

A functional nonlinear mixed effects modeling framework for longitudinal functional responses

Linglong Kong¹, Xinchao Luo², Jinhan Xie^{3,1},
Lixing Zhu⁴, and Hongtu Zhu⁵

¹*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada*
e-mail: lkong@ualberta.ca; jinhanxie@163.com

²*School of Finance and Statistics, East China Normal University, Shanghai 200062, China*
e-mail: fallenstar0909@gmail.com

³*Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming 650091, People's Republic of China*
e-mail: jinhanxie@163.com

⁴*Department of Statistics, Beijing Normal University at Zhuhai, Zhuhai, China*
e-mail: lzhu@bnu.edu.cn

⁵*Department of Biostatistics and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, U.S.A.*
e-mail: htzhu@email.unc.edu

Abstract: In this paper, we introduce a functional nonlinear mixed effects modeling framework designed to quantify the random, nonlinear relationship between individual spatiotemporal functional trajectories and longitudinal responses. Our proposed framework accounts for within-individual variability through a spatiotemporal process. We detail an estimation method for determining fixed and random effect functions and spatiotemporal covariance operators and establish their asymptotic properties, including uniform consistency and weak convergence. We also develop global linear hypothesis tests and bootstrap-based simultaneous confidence bands for fixed effect functions. To assess the finite-sample performance of our method, we perform a numerical analysis using both simulated and real-world datasets. Our results demonstrate that the proposed model class is significantly more flexible and effective in detecting functional fixed effects compared to existing nonlinear mixed effects models. We apply our approach to an autism research database to investigate the impact of age and spatial dynamics on fractional anisotropy along the corpus callosum white matter fiber skeleton.

MSC2020 subject classifications: 62G05, 60J05, 62G08, 62G20.

Keywords and phrases: Functional random effect, functional response, global test statistic, longitudinal response, simultaneous confidence band, weak convergence.

Received May 2023.

1. Introduction

There is a growing need for analyzing extensive, curated longitudinal data obtained from large-scale neuroimaging studies [2, 36, 45, 38, 43], such as the Alzheimer’s Disease Neuroimaging Initiative [27], the Baby Connectome project [16], the Adolescent Brain Cognitive Development Study [5], and the National Database for Autism Research [15]. Gaining insights into the correlations between clinical covariates, changes in brain function across spatial locations and time, and the progression of neurodegenerative and neuropsychiatric diseases will contribute to enhanced diagnostic and treatment methods for these conditions [19, 9, 30]. Furthermore, there is considerable interest in utilizing longitudinal neuroimaging data to accurately depict the changes and development of brain cortical and subcortical structures (e.g., hippocampus) over time and among different groups [2, 22]. Developing and implementing innovative statistical methods for such data is a critical prerequisite for advancing our understanding of these complex relationships.

To formalize this setting, let’s consider longitudinal functional data from n different subjects. Typically, these data are either observed or registered to a large number of locations in a common compact set \mathcal{S} , across multiple time points $\{t_{i,j} : i = 1, \dots, n; j = 1, \dots, n_i\}$, where n_i represents the total number of longitudinal measurements for the i th subject and $t_{i,j}$ denotes the j th measurement time point for the i th subject. This dataset consists of observed time-varying clinical variables $x_{i,j} = x(t_{i,j}) \in \mathbb{R}^q$ and longitudinal functional responses $y_{i,j}(s_m) = y(t_{i,j}, s_m)$, with grid points s_m in \mathcal{S} . Throughout this paper, functional data are measured densely on \mathcal{S} and we focus on a fixed number of time points with sparse longitudinal data across the set $\{t_{i,j} : i = 1, \dots, n; j = 1, \dots, n_i\}$, that is, $\max_{i \leq n} n_i < n_0 < \infty$. Spatial and temporal correlation refers to correlations across s and t , respectively. For simplicity in notation, we assume that $\mathcal{S} = [0, 1]$ and $\mathcal{S}_0 = \{s_m : 0 = s_1 < \dots < s_M = 1\} \subset \mathcal{S} = [0, 1]$. Our results can be readily extended to two and three dimensions.

In many situations, fitting nonlinear models becomes particularly challenging, as their applications inherently encompass more intricate patterns of change [21, 3]. Considering the intricacies inherent in certain longitudinal data, articulating a growth model that precisely captures the longitudinal process can be demanding. While a quadratic growth model might provide a decent approximation for some nonlinear changes, it might falter in representing processes that stabilize over time. In such cases, an exponential growth function could provide a more accurate depiction of the process. As data complexity increases, there’s a foreseeable need for more sophisticated models. For example, the double exponential function is commonly used to model human population growth [35]; The three-parameter logistic function is well employed to describe the S-shaped pattern that is observed in the growth curve of soybean data [8]; The Gompertz function has been used to characterize longitudinal white matter development during early childhood [9, 19]. Indeed, nonlinear models often require fewer parameters than competing linear models, such as a polynomial, leading to a more parsimonious description of the data. We refer to [28] for more comprehensive

applications on nonlinear mixed effects functions.

In this paper, we develop a functional nonlinear mixed effects modeling framework to elucidate dynamic changes in longitudinal functional data and to characterize spatiotemporal variation and nonlinear associations with other covariates of interest. Specifically, we consider the class of functional nonlinear random effects models of the form:

$$y_{i,j}(s) = f(\phi_i(s), x_{i,j}) + \varepsilon_{i,j}(s) = f(\beta(s) + b_i(s), x_{i,j}) + \varepsilon_{i,j}(s), \quad (1)$$

where $f : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$ is a differentiable link function encompassing various forms such as exponential, logistic, double exponential functions, among others, $\phi_i : \mathcal{S} \rightarrow \mathbb{R}^p$ is a vector of random functions, and $\varepsilon_{i,j} : \mathcal{S} \rightarrow \mathbb{R}$ is a random error term. In particular, when f takes an identical link function, the model (1) simplifies to the functional linear mixed effects models as described by [41]. We further specify $\phi_i(s) = \beta(s) + b_i(s)$, where $\beta(s) = (\beta_1(s), \dots, \beta_p(s))^\top$ and $b_i = (b_{i,1}(s), \dots, b_{i,p}(s))^\top$ are $p \times 1$ vectors of fixed and random effect functions, respectively. The $b_i(s)$ and $\varepsilon_{i,j}$ are assumed to be independent and identical copies of $G(0, \Sigma^b)$ and $G(0, \Sigma^\varepsilon)$, respectively, where $G(\mu, \Sigma)$ denotes a Gaussian process with mean function $\mu : \mathcal{S} \rightarrow \mathbb{R}$ and covariance function $\Sigma : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$. Furthermore, we assume $\Sigma^\varepsilon(s, s') = 1(s = s')\sigma_\varepsilon^2(s)$ and express the covariance matrix of $(\Sigma^b)(s, s)$ in the form of a relative precision factor, $\Delta(s)$, which is any matrix that satisfies $(\Sigma^b)^{-1}(s, s) = \sigma^{-2}(s)\Delta(s)^\top \Delta(s)$.

Model (1) can be viewed as a novel extension of numerous popular nonlinear random effects models [4, 24]. Like most other functional mixed effects models [6, 45, 13, 26, 14], the proposed model includes both fixed and random effect functions. The former characterizes varying associations between the longitudinal functional response and covariates of interest, while the latter captures medium-to-long-range spatiotemporal covariance and short-range covariance structures associated with certain repeated experimental factor levels. The concept of functional mixed effects models for correlated functional data was first introduced in [14]. Following this, [26] and subsequent work have expanded upon general functional mixed effects models with multiple levels of random effect functions as well as curve-to-curve deviations. Recently, several functional mixed effects modeling methods for longitudinal functional data have been developed; see [25, 41, 7]. To the best of our knowledge, little work has been done on the theoretical properties of statistical estimators or related inference procedures for this class of nonlinear models.

In this paper, we develop the functional nonlinear mixed effects models and related tools for statistical inference mentioned above. Unlike most existing approaches, the model (1) and the framework proposed in this paper allow both random and fixed effect functions to enter in a nonlinear fashion. This setup is motivated by real-world data analytic needs: nonlinear link functions are relevant to many disciplines such as forestry, agriculture, ecology, biomedicine, and pharmacokinetics [18, 32, 24], and are commonly applied to the analysis of nonlinear growth data [19, 9, 30]. Our main contributions can be summarized in three aspects. First, our proposed model offers greater flexibility by including

a nonlinear link function, which results in higher statistical power in detecting fixed effect significance based on the proposed global hypothesis test. Second, our formulation explicitly incorporates spatial smoothness into the estimators for the fixed effect functions and spatiotemporal covariance operators, allowing for a more accurate representation of the underlying data structure. Third, we comprehensively establish the theoretical properties of the proposed estimators, including the weak convergence of the functional fixed effect estimator, the uniform convergence rate of the estimator for the spatiotemporal covariance operator, the asymptotic distribution of the test statistic in the proposed global linear hypothesis test of fixed effect significance, and an asymptotic simultaneous confidence band for each fixed effect.

The rest of this paper is organized as follows. In Section 2, we present the proposed estimation and inference procedures. In Section 3, we study the asymptotic properties of the proposed estimators and the test statistic. We evaluate the performance of the proposed procedure through extensive simulation studies in Section 4.1 and a real data application in Section 4.2. Technical details are deferred to the Appendix.

2. Methodology

2.1. Estimation procedure

Our estimation procedure for model (1) consists of three key steps.

- First, starting from consistent point estimators $\hat{\beta}(s_m)$ of $\beta(s_m)$ for each $s_m \in \mathcal{S}_0$, we construct a consistent functional estimator $\tilde{\beta}(s)$ for the fixed effect function $\beta(s)$.
- Second, estimate the individual random-effect functions $b_i(s)$.
- Third, obtain estimates of the covariance operators for each component and their spectral decompositions.

Each of these three steps is explained in detail below.

In the first step, we estimate the fixed-effect functions in $\beta(s)$ at each observed spatial grid point using a consistent estimator. For each $s_m \in \mathcal{S}_0$, model (1) is the nonlinear mixed effects model $y_{i,j}(s_m) = f(\beta(s_m) + b_i(s_m), x_{i,j}) + \varepsilon_{i,j}(s_m)$, where $b_i(s_m) \sim N(0, \Sigma^b(s_m, s_m))$ and $\varepsilon_{i,j}(s_m) \sim N(0, \sigma_\varepsilon^2(s_m))$. In this step, we use the maximum likelihood estimator $\hat{\beta}(s_m)$ for $\beta(s_m)$. Alternatively, one may consider other consistent estimators such as the two-stage or Lindstrom–Bates estimators [24, 10].

Assuming certain smoothness restrictions on $\beta(s)$, given as Assumption 8 in Section 3, we calculate a kernel density estimator of $\beta(s)$ using the collection of point estimators $\{\hat{\beta}(s_m) : s_m \in \mathcal{S}_0\}$. Specifically, given a normalized kernel function \tilde{K}_h with bandwidth parameter h , the kernel density estimator of $\beta(s)$ is $\tilde{\beta}(s) = \tilde{\beta}(s, h_1) = \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) \hat{\beta}(s_m)$ for any $s \in \mathcal{S}$ [37]. The bandwidth parameter h_1 can be set via leave-one-out cross validation as a minimizer of $\text{CV}(h_1) = \sum_{m=1}^M \|\hat{\beta}(s_m) - \tilde{\beta}^{(-m)}(s_m, h_1)\|^2 / M$ over h_1 , where $\|\cdot\|^2$ denotes the

l_2 norm and $\tilde{\beta}^{(-m_0)}(s_{m_0}, h_1)$ is the kernel density estimator, after the exclusion of the $m = m_0$ term, evaluated at $s = s_{m_0}$.

In the second step, we employ a local linear smoother [12] to estimate the individual $b_i(s)$ values. The Taylor expansion of $b_i(s_m)$ around s gives us the approximation $b_i(s_m) \approx b_i(s) + \dot{b}_i(s)(s_m - s) = B_i(s)Z(s_m - s)$, where $Z(s_m - s) = (1, s_m - s)^\top \in \mathbb{R}^2$ and $B_i(s) = (b_i(s), \dot{b}_i(s)) \in \mathbb{R}^{p \times 2}$ with $\dot{b}_i(s) = (\dot{b}_{i,1}(s), \dots, \dot{b}_{i,p}(s))^\top$, and $\dot{b}_{i,l}(s) = \partial b_{i,l}(s) / \partial s$ for $l = 1, \dots, p$. To estimate $B_i(s)$ for each $i = 1, \dots, n$ and a given s , we minimize the weighted nonlinear least squares objective function [42]

$$S_M(B_i(s)) = \sum_{j=1}^{n_i} \sum_{m=1}^M \left\{ y_{i,j}(s_m) - f(\hat{\beta}(s_m) + B_i(s)Z(s_m - s), x_{i,j}) \right\}^2 \tilde{K}_{h_2}(s_m - s)$$

over $B_i(s)$. To find the minimizer $\hat{B}_i(s) = (\hat{b}_i(s), \hat{\dot{b}}_i(s))$, we propose an iterative procedure.

Define $a \otimes a = aa^\top$ for any vector a and let $C \otimes D$ denote the Kronecker product of two matrices C and D . For a matrix C of size $M_1 \times M_2$, we use $\text{vec}(C)$ to denote its vectorization, which is a vector of size $M_1 M_2$ with elements $(c_{1,1}, \dots, c_{M_1,1}, c_{2,1}, \dots, c_{1,M_2}, \dots, c_{M_1,M_2})^\top$. Given a current estimate $B_i(s)^{(r)}$ of $B_i(s)$, we further define $W_{i,j}(s_m)^{(r)}$ as

$$y_{i,j}(s_m) - f(\hat{\beta}(s_m) + B_i(s)^{(r)}Z(s_m - s), x_{i,j}) + A_{i,j}(B_i(s)^{(r)})^\top \text{vec}(B_i(s)^{(r)}),$$

where $A_{i,j}(B_i(s)) = \partial f(\hat{\beta}(s_m) + B_i(s)Z(s_m - s), x_{i,j}) / \partial \text{vec}(B_i(s))$. An updated estimate $B_i(s)^{(r+1)}$ can be obtained as the minimizer of

$$\sum_{j=1}^{n_i} \sum_{m=1}^M \left\{ W_{i,j}(s_m)^{(r)} - A_{i,j}(B_i(s)^{(r)})^\top \text{vec}(B_i(s)) \right\}^2 \tilde{K}_{h_2}(s_m - s)$$

over $B_i(s)$. The bandwidth parameter h_2 can be determined through cross-validation. An estimate of $b_i(s)$, as the first column of $B_i^{(r+1)}(s)$, is given by $b_i(s)^{(r+1)} = \hat{B}_i(s)^{(r+1)}(1, 0)^\top$. The sequence $b_i(s)^{(r)}$ converges to $\hat{b}_i(s)$ as $r \rightarrow \infty$.

In the third step, we estimate the covariance operator Σ^b using an empirical estimator with $\hat{\Sigma}^b(s, s') = (\hat{\Sigma}_{l,l'}^b(s, s'))_{l,l'} = N^{-1} \sum_{i=1}^n n_i \hat{b}_i(s) \hat{b}_i(s')^\top$, where $N = \sum_{i=1}^n n_i$. Asymptotic properties of the estimators in the second and third steps rely on a Karhunen–Loeve expansion for Σ^b , presented in Section 3.

Remark 2.1. *To extend our estimation procedure from $\mathcal{S} = [0, 1]$ to a two-dimensional compact set Ω in \mathbb{R}^2 , model (1) has the following formula:*

$$y_{i,j}(\tilde{\mathbf{s}}) = f(\phi_i(\tilde{\mathbf{s}}), x_{i,j}) + \varepsilon_{i,j}(\tilde{\mathbf{s}}) = f(\beta(\tilde{\mathbf{s}}) + b_i(\tilde{\mathbf{s}}), x_{i,j}) + \varepsilon_{i,j}(\tilde{\mathbf{s}}),$$

where $\tilde{\mathbf{s}} = (\tilde{s}_1, \tilde{s}_2)^\top$ denotes a point in Ω . For this model, we only need to modify the above first and second estimation steps by changing $b_i(s)$ and $s - s_m$ into

2×1 vectors. Except this, we can also adopt some existing works to approximate the functions $\beta(\tilde{\mathbf{s}})$ and $b_i(\tilde{\mathbf{s}})$ by splines which are piecewise polynomial bivariate functions over a two-dimensional triangulated domain; see [23].

2.2. Inference procedure

We next propose an inference procedure consisting of four steps as follows. We define $\hat{f}_{i,j}(s) = f(\hat{\beta}(s) + \hat{b}_i(s), x_{i,j})$, $\hat{X}_{i,j}(s) = \partial \hat{f}_{i,j}(s) / \partial \hat{\beta}(s)$, and $\hat{Z}_{i,j}(s) = \partial \hat{f}_{i,j}(s) / \partial \hat{b}_i(s)$.

