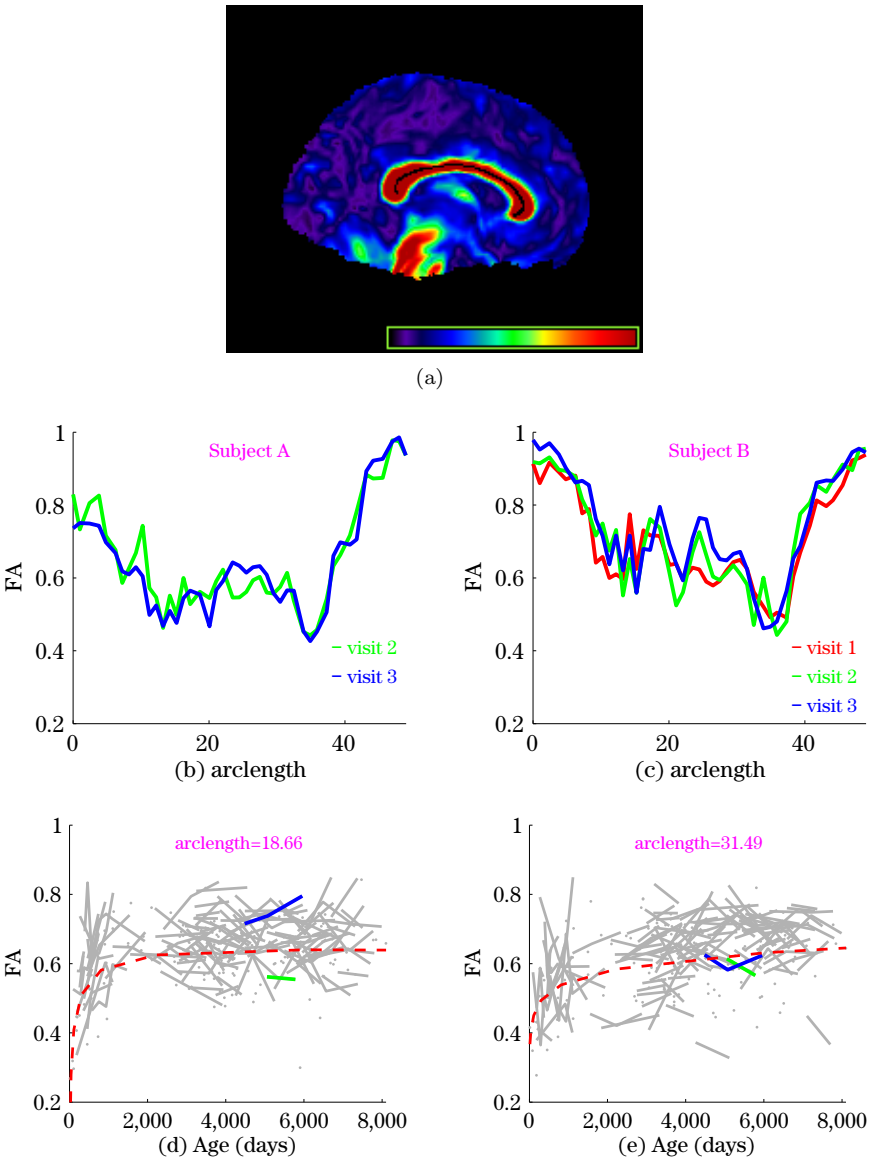


Figure 5. Data analysis: (a) 3D visualization of the corpus callosum in the sagittal view, with the FA skeleton template overlaid. (b) and (c) FAs along the corpus callosum obtained from two selected subjects A (b) and B (c) with two or three visits. Different visits for the same subjects are indicated by color. (d) and (e) FA values varying by age at selected locations: arclength = 18.66 (d) and arclength = 31.49 (e) along the corpus callosum for all 253 subjects, with green and blue lines corresponding to subjects A and B, respectively. Red dashed lines represent the fitted lines for the male group.

