Received: 29 July 2021 | Accepted: 11 August 2022

DOI: 10.1111/biom.13744

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Functional data analysis with covariate-dependent mean and covariance structures

Chenlin Zhang[1] | **Huazhen Lin**[1] | **Li Liu**[2] | **Jin Liu**[3] | **Yi Li**[4]

[1]Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, China

[2]School of Mathematics and Statistics, Wuhan University, Wuhan, China

[3]Centre for Quantitative Medicine, Program in Health Services & Systems Research, Duke-NUS Medical School, Singapore

[4]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
Huazhen Lin, Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, China.
Email: linhz@swufe.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 11829101, 11931014, 12171374

Abstract

Functional data analysis has emerged as a powerful tool in response to the ever-increasing resources and efforts devoted to collecting information about response curves or anything that varies over a continuum. However, limited progress has been made with regard to linking the covariance structures of response curves to external covariates, as most functional models assume a common covariance structure. We propose a new functional regression model with covariate-dependent mean and covariance structures. Particularly, by allowing variances of random scores to be covariate-dependent, we identify eigenfunctions for each individual from the set of eigenfunctions that govern the variation patterns across all individuals, resulting in high interpretability and prediction power. We further propose a new penalized quasi-likelihood procedure that combines regularization and B-spline smoothing for model selection and estimation and establish the convergence rate and asymptotic normality of the proposed estimators. The utility of the developed method is demonstrated via simulations, as well as an analysis of the Avon Longitudinal Study of Parents and Children concerning parental effects on the growth curves of their offspring, which yields biologically interesting results.

KEYWORDS

B-spline approximation, functional principal component analysis (FPCA), functional response regression analysis, individual-specific mean and covariance structure, penalized maximum quasi-likelihood estimator

1 | INTRODUCTION

The last two decades have witnessed the emergence of functional data models (Li & Hsing, 2010; Li et al., 2010; Yao & Müller, 2010; Zhong et al., 2020) as powerful tools for analyzing large volumes of functional data collected from diverse fields, ranging from medical studies, speech recognition, biological development, and climatology to online auctions. Heterogeneity has also been commonly observed in functional data, as the means and variations of the observed curves have often been manifested to depend on subjects' characteristics (Jiang & Wang, 2011; Li et al., 2017). Identifying and modeling such heterogeneity has sparked much research interest. For example, several works have modeled heterogeneous functional responses using traditional regression models; Jiang and Wang (2011) proposed a single-index model, and Li et al. (2017) proposed a functional varying-coefficient single-index model (FVCSIM). While both works modeled the covariate-dependent means of functional responses, they

 © 2022 The International Biometric Society.

treated the covariances of functional responses as negligible, ignoring that they can be essential components of functional responses.

However, since functional data analysis is an infinite-dimensional process, dimensionality reduction is the key to effectively analyzing the given data, while capturing the covariances among functional responses embedded in the curves. Functional principal component analysis (FPCA), a covariance-based dimensionality reduction method, has demonstrated its capability of providing parsimonious representations of infinite-dimensional processes and has become the cornerstone for most functional data methods, including those presented in Yao et al. (2005), Li and Hsing (2010), and Zhou et al. (2018). However, all of these methods ignore the heterogeneity that is inherent in functional data and fail to consider external covariates when modeling the covariance structures of functional data.