Step (I) is to approximate model (1) as follows:

$$y_{i,j}(s_m) \approx \hat{f}_{i,j}(s_m) + \hat{X}_{i,j}(s_m)^\top \{\beta(s_m) - \hat{\beta}(s_m)\} + \hat{Z}_{i,j}(s_m)^\top \{b_i(s_m) - \hat{b}_i(s_m)\} + \varepsilon_{i,j}(s_m).$$

Thus, let $\hat{\omega}_{i,j}(s) = y_{i,j}(s) - \hat{f}_{i,j}(s) + \hat{X}_{i,j}(s)^\top \hat{\beta}(s) + \hat{Z}_{i,j}(s)^\top \hat{b}_i(s)$, we have

$$\hat{\omega}_i(s_m) = (\hat{\omega}_{i,1}(s_m), \dots, \hat{\omega}_{i,n_i}(s_m))^\top = \hat{X}_i(s_m) \beta(s_m) + e_i(s_m), \quad (2)$$

where $e_i(s_m) = (\hat{Z}_{i,1}(s_m)^\top b_i(s_m) + \varepsilon_{i,1}(s_m), \dots, \hat{Z}_{i,j}(s_m)^\top b_i(s_m) + \varepsilon_{i,n_i}(s_m))^\top$ and $\hat{X}_i(s_m) = (\hat{X}_{i,1}(s_m), \dots, \hat{X}_{i,n_i}(s_m))^\top$. Furthermore, $e_i(s_m)$ can be approximated as a $N(0, \Sigma_i(s_m))$ distribution with $\Sigma_i(s) = \sigma_\varepsilon^2(s) I_{n_i} + \hat{Z}_i(s) \Sigma^b(s, s) \hat{Z}_i(s)^\top$ for $s \in \mathcal{S}$, in which $\hat{Z}_i(s_m) = (\hat{Z}_{i,1}(s_m), \dots, \hat{Z}_{i,n_i}(s_m))^\top$.

Step (II) is to develop a global test statistic, denoted as S_N , for the global linear hypotheses

$$H_0 : R\beta(s) = b_0(s) \text{ for all } s \in \mathcal{S} \text{ versus } H_1 : R\beta(s) \neq b_0(s) \text{ for some } s \in \mathcal{S},$$

where $R \in \mathbb{R}^{p_0 \times p}$ has rank p_0 , and $b_0 : \mathcal{S} \rightarrow \mathbb{R}^{p_0}$. We first use $d(s) = R[\tilde{\beta}(s) - \text{BIAS}\{\tilde{\beta}(s)\}] - b_0(s)$ to approximate $R\beta(s) - b_0(s)$. As shown in Section 3, the bias term $\text{BIAS}\{\tilde{\beta}(s)\}$ can be approximated as $\{0.5\ddot{\beta}(s) + \dot{\beta}(s)\dot{\pi}(s)\pi(s)^{-1}\}h_1^2\mu_2(K)$, where $\mu_2(K) = \int K(t)t^2 dt$. Second, we approximate the covariance matrix of $\tilde{\beta}(s)$ to be $\hat{\Sigma}(s, s) \approx H_N(s)^{-1} \left\{ \sum_{i=1}^n \hat{S}_i(s)^\top \hat{S}_i(s) \right\} H_N(s)^{-1}$, where $H_N(s) = \sum_{i=1}^n \hat{X}_i(s)^\top \Sigma_i(s)^{-1} \hat{X}_i(s)$ and $\hat{S}_i(s) = S_i(s, \hat{\beta}(s)) = \hat{X}_i(s)^\top \Sigma_i(s)^{-1} \{\hat{\omega}_i(s) - \hat{X}_i(s) \hat{\beta}(s)\}$. Therefore, let $\pi_m = s_{m+1} - s_m$, the global test statistic S_N can be approximated as

$$\int_0^1 d(s)^\top \{R\hat{\Sigma}(s, s)R^\top\}^{-1} d(s) ds \approx \sum_{m=1}^{M-1} d(s_m)^\top \{R\hat{\Sigma}(s_m, s_m)R^\top\}^{-1} d(s_m) \pi_m.$$

Step (III) is to propose a score-based bootstrap method to obtain the p -value for the proposed test [17]. The complicated form of the asymptotic distribution of S_N makes it difficult to directly approximate the percentiles of S_N under H_0 . We proceed as follows. First, fit model (1) under H_0 and estimate $\hat{\beta}^*(s)$. For each $g = 1, \dots, G$, independently generate $\tau_i^{(g)} \sim N(0, 1)$ and construct

$S_i^*(s_m)^{(g)} = \widehat{X}_i(s_m)^\top \Sigma_i(s_m)^{-1} \{\widehat{\omega}_i(s_m) - \widehat{X}_i(s_m) \widehat{\beta}^*(s_m)\} \tau_i^{(g)}$ and $S_i(s_m)^{(g)} = \widehat{X}_i(s_m)^\top \Sigma_i(s_m)^{-1} \{\widehat{\omega}_i(s_m) - \widehat{X}_i(s_m) \widehat{\beta}(s_m)\} \tau_i^{(g)}$ for $i = 1, \dots, n$. Second, calculate $\widehat{\Sigma}(s_m, s_m)^{(g)} = H_N(s_m)^{-1} \{\sum_{i=1}^n S_i(s_m)^{(g)T} S_i(s_m)^{(g)}\} H_N(s_m)^{-1}$ and compute

$$S_N^{(g)} = \sum_{m=1}^{M-1} \widetilde{S}(s_m)^\top \{R \widehat{\Sigma}(s_m, s_m)^{(g)} R^\top\}^{-1} \widetilde{S}(s_m) \pi_m,$$

yielding $\{S_N^{(g)} : g = 1, \dots, G\}$, where $\widetilde{S}(s_m) = R H_N(s_m)^{-1} \sum_{i=1}^n S_i^*(s_m)^{(g)}$. Finally, calculate the p -value for the proposed test as $p = G^{-1} \sum_{g=1}^G 1\{S_N^{(g)} > S_N\}$ and reject the null hypothesis H_0 if p is smaller than some given significance level α .

Step (IV) is to construct a $100(1 - \alpha)\%$ simultaneous confidence band for $\beta_l(s)$ such that $\text{pr}(\widetilde{\beta}_l^{L,\alpha}(s) < \beta_l(s) < \widetilde{\beta}_l^{U,\alpha}(s) \text{ for all } s \in \mathcal{S}) = 1 - \alpha$, where $\widetilde{\beta}_l^{L,\alpha}$ and $\widetilde{\beta}_l^{U,\alpha}$ are lower and upper limit functions, respectively. Following the procedure of [42], we construct simultaneous confidence bands as

$$\left(\widetilde{\beta}_l(s) - \text{BIAS}\{\widetilde{\beta}_l(s)\} - n^{-1/2} C_l(\alpha), \widetilde{\beta}_l(s) - \text{BIAS}\{\widetilde{\beta}_l(s)\} + n^{-1/2} C_l(\alpha) \right) \quad (3)$$

over s , where $C_l(\alpha)$ is approximated via wild bootstrap. More specifically, consider the stochastic process $G(s)^{(g)} = n^{1/2} \sum_{m=1}^M \widetilde{K}_{h_1}(s_m - s) \{H_N(s_m)^{-1} \sum_{i=1}^n S_i(s_m) \tau_i^{(g)}\}$. The empirical $(1 - \alpha)$ -level quantile of $\sup_{s \in \mathcal{S}} |e_l G(s)^{(g)}|$ is an estimator of $C_l(\alpha)$, where $e_l \in \mathbb{R}^p$ has its l th element equal to $1\{i = l\}$.

3. Asymptotic properties

To further explore the asymptotic properties of the proposed estimators and the test statistic S_N , we present some relevant notation and regularity assumptions. First, we consider the spectral decomposition of $\Sigma_{l,l}^b$ for $l = 1, \dots, p$ [29]. Assuming that $\Sigma_{l,l}^b$ is continuous on \mathcal{S}^2 , and that $\Sigma_{l,l}^b$ admits a decomposition of the form $\Sigma_{l,l}^b(s, s') = \sum_{k=1}^\infty \lambda_{k,l} \psi_{k,l}(s) \psi_{k,l}(s')$, where $\{(\lambda_{k,l}, \psi_{k,l}(s)) : k = 1, \dots\}$ is the set of ordered eigenvalue-eigenfunction pairs, i.e., $\lambda_{k,l} \geq \lambda_{k+1,l}$. By Mercer's theorem, the eigenfunctions form an orthonormal basis for the space of square-integrable functions on \mathcal{S} . Consequently, $b_{i,l}$ admits the Karhunen–Loeve expansion $b_{i,l}(s) = \sum_{k=1}^\infty \xi_{i,k,l} \psi_{k,l}(s)$, where $\xi_{i,k,l} = \int_0^1 b_{i,l}(s) \psi_{k,l}(s) ds$ is the k th functional principal component score of $b_{i,l}$. For a fixed i , the scores $\{\xi_{i,k,l}\}_{k,l}$ are uncorrelated random variables with mean zero and variance $\lambda_{k,l}$.

Next, we define $\theta_0(s) = \{\beta(s), \sigma_\epsilon^2(s), \Sigma^b(s, s)\}$ and let $\theta(s)$ be the corresponding collection of estimates. We define $\Theta = \prod_{s \in \mathcal{S}} \mathcal{B}$ as the parameter space for $\theta(s)_{s \in \mathcal{S}}$, where \mathcal{B} is a compact set in \mathbb{R}^d with $d = (p^2 + 3p + 2)/2$. We define $y_i = (y_{i,1}, \dots, y_{i,n_i})^\top$ and let $P_\theta(i, s) = \text{pr}(y_i(s) \mid x_i, \theta(s)) = \int P(y_i(s) \mid x_i, \beta(s), b_i(s), \sigma^2(s)) P(b_i(s) \mid \Delta(s)) db_i(s)$ be the marginal density of $y_i(s)$ conditional on x_i . Similarly, we define $P_\theta(s) = \text{pr}(y(s) \mid x, \theta(s)) = \int P(y(s) \mid$

$x, \beta(s), b(s), \sigma^2(s)P(b(s) | \Delta(s))db(s)$. At each s , the model likelihood [28] can be expressed as

$$L_{\theta(s)} = \frac{|\Delta(s)|^n}{\{2\pi\sigma^2(s)\}^{(N+nq)/2}} \prod_{i=1}^n \int \exp \left\{ -\frac{\|y_i - f(\phi_i(s), x_i)\|^2 + \|\Delta(s)b_i(s)\|^2}{2\sigma^2(s)} \right\} db_i(s).$$

We make the following assumptions to facilitate the technical details throughout the paper. The current set of assumptions simplifies the proof even though some of them might be weakened.

Assumption 3.1. For some $\kappa > 4$ and for all $s_m \in \mathcal{S}_0$, $\sup_{s_m \in \mathcal{S}_0} E\{|\varepsilon_{i,j}(s_m)|^\kappa\} < \infty$.

Assumption 3.2. The covariate vector $x_{i,j} = (x_{i,j,1}, \dots, x_{i,j,q})^\top \in \mathbb{R}^q$ may or may not be time-dependent with $\|x_{i,j}\|_\infty \leq c_0 < \infty$ for some universal positive constant c_0 .

Assumption 3.3. For all $s \in \mathcal{S}$, suppose that $\theta_0(s) \in \mathcal{B}$ is the unique maximizer of $E\{\log P_\theta(s) | x, \theta(s)\}$, where the expectation is taken with respect to the true distribution of $y(s)$ given x .

Assumption 3.4. The set Θ is compact and, for all $s \in \mathcal{S}$ and $\theta(s) \in \mathcal{B}$, suppose that $\log\{P_\theta(s)\}$ is twice continuously differentiable on Θ . For all $j, k = 1, \dots, p$, $E\{|\partial_j \log\{P_\theta(s)\}|\} < \infty$ and $\log\{P_\theta(s)\}$, $|\partial_j \log\{P_\theta(s)\}|$, and $|\partial_j \partial_k \log\{P_\theta(s)\}|$ are dominated by an integral function $G(y(s), x)$ such that $E\{\sup_{s \in \mathcal{S}} |G(y(s), x)|^r\} < \infty$ for some $r \geq 1$, where $\partial_j = \partial/\partial\theta_j(s)$.

Assumption 3.5. The set of grid points \mathcal{S}_0 is randomly generated from a density function $\pi(s)$ that is positive for all $s \in \mathcal{S}$ and has bounded support and a continuous second-order derivative.

Assumption 3.6. Assume that each component of the functions in the classes $\{I_{\beta(s)}^{-1} \partial \log P_\theta(s) / \partial \beta(s) : s \in \mathcal{S}\}$ and $\{b(s)b(t)^\top : (s, t) \in \mathcal{S}^2\}$ is P -Donsker, where $I_{\beta(s)}$ denotes an information matrix.

Assumption 3.7. The kernel function K is a continuous, symmetric, bounded density function with $[-1, 1]$ as its support and $\int_{-1}^1 K(\mu)d\mu = 1$, $\int_{-1}^1 \mu K(\mu)d\mu = 0$, and $\mu_2(K) = \int_{-1}^1 \mu^2 K(\mu)d\mu < \infty$.

Assumption 3.8. Each component of $\beta(s)$ has a finite, continuous second-order derivative on \mathcal{S} .

Assumption 3.9. As both M and n tend to infinity, $\log M \leq n^\nu$ for some $0 < \nu < 1$, $h_1 \rightarrow 0$, $h_2 \rightarrow 0$, $Mh_1 \rightarrow \infty$, $Mh_2 \rightarrow \infty$, and $h_1^{-1} |\log h_1|^{1-2/q_1} \leq M^{1-2/q_1}$, where $q_1 \in (2, 4)$.

Assumption 3.10. The link function f satisfies $E\{\sup_{\theta \in \Theta} \|f^{(k_1)}\|_2\} < \infty$ for derivative orders $k_1 = 1, 2$, where $f^{(k_1)}(\phi, x) = \partial^{k_1} f(\phi, x) / \partial \phi^{k_1}$. Similarly, for $k_2 = 1, 2, 3$, $E\{\|b_i(s)\|_2^{r_1}\} < \infty$ and $E\{\sup_{s \in \mathcal{S}} \|b_i(s)^{(k_2)}\|_2^{r_2}\} < \infty$ for some $r_1, r_2 \in (2, \infty)$.

Assumption 3.11. For each l , there is a finite, positive integer E_l such that $\lambda_{1,l} > \dots > \lambda_{E_l,l} \geq \lambda_{E_l+1,l} \geq \dots \geq 0$.

Remark 3.1. Assumption 1 requires a uniform bound on a high-order moment of $\varepsilon_{i,j}(s_m)$ for all grid points $s_m \in \mathcal{S}_0$. Assumption 2 can be relaxed for equality of the distribution of covariate vectors $x_{i,j}$. Assumptions 3 and 4 are generalizations of standard conditions used to guarantee first-order asymptotic properties of M-estimators [33]. Assumption 3 is an identifiability condition, while Assumption 4 is a uniform smoothness and integrability condition. Assumption 5 is a weak condition on the random grid points [42]. In many neuroimaging applications, M is often much larger than n so, for sufficiently large M , a regular grid of voxels is well-approximated by voxels generated by a uniform distribution in a compact subset of a Euclidean space. Assumption 6 is used to avoid smoothness conditions on the sample path, while Assumption 7 is a common assumption for kernel density methods. Assumption 8 is a standard smoothness condition on β . The portion of Assumption 9 pertaining to bandwidth is similar to that in [42]. Assumption 10 requires a uniform bound on the derivative of the nonlinear link and random effect functions. Finally, Assumption 11 regards the multiplicity of the largest E_l eigenvalues for each l and is only required for investigating asymptotic properties of the corresponding eigenfunctions.

As shown by [33], for each grid point $s_m \in \mathcal{S}_0$ and under Assumptions 1-3, the maximum likelihood estimator has consistency and asymptotic normality properties. Additionally, [11] calculated the asymptotic variance of the maximum likelihood estimator in a one-parameter balanced exponential model and established that the estimator attains its absolute lower bound. Our work builds on these asymptotic results by investigating the properties of $\hat{\beta}$ and the corresponding kernel density estimator $\hat{\beta}$.

Theorem 3.1. Under Assumptions 1-9, (i) $\sup_{s_m \in \mathcal{S}_0} \|\hat{\beta}(s_m) - \beta(s_m)\|_2 = O_p(n^{-1/2})$ and (ii) the process $\{n^{1/2}[\tilde{\beta}(s) - \beta(s) - \{0.5\tilde{\beta}(s) + \dot{\beta}(s)\dot{\pi}(s)\pi(s)^{-1}\}h_1^2 \mu_2(K)] : s \in \mathcal{S}\}$ converges weakly to a centered Gaussian process with covariance function Σ . Precise definitions of $\hat{\beta}$, $\tilde{\beta}$, Σ , π , and $\dot{\pi}$ are given in the accompanying supplementary materials.