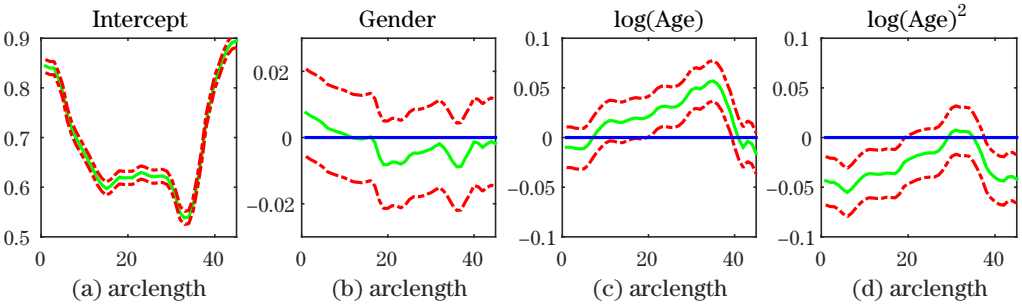


Figure 6. 95% simultaneous confidence bands for the coefficient functions. The solid curves are the estimated coefficient functions and the dashed curves are the 95% simultaneous confidence bands. The thin horizontal line is the line crossing the origin (0, 0).

created and thinned to obtain a mean FA skeleton, which represents the centers of all white matter tracts common to the group. Subsequently, each subject's aligned FA data sets were projected onto this skeleton. Although several DTI fiber tracts were tracked, we chose to focus on the corpus callosum (see Fig. 5 (a)) to illustrate the applicability of our method in assessing the effects of covariates of interest, such as patient age and gender. In this case, there are $M = 45$ grid points along each fiber tract. The FA values were extracted at each grid point multiple times (one to nine times) along the selected fiber tracts for all 253 infants.

The goal of the data analysis is to delineate the development of skeleton diffusion properties over time. We fitted FMEM (2.1) and (2.2) with $x_i = (1, \text{Gender}, \log(\text{Age}), \{\log(\text{Age})\}^2)^T$ and $z_i = (1, \log(\text{Age}))^T$ to the selected FA tracts obtained from all 253 subjects. The coefficient functions associated with $\log(\text{Age})$ and $\{\log(\text{Age})\}^2$ were included to detect age effects in the FA changes. In addition, as shown in Fig. 5, there are random subject-to-subject variations in the FA measures at each grid point along this tract, as well as in the age effect on FA measures. We included random intercept and age effects in the model to account for the inter-subject variations.

We applied the FMEM, WFMM, and FAMM to this data set and estimated all unknown quantities, but only discuss the results based on the FMEM here. The results for the WFMM and FAMM are provided in the Supplementary Material. The computation times for the FMEM, WFMM, and FAMM are, respectively, 55.8 seconds, 7.9 seconds, and 6.078 hours.

For the FMEM, we constructed the estimated functional coefficients of $\beta(s)$

and their 95% SCBs, along with the global test statistic $S_n$, to test for the significance of the gender and age effects on the FA values. The $p$-value of $S_n$ was approximated using the resampling method with $Q = 1,000$ replications. Figure 6 presents the estimated coefficient functions corresponding to the intercept, gender, log(Age), and $\{\log(\text{Age})\}^2$, along with their 95% SCBs. The intercept function describes the overall trend of the FA along the corpus callosum. In general, the central regions of the corpus callosum show smaller FA values, whereas the peripheral regions show larger FA values. In Figure 6, the SCB contains the horizontal line crossing $(0, 0)$ for the gender effect, whereas the horizontal line is out of the 95% SCB for the age effect, indicating a significant age effect. This agrees with our analysis results based on $S_n$ for the gender and age effects. We obtained $p$-values of 0.215 and $< 0.0001$ for the gender and age effects, respectively, indicating a significant age effect but no gender effect.

## Supplementary materials

available in the attached file include the proofs of Lemmas 1–13, Theorems 1–3, and Corollary 1.

## Acknowledgments

The research of Dr. Zhu was supported by NSF grants SES-1357666 and DMS-1407655, NIH grants MH086633 and MH116527, a grant from the Cancer Prevention Research Institute of Texas, and the endowed Bao-Shan Jing Professorship in Diagnostic Imaging. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We would like to thank Drs. Morris and Herrick for helping with WFMM.