In fact, as it is difficult to incorporate covariate information into the eigendecomposition procedure, even the studies on covariate-adjusted PCA are limited (Li et al., 2015). Some works have accounted for the effects of covariates on response curves via FPCA. For example, Chiou et al. (2003a) and Chiou et al. (2003b) proposed a functional smooth random effects model (FSREM), while conducting FPCA independently of covariates. Li et al. (2016), Chen et al. (2019), and Backenroth et al. (2018) considered a covariate-adjusted functional PCA by establishing relationships between scores and covariates. Li et al. (2016) and Chen et al. (2019) considered the dependence of a score ξ_{ik} on covariates by assuming that $\xi_{ik} = \boldsymbol{\alpha}_k'\mathbf{X}_i + \epsilon_{ik}$, whereas Backenroth et al. (2018) supposed that $\text{var}(\xi_{ik}|\mathbf{X}_i) = \exp(\mathbf{X}_i'\boldsymbol{\alpha}_k)$. However, these specified links, including linear or exponential links, make the corresponding eigenfunctions either important or unimportant for all individuals by identifying $\boldsymbol{\alpha}_k$ as zero or nonzero; that is, the eigenfunctions are shared by all individuals. To study individual-specific eigenfunctions, Cardot (2006) assumed the dependencies of mean functions, eigenfunctions, and their corresponding FPC scores on a single covariate. Extending Cardot (2006) to accommodate multiple covariates is challenging due to the curse of dimensionality. In addition, modeling covariances using the method proposed by Cardot (2006) may incur considerable instability due to the involvement of a large number of unknown eigenfunctions and score functions that are covariate-dependent.

To address these issues, we propose a new FPCA framework to model covariate-dependent mean and covariance functions. By allowing the covariate-dependent variances of the random scores that are associated with eigenfunctions into the model, for each individual, our procedure determines the magnitude of influence for a common set of eigenfunctions; with further regularization, we identify FPCs for each individual and construct covariate-dependent covariances for the functional responses. Compared with the existing works, our proposal exhibits several strengths. First, with covariate-dependent means and scores, our model allows the score of an eigenfunction to be exactly 0 for a particular subject, providing a more parsimonious and interpretable representation for each individual. Individuals with different eigenfunction sets enable us to better explore their interindividual differences. Second, based again on covariate-dependent means and scores, our model can identify the subsets of covariates that are significant for the mean and covariance, respectively. Third, our method allows for arbitrary functions that link multiple covariates to means and covariances, and therefore, possesses sufficient flexibility and robustness. Fourth, our method naturally accommodates an efficient algorithm combining the alternating direction method of multipliers (ADMMs) and linear approximation techniques, wherein each iterative step presents closed-form expressions or can be implemented using existing packages. Finally, we establish uniform consistency and asymptotic normality, justifying the utility of the proposed estimator in a large sample setting.

By comparing our method with several competing methods on finite samples, we demonstrate its advantageous performance, especially with respect to interpretation. We also apply it to analyze data obtained from the second generation of the Avon Longitudinal Study of Parents and Children (ALSPAC). We demonstrate that our method provides more predictive (Figure 2F) and insightful results than the existing methods; the proposed method identifies that covariates, such as birth weight and diabetes, are associated with the mean and covariance structures of functional responses (Table 4), and eight variables are identified as being associated with the mean (Table 3), most of which are not detected by the existing methods.

The remainder of the paper is organized as follows. Section 2 introduces the proposed method and estimation approach, and Section 3 establishes the uniform consistency and asymptotic normality of this technique. Sections 4 and 5 demonstrate the utility of the proposed method via simulations and an application to the ALSPAC. Section 6 concludes the paper with a brief discussion on further research. Technical proofs and more notations are relegated to the Supporting Information. The R code is available online as Supporting Information.

2 | MODEL AND ESTIMATION PROCEDURE

2.1 | Model

Denote by $\{\mathbf{X}_i, Z_i(\cdot)\}_{(i=1,\ldots,n)}$ n independent and identically distributed (i.i.d.) realizations of $\{\mathbf{X}, Z(\cdot)\}$, where $Z(\cdot)$ is a random function and $\mathbf{X}$ is a p-dimensional vector of covariates. Our goal is to estimate the conditional mean and covariance of $Z_i(\cdot)$ given $\mathbf{X}_i$. To proceed, we note that, without considering $\mathbf{X}_i$, a Karhunen–Loéve expansion (Ash & Gardner, 1978) for $Z_i(t)$ would be