Remark 3.2. The asymptotic bias of $\tilde{\beta}(s)$ is of the order h_1^2 as in nonparametric regression. Despite the complex form of the asymptotic conditional covariance matrix of $\tilde{\beta}(s)$ due to within-curve dependence, it converges to the covariance matrix of the maximum likelihood estimator with order n^{-1} as $M \rightarrow \infty$. Selecting an optimal bandwidth h_1 for model (1) can be a challenging task, however, any bandwidth h_1 that satisfies the conditions of $h_1 \rightarrow 0$ and $Mh_1 \rightarrow \infty$ can ensure the weak convergence of $\tilde{\beta}(s) : s \in \mathcal{S}$.

We continue by investigating the asymptotic properties of $\hat{b}_i(s)$. We define $v_0(K) = \int K^2(\mu)d\mu$, $K^*(u) = \int K(\mu)K(\mu + u)d\mu$, $X = \{x_{i,j} : i = 1, \dots, n; j = 1, \dots, n_i\}$, and $b = \{b_1, \dots, b_n\}$. Further define $f_{i,j}(s, s') = f(\beta(s) + b_i(s'), x_{i,j})$, $f_{i,j}^{k_1, k_2}(s, s') = \partial^{k_1+k_2} f_{i,j}(s, s') / \partial \beta(s)^{k_1} \partial b_i(s')^{k_2}$, and $f_{i,j}^{k_1, k_2^*}(s, s') = \partial^{k_1+k_2} f(\beta(s) + B_i(s')Z(s - s'), x_{i,j}) / \partial \beta(s)^{k_1} \partial B_i(s')^{k_2}$

Theorem 3.2. Under Assumptions 1–10, (i) $\hat{b}_i(s)$ is uniformly consistent, with

$$\sup_{s \in \mathcal{S}} \|\hat{b}_i(s) - b_i(s)\|_2 = O_p \left[h_2^2 + n^{-1/2} + \{|\log h_2| / (Mh_2)\}^{1/2} \right].$$

(ii) The asymptotic bias, $\text{bias}\{\hat{b}_i(s) \mid \mathcal{S}_0, X, b(s)\}$ and covariance, $\text{cov}\{\hat{b}_i(s), \hat{b}_i(s') \mid \mathcal{S}_0, X, b(s)\}$, are, respectively, given by

$$\begin{aligned} & \mathcal{F}_i(s) \sum_{j=1}^{n_i} \{0.5 f_{i,j}^{0,1}(s, s)^\top \ddot{b}_i(s) h_2^2 + f_{i,j}^{1,0}(s, s)^\top \bar{O}_p(n^{-1/2})\} f_{i,j*}^{0,1}(s, s) \{1 + o_p(1)\}, \\ & (Mh_2)^{-1} K^*((s - s')/h) \pi(s')^{-1} \mathcal{F}_i(s) \sum_{j=1}^{n_i} f_{i,j*}^{0,1}(s, s) f_{i,j*}^{0,1}(s', s')^\top \mathcal{F}_i(s') O_p(1), \end{aligned}$$

where $\mathcal{F}_i(s) = \{\sum_{j=1}^{n_i} f_{i,j}^{0,1*}(s, s)^2\}^{-1}$. A full definition of $f_{i,j*}^{0,1}(s, s)$ is given in the accompanying supplementary materials.

(iii) The mean integrated square error, conditional on (X, b) , is

$$\begin{aligned} & \int E[\{\hat{b}_i(s) - b_i(s)\}^\otimes \mid X, b] \pi(s) ds \\ &= (Mh_2)^{-1} v_0(K) O_p(1) \int \mathcal{F}_i(s) \sum_{j=1}^{n_i} f_{i,j}^{0,1*}(s, s)^\otimes \mathcal{F}_i(s) ds + \{1 + o_p(1)\} \cdot \\ & \int \left[\mathcal{F}_i(s) \sum_{j=1}^{n_i} \{0.5 f_{i,j}^{0,1}(s, s)^\top \ddot{b}_i(s) h_2^2 + f_{i,j}^{1,0}(s, s)^\top \bar{O}_p(n^{-1/2})\} f_{i,j*}^{0,1}(s, s) \right]^\otimes \pi(s) ds. \end{aligned}$$

(iv) The optimal bandwidth h_2^* for minimizing the mean integrated square error in (iii) is $O(M^{-1/5})$ if $M^{4/5} = o(n)$ and $O(nM^{-1})$ if $n = o(M^{4/5})$.

Remark 3.3. Theorem 3.2 provides insights into the statistical properties of the smoothed random effect estimates $\hat{b}_i(s)$. Theorem 3.2 (i) establishes the uniform consistency of the kernel density estimator $\hat{b}_i(s)$. Given individual random effects $b_i(s)$, Theorem 3.2 (ii) shows that the bias of $\hat{b}_i(s)$ is made up of two terms, each of order $n^{-1/2}$, which are due to the maximum likelihood estimation and random effects smoothing. The mean integrated square error of $\hat{b}_i(s)$ in Theorem 3.2 (iii) can be decomposed into an $O_p(n^{-1})$ term introduced by the maximum likelihood estimation of β and an $O_p(h_2^4 + M^{-1}h_2^{-1})$ term introduced by the reconstruction of $b_i(s)$. The optimal bandwidth is of the same order as that for local polynomial kernel smoothers. Under the optimal bandwidth described in Theorem 3.2 (iv), the mean integrated square error can be of order n^{-1} or $M^{-4/5}$, depending on the relationship between n and M .

The following theorem states that as the sample size n grows, the estimator $\hat{\Sigma}^b$ converges in probability to the true covariance matrix Σ^b . Additionally, the eigenfunctions $\hat{\psi}_{l,k}$ and eigenvalues $\hat{\lambda}_{l,k}$ of $\hat{\Sigma}^b$ converge in probability to the corresponding eigenfunctions $\psi_{l,k}$ and eigenvalues $\lambda_{l,k}$ of Σ^b . These results hold under appropriate regularity conditions on Σ^b and the estimation procedure.

Theorem 3.3. (i) Under Assumptions 1–10, with h_2^* denoting the optimal bandwidth described in Remark 3,

$$\sup_{(s,s') \in \mathcal{S}^2} |\widehat{\Sigma}^b(s, s') - \Sigma^b(s, s')| = O_p((h_2^*)^2 + (Mh_2^*)^{-1} + (\log n/n)^{1/2}).$$

(ii) Under Assumptions 1–11, for $k = 1, 2, \dots$,

$$\left\{ \int_0^1 [\widehat{\psi}_{l,k}(s) - \psi_{l,k}(s)]^2 ds \right\}^{1/2} = O_p\{(h_2^*)^2 + (M\widehat{h}_2^*)^{-1} + (\log n/n)^{1/2}\}$$

and $|\widehat{\lambda}_{l,k} - \lambda_{l,k}| = O_p\{(h_2^*)^2 + (Mh_2^*)^{-1} + (\log n/n)^{1/2}\}.$

Remark 3.4. Theorem 3.3 characterizes the uniform weak convergence rate of $\widehat{\Sigma}^b$, $\widehat{\psi}_k$, and $\widehat{\lambda}_k$, similar to Theorem 3 in [42]. In contrast, none of the rates in this result include an $O_p(h_1^2)$ term since only an estimate of β is used, without any smoothing. For further discussion, see [42].

Let \Rightarrow denote weak convergence of a sequence of stochastic process and $G(s)$ a centered Gaussian process indexed by $s \in \mathcal{S}$. Theorem 3.1 establishes that $n^{1/2}[\widehat{\beta}_l(s) - \beta(s) - \text{BIAS}\{\widehat{\beta}_l(s)\}] \Rightarrow G(s)$. Consequently, we can construct a simultaneous confidence band for each fixed effect function, namely, each component of β . For $C_l(\alpha)$ satisfying $\text{pr}(\sup_{s \in \mathcal{S}} |G(s)| \leq C_l(\alpha)) = 1 - \alpha$, where $\sup_{s \in \mathcal{S}} |n^{1/2}[\widehat{\beta}_l(s) - \beta(s) - \text{BIAS}\{\widehat{\beta}_l(s)\}]| \Rightarrow \sup_{s \in \mathcal{S}} |G(s)|$, the confidence band given in (3) is a $1 - \alpha$ simultaneous confidence band for β_l .

Theorem 3.4. Under Assumptions 1–10, we have the following results:

(i) $S_N \Rightarrow \int_0^1 G(s)^\top G(s) ds$ under the null hypothesis H_0 ,

(ii) $P(S_N \geq S_{N,\alpha} | H_{1n}) \rightarrow 1$ as $n \rightarrow \infty$ for a sequence of local alternatives $H_{1n} : R\beta(s) - b_0(s) = n^{-\kappa/2}d(s)$, where κ is any scalar in $[0, 1)$, $S_{N,\alpha}$ is the upper 100α percentile of S_N under H_0 , and $0 < \int_{\mathcal{S}} \|d(s)\|^2 ds < \infty$.

Remark 3.5. Theorem 7 of [40] and Theorem 2 of [39] characterize the asymptotic distribution of S_N : the accompanying discussion is also valid for Theorem 3.4.

As a final asymptotic result, we verify that the bootstrapped process $\{G(s)^{(g)} : s \in \mathcal{S}\}$ can be used to approximate the null distribution of S_N .

Theorem 3.5. Under Assumptions 1–9, $G(s)^{(g)} \Rightarrow G(s)$, conditional on the data.

Remark 3.6. Theorem 3.5, which follows immediately from Theorem 5 of [42], validates the use of the bootstrapped process $G(s)^{(g)}$. Again, $\beta(s)$ is estimated solely via maximum likelihood, so the bias correction is unnecessary, resulting in a substantial reduction in computational load.

4. Numerical studies

4.1. Simulation results

To provide an initial demonstration of the finite performance of estimation and inference procedures for the model (1), we conduct a series of Monte Carlo simulation studies. In this subsection, we consider simulated longitudinal functional data and clinical variables from n independent samples. Let n_i be the total number of longitudinal measurements for the i th sample and $t_{i,j}$ be the j th measurement time point for the i th subject, so $j = 1, \dots, n_i$, $i = 1, \dots, n$. Specifically, for the i th sample at time $t_{i,j}$, $j = 1, \dots, n_i$, $i = 1, \dots, n$, we generate data from the following models:

1. $y_{i,j}(s) = 1 + 0.1\{x_{i,j,1}\phi_{1,i}(s) + x_{i,j,2}\phi_{2,i}(s) + 2\}^3 + \varepsilon_{i,j}(s)$;
2. $y_{i,j}(s) = 5\{\sin(x_{i,j,1}\phi_{1,i}(s) + x_{i,j,2}\phi_{2,i}(s))\}^3 + \varepsilon_{i,j}(s)$;
3. $y_{i,j}(s) = 1 - 2.5 \exp\{-\exp\{x_{i,j,1}\phi_{1,i}(s) + x_{i,j,2}\phi_{2,i}(s)\}\} + \varepsilon_{i,j}(s)$,

where $\phi_{l,i}(s) = \beta_l(s) + b_{l,i}(s)$ for $l = 1, 2$. Notice that each sample is observed n_i times. Model 3 is the double exponential function, which is commonly employed in population dynamics to model human population growth; see [35]. Let s_m be equidistant points in $[0, 1]$ with $s_1 = 0$ and $s_M = 1$, $\varepsilon_{i,j}(s) \sim N(0, 0.1)$, and $(x_{i,j,1}, x_{i,j,2})^\top \sim N((0, 0)^\top, \Sigma)$, where $\Sigma = (\sigma_{j,k})_{j,k \in \mathbb{R}^{p \times p}}$ and $\sigma_{j,k} = 0.3^{|k-j|}$. The functional random effects are defined as $b_i(s) = \sin(2\pi s)N((0, 0)^\top, 0.1\Sigma) + \cos(2\pi s)N((0, 0)^\top, 0.2\Sigma)$, and the functional fixed effects are given by $\beta_1(s) = cs^2$ and $\beta_2(s) = (1-s)^2$. Setting $c = 0$ results in $\beta_1(s) = 0$ for all s . We employ two simulation studies to examine the proposed estimation and inference procedures. In each simulation study, we consider the Epanechnikov kernel function, i.e., $\tilde{K}(u) = 0.75(1-u^2)I(|u| \leq 1)$, where $I(\cdot)$ is an indicator function. The codes for this paper are written in R and can be obtained at <https://github.com/statisticalxjh/FNMEM>.

In the first simulation study, we evaluate the proposed S_N -based test for the null hypothesis $H_0 : \beta_1(s) = 0$ across all s , as opposed to the alternative hypothesis $H_1 : \beta_1(s) \neq 0$ for at least one s . Our primary focus is the test's rejection rate, specifically its Type I error rate when $c = 0$ and its power for values of $c = 0.05, 0.1, 0.15$, and 0.2 . Additionally, we explore varying sample sizes with $n = 50$ and 100 for each c and set $M = 25$ and $n_i = 5$. We then compute the rejection rate for the score bootstrap method using $G = 500$ in each scenario, applying the significance levels $\alpha = 0.05$ and 0.01 , and conducting 200 replications. In this simulation study, we also evaluate an alternative method: the standard nonlinear mixed-effects model which only uses longitudinal data, without any functional components, denoted by NMEM [28];

Figure 1 displays power curves at the specified significance levels α and sample sizes n for both a functional nonlinear mixed effects model and a nonlinear mixed effects model. For both models, Σ is estimated using the asymptotic covariance matrix without any smoothing. We observe that the Type I error rates derived from the score bootstrap are well-controlled at both significance levels. It is also evident that the NMEM is much less powerful compared to the

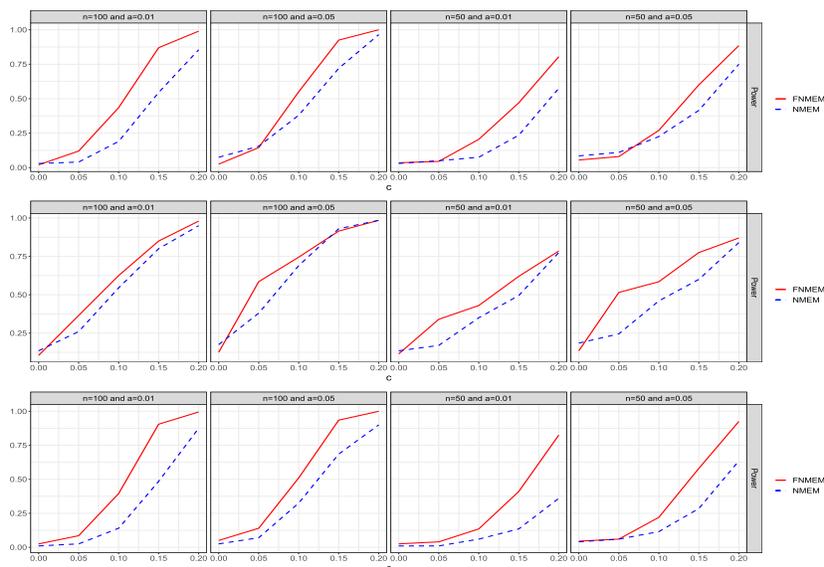


FIG 1. Power curves illustrating the results of the first simulation study, calculated at five values of c for a functional nonlinear mixed effects model (FNMEM) and nonlinear mixed effects model (NMEM). The S_N -based rejection rates are obtained using the wild bootstrap method. Upper panel: Model 1; Middle panel: Model 2; Lower panel: Model 3. The sample size n and significance level α are noted in each subplot.

FNMEM. The analogy here is the standard NMEM with just longitudinal data (i.e. no functional components), since FNMEM is an extension of NMEM. As anticipated, the power increases with a larger sample size.

In the second simulation study, we investigate the finite-sample performance of the proposed simultaneous confidence bands. Employing the same data generation procedure as before, we consider different sample sizes $n = 50$ and 100 and fix $c = 1$, $n_i = 5$, and $M = 25, 50$, and 75 . For each of the 200 replications, we use a wild bootstrap with $G = 500$ to compute simultaneous confidence bands for each component of β . In Table 1, we present empirical coverage probabilities for two α values. We observe that increasing the number of grid points M enhances the coverage probability, and the coverage probabilities are close to the corresponding confidence level $1 - \alpha$. The Monte Carlo errors are approximately 0.015 for $\alpha = 0.05$, given by the formula $(0.95 \times 0.05/200)^{1/2}$. Figures depicting typical 95% and 99% simultaneous confidence bands for $n = 50$, $M = 75$, and $n = 100$, $M = 75$ are presented in Figures 2-4.

4.2. Real data analysis

In the second numerical study, we examine a real-world dataset acquired from the National Database for Autism Research (NDAR) (<http://ndar.nih.gov/>),

TABLE 1

Empirical coverage probabilities of the simultaneous $100(1 - \alpha)\%$ confidence bands for each component of β in the second simulation, based on 200 simulated data sets. The number of uniform grid points M and the significance level α are noted for each result.