## Appendix

### Wild bootstrap method for critical values of $S_n$

We have shown that the asymptotic distribution of $S_n$ is very complicated hence 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 (2.1) and (2.2) under the null hypothesis $H_0$, which yields $\hat{\beta}^*(s_m)$, $\hat{u}^*_{ij,G}(s_m)$ and $\hat{\epsilon}^*_{ij}(s_m) = y_{ij}(s_m) - x_{ij}^T\hat{\beta}^*(s_m) - \hat{u}^*_{ij,G}(s_m)$ for all $i, j$ and $m = 1, \ldots, M$.

Step 2. Generate a random sample $\tau_i^{(q)}$ and $\tau_{ij}(s_m)^{(q)}$ from a $N(0,1)$ generator for all $i, j$ and $m = 1, \ldots, M$ and then construct

$$\hat{y}_{ij}(s_m)^{(q)} = x_{ij}^T \hat{\beta}^*(s_m) + \tau_i^{(q)} \hat{u}_{ij,G}^*(s_m) + \tau_{ij}(s_m)^{(q)} \hat{\epsilon}_{ij}^*(s_m).$$

Then, based on $\hat{y}_{ij}(s_m)^{(q)}$, we recalculate $\hat{\beta}(s)^{(q)}$, and $d(s)^{(q)} = C\hat{\beta}(s)^{(q)} - \beta_0(s)$. Subsequently, we compute

$$S_n^{(q)} = n \int_0^1 d(s)^{(q)T} \left[ C \left\{ \sum_{i=1}^n X_i \hat{\Sigma}_{y_i,G}(s,s)^{-1} X_i^T \right\}^{-1} C^T \right]^{-1} d(s)^{(q)} ds.$$

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

### Wild bootstrap methods for simultaneous confidence bands of $\beta(\cdot)$

Although there are several methods of determining $C_l(\alpha)$ including random field theory (Worsley et al. (2004)), we develop an efficient resampling method to approximate $C_l(\alpha)$ as follows (Kosorok (2003)).

- We calculate $\hat{r}_i(s_m) = y_i(s_m) - X_i^T \tilde{\beta}(s_m)$ for all $i, j$, and $m$.

- For $q = 1, \ldots, Q$, we independently simulate $\{\tau_i^{(q)} : i = 1, \ldots, n\}$ from $N(0,1)$ and calculate a stochastic process $G(s)^{(q)}$ given by

$$\sqrt{n}[I_{p_x} \otimes (1,0)]\text{vec}(\Sigma(s,h_1)^{-1}$$
$$\sum_{i=1}^n \tau_i^{(q)} \sum_{m=1}^M K_h(s_m - s)\{s_h(s_m - s) \otimes X_i\} \hat{\Sigma}_{y_i,G}(s,s)^{-1} \hat{r}_i(s_m)).$$

- We calculate $\sup_{s \in [0,1]} |e_l G(s)^{(q)}|$ for all $q$, where $e_l$ is a $p_x \times 1$ vector with the $l$-th element 1 and 0 otherwise, and use their $1 - \alpha$ empirical percentile to estimate $C_l(\alpha)$.

## References

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 Series B (Statistical Methodology)* **103**, 247–254.

Cao, G., Yang, L. and Todem, D. (2012). Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics* **24**, 359–377.

Cederbaum, J., Pouplier, M., Hoole, P. and Greven, S. (2016). Functional linear mixed models for irregularly or sparsely sampled data. *Statistical Modelling* **16**, 67–88.

Chen, K. and Müller, H.-G. (2012). Modeling repeated functional observations. *Journal of the American Statistical Association* **107**, 1599–1609.

Di, C., Crainiceanu, C. M., Caffo, B. S. and Punjabi, N. M. (2009). Multilevel functional principal component analysis. *Annals of Applied Statistics* **3**, 458–488.

Di, C., Crainiceanu, C. M. and Jank, W. (2014). Multilevel sparse functional principal component analysis. *Stat* **3**, 126–143.

Diggle, P., Heagerty, P., Liang, K. Y. and Zeger, S. (2002). *Analysis of Longitudinal Data (2nd Edition.)*. Oxford University Press: New York.

Evans, A. C. and Group, B. D. C. (2006). The NIH MRI study of normal brain development. *NeuroImage* **30**, 184–202.