$$Z_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t), \tag{1}$$

where $\mu(t) = E\{Z_i(t)\}$ is the overall mean function; $\phi_k(t)$ is the kth orthonormal eigenfunction of the covariance function $C(s,t) = \mathrm{cov}\{Z_i(s), Z_i(t)\}$, satisfying $\int \phi_k(t)\phi_j(t)dt = 1$ if $j = k$ and 0 otherwise; ξ_{ik} denotes FPC scores with $E(\xi_{ik}) = 0$, $\mathrm{var}(\xi_{ik}) = \rho_k$ and $\mathrm{cov}(\xi_{ij}, \xi_{ik}) = 0$ if $j \neq k$; and ρ_k is the eigenvalue corresponding to the eigenfunction $\phi_k(\cdot)$. Since $\sup_{t\in[0,1]} E\{\sum_{k=1}^{\infty} \xi_{ik}\phi_k(t) - \sum_{k=1}^{K_n} \xi_{ik}\phi_k(t)\}^2 \to 0$, one can reasonably suppose that

$$Z_i(t) \approx \mu(t) + \sum_{k=1}^{K_n} \xi_{ik}\phi_k(t), \tag{2}$$

as $K_n \to \infty$. The approximation of (2) with a fixed K_n is commonly adopted for longitudinal and functional data analysis (see, e.g., Hall and Mohammad, 2006; Yao et al., 2005). To be more flexible, Hall and Mohammad (2006) and Lin et al. (2018) considered model (2) with $K_n \to \infty$ as $n \to \infty$, which is also adopted in this paper.

Then, with $\mathbf{X}_i$, we propose modifying (2) by modeling the dependence of the functional response on covariates via

$$Z_i(t) = \mu(t, \mathbf{X}_i'\boldsymbol{\beta}) + \sum_{k=1}^{K_n} \xi_{ik}\phi_k(t), \tag{3}$$

$$\mathrm{var}(\xi_{ik}|\mathbf{X}_i) = \rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k),\ i = 1, \ldots, n,$$

where $\mu(\cdot,\cdot)$, $\phi_k(\cdot)$ and $\rho_k(\cdot)$ are unknown functions, and $\boldsymbol{\xi}_i = (\xi_{ik}, k = 1, \ldots, K_n)'$ are independent with $E(\boldsymbol{\xi}_i|\mathbf{X}_i) = \mathbf{0}$, $\mathrm{cov}(\boldsymbol{\xi}_i|\mathbf{X}_i) = \boldsymbol{\Lambda}(\mathbf{X}_i)$ and $\boldsymbol{\Lambda}(\mathbf{X}_i) = \mathrm{diag}\{\mathrm{var}(\xi_{ik}|\mathbf{X}_i), k = 1, \ldots, K_n\}$. To maintain interpretability, we require the unknown $\phi_k(t)$, $k = 1, \ldots, K_n$ to be the same across individuals. The model is intuitive because it links the contribution of each functional direction, $\phi_k(\cdot)$, to $Z_i(\cdot)$ via a single-index model $\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)$. Its magnitude may govern the selection of principal components for $Z_i(\cdot)$. When conducting a traditional FPCA, we assume that the eigenfunctions $\phi_k(t)$, $k = 1, \ldots, K_n$ are to be shared by all individuals. In contrast, the proposed model (3) allows individuals to have different sets of eigenfunctions by identifying whether $\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)$ is zero or not. For example, if $|\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)|$ is large, the corresponding component $\phi_k(\cdot)$ is important for individual i when explaining the proportion of variation that is attributable to that direction; if $\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k) = 0$, the component $\phi_k(\cdot)$ is not selected for individual i, indicating one fewer principal component for $Z_i(\cdot)$. Thus, each individual's trajectory can be more fully and parsimoniously represented by its projections onto the functional components selected via $\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)$. This is also observed in our real data analysis, which shows that the individuals satisfying $\mathbf{X}_i'\boldsymbol{\alpha}_2 < -0.3$ are represented by two eigenfunctions $\phi_1(t)$ and $\phi_2(t)$, and the rest of the individuals are represented only by $\phi_1(t)$, where $\boldsymbol{\alpha}$ is displayed in Table 4. Model (3) is termed the functional regression model with individual-specific mean and covariance structures (FRIS).