Model 1					
M	β_1	β_2	β_1	β_2	
	$n = 50, \alpha = 0.05$		$n = 50, \alpha = 0.01$		
25	0.940	0.965	0.995	0.985	
50	0.970	0.955	0.990	0.980	
75	0.960	0.965	0.990	0.995	
	$n = 100, \alpha = 0.05$		$n = 100, \alpha = 0.01$		
25	0.955	0.940	0.970	0.980	
50	0.950	0.975	0.985	0.995	
75	0.960	0.965	0.995	0.995	
Model 2					
	$n = 50, \alpha = 0.05$		$n = 50, \alpha = 0.01$		
25	0.955	0.965	0.990	0.990	
50	0.945	0.970	0.985	0.990	
75	0.970	0.940	0.990	0.985	
	$n = 100, \alpha = 0.05$		$n = 100, \alpha = 0.01$		
25	0.945	0.935	0.975	0.970	
50	0.930	0.960	0.985	0.995	
75	0.940	0.940	0.995	0.980	
Model 3					
	$n = 50, \alpha = 0.05$		$n = 50, \alpha = 0.01$		
25	0.920	0.950	0.980	0.990	
50	0.960	0.970	0.985	0.990	
75	0.950	0.950	0.985	1.000	
	$n = 100, \alpha = 0.05$		$n = 100, \alpha = 0.01$		
25	0.940	0.960	0.980	0.980	
50	0.960	0.965	0.995	0.985	
75	0.960	0.960	0.995	0.990	

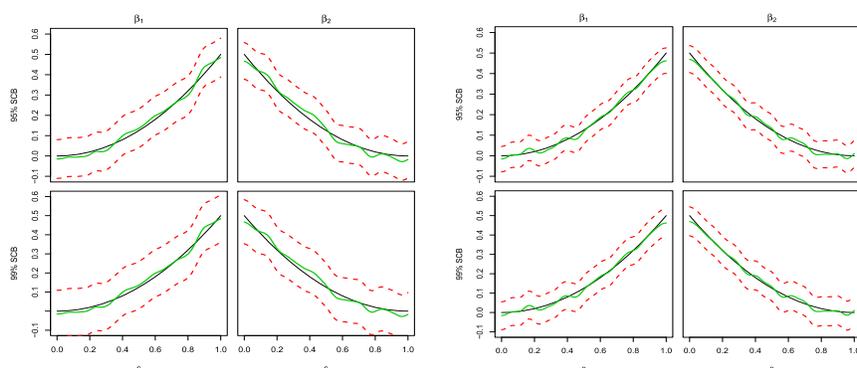


FIG 2. Typical 95% and 99% simultaneous confidence bands in the first and second rows, respectively, for $n = 50, M = 75$ (left panel) and $n = 100, M = 75$ (right panel) in the second simulation study under Model 1. The solid black, solid green, and dashed red curves represent, respectively, the true curves, estimated curves, and corresponding simultaneous confidence bands for β_j in the j th column ($j = 1, 2$).

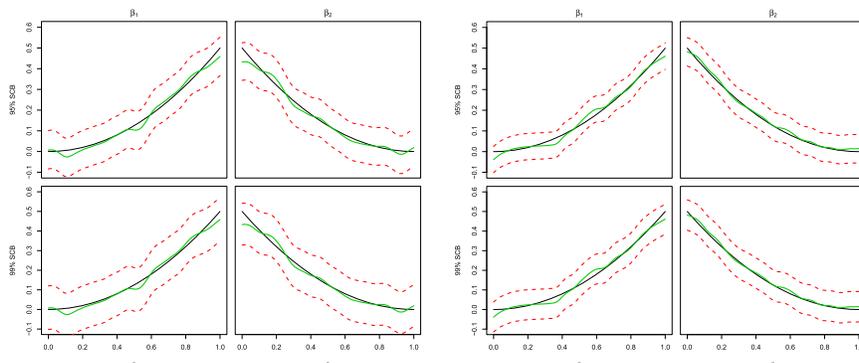


FIG 3. Typical 95% and 99% simultaneous confidence bands in the first and second rows, respectively, for $n = 50, M = 75$ (left panel) and $n = 100, M = 75$ (right panel) in the second simulation study under Model 2. The solid black, solid green, and dashed red curves represent, respectively, the true curves, estimated curves, and corresponding simultaneous confidence bands for β_j in the j th column ($j = 1, 2$).

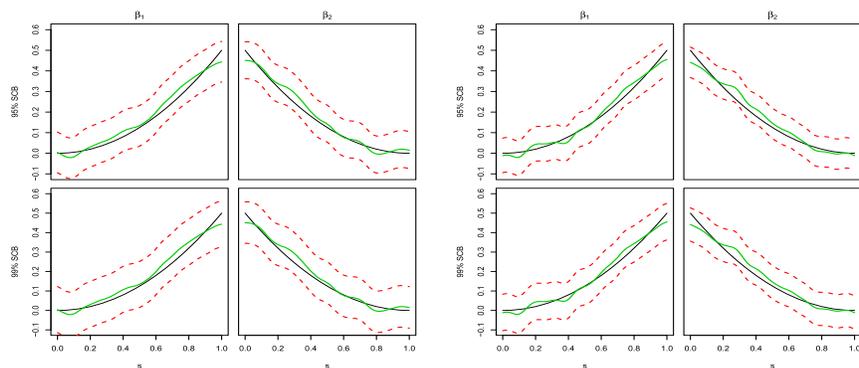


FIG 4. Typical 95% and 99% simultaneous confidence bands in the first and second rows, respectively, for $n = 50, M = 75$ (left panel) and $n = 100, M = 75$ (right panel) in the second simulation study under Model 3. The solid black, solid green, and dashed red curves represent, respectively, the true curves, estimated curves, and corresponding simultaneous confidence bands for β_j in the j th column ($j = 1, 2$).

TABLE 2

A summary of age at each repeated visit. The mean, standard deviation (SD), and range of subject ages, measured in years, are presented for each group. In total, $n = 253$ unique subjects are represented.

Visit number	Number of scans	Age (SD)	Age range
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]

a research data repository funded by the National Institutes of Health (NIH). The dataset consists of 416 high-quality MRI scans for 253 children, including 126 males. Summaries of subject age, stratified by the number of visits per subject, are presented in Table 2.

The data processing involved two stages: a weighted least squares procedure for constructing diffusion tensors [1, 44], and a tract-based statistics pipeline implemented in the Functional Magnetic Resonance Imaging of the Brain Software Library (FSL) [31]. Specifically, fractional anisotropy maps were computed for each subject from diffusion tensors after eddy current correction and automatic brain extraction using FSL. The aforementioned tract-based pipeline aligned fractional anisotropy maps from each subject into a common space through nonlinear registration and generated a mean fractional anisotropy skeleton representing the centers of all white matter tracts shared by the entire sample. The aligned fractional anisotropy data for each subject was then projected onto this skeleton. Our analysis focuses solely on the midsagittal corpus callosum skeleton, where fractional anisotropy is measured at $M = 45$ spatial grid points. The corpus callosum, the largest fiber tract in the human brain, is a topologically organized structure responsible for communication between the two hemispheres.

Recently, nonlinear mixed effects models based on the Gompertz function have been employed to characterize longitudinal white matter development during early childhood [19, 30]. These models take the form $y = \phi_1 \exp\{-\phi_2 \phi_3^\top\}$, where ϕ_1 represents the asymptote, ϕ_2 the delay, and ϕ_3 the reciprocal of exponentiated speed. Due to its asymptotic nature, the Gompertz function described in previous works can only account for developmental changes observed in adolescence and adulthood [20]. Following [9], we utilize the modified Gompertz function that allows for continued growth and extend it to the functional nonlinear mixed effects model. Specifically, we consider the following model:

$$y_{i,j}(s) = \phi_{1,i}(s) \exp\{-\phi_{2,i}(s)\phi_{3,i}(s)^{t_{i,j}} + \phi_{4,i}t_{i,j}\} + \varepsilon_{i,j}(s), \quad (4)$$

where $\phi_i = (\phi_{1,i}, \phi_{2,i}, \phi_{3,i}, \phi_{4,i})^\top = \beta + b_i$, $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^\top$ contains the fixed effects, $b_i = (b_{1,i}, b_{2,i}, 0, 0)^\top$ contains the random effects for the i th subject,

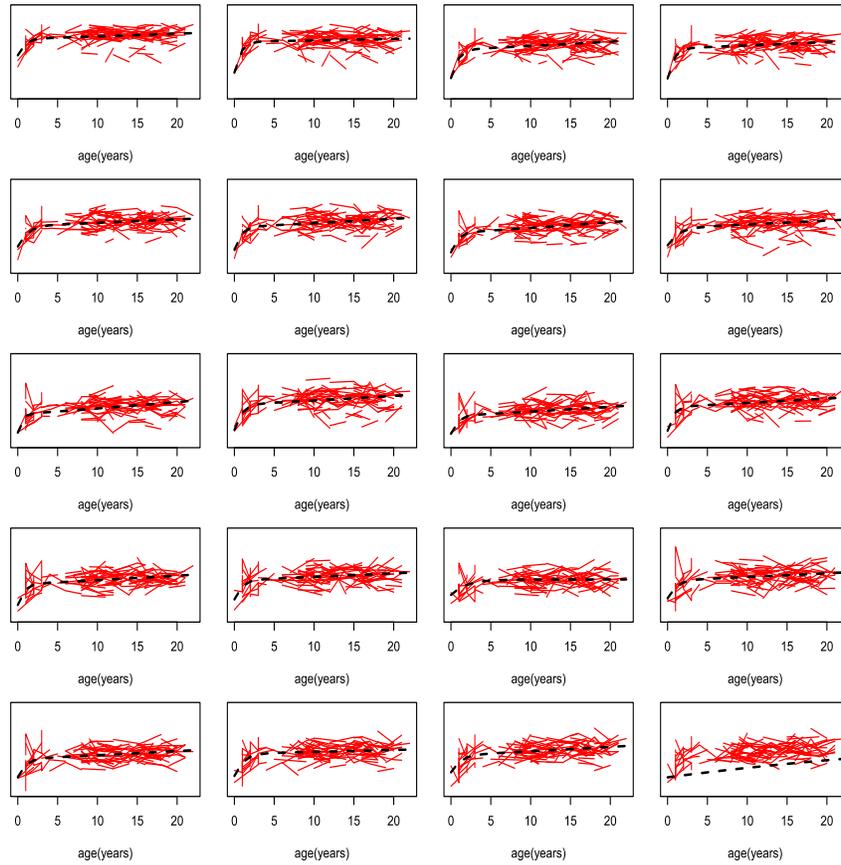


FIG 5. Observed fractional anisotropy for each subject (solid red lines), varying as a function of age t at a fixed spatial location s_m , with $m = 5, \dots, 24$ varying across the subplots. Estimated fractional anisotropy values obtained from the proposed model are superimposed (dotted black lines).

and the temporal variable t represents subject age at observation, in years. Note that $b_{3,i}$ and $b_{4,i}$ are set to zero in order to reduce the number of random effects in model (4).

We selected 116 children, who underwent at least two repeated MRI scans, resulting in a total of 279 scans. The variation of fractional anisotropy (y) with age (t) is illustrated in Figure 5 at fixed spatial positions (s_m) along the corpus callosum skeleton, with $m = 5, \dots, 24$. The figure shows significant subject-level variability in the data at each grid point and in the effect of age.

We used the estimation procedure outlined in Section 2.1 to fit the model (4). The population-level predictions for y , with $b_i = 0$, are also displayed in Figure 5. Figure 6 shows the functional fixed effect estimates and corresponding 95% and 99% simultaneous confidence bands, which were constructed using a

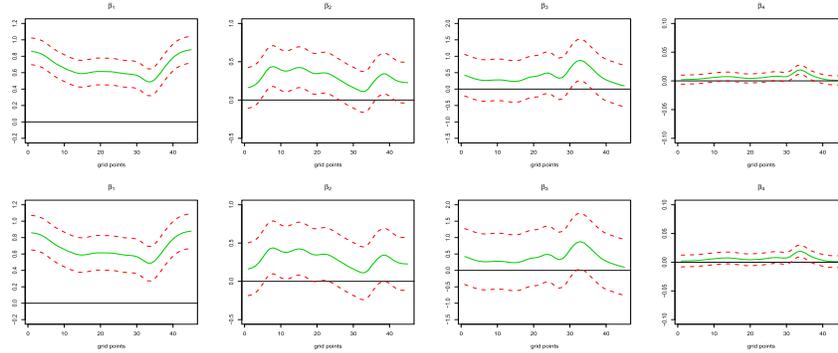


FIG 6. Estimates (solid green lines) and $100(1 - \alpha)\%$ simultaneous confidence bands (dotted red lines) for the fixed effect β_j from the real-world data neuroimaging analysis in the j th column ($j = 1, 2, 3, 4$). In the first and second rows, $\alpha = 0.05$ and $\alpha = 0.01$, respectively.

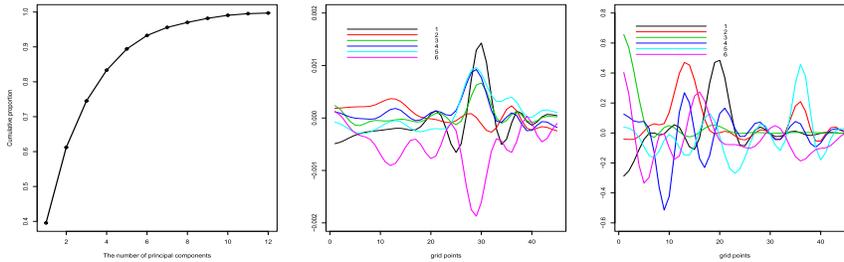


FIG 7. Left panel: the $100(1 - \alpha)\%$ cumulative proportion from one to twelve eigenvalues; Middle panel: the first six eigenfunctions corresponding to $b_{1,i}$; Right panel: the first six eigenfunctions corresponding to $b_{2,i}$.

wild bootstrap with 500 replications. The horizontal line at $\beta_j = 0$ serves as a reference, allowing us to conclude that the first three fixed effect functions ($\beta_j, j = 1, 2, 3$) are non-zero, while the last function (β_4) is non-zero only at the spatial location s_m with $m = 30, \dots, 40$. The remaining locations have a close-to-zero value for β_4 . Additionally, the first twelve eigenvalues and six eigenfunctions of $\widehat{\Sigma}^b(s, s')$ are displayed in Figure 7. We found that the first six eigenvalues explain 93.29% of the total variance, while the remaining eigenvalues explain close to 0%.

While the age effect, as shown in Figure 5, appears to level off quickly at some spatial positions, the proposed global test statistic ($S_N = 455.93, p\text{-value} < 0.001$) with $R = (0, 0, 1, 1)$ suggests a global age effect.

Appendix

We present the technique proofs of the theorems and lemmas in Section 3 in this paper.

A.1. Proofs of supporting lemmas

We first introduce some notation. Define $K_0(s, h_1) = \int K_{h_1}(t - s)\pi(t)dt$,

$$V(s) = \left\{ \sum_{i=1}^n \hat{X}_i(s)^\top \Sigma_i(s)^{-1} \hat{X}_i(s) \right\}^{-1} \sum_{i=1}^n \hat{X}_i(s)^\top \Sigma_i(s)^{-1} \hat{Z}_i(s) b_i(s), \quad (5)$$

$$R(s) = \left\{ \sum_{i=1}^n \hat{X}_i(s)^\top \Sigma_i(s)^{-1} \hat{X}_i(s) \right\}^{-1} \sum_{i=1}^n \hat{X}_i(s)^\top \Sigma_i(s)^{-1} \varepsilon_{i,j}(s), \quad (6)$$

and

$$\Delta(s, h_1) = \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) V(s_m) - \frac{1}{K_0(s, h_1)} \int K_{h_1}(t - s) V(t) \pi(t) dt. \quad (7)$$

Further define $\mathbb{M}\{\theta(s)\} = E\{\log P_\theta(i, s)\}$, $\mathbb{M}_n\{\theta(s)\} = n^{-1} \sum_{i=1}^n \log P_\theta(i, s)$, and

$$\bar{\varepsilon}_i(s) = n_i^{-1} \sum_{j=1}^{n_i} \sum_{m=1}^M \varepsilon_{i,j}(s_m) \tilde{K}_{h_2}(s_m - s).$$

Lemma A.1. *Under Assumptions 2-4 and 9, $\sup_{s_m \in \mathcal{S}_0} d(\hat{\theta}(s_m), \theta_0(s_m))$ converges in probability to 0.*