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

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its Interface* **1**, 179–195.

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2004). *Applied Longitudinal Analysis*. Wiley, New York.

Greven, S., Crainiceanu, S., Caffo, B. S. and Reich, D. (2010). Longitudinal functional principal component analysis. *Electronic Journal of Statistics* **4**, 1022–1054.

Guo, W. (2002). Functional mixed effects models. *Biometrics* **58**, 121–128.

Horvath, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. New York, N. Y.: Springer.

Kosorok, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *Journal of Multivariate Analysis* **84**, 299–318.

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.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association* **96**, 1045–1056.

Meyer, M. J., Coull, B. A., Versace, F., Cinciripini, P. and Morris, J. S. (2015). Bayesian function-on-function regression for multilevel functional data. *Biometrics* **71**, 563–574.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society. Series B. (Statistical Methodology)* **68**, 179–199.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., Trojanowski, J. Q., Toga, A. W. and Beckett, L. (2005). Ways toward an early diagnosis in alzheimer's disease: The alzheimer's disease neuroimaging initiative (ADNI). *Alzheimer's & Dementia* **1**, 55–66.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd Edition. Springer-Verlag: New York.

Reiss, P. T., Huang, L., Chen, H. and Colcombe, S. (2014). Varying-smoother models for functional responses. *arXiv preprint arXiv:1412.0778*.

Scheipl, F., Staicu, A. and Greven, S. (2015), Functional additive mixed models. *Journal of Computational and Graphic Statistics* **24**, 477–501.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. John Wiley & Sons, New York, N.Y.

Shi, J. Q. and Choi, T. (2011). *Gaussian Process Regression Analysis for Functional Data.* Chapman & Hall/CRC.

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M., Matthews, P. and Behrens, T. E. (2006). Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage* **31**, 1487–1505.

Staicu, A. M., Lahiri, S. and Carroll, R. J. (2015). Significance tests for functional data with complex dependence structure. *Journal of Statistical Planning and Inference* **156**, 1–13.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.

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

Wu, H. and Zhang, J. (2002). Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association* **97**, 883–889.

Wu, H. and Zhang, J. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. Hoboken, John Wiley & Sons, Inc, New Jersey.

Xiao, L., Zipunnikov, V., Ruppert, D. and Crainiceanu, C. (2016). Fast covariance estimation for high-dimensional functional data. *Statistics and Computing* **26**, 409–421.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

Yuan, Y., Gilmore, J. H., Geng, X., Styner, M., Chen, K., Wang, J. L. and Zhu, H. (2014). Fmem: Functional mixed effects modeling for the analysis of longitudinal white matter tract data. *NeuroImage* **84**, 753–764.

Zhang, J. (2011). Statistical inferences for linear models with functional responses. *Statistica Sinica* **21**, 1431–1451.

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

Zhou, L., Huang, J. Z., Martinez, J. G., Maity, A., Baladandayuthapani, V. and Carroll., R. J. (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of American Statistical Association* **105**, 390–400.

Zhu, H., Brown, P. and Morris, J. (2011). Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Asssociation* **106**, 1167–1179.

Zhu, H., Li, R. and Kong, L. (2012). Multivariate varying coefficient model for functional responses. *The Annals of Statistics* **40**, 2634–2666.

Zipunnikov, V., Greven, S., Shou, H., Caffo, B., Reich, D. S. and Crainiceanu, C. (2014). Longitudinal high-dimensional principal components analysis with application to diffusion tensor imaging of multiple sclerosis. *Annals of Applied Statistics* **8**, 2175–2202.

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.
E-mail: hzhu5@mdanderson.org
Department of Statistics, University of Pittsburgh, PA 15260, USA.
E-mail: khchen@pitt.edu
Statistics & Decision Sciences, Janssen R&D, LLC, Shanghai 200233, China.
E-mail: xluo27@its.jnj.com
Takeda Pharmaceuticals, 35 Landsdowne St, Cambridge, MA, 02139, USA.
E-mail: yy9615@hotmail.com
Department of Statistics, University of California at Davis CA 95616, USA.
E-mail: janelwang@ucdavis.edu

(Received May 2017; accepted February 2018)