Proof. Define $\mathcal{F} = \{\log P_\theta(s_m) : s_m \in \mathcal{S}_0, \theta(s_m) \in \mathcal{B}\}$ and the envelope function $F = \sup_{s_m \in \mathcal{S}_0} G(y(s_m), x)$. We first show that \mathcal{F} is pr-Glivenko-Cantelli, that is, that

$$\sup_{\theta \in \Theta} \sup_{s_m \in \mathcal{S}_0} |\mathbb{M}_n(\theta(s_m)) - \mathbb{M}(\theta(s_m))| \rightarrow 0 \quad (8)$$

almost surely. Since $E^*(F) < \infty$ under Assumption 4, we need to prove that $\log N(\epsilon, \mathcal{F}_K, L_1(\mathbb{M}_n)) = o_p^*(n)$ for all $K < \infty$ and $\epsilon > 0$, where \mathcal{F}_K is the class of functions $\{f(\cdot)1(F \leq K) : f \in \mathcal{F}\}$. For each $s_0 \in \mathcal{S}_0$, define $\mathcal{F}_K^{s_0} = \{f \in \mathcal{F}_K : f(s_m) = f(s_0)\}$. Theorem 2.6.7 in [34] implies that $N(\epsilon, \mathcal{F}_K^{s_0}, L_1(\mathbb{M})) \leq C_1 \left(\frac{C_2}{\epsilon}\right)^p$ for some positive constants C_1 and C_2 such that

$$\begin{aligned} N(\epsilon, \mathcal{F}_K, L_1(\mathbb{M}_n)) &\leq M \times N(\epsilon, \mathcal{F}_K^{s_0}, L_1(\mathbb{M}_n)) \\ &\rightarrow M \times N(\epsilon, \mathcal{F}_K^{s_0}, L_1(\mathbb{M})) \leq MC_1 \left(\frac{C_2}{\epsilon}\right)^p. \end{aligned}$$

in probability. It follows that

$$\begin{aligned} &\log N(\epsilon, \mathcal{F}_K, L_1(\mathbb{M}_n)) \\ &= O_p^*(\log(M) + \log(C_1) + p \log(C_2/\epsilon)) \\ &= O_p^*\{n^\nu + O(1)\} \\ &= o_p^*(n) \end{aligned}$$

under Assumption 9. By Assumptions 3 and 4,

$$\sup_{d(\theta(s_m), \theta_0(s_m)) > \epsilon} \sup_{s_m \in \mathcal{S}_0} \mathbb{M}\{\theta(s_m)\} < \sup_{s_m \in \mathcal{S}_0} \mathbb{M}\{\theta_0(s_m)\}. \tag{9}$$

Finally, we follow the arguments in Theorem 5.7 in [33]. Since (8) implies that $\sup_{s_m \in \mathcal{S}_0} |\mathbb{M}_n\{\theta_0(s_m)\} - \mathbb{M}\{\theta_0(s_m)\}| \rightarrow 0$, then for each $s_m \in \mathcal{S}_0$, $\mathbb{M}_n\{\theta_0(s_m)\} > \mathbb{M}\{\hat{\theta}(s_m)\}$, and so

$$\begin{aligned} & \sup_{s_m \in \mathcal{S}_0} [\mathbb{M}\{\theta_0(s_m)\} - \mathbb{M}\{\hat{\theta}(s_m)\}] & (10) \\ & \leq \sup_{s_m \in \mathcal{S}_0} [\mathbb{M}\{\theta_0(s_m)\} - \mathbb{M}_n\{\theta_0(s_m)\}] + \sup_{s_m \in \mathcal{S}_0} [\mathbb{M}_n\{\theta_0(s_m)\} - \mathbb{M}\{\hat{\theta}(s_m)\}] \\ & \leq o_p(1) + \sup_{s_m \in \mathcal{S}_0} [\mathbb{M}_n\{\hat{\theta}(s_m)\} - \mathbb{M}\{\hat{\theta}(s_m)\}] \\ & \leq o_p(1) + \sup_{\theta \in \Theta} \sup_{s_m \in \mathcal{S}_0} |\mathbb{M}_n\{\theta(s_m)\} - \mathbb{M}\{\theta(s_m)\}| \\ & \rightarrow 0. & (11) \end{aligned}$$

By (9), for any arbitrary $\epsilon > 0$, there exists a positive constant δ depending only on ϵ such that $\sup_{s_m \in \mathcal{S}_0} \mathbb{M}\{\theta(s_m)\} \leq \sup_{s_m \in \mathcal{S}_0} \mathbb{M}\{\theta_0(s_m)\} - \delta(\epsilon)$ for every $\theta(s_m)$ when $\sup_{s_m \in \mathcal{S}_0} d(\theta(s_m), \theta_0(s_m)) > \epsilon$. Consequently, by (10),

$$\begin{aligned} & P\left(\sup_{s_m \in \mathcal{S}_0} d(\hat{\theta}(s_m), \theta_0(s_m)) > \epsilon\right) \\ & \leq P\left(\sup_{s_m \in \mathcal{S}_0} \mathbb{M}\{\hat{\theta}(s_m)\} < \sup_{s_m \in \mathcal{S}_0} \mathbb{M}\{\theta_0(s_m)\} - \delta(\epsilon)\right) \\ & \leq P\left(\sup_{s_m \in \mathcal{S}_0} [\mathbb{M}\{\theta_0(s_m)\} - \mathbb{M}\{\hat{\theta}(s_m)\}] > \delta(\epsilon)\right) \\ & \rightarrow 0. \end{aligned}$$

□

Lemma A.2. Under Assumptions, 1, 5, 7, and 9, for any $r \geq 0$,

$$\sup_{s \in \mathcal{S}} \left| \int K_{h_1}(t-s) \frac{(t-s)^r}{h_1^r} d\{\Pi_M(t) - \Pi(t)\} \right| = O_p\{(Mh_1)^{-1/2}\}$$

and

$$\sup_{s \in \mathcal{S}} \left| \int K_{h_1}(t-s) \frac{(t-s)^r}{h_1^r} \varepsilon_{i,j}(t) d\Pi_M(t) \right| = O_p\{(Mh_1)^{-1/2} |\log h_1|^{1/2}\},$$

where Π_M is the sampling distribution function based on \mathcal{S}_0 and Π is the distribution function of s .

Proof. The proof follows that of Lemma 2 in [42]. □

Lemma A.3. Under Assumptions 1, 5, 6, 7, and 9, $\sup_{s \in \mathcal{S}} |\Delta(s, h_1)| = o_p(1)$.

Proof. By Lemma 2 with $r = 0$,

$$\begin{aligned} & \sup_{s \in \mathcal{S}_0} \left| \frac{1}{M} \sum_{m=1}^M K_{h_1}(s_m - s) - K_0(s, h_1) \right| \\ &= \sup_{s \in \mathcal{S}_0} \left| \int K_{h_1}(t - s) d\{\Pi_M(t) - \Pi(t)\} \right| = O_p\{(Mh_1)^{-1/2}\}, \end{aligned}$$

and so we can write

$$\begin{aligned} & \Delta(s, h_1) \\ &= \frac{1}{K_0(s, h_1) + O_p\{(Mh_1)^{-1/2}\}} \left\{ \frac{1}{M} \sum_{m=1}^M K_{h_1}(s_m - s)V(s_m) \right. \\ & \quad \left. - \int K_{h_1}(t - s)V(t)d\Pi(t) \right\} \\ &= \frac{1 + O_p\{(Mh_1)^{-1/2}\}}{K_0(s, h_1)} \left[\frac{1}{M} \sum_{m=1}^M K_{h_1}(s_m - s)\{V(s_m) - V(s)\} \right. \\ & \quad \left. + \left\{ \frac{1}{M} \sum_{m=1}^M K_{h_1}(s_m - s) - \int K_{h_1}(t - s)d\Pi(t) \right\} V(s) \right. \\ & \quad \left. + \int K_{h_1}(t - s)\{V(s) - V(t)\}d\Pi(t) \right] \\ &= \frac{1 + o_p(1)}{K_0(s, h_1)} \{(I) + (II) + (III)\}. \end{aligned}$$

By Assumption 6, V converges weakly to a Gaussian process. It then follows from the Donsker Theorem [34] that $\sup_{s \in \mathcal{S}} \|V(s)\|_2 = O_p(1)$. Thus, examining the three terms (I), (II), and (III), we can conclude that

$$\begin{aligned} \frac{(I)}{K_0(s, h_1)} &\leq \frac{1}{K_0(s, h_1)} \frac{1}{M} \sum_{m=1}^M K_{h_1}(s_m - s) |V(s_m) - V(s)| \\ &\leq \sup_{|s' - s| \leq h_1} |V(s') - V(s)| \sup_{s \in \mathcal{S}} \frac{1}{MK_0(s, h_1)} \sum_{m=1}^M K_{h_1}(s_m - s) \\ &= o_p(1) \times \frac{K_0(s, h_1) + O_p\{(Mh_1)^{-1/2}\}}{K_0(s, h_1)} \\ &= o_p(1), \\ \frac{(II)}{K_0(s, h_1)} &\leq \frac{1}{K_0(s, h_1)} \sup_{s \in \mathcal{S}} |V(s)| \sup_{s \in \mathcal{S}} \left| \int K_{h_1}(t - s) d\{\Pi_M(t) - \Pi(t)\} \right| \\ &= O_p(h_1) \times O_p(1) \times O_p((Mh_1)^{-1/2}) \\ &= o_p(1), \end{aligned}$$

and

$$\begin{aligned} \frac{\text{(III)}}{K_0(s, h_1)} &\leq \frac{1}{K_0(s, h_1)} \int K_{h_1}(t-s) |V(s) - V(t)| d\Pi(t) \\ &\leq \sup_{|s'-s| \leq h_1} |V(s') - V(s)| \sup_{s \in \mathcal{S}} \frac{1}{K_0(s, h_1)} \int K_{h_1}(t-s) d\Pi(t) \\ &= o_p(1) \times 1 \\ &= o_p(1). \end{aligned}$$

This completes the proof of Lemma A.3. \square

Lemma A.4. Under Assumptions 1, 2, 5, 7, and 9,

$$\begin{aligned} &\sup_{s \in \mathcal{S}} n^{-1/2} h_1 \left| \sum_{i=1}^n \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) \hat{X}_i(s_m)^\top \Sigma_i(s_m)^{-1} \varepsilon_{i,j}(s_m) \right| \\ &= O_p\{(Mh_1 |\log h_1|)^{1/2}\} = o_p(Mh_1). \end{aligned}$$

Proof. The proof follows that of Lemma 1 in [41]. \square

Lemma A.5. Under Assumptions 1–10,

$$\begin{aligned} &\sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i \bar{\varepsilon}_i(s) b_i(t) \right| = O_p\{(\log n/n)^{1/2}\}, \\ &\sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i \bar{\varepsilon}_i(s) \bar{\varepsilon}_i(t) \right| = O_p\{(Mh_2)^{-1} + (\log n/n)^{1/2}\}, \end{aligned}$$

and

$$\sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i \bar{\varepsilon}_i(s) \ddot{b}_i(t) h_2^2 \right| = O_p\{(\log n/n)^{1/2}\}.$$

Proof. The proof follows from those of Lemmas 6 and 7 in [42]. \square

A.2. Proof of Theorem 1

Proof of Theorem 1 (i). By Lemma A.1, maximum likelihood estimation is uniformly consistent. We can apply a Taylor expansion to obtain

$$0 = \frac{\partial \log L_{\hat{\theta}(s_m)}}{\partial \beta(s_m)} = \frac{\partial \log L_{\theta(s_m)}}{\partial \beta(s_m)} + \frac{\partial^2 \log L_{\theta^*(s_m)}}{\partial \beta(s_m) \partial \beta(s_m)^\top} \{\hat{\beta}(s_m) - \beta(s_m)\},$$

where each element of the $\beta(s_m)$ component of $\theta^*(s_m)$ is between $\beta(s_m)$ and $\hat{\beta}(s_m)$. Rewriting the above result,

$$n^{1/2} \{\hat{\beta}(s_m) - \beta(s_m)\} = - \left\{ \frac{1}{n} \frac{\partial^2 \log L_{\theta^*(s_m)}}{\partial \beta(s_m) \partial \beta(s_m)^\top} \right\}^{-1} \left\{ n^{-1/2} \frac{\partial \log L_{\theta(s_m)}}{\partial \beta(s_m)} \right\}.$$

By Assumption 4 and the uniform strong law of large numbers,

$$\sup_{s_m \in \mathcal{S}_0} \left| -\frac{1}{n} \frac{\partial^2 \log L_{\theta^*(s_m)}}{\partial \beta(s_m) \partial \beta(s_m)^\top} - I_{\beta^*(s_m)} \right| \rightarrow 0$$

and the information matrix $I_{\beta^*(s_m)}$ converges to $I_{\beta(s_m)}$ by the continuity of $I_{\beta(s_m)}$. Then as $n \rightarrow \infty$, $n^{1/2}\{\hat{\beta}(s_m) - \beta(s_m)\}$ converges to a Gaussian process by Assumption 6 [34] with mean function zero and covariance operator

$$\left[\begin{array}{cc} I_{\beta(s)}^{-1} & I_{\beta(s)}^{-1} E \left\{ \frac{\partial \log P_{\theta(s)}}{\partial \beta} \left(\frac{\partial \log P_{\theta(t)}}{\partial \beta} \right)^\top \right\} I_{\beta(t)}^{-1} \\ I_{\beta(t)}^{-1} E \left\{ \frac{\partial \log P_{\theta(t)}}{\partial \beta} \left(\frac{\partial \log P_{\theta(s)}}{\partial \beta} \right)^\top \right\} I_{\beta(s)}^{-1} & I_{\beta(t)}^{-1} \end{array} \right].$$

Consequently, $\sup_{s_m \in \mathcal{S}_0} \left\| n^{1/2}\{\hat{\beta}(s_m) - \beta(s_m)\} \right\|_2 = O_p(1)$, completing the proof of Theorem 1 (i). \square

Proof of Theorem 1 (ii). Using the smoothed estimator of β ,

$$\begin{aligned} n^{1/2}\{\tilde{\beta}(s) - \beta(s)\} &= n^{1/2} \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) \hat{\beta}(s_m) - n^{1/2}\beta(s) \\ &= n^{1/2} \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) \{\hat{\beta}(s_m) - \beta(s_m)\} \\ &\quad + n^{1/2} \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) \{\beta(s_m) - \beta(s)\} \\ &= U_1(s) + U_2(s), \end{aligned}$$

where the last equality defines U_1 and U_2 . We first prove that U_1 converges weakly to a centered Gaussian Process. We may write that

$$\begin{aligned} &\hat{\beta}(s_m) - \beta(s_m) \\ &= \left\{ \sum_{i=1}^n \hat{X}_i(s_m)^\top \Sigma_i(s_m)^{-1} X_i(s_m) \right\}^{-1} \sum_{i=1}^n \hat{X}_i(s_m)^\top \Sigma_i(s_m)^{-1} \hat{Z}_i(s_m) b_i(s_m) \\ &\quad + \left\{ \sum_{i=1}^n \hat{X}_i(s_m)^\top \Sigma_i(s_m)^{-1} X_i(s_m) \right\}^{-1} \sum_{i=1}^n \hat{X}_i(s_m)^\top \Sigma_i(s_m)^{-1} \varepsilon_{i,j}(s_m) \\ &= V(s_m) + R(s_m). \end{aligned}$$

By (5) and (7), we can write

$$\begin{aligned} \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) V(s_m) &= \Delta(s, h_1) + \frac{1}{K_0(s, h_1)} \int K_{h_1}(t - s) V(t) \pi(t) d(t) \\ &= \Delta(s, h_1) + \frac{1}{K_0(s, h_1)} V(s) \int K_{h_1}(t - s) \pi(t) d(t) \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{K_0(s, h_1)} \int K_{h_1}(t - s)[V(t) - V(s)]\pi(t)d(t) \\
 & = \Delta(s, h_1) + \text{(I)} + \text{(II)}.
 \end{aligned}$$

Lemma 3 implies that $\Delta(s, h_1)$ converges uniformly to zero. (I) is asymptotically tight since

$$\text{(I)} = \frac{1}{K_0(s, h_1)}V(s)K_0(s, h_1) = V(s),$$

where $V(s)$ converges weakly to a Gaussian process. It follows that, when $h_1 \rightarrow 0$,

$$\text{(II)} \leq \sup_{|t-s| \leq h_1} |V(t) - V(s)| \frac{1}{K_0(s, h_1)} \int K_{h_1}(t - s)\pi(t)d(t) = o_p(1) \times 1 = o_p(1).$$

Consequently, $\sum_{m=1}^M \tilde{K}_{h_1}(s_m - s)V(s_m) \rightarrow V(s)$.

Moreover, $n^{1/2} \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s)R(s_m) = o_p(1)$ according to Lemma A.4, so U_1 converges to a Gaussian process with mean of zero and a covariance function of

$$\begin{aligned}
 & \Sigma(s, t) \\
 & = \lim_{n \rightarrow \infty} \Sigma_X(s) \frac{1}{n} \sum_{i=1}^n \hat{X}_i(s)^\top \Sigma_i(s)^{-1} \begin{bmatrix} \hat{Z}_1(s)\Sigma^b(s, t)\hat{Z}_1(t) & \dots \\ \vdots & \ddots \\ \dots & \hat{Z}_{n_i}(s)\Sigma^b(s, t)\hat{Z}_{n_i}(t) \end{bmatrix} \\
 & \Sigma_i(t)^{-1} \hat{X}_i(t)\Sigma_X(t),
 \end{aligned}$$

where $\Sigma_X(s) = \{ \frac{1}{n} \sum_{i=1}^n \hat{X}_i(s)^\top \Sigma_i(s)^{-1} \hat{X}_i(s) \}^{-1}$.

We next calculate $U_2(s)$. By a Taylor expansion,

$$\begin{aligned}
 & U_2(s) \tag{12} \\
 & = n^{1/2} \frac{\sum_{m=1}^M K_{h_1}(s_m - s) \{ \beta(s_m) - \beta(s) \}}{\sum_{m=1}^M K_{h_1}(s_m - s)} \\
 & = n^{1/2} \frac{Mh_1 \int K(\mu) \{ \beta(s + h_1\mu) - \beta(s) \} \pi(s + h_1\mu) d\mu [1 + O_p\{(Mh_1)^{-1/2}\}]}{Mh_1 \int K(\mu) \pi(s + h_1\mu) d\mu \{1 + O_p(Mh_1^{-1/2})\}} \\
 & = n^{1/2} \frac{\int K(\mu) \{ \dot{\beta}(s)h_1\mu + 0.5\ddot{\beta}(s)h_1^2\mu^2 + o(h_1^2) \} \{ \pi(s) + \dot{\pi}(s)h_1\mu + o(h_1) \} d\mu}{\int K(\mu) \{ \pi(s) + \dot{\pi}(s)h_1\mu + o(h_1) \} d\mu} \\
 & \quad \{1 + o_p(1)\} \\
 & = n^{1/2} \frac{\{0.5\ddot{\beta}(s)\pi(s) + \dot{\beta}(s)\dot{\pi}(s) + o(h_1)\} h_1^2 \mu_2(K)}{\pi(s) + o(h_1)} \{1 + o_p(1)\} \\
 & = n^{1/2} \{0.5\ddot{\beta}(s) + \dot{\beta}(s)\dot{\pi}(s)/\pi(s)\} h_1^2 \mu_2(K) \{1 + o_p(1)\}, \tag{13}
 \end{aligned}$$

where $\dot{\pi}(s) = \partial\pi(s)/\partial s$, $\dot{\beta}(s) = (\partial\beta_1(s)/\partial s, \dots, \partial\beta_p(s)/\partial s)^\top$, and $\ddot{\beta}(s) = (\partial\beta_1^2(s)/\partial s^2, \dots, \partial\beta_p^2(s)/\partial s^2)^\top$. This completes the proof of Theorem 1 (ii). \square

A.3. Proof of Theorem 2

Proof of Theorem 2 (i). Define $f_{i,j}^{0,1*}(s, s) = f_{i,j}^{0,1*}(s, s)(1, 0)^\top$. Note that minimizing $S_M(B_i(s))$ is equivalent to minimizing

$$S_M^*(B_i(s)) = \frac{1}{2} \sum_{j=1}^{n_i} \sum_{m=1}^M \left\{ y_{i,j}(s_m) - f(\hat{\beta}(s_m) + B_i(s)Z(s_m - s), x_{i,j}) \right\}^2 \tilde{K}_{h_2}(s_m - s).$$

The desired estimator is thus the solution to $S'_M(B_i(s)) = 0$.

Because S'_M is a continuous function, by the mean-value theorem, there exists $B_i^*(s)$ whose elements are between those of $B_i(s)$ and $\hat{B}_i(s)$, satisfying $S'_M(\hat{B}_i(s)) = S'_M(B_i(s)) + S''_M(B_i^*(s))(\hat{B}_i(s) - B_i(s))$. As $S'_M(\hat{B}_i(s)) = 0$, we have, by definition, that

$$\hat{B}_i(s) - B_i(s) = -S''_M(B_i^*(s))^{-1} S'_M(B_i(s)). \tag{14}$$

We first calculate $S^{*'}_M(B_i(s))$. Recall that $y_{i,j}(s_m) = f(\beta(s_m) + b_i(s_m), x_{i,j}) + \varepsilon_{i,j}(s_m)$ and $f(\beta(s_m) + B_i(s)Z(s_m - s), x_{i,j}) = f(\beta(s_m) + b_i(s) + \hat{b}_i(s)(s_m - s), x_{i,j})$. Under Assumptions 7–10 and by a Taylor expansion,

$$\begin{aligned} S^{*'}_M(B_i(s)) &= \sum_{j=1}^{n_i} \sum_{m=1}^M \left\{ y_{i,j}(s_m) - f(\hat{\beta}(s_m) + B_i(s)Z(s_m - s), x_{i,j}) \right\} \\ &\quad \partial f(\hat{\beta}(s_m) + B_i(s)Z(s_m - s), x_{i,j}) / \partial B_i(s) \cdot \tilde{K}_{h_2}(s_m - s) \\ &= \sum_{j=1}^{n_i} \sum_{m=1}^M \left\{ f(\beta(s_m) + b_i(s_m), x_{i,j}) - f(\hat{\beta}(s_m) + B_i(s)Z(s_m - s), x_{i,j}) \right. \\ &\quad \left. + \varepsilon_{i,j}(s_m) \right\} \times f_{i,j}^{0,1*}(s_m, s_m) \tilde{K}_{h_2}(s_m - s) \{1 + O_p(h_2^2)\} \{1 + O_p(n^{-1/2})\} \\ &= \left[\sum_{j=1}^{n_i} \sum_{m=1}^M \left\{ f(\beta(s_m) + b_i(s_m), x_{i,j}) - f(\beta(s_m) + B_i(s)Z(s_m - s)) \right\} \right. \\ &\quad \left. f_{i,j}^{0,1*}(s_m, s_m) \tilde{K}_{h_2}(s_m - s) + \sum_{j=1}^{n_i} \sum_{m=1}^M \left\{ f(\beta(s_m) + B_i(s)Z(s_m - s), x_{i,j}) \right. \right. \\ &\quad \left. \left. - f(\hat{\beta}(s_m) + B_i(s)Z(s_m - s), x_{i,j}) \right\} f_{i,j}^{0,1*}(s_m, s_m) \tilde{K}_{h_2}(s_m - s) \right. \\ &\quad \left. + \sum_{j=1}^{n_i} \sum_{m=1}^M \varepsilon_{i,j}(s_m) f_{i,j}^{0,1*}(s_m, s_m) \tilde{K}_{h_2}(s_m - s) \right] \{1 + O_p(h_2^2) + O_p(n^{-1/2})\} \\ &= \{S'_{i,1}(s) + S'_{i,2}(s) + S'_{i,3}(s)\} \{1 + O_p(h_2^2) + O_p(n^{-1/2})\}. \tag{15} \end{aligned}$$

Similar to calculating U_2 in (12), the three terms above can be bounded, with

$$\begin{aligned}
& S'_{i,1}(s) \\
&= \sum_{j=1}^{n_i} \sum_{m=1}^M \{f_{i,j}^{0,1}(s_m, s)^\top + O_p(h_2)\} \{0.5\ddot{b}_i(s)(s_m - s)^2 + o_p(h_2^2)\} f_{i,j}^{0,1*}(s_m, s_m) \cdot \\
&\quad \tilde{K}_{h_2}(s_m - s) \\
&= \sum_{j=1}^{n_i} \sum_{m=1}^M \{f_{i,j}^{0,1}(s, s)^\top + O_p(h_2)\} \{0.5\ddot{b}_i(s)(s_m - s)^2 + o_p(h_2^2)\} \cdot \\
&\quad \{f_{i,j}^{0,1*}(s, s) + O_p(h_2)\} \tilde{K}_{h_2}(s_m - s) \\
&\rightarrow \left[\int \{f_{i,j}^{0,1}(s, s)^\top + O_p(h_2)\} \{0.5\ddot{b}_i(s)h_2^2\mu^2 + o_p(h_2^2)\} \{f_{i,j}^{0,1*}(s, s) + O_p(h_2)\} \cdot \right. \\
&\quad \left. K(\mu)\pi(s + h_2\mu)d\mu \right] / \{\pi(s) + O_p(h_2^2)\} \times [1 + O_p\{(Mh_2)^{-1/2}\}] \\
&= \sum_{j=1}^{n_i} 0.5f_{i,j}^{0,1}(s, s)^\top \ddot{b}_i(s)h_2^2 f_{i,j}^{0,1*}(s, s) [1 + O_p(h_2^2) + O_p\{(Mh_2)^{-1/2}\}]. \quad (16)
\end{aligned}$$

$$\begin{aligned}
& S'_{i,2}(s) \\
&= \sum_{j=1}^{n_i} \sum_{m=1}^M \{f_{i,j}^{1,0}(s, s)^\top + \vec{O}_p^\top(h_2)\} \vec{O}_p(n^{-1/2}) \{f_{i,j}^{0,1*}(s, s) + O_p(h_2)\} \tilde{K}_{h_2}(s_m - s) \\
&\rightarrow \sum_{j=1}^{n_i} \frac{f_{i,j}^{1,0}(s, s)^\top \vec{O}_p(n^{-1/2}) f_{i,j}^{0,1*}(s, s) \{1 + O_p(h_2)\} \pi(s)}{\pi(s) + O_p(h_2^2)} [1 + O_p\{(Mh_2)^{-1/2}\}], \quad (17)
\end{aligned}$$

where $\vec{O}_p(r_n)$ denotes a vector whose elements are all $O_p(r_n)$; and

$$\begin{aligned}
S'_{i,3}(s) &= \sum_{j=1}^{n_i} \sum_{m=1}^M \varepsilon_{i,j}(s_m) \{f_{i,j}^{0,1*}(s, s) + O_p(h_2)\} \tilde{K}_{h_2}(s_m - s) \\
&= \sum_{j=1}^{n_i} f_{i,j}^{0,1*}(s, s) \{1 + O_p(h_2)\} \sum_{m=1}^M \varepsilon_{i,j}(s_m) \tilde{K}_{h_2}(s_m - s). \quad (18)
\end{aligned}$$

By Assumption 10 and Theorem 1 (i),

$$\begin{aligned}
& \sup_{s \in \mathcal{S}} |S'_{i,1}(s)| \\
&\leq \sum_{j=1}^{n_i} \sup_{s \in \mathcal{S}} \left| 0.5f_{i,j}^{0,1}(s, s)^\top \ddot{b}_i(s)h_2^2 f_{i,j}^{0,1*}(s, s) [1 + O_p(h_2^2) + O_p\{(Mh_2)^{-1/2}\}] \right| \\
&= O_p(n_i h_2^2), \quad (19)
\end{aligned}$$

$$\begin{aligned} & \sup_{s \in \mathcal{S}} |S'_{i,2}(s)| \\ & \leq \sum_{j=1}^{n_i} \sup_{s \in \mathcal{S}} \left| f_{i,j}^{1,0}(s, s)^\top \bar{O}_p(n^{-1/2}) f_{i,j}^{0,1*}(s, s) \{1 + O_p(h_2)\} [1 + O_p\{(Mh_2)^{-1/2}\}] \right| \\ & = O_p(n_i n^{-1/2}), \end{aligned} \tag{20}$$

and

$$\begin{aligned} & \sup_{s \in \mathcal{S}} |S'_{i,3}(s)| \\ & \leq \sum_{j=1}^{n_i} \sup_{s \in \mathcal{S}} \left| \frac{f_{i,j}^{0,1*}(s, s) \{1 + O_p(h_2)\}}{M^{-1} \sum_{m=1}^M K_{h_2}(s_m - s)} \right| \sup_{s \in \mathcal{S}} \left| M^{-1} \sum_{m=1}^M \varepsilon_{i,j}(s_m) K_{h_2}(s_m - s) \right| \\ & \leq \sum_{j=1}^{n_i} \sup_{s \in \mathcal{S}} \left| \frac{f_{i,j}^{0,1*}(s, s) \{1 + O_p(h_2)\}}{\pi(s) + O_p(h_2^2)} \right| O_p\{(Mh_2)^{-1/2} |\log h_2|^{1/2}\} \\ & \rightarrow O_p\{n_i (Mh_2)^{-1/2} |\log h_2|^{1/2}\}. \end{aligned} \tag{21}$$

We next calculate $S''_M(B_i^*(s))$. Define

$$\begin{aligned} b_i^*(s) &= B_i^*(s)(1, 0)^\top, \\ f_{i,j}^{0,k*}(\hat{\beta}(s_m), B_i^*(s)Z(s_m - s)) &= \partial^k f(\hat{\beta}(s_m) + B_i^*(s)Z(s_m - s), x_{i,j}) / \partial B_i^*(s)^k, \\ f_{i,j}^{0,k}(\beta(s), b_i^*(s)) &= \partial^k f(\beta(s) + b_i^*(s), x_{i,j}) / \partial b_i^*(s)^k, \end{aligned}$$

we have

$$f_{i,j}^{0,k}(\hat{\beta}(s_m), B_i^*(s)Z(s_m - s)) = f_{i,j}^{0,k}(\beta(s), b_i^*(s)) \{1 + O_p(h_2) + O_p(n^{-1/2})\}.$$

Similar to (15), we have that

$$\begin{aligned} & S''_M(B_i^*(s)) \\ &= \sum_{j=1}^{n_i} \sum_{m=1}^M \left[-f_{i,j}^{0,1*}(\hat{\beta}(s_m), B_i^*(s)Z(s_m - s)) f_{i,j}^{0,1*}(\hat{\beta}(s_m), B_i^*(s)Z(s_m - s))^\top \right. \\ & \quad \left. + \{y_{i,j}(s_m) - f(\hat{\beta}(s_m) + B_i^*(s)Z(s_m - s), x_{i,j})\} f_{i,j}^{0,2*}(\hat{\beta}(s_m), B_i^*(s)Z(s_m - s)) \right] \\ & \quad \tilde{K}_{h_2}(s_m - s) \\ &= \left[-\sum_{j=1}^{n_i} \sum_{m=1}^M f_{i,j}^{0,1*}(\beta(s), b_i^*(s))^2 \tilde{K}_{h_2}(s_m - s) \right. \\ & \quad \left. + \sum_{j=1}^{n_i} \sum_{m=1}^M \left\{ f(\beta(s_m) + b_i(s_m), x_{i,j}) - f(\beta(s_m) + B_i(s)Z(s_m - s), x_{i,j}) \right\} \right. \\ & \quad \left. f_{i,j}^{0,2*}(\beta(s), b_i^*(s)) \tilde{K}_{h_2}(s_m - s) \right. \\ & \quad \left. + \sum_{j=1}^{n_i} \sum_{m=1}^M \left\{ f(\beta(s_m) + B_i(s)Z(s_m - s), x_{i,j}) \right\} \right] \end{aligned}$$

$$\begin{aligned}
& - f(\beta(s_m) + B_i^*(s)Z(s_m - s), x_{i,j}) \Big\} f_{i,j}^{0,2*}(\beta(s), b_i^*(s)) \tilde{K}_{h_2}(s_m - s) \\
& + \sum_{j=1}^{n_i} \sum_{m=1}^M \left\{ f(\beta(s_m) + B_i^*(s)Z(s_m - s), x_{i,j}) \right. \\
& \quad \left. - f(\hat{\beta}(s_m) + B_i^*(s)Z(s_m - s), x_{i,j}) \Big\} f_{i,j}^{0,2*}(\beta(s), b_i^*(s)) \tilde{K}_{h_2}(s_m - s) \\
& + \sum_{j=1}^{n_i} \sum_{m=1}^M \varepsilon_{i,j}(s_m) f_{i,j}^{0,2*}(\beta(s), b_i^*(s)) \tilde{K}_{h_2}(s_m - s) \Big] \{1 + O_p(h_2) + O_p(n^{-1/2})\} \\
& = \{S''_{i,1}(s) + S''_{i,2}(s) + S''_{i,3}(s) + S''_{i,4}(s) + S''_{i,5}(s)\} \{1 + O_p(h_2) + O_p(n^{-1/2})\}. \tag{22}
\end{aligned}$$

It follows that

$$S''_{i,1}(s) = - \sum_{j=1}^{n_i} f_{i,j}^{0,1*}(\beta(s), b_i^*(s))^2, \tag{23}$$

so by a Taylor expansion,

$$\begin{aligned}
& S''_{i,3}(s) \\
& = \sum_{j=1}^{n_i} \sum_{m=1}^M f_{i,j}^{0,1*}(\beta(s_m), B_i^{**}(s)Z(s_m - s)) \{B_i^{**}(s) - B_i^*(s)\} \cdot \\
& \quad f_{i,j}^{0,2*}(\beta(s), b_i^*(s)) \tilde{K}_{h_2}(s_m - s) \\
& \rightarrow \sum_{j=1}^{n_i} f_{i,j}^{0,1*}(\beta(s), b_i^{**}(s)) \{B_i^{**}(s) - B_i^*(s)\} f_{i,j}^{0,2*}(\beta(s), b_i^*(s)) \cdot \\
& \quad [1 + O_p(h_2) + O_p\{(Mh_2)^{-1/2}\}], \tag{24}
\end{aligned}$$

where all elements of $B_i^{**}(s)$ are between those of $B_i^*(s)$ and $B_i(s)$, and $b_i^{**}(s) = B_i^{**}(s)(1, 0)^\top$. Furthermore, similar to (19)–(21), we have that

$$\sup_{s \in \mathcal{S}} |S''_{i,2}(s)| \rightarrow O_p(n_i h_2^2), \tag{25}$$

$$\sup_{s \in \mathcal{S}} |S''_{i,4}(s)| \rightarrow O_p(n_i n^{-1/2}), \tag{26}$$

and

$$\sup_{s \in \mathcal{S}} |S''_{i,5}(s)| \rightarrow O_p\{n_i (Mh_2)^{-1/2} |\log h_2|^{1/2}\}. \tag{27}$$

We next show the uniform consistency of $\hat{b}_i(s)$. From (23), we know that $S''_{i,1}(s) = O_p(1)$ under Assumption 10. Combining (14), (19)–(21), and (23)–(27),

$$\begin{aligned}
& \sup_{s \in \mathcal{S}} |\hat{b}_i(s) - b_i(s)| \\
& = \{O_p(1) + o_p(1)\}^{-1} [O_p(h_2^2) + O_p(n^{-1/2}) + O_p\{(Mh_2)^{-1/2} |\log h_2|^{1/2}\}].
\end{aligned}$$

$$\begin{aligned} & \{1 + O_p(h_2^2) + O_p(n^{-1/2})\} \\ & = O_p\{h_2^2 + n^{-1/2} + (Mh_2)^{-1/2}|\log h_2|^{1/2}\}, \end{aligned}$$

thus completing the proof of Theorem 2 (i). □

Proof of Theorem 2 (ii). By (i),

$$\sup_{s \in \mathcal{S}} |S''_{i,3}(s)| = n_i O_p\{h_2^2 + n^{-1/2} + (Mh_2)^{-1/2}|\log h_2|^{1/2}\}.$$

Consequently,

$$\begin{aligned} & S''_M(B_i^*(s)) \\ & \rightarrow - \sum_{j=1}^{n_i} f_{i,j}^{0,1*}(s, s)^2 \{1 + O_p(h_2^2) + O_p(n^{-1/2}) + (Mh_2)^{-1/2}|\log h_2|^{1/2}\}. \end{aligned} \quad (28)$$

On the other hand, since $E(-S_M^{-1} S'_{i,3}(s) \mid \mathcal{S}_0, X, b) = 0$, by (16)–(18),

$$\begin{aligned} & S'_M(B_i(s)) \\ & \rightarrow \sum_{j=1}^{n_i} [0.5 f_{i,j}^{0,1}(s, s)^\top \ddot{b}_i(s) h_2^2 \{1 + o_p(1)\} + f_{i,j}^{1,0}(s, s)^\top \vec{O}_p(n^{-1/2})] f_{i,j}^{0,1*}(s, s) \\ & \quad \times [1 + O_p(h_2^2) + O_p(n^{-1/2})]. \end{aligned} \quad (29)$$

Combining (29) and (28), it can be shown that

$$\begin{aligned} & E\{\hat{b}_i(s) \mid \mathcal{S}_0, X, b\} - b_i(s) \\ & = [E\{\hat{B}_i(s) \mid \mathcal{S}_0, X, b\} - B_i(s)](1, 0)^\top \\ & = \mathcal{F}_i(s) \sum_{j=1}^{n_i} \{0.5 f_{i,j}^{0,1}(s, s)^\top \ddot{b}_i(s) h_2^2 + f_{i,j}^{1,0}(s, s)^\top \vec{O}_p(n^{-1/2})\} f_{i,j*}^{0,1}(s, s) \{1 + o_p(1)\}, \end{aligned}$$

where $\mathcal{F}_i(s) = \left\{ \sum_{j=1}^{n_i} f_{i,j}^{0,1*}(s, s)^2 \right\}^{-1}$.

Furthermore,

$$\begin{aligned} & \hat{b}_i(s) - E\{\hat{b}_i(s) \mid \mathcal{S}, X, b\} = S''_M(B_i^*(s))^{-1} S'_{i,3}(s)(1, 0)^\top \{1 + o_p(1)\} \\ & = \mathcal{F}_i(s) \sum_{j=1}^{n_i} f_{i,j*}^{0,1}(s, s) \sum_{m=1}^M \varepsilon_{i,j}(s_m) \tilde{K}_{h_2}(s_m - s) \{1 + o_p(1)\}. \end{aligned}$$

Consequently,

$$\begin{aligned} & cov\{\hat{b}_i(s), \hat{b}_i(t) \mid \mathcal{S}_0, X, b\} \\ & = \mathcal{F}_i(s) \sum_{j=1}^{n_i} f_{i,j*}^{0,1}(s, s) f_{i,j*}^{0,1}(t, t)^\top \frac{\sum_{m=1}^M \sigma_\varepsilon^2(s_m) K_{h_2}(s_m - s) K_{h_2}(s_m - t)}{\{\sum_{m=1}^M K_{h_2}(s_m - s)\} \{\sum_{m=1}^M K_{h_2}(s_m - t)\}}. \end{aligned}$$

$$\begin{aligned}
 & \mathcal{F}_i(t)\{1 + o_p(1)\} \\
 = & \mathcal{F}_i(s) \sum_{j=1}^{n_i} \frac{(Mh_2)^{-1} \int \sigma_\varepsilon^2(s + h_2\mu)K(\mu)K(\mu + \frac{s-t}{h_2})\pi(s + h_2\mu)d\mu}{\int K(\mu)\pi(s + h_2\mu)d\mu \int K(\mu)\pi(t + h_2\mu)d\mu} \\
 & f_{i,j^*}^{0,1}(s, s)f_{i,j^*}^{0,1}(t, t)^\top [1 + O_p\{(Mh_2)^{-1/2}\}] \left\{ \sum_{j=1}^{n_i} f_{i,j^*}^{0,1}(t, t)^2 \right\}^{-1} \{1 + o_p(1)\} \\
 = & (Mh_2)^{-1} \mathcal{F}_i(s) \sum_{j=1}^{n_i} f_{i,j^*}^{0,1}(s, s)f_{i,j^*}^{0,1}(t, t)^\top \frac{\pi(s)O_p(1)K^*((s-t)/h)}{\pi(s)\pi(t)} \mathcal{F}_i(t) \cdot \\
 & \{1 + o_p(1)\},
 \end{aligned}$$

thus completing the proof of Theorem 2 (ii). □

Proof of Theorem 2 (iii)–(iv). By (i) and (ii),

$$\begin{aligned}
 & E[\{\hat{b}_i(s) - b_i(s)\}^\otimes | \mathcal{S}_0, X, b] \\
 = & [E\{\hat{b}_i(s) - b_i(s) | \mathcal{S}_0, X, b\}]^\otimes + \text{var}[\{\hat{b}_i(s) - b_i(s)\} | \mathcal{S}_0, X, b] \\
 = & \left[\mathcal{F}_i(s) \sum_{j=1}^{n_i} \{0.5f_{i,j}^{0,1}(s, s)^\top \ddot{b}_i(s)h_2^2 + f_{i,j}^{1,0}(s, s)^\top \vec{O}_p(n^{-1/2})\} f_{i,j^*}^{0,1}(s, s) \right]^\otimes \\
 & \{1 + o_p(1)\} \\
 & + (Mh_2)^{-1} v_0(K)\pi(s)^{-1} \mathcal{F}_i(s) \sum_{j=1}^{n_i} f_{i,j^*}^{0,1}(s, s)^\otimes \mathcal{F}_i(s) O_p(1).
 \end{aligned}$$

As $E[\{\hat{b}_i(s) - b_i(s)\}^\otimes | X, b] = E(E[\{\hat{b}_i(s) - b_i(s)\}^\otimes | \mathcal{S}, X, b] | \mathcal{S})$, we conclude that (iii) and (iv) hold. □

A.4. Proof of Theorem 3

Proof of Theorem 3. Since (ii) follows immediately from (i) [42], we only prove (i) here. Defining $\Delta_i(s) = \hat{b}_i(s) - b_i(s)$, it can be shown that

$$\begin{aligned}
 \hat{\Sigma}_b(s, t) &= N^{-1} \sum_{i=1}^n n_i \hat{b}_i(s) \hat{b}_i(t)^\top \\
 &= N^{-1} \sum_{i=1}^n n_i b_i(s) b_i(t)^\top + N^{-1} \sum_{i=1}^n n_i b_i(s) \Delta_i(t)^\top \\
 &\quad + N^{-1} \sum_{i=1}^n n_i \Delta_i(s) b_i(t)^\top + N^{-1} \sum_{i=1}^n n_i \Delta_i(s) \Delta_i(t)^\top \\
 &= \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)}. \tag{30}
 \end{aligned}$$

We first examine (I). By Assumption 6, $\{b_i(s)b_i(t)^\top\}$ is P -Donsker so that $N^{-1} \sum_{i=1}^n n_i b_i(s)b_i(t)^\top$ converges to a centered Gaussian process with a covariance matrix of $\Sigma^b(s, t)$. Therefore,

$$\sup_{(s,t)} |N^{-1} \sum_{i=1}^n n_i b_i(s)b_i(t)^\top - \Sigma^b(s, t)| = O_p(n^{-1/2}). \tag{31}$$

We next examine (II). Define $S_{i,k}^*(s) = \left\{ \sum_{j=1}^{n_i} f_{i,j}^{0,1*}(s, s)^2 \right\}^{-1} S'_{i,k}(s)(1, 0)^\top$ ($k = 1, 2, 3$). From the proof of Theorem 1, $\Delta_i(s) = \{S_{i,1}^*(s) + S_{i,2}^*(s) + S_{i,3}^*(s)\}\{1 + o_p(1)\}$. Consequently,

$$\begin{aligned} & N^{-1} \sum_{i=1}^n n_i b_i(s)\Delta_i(t)^\top \\ & \leq N^{-1} \left\{ \left| \sum_{i=1}^n n_i b_i(s)S_{i,1}^*(t)^\top \right| + \left| \sum_{i=1}^n n_i b_i(s)S_{i,2}^*(t)^\top \right| + \left| \sum_{i=1}^n n_i b_i(t)S_{i,3}^*(t)^\top \right| \right\} \\ & \quad \{1 + o_p(1)\}, \end{aligned}$$

where

$$\begin{aligned} & N^{-1} \left| \sum_{i=1}^n n_i b_i(s)S_{i,1}^*(t)^\top \right| \\ & \leq N^{-1} \sum_{i=1}^n n_i \sup_{s \in \mathcal{S}} |b_i(s)|(1, 0) \sup_{t \in \mathcal{S}} |S'_{i,1}(t)^\top| \mathcal{F}_i(t) = O_p(h_2^2). \end{aligned}$$

Similarly,

$$\begin{aligned} & N^{-1} \left| \sum_{i=1}^n n_i b_i(s)S_{i,2}^*(t)^\top \right| \\ & \leq N^{-1} \sum_{i=1}^n n_i \sup_{s \in \mathcal{S}} |b_i(s)|(1, 0) \sup_{s \in \mathcal{S}} |S'_{i,2}(t)^\top| \mathcal{F}_i(t) = O_p(n^{-1/2}). \end{aligned}$$

It follows from Lemma A.5 that

$$\begin{aligned} & \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i b_i(s)S_{i,3}^*(t)^\top \right| \\ & \leq \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i \bar{\varepsilon}_i(t)b_i(s)|(1, 0) \sup_{t, x_{i,j}} |f_{i,j}^{0,1*}(t, t)^\top| \sup_{t, x_{i,j}} n_i \mathcal{F}_i(t) \{1 + o_p(1)\} \right| \\ & = O_p\{(\log n/n)^{1/2}\}, \end{aligned}$$

from which we obtain that

$$\sup_{(s,t)} \text{(II)} = O_p\{h_2^2 + (\log n/n)^{1/2}\}. \tag{32}$$

Similarly, for (III),

$$\sup_{(s,t)} \text{(III)} = O_p\{h_2^2 + (\log n/n)^{1/2}\}. \tag{33}$$

Lastly, express (IV) as

$$\begin{aligned} & \left| \sum_{i=1}^n n_i \Delta_i(s) \Delta_i(t)^\top \right| \\ & \leq \sup_{(s,t)} \left| \sum_{i=1}^n n_i S_{i,1}^*(s) S_{i,1}^*(t)^\top \right| + \sup_{(s,t)} \left| \sum_{i=1}^n n_i S_{i,2}^*(s) S_{i,2}^*(t)^\top \right| \\ & \quad + \sup_{(s,t)} \left| \sum_{i=1}^n n_i S_{i,3}^*(s) S_{i,3}^*(t)^\top \right| + 2 \sup_{(s,t)} \left\{ \left| \sum_{i=1}^n n_i S_{i,1}^*(s) S_{i,2}^*(t)^\top \right| \right. \\ & \quad \left. + \left| \sum_{i=1}^n n_i S_{i,1}^*(s) S_{i,3}^*(t)^\top \right| + \left| \sum_{i=1}^n n_i S_{i,1}^*(s) S_{i,3}^*(t)^\top \right| \right\}. \end{aligned}$$

Since

$$\begin{aligned} & \sup_{s \in \mathcal{S}} |S_{i,1}^*(s)| \\ & = \sup_{s \in \mathcal{S}} \left| \mathcal{F}_i(s) \sum_{j=1}^{n_i} 0.5 f_{i,j}^{0,1}(s, s)^\top \ddot{b}_i(s) h_2^2 f_{i,j}^{0,1*}(s, s) [1 + o_p(1)] \right| = O_p(h_2^2), \end{aligned}$$

we have that

$$\sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i S_{i,1}^*(s) S_{i,1}^*(t)^\top \right| = O_p(h_2^4). \tag{34}$$

Similarly,

$$\begin{aligned} & \sup_{s \in \mathcal{S}} |S_{i,2}^*(s)| \\ & = \sup_{s \in \mathcal{S}} \left| \mathcal{F}_i(s) \sum_{j=1}^{n_i} f_{i,j}^{1,0}(s, s)^\top \bar{O}_p(n^{-1/2}) f_{i,j}^{0,1*}(s, s) \{1 + o_p(1)\} \right| = O_p(n^{-1/2}), \end{aligned}$$

and so

$$\sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i S_{i,2}^*(s) S_{i,2}^*(t)^\top \right| = O_p(n^{-1}). \tag{35}$$

By Lemma A.5,

$$\begin{aligned} & \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i S_{i,3}^*(s) S_{i,3}^*(t)^\top \right| \\ & = \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i \bar{\varepsilon}_i(s) \bar{\varepsilon}_i(t) n \mathcal{F}_i(s) n \mathcal{F}_i(t) \{1 + o_p(1)\} \right| \\ & = O_p(1) \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i \bar{\varepsilon}_i(s) \bar{\varepsilon}_i(t) \right| = O_p\{(Mh_2)^{-1} + (\log n/n)^{1/2}\}, \end{aligned} \tag{36}$$

$$\begin{aligned} & \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i S_{i,1}^*(s) S_{i,2}^*(t)^\top \right| \\ &= \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i \mathcal{F}_i(s) \sum_{j=1}^{n_i} 0.5 f_{i,j}^{0,1}(s, s)^\top \ddot{b}_i(s) h_2^2 f_{i,j}^{0,1*}(s, s) O_p(n^{-1/2}) \right| \\ &= O_p(h_2^2 n^{-1/2}), \end{aligned} \tag{37}$$

and

$$\begin{aligned} & \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i S_{i,1}^*(s) S_{i,3}^*(t)^\top \right| \\ &= \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i \bar{\varepsilon}_i(t) \left\{ \mathcal{F}_i(s) \sum_{j=1}^{n_i} 0.5 f_{i,j}^{0,1}(s, s)^\top \ddot{b}_i(s) h_2^2 f_{i,j}^{0,1*}(s, s) \right\} n \mathcal{F}_i(t) \cdot \right. \\ & \quad \left. \{1 + o_p(1)\} \right| \\ &= O_p(1) \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i \bar{\varepsilon}_i(t) \ddot{b}_i(s) h_2^2 \right| \\ &= O_p\{(\log n/n)^{1/2}\}. \end{aligned} \tag{38}$$

Then by Lemma A.2,

$$\begin{aligned} & \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i S_{i,2}^*(s) S_{i,3}^*(t)^\top \right| = \sup_{(s,t)} N^{-1} \left| \sum_{i=1}^n n_i \bar{\varepsilon}_i(t) O_p(n^{-1/2}) n \mathcal{F}_i(t) \right| \\ &= O_p\{(Mnh_2)^{-1/2} |\log h_1|^{1/2}\}. \end{aligned} \tag{39}$$

Combining (34)–(39),

$$\sup_{(s,t)} \text{(IV)} = O_p\{h_2^4 + (Mh_2)^{-1} + (\log n/n)^{1/2}\}. \tag{40}$$

Consequently, combining (31)–(33) and (40) with (30),

$$\sup_{(s,t)} |\hat{\Sigma}^b(s, t) - \Sigma^b(s, t)| = O_p\{h_2^2 + (Mh_2)^{-1} + (\log n/n)^{1/2}\}.$$

This completes the proof of Theorem 3. □

A.5. Proof of Theorem 4

Proof. Using similar arguments as in Theorem 7 of Zhang & Chen (2007), the conclusion of Theorem 4(i) holds. Therefore we omit the proof of Theorem 4(i) here.

We next show Theorem 4(ii). Recall that $d(s) = R[\tilde{\beta}(s) - \text{BIAS}\{\tilde{\beta}(s)\}] - b_0(s)$. Let $\nu(s) = \{R\hat{\Sigma}(s, s)R^\top\}^{-1/2}d(s)$ and $\text{AGP}(\eta, \gamma)$ denotes an asymptotic

Gaussian process with mean function $\eta(t)$ and covariance function $\gamma(s, t)$. Under H_{1n} , from the proof of Theorem 1(ii), we know $\nu(s) \sim \text{AGP}(\eta_\nu, \gamma_\nu)$, where $\eta_\nu(s) = \{R\widehat{\Sigma}(s, s)R^\top\}^{-1/2}n^{-\kappa/2}d(s)$ and $\gamma_\nu(s, t) = \{\gamma_{\nu, ij}(s, t)\}_{1 \leq i, j \leq p_0}$ with $\gamma_{\nu, ij}(s, t) = \text{cov}(\nu_i(s), \nu_j(t))$. Using Mercer's theorem, then there exists a set of orthonormal basis functions $\zeta_k = (\zeta_{k1}, \dots, \zeta_{kp_0})$ in \mathcal{H} such that

$$\gamma_\nu(s, t) = \sum_{k=1}^{\infty} \lambda_k \zeta_k(s) \zeta_k^\top(t),$$

where \mathcal{H} is a Hilbert space of p_0 -dimensional vectors of functions in $L_2(\mathcal{S})$ with $\langle \zeta_k, \zeta_{k'} \rangle_{\mathcal{H}} = \delta_{kk'}$ and $\lambda_1, \dots, \lambda_k, \dots$ are the decreasingly-ordered eigenvalues of $\gamma_\nu(s, t)$. Without loss of generality, let m denote the number of positive eigenvalues. When all the eigenvalues are positive, we let $m = \infty$. Let $\xi_k = \langle \nu, \zeta_k \rangle_{\mathcal{H}}$, we know $\xi_k \sim N(\mu_k, \lambda_k)$, where $\mu_k = \langle \eta_\nu, \zeta_k \rangle_{\mathcal{H}}$. It then follows Theorem 1 of [39], we have

$$S_N = \int_0^1 \nu(s)^\top \nu(s) ds = \sum_{k=1}^m \lambda_k A_k + \sum_{k=m+1}^{\infty} \mu_k^2,$$

where $A_k \sim \chi^2(\lambda_k^{-1} \mu_k^2)$. Note that the null distribution of S_N is mixture of χ^2 since $\mu_k = 0$ holds under H_0 . Under H_{1n} , we have $\mu_k^2 = n^{1-\kappa} \phi_k^2$, where $\phi_k = \sum_{j=1}^{p_0} \int_0^1 \{Rn\widehat{\Sigma}(s, s)R^\top\}^{-1/2} d_j(s) \zeta_{kj}(s) ds$. Therefore, we have

$$S_N = \sum_{k=1}^m \lambda_k Z_k^2 + 2n^{(1-\kappa)/2} \phi_\lambda Z_k + n^{1-\kappa} \phi^2,$$

where $\phi_\lambda = \sum_{k=1}^m \lambda_k^{1/2} \phi_k$, $\phi^2 = \sum_{k=1}^m \phi_k^2 > 0$, and $Z_K \sim N(0, 1)$. Using similar arguments in Theorem 3 of [39], we know that $S_N \sim AN(n^{1-\kappa} \phi^2, 4n^{1-\kappa} \phi_k^2)$ under H_{1n} . Therefore, we have $P(S_N > S_{N,\alpha} | H_{1n}) = \Phi\{n^{(1-\kappa)/2} \phi^2 / (2\phi_k)\} + o(1) \rightarrow 1$ as $n \rightarrow \infty$. Hence, we have completed the proof of Theorem 4. \square

A.6. Proof of Theorem 5

Proof. Recalling the uniform consistency of $\hat{\beta}$ and \hat{b}_i established in Theorems 1 and 2, we approximate $\log L_{\theta(s)}$ as

$$\begin{aligned} & \log L_{\theta(s)}^* \\ &= -\frac{1}{2} \sum_{i=1}^n [\log |\Sigma_i(s)| + \sigma_\varepsilon^2(s) \{\hat{\omega}_i(s) - \hat{X}_i(s)\beta(s)\}^\top \Sigma_i(s)^{-1} \{\hat{\omega}_i(s) - \hat{X}_i(s)\beta(s)\}] \\ & \quad - \frac{N}{2} \log(2\pi\sigma^2). \end{aligned}$$

Define $r_i(s) = \hat{\omega}_i(s) - \hat{X}_i(s)\beta(s)$ and

$$\tilde{G}(s)^{(g)} = n^{1/2} \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) \left\{ \Sigma_X(s_m) \sum_{i=1}^n \hat{X}_i(s_m)^\top \Sigma_i(s_m)^{-1} r_i(s_m) \tau_i^{(g)} \right\}.$$

We will prove Theorem 5 in three steps. First, we establish the unconditional weak convergence of $\tilde{G}(s)^{(g)}$. Second, we prove the weak convergence of $\tilde{G}(s)^{(g)}$ conditional on the data. Third, we prove the weak convergence of $G(s)^{(g)}$ conditional on the data by showing that $\tilde{G}(s)^{(g)}$ and $G(s)^{(g)}$ are asymptotically equivalent as $n \rightarrow \infty$.

First, noting that

$$\frac{\partial \log L_{\theta(s)}^*}{\partial \beta(s)} = -\sigma_\varepsilon^2(s) \sum_{i=1}^n \hat{X}_i(s)^\top \Sigma_i(s)^{-1} \{\hat{\omega}_i(s) - \hat{X}_i(s)\beta(s)\}$$

and

$$\frac{\partial^2 \log L_{\theta(s)}^*}{\partial \beta(s)^2} = \sigma_\varepsilon^2(s) \sum_{i=1}^n \hat{X}_i(s)^\top \Sigma_i(s)^{-1} \hat{X}_i(s),$$

we have that

$$\tilde{G}(s)^{(g)} \approx n^{1/2} \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) \left[\left\{ \frac{\partial^2 \log L_{\theta(s_m)}^*}{\partial \beta(s_m)^2} \right\}^{-1} \sum_{i=1}^n \tau_i^{(g)} \frac{\partial \log P_\theta(i, s_m)}{\partial \beta(s_m)} \right].$$

Treating $\tau_i^{(g)} \partial \log P_\theta(i, s_m) / \partial \beta(s_m)$ as a new covariate vector, we can apply the arguments in the proof of Theorem 1 (i) to conclude that $\tilde{G}(s)^{(g)}$ converges weakly to $\tilde{G}(s)$ and, thus, $\tilde{G}(s)^{(g)}$ is asymptotically measurable.

Second, rewrite $\tilde{G}(s)^{(g)}$ as

$$\tilde{G}(s)^{(g)} \approx n^{1/2} \sum_{i=1}^n \tau_i^{(g)} \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) \left\{ \frac{\partial^2 \log L_{\theta(s_m)}^*}{\partial \beta(s_m)^2} \right\}^{-1} \frac{\partial \log P_\theta(i, s_m)}{\partial \beta(s_m)}.$$

Therefore, conditional on the data, $\tilde{G}(s)^{(g)}$ is a normal random vector with a mean of zero. To calculate its variance, we take an expectation with respect to $\tau_i^{(g)}$ and obtain

$$\begin{aligned} \text{var}_\tau[\tilde{G}(s)^{(g)}] &\approx n \sum_{i=1}^n \left\{ \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) \left(\frac{\partial^2 \log L_{\theta(s_m)}^*}{\partial \beta(s_m)^2} \right)^{-1} \frac{\partial \log P_\theta(i, s_m)}{\partial \beta(s_m)} \right\}^{\otimes 2} \\ &\rightarrow n^{-1} \sum_{i=1}^n \left\{ \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) I_{\beta(s_m)}^{-1} \frac{\partial \log P_\theta(i, s_m)}{\partial \beta(s_m)} \right\}^{\otimes 2}. \end{aligned}$$

Since $E\{\partial \log P_\theta(i, s_m) / \partial \beta(s_m)\} = 0$,

$$\begin{aligned} \text{var}_\tau\{\tilde{G}(s)^{(g)}\} &\rightarrow E \left\{ \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) I_{\beta(s_m)}^{-1} \frac{\partial \log P_\theta(i, s_m)}{\partial \beta(s_m)} \right\}^{\otimes 2} \\ &= \text{var} \left\{ \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) I_{\beta(s_m)}^{-1} \frac{\partial \log P_\theta(i, s_m)}{\partial \beta(s_m)} \right\} \end{aligned}$$

$$\begin{aligned}
&= \text{var} \left[\frac{V(s)\pi(s) + O_p(h_2^2)}{\pi(s) + O_p(h_2^2)} \{1 + (Mh_2)^{-1/2}\} \right] \\
&= \Sigma(s, s) + O_p\{h_2^2 + (Mh_2)^{-1/2}\}.
\end{aligned}$$

Similarly, we can show that

$$\text{cov}_\tau\{\tilde{G}(s)^{(g)}, \tilde{G}(t)^{(g)}\} = \Sigma(s, t) + O_p\{h_2^2 + (Mh_2)^{-1/2}\}.$$

Following the proof of Theorem 5 in [42], we can obtain the weak convergence of $\tilde{G}(s)^{(g)}$ via the Cramer-Wald method. From these arguments, we can conclude that

$$\begin{aligned}
\Delta_{n,\beta} &= n^{-1} \sup_{s \in \mathcal{S}} \sum_{i=1}^n \text{tr} \left\{ \sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) \left(\Sigma_X(s_m) [\hat{X}_i(s_m)^\top \Sigma_i(s_m)^{-1} \right. \right. \\
&\quad \left. \left. \{ \hat{\beta}(s_m) - \beta(s_m) \} \}^\otimes \right) \right\} \\
&= n^{-1} \sup_{s \in \mathcal{S}} \text{tr} \left[\sum_{m=1}^M \tilde{K}_{h_1}(s_m - s) \{ \Sigma_X(s_m) O_p(1) \} \right] \\
&= O_p(n^{-1}).
\end{aligned}$$

This completes the proof of Theorem 5. \square

Acknowledgments

The authors would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that improved the quality of this paper.

Funding

Linglong Kong was partially supported by grants from the Canada CIFAR AI Chairs program, the Alberta Machine Intelligence Institute (AMII), and Natural Sciences and Engineering Council of Canada (NSERC), and the Canada Research Chair program from NSERC. Lixing Zhu was supported by the National Scientific Foundation of China (NSFC12131006). The research of Hongtu Zhu was partially supported by the National Institute On Aging (NIA) of the National Institutes of Health (NIH) under Award Numbers RF1AG082938.

References

- [1] Bassler, 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* **103**, 247–254.

- [2] Bethlehem, R. A., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., et al. (2022). Brain charts for the human lifespan. *Nature* pages 1–11.
- [3] Blozis, S. A. and Haring, J. R. (2021). Fitting nonlinear mixed-effects models with alternative residual covariance structures. *Sociological Methods & Research* **50**, 531–566. [MR4243229](#)
- [4] Cao, J., Huang, J. Z., and Wu, H. (2011). Penalized nonlinear least squares estimation of time-varying parameters in ordinary differential equations. *Journal of Computational and Graphical Statistics* **21**, 1–15. [MR2913355](#)
- [5] Casey, B., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., et al. (2018). The adolescent brain cognitive development (abcd) study: imaging acquisition across 21 sites. *Developmental cognitive neuroscience* **32**, 43–54.
- [6] 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. [MR3457688](#)
- [7] Cheng, Y., Shi, J. Q., and Eyre, J. (2019). Nonlinear mixed-effects scalar-on-function models and variable selection. *Statistics and Computing* **30**, 129–140. [MR4057475](#)
- [8] Davidian, M. and Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: an overview and update. *Journal of agricultural, biological, and environmental statistics* **8**, 387–419.
- [9] Dean, D. C., O’Muircheartaigh, J., Dirks, H., Waskiewicz, N., Walker, L., Doernberg, E., Piryatinsky, I., and Deoni, S. C. (2015). Characterizing longitudinal white matter development during early childhood. *Brain Structure and Function* **220**, 1921–1933.
- [10] Demidenko, E. (1997). *Comparison of Methods for General Nonlinear Mixed-Effects Models*. Springer-Verlag, New York.
- [11] Demidenko, E. (2004). *Mixed Models: Theory and Applications*. Wiley-Interscience. [MR2077875](#)
- [12] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London. [MR1383587](#)
- [13] Greven, S., Crainiceanu, S., Caffo, B. S., and Reich, D. (2010). Longitudinal functional principal component analysis. *Electronic Journal of Statistics* **4**, 1022–1054. [MR2727452](#)
- [14] Guo, W. (2002). Functional mixed effects models. *Biometrics* **58**, 121–128. [MR1891050](#)
- [15] Hall, D., Huerta, M., Mcauliffe, M., and Farber, G. (2012). Sharing heterogeneous data: The national database for autism research. *Neuroinformatics* **10**, 331–9.
- [16] Howell, B. R., Styner, M. A., Gao, W., Yap, P.-T., Wang, L., Baluyot, K., Yacoub, E., Chen, G., Potts, T., Salzwedel, A., et al. (2019). The unc/umn baby connectome project (bcp): An overview of the study design and protocol development. *NeuroImage* **185**, 891–905.

- [17] Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods* **1**, 23–41.
- [18] Kuhn, E. and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics and Data Analysis* **49**, 1020–1038. [MR2143055](#)
- [19] Kulikova, S., Hertz-Pannier, L., Dehaene-Lambertz, G., Buzmakov, A., Poupon, C., and Dubois, J. (2015). Multi-parametric evaluation of the white matter maturation. *Brain Structure and Function* **220**, 3657–3672.
- [20] Lebel, C. and Beaulieu, C. (2011). Longitudinal development of human brain wiring continues from childhood into adulthood. *Journal of neuroscience* **31**, 10937–10947.
- [21] Lee, I. H. and Rojewski, J. W. (2009). Development of occupational aspiration prestige: A piecewise latent growth model of selected influences. *Journal of Vocational Behavior* **75**, 82–90.
- [22] Li, G., Wang, L., Yap, P.-T., Wang, F., Wu, Z., Meng, Y., Dong, P., Kim, J., Shi, F., Reikik, I., et al. (2019). Computational neuroanatomy of baby brains: A review. *NeuroImage* **185**, 906–925.
- [23] Li, X., Wang, L., Wang, H. J., and Initiative, A. D. N. (2021). Sparse learning and structure identification for ultrahigh-dimensional image-on-scalar regression. *Journal of the American Statistical Association* **116**, 1994–2008. [MR4353728](#)
- [24] Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**, 673–687. [MR1085815](#)
- [25] Luo, X., Zhu, L., Kong, L., and Zhu, H. (2015). Functional nonlinear mixed effects models for longitudinal image data. In *International Conference on Information Processing in Medical Imaging*, pages 794–805. Springer.
- [26] 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. [MR2188981](#)
- [27] 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.
- [28] Pinheiro, J. and Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer science & business media.
- [29] Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **53**, 233–243. [MR1094283](#)
- [30] Sadeghi, N., Prastawa, M., Fletcher, P. T., Wolff, J., Gilmore, J. H., and Gerig, G. (2013). Regional characterization of longitudinal dt-mri to study white matter maturation of the early developing brain. *NeuroImage* **68**, 236–247.
- [31] Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., et al. (2006). Tract-based spatial statistics: Voxelwise

- analysis of multi-subject diffusion data. *NeuroImage* **31**, 1487–1505.
- [32] Strogatz, S. H. (1994). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Westview, Cambridge, MA. [MR3837141](#)
- [33] Van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press. [MR1652247](#)
- [34] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York. [MR1385671](#)
- [35] Varfolomeyev, S. and Gurevich, K. (2001). The hyperexponential growth of the human population on a macrohistorical scale. *Journal of Theoretical Biology* **212**, 367–372.
- [36] Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application* **3**, 257–295.
- [37] Wied, D. and Weißbach, R. (2010). Consistency of the kernel density estimator: A survey. *Statistical Papers* **53**, 1–21. [MR2878587](#)
- [38] 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.
- [39] Zhang, J.-T. (2011). Statistical inferences for linear models with functional responses. *Statistica Sinica* pages 1431–1451. [MR2827530](#)
- [40] Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *Annals of Statistics* **35**, 1052–1079. [MR2341698](#)
- [41] Zhu, H., Chen, K., Luo, X., Yuan, Y., and Wang, J.-L. (2019). FMEM: Functional mixed effects models for longitudinal functional responses. *Statistica Sinica* **29**, 2007–2033. [MR3970345](#)
- [42] Zhu, H., Li, R., and Kong, L. (2012). Multivariate varying coefficient model for functional responses. *Annals of Statistics* **40**, 2634–2666. [MR3097615](#)
- [43] Zhu, H., Li, T., and Zhao, B. (2023). Statistical learning methods for neuroimaging data analysis with applications. *Annual Review of Biomedical Data Science* **6**, in press. [MR3653719](#)
- [44] Zhu, H., Zhang, H., Ibrahim, J. G., and Peterson, B. S. (2007). Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data. *Journal of the American Statistical Association* **102**, 1085–1102. [MR2412530](#)
- [45] 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. [MR3292493](#)