To make model (3) identifiable, we assume that

(IC) $\|\boldsymbol{\beta}\| = 1$, $\|\boldsymbol{\alpha}_k\| = 1$, and the first nonzero elements of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_k$ are positive for $k = 1, \ldots, K_n$. Denoting $\boldsymbol{\phi}(t) = \{\phi_1(t), \ldots, \phi_{K_n}(t)\}'$, $\phi_k(0) > 0$ for $k = 1, \ldots K_n$. We further assume that $\int \boldsymbol{\phi}(t)\boldsymbol{\phi}(t)'dt = \mathbf{I}_{K_n}$, where $\mathbf{I}_d$ is a $d \times d$ identity matrix and $\|\cdot\|$ denotes the ℓ_2 norm of a vector.

To reflect that a typical functional dataset consists of observations at irregularly spaced locations or time points, we assume that n_i measurements are taken for $Z_i(\cdot)$ at random time points $t_{i1}, \ldots, t_{i,n_i}$, and we observe the random functions $Z_i(\cdot)$ with their measurement errors; that is,

$$Y_i(t_{ij}) = Z_i(t_{ij}) + \epsilon_i(t_{ij}),\ j = 1, \ldots, n_i;\ i = 1, \ldots, n, \tag{4}$$

where $\epsilon_{ij} = \epsilon_i(t_{ij})$ are i.i.d. measurement errors with $E(\epsilon_{ij}) = 0$ and $\mathrm{var}(\epsilon_{ij}) = \sigma^2$. The number of measurements n_i made on the i-th subject is also determined at random to reflect sparse and irregular designs, which are assumed to be i.i.d. and independent of all other random variables.

2.2 | Estimation

Writing $\mathbf{Y}_i = \{Y_i(t_{i1}), \ldots, Y_i(t_{i,n_i})\}'$ and $\mathbf{t}_i = (t_{i1}, \ldots, t_{i,n_i})'$, models (3) and (4) can be expressed as $E(\mathbf{Y}_i) = \boldsymbol{\mu}_i = \mu(\mathbf{t}_i, \mathbf{X}_i'\boldsymbol{\beta}) \hat{=} \{\mu(t_{i1}, \mathbf{X}_i'\boldsymbol{\beta}), \ldots, \mu(t_{i,n_i}, \mathbf{X}_i'\boldsymbol{\beta})\}'$ and $\mathrm{cov}(\mathbf{Y}_i) = \boldsymbol{\Sigma}_i = \sum_{k=1}^{K_n} \phi_k(\mathbf{t}_i)\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)\phi_k(\mathbf{t}_i)' + \sigma^2\mathbf{I}_{n_i}$, where $\phi_k(\mathbf{t}_i) = \{\phi_k(t_{i1}), \ldots, \phi_k(t_{i,n_i})\}'$. Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_{K_n}')'$ and $\boldsymbol{\rho}(u) = \{\rho_1(u), \ldots, \rho_{K_n}(u)\}'$. Denote all of the unknown parameters

and functions by $\pi = (\beta', \alpha', \sigma^2, \mu, \phi', \rho')'$. One might attempt to estimate π by maximizing the log quasi-likelihood function of $\mathbf{Y}_i = \{Y_i(t_{i1}), \ldots, Y_i(t_{i,n_i})\}'$, with $i = 1, \ldots, n$, which is given by

$$L_n(\pi) = -\frac{1}{2n}\sum_{i=1}^{n} \log|\Sigma_i| - \frac{1}{2n}\sum_{i=1}^{n}(\mathbf{Y}_i - \mu_i)'\Sigma_i^{-1}(\mathbf{Y}_i - \mu_i), \quad (5)$$

up to a constant.

However, because $\mu(\cdot)$, $\phi(\cdot)$, and $\rho(\cdot)$ are infinite-dimensional functions, a direct maximization of (5) is difficult. Instead, we propose to adapt more implementable smooth spline techniques to estimate these functions (Chen & Tong, 2010). We approximate them by $\mu(t,u) \approx \gamma'\mathbf{B}_n(t,u)$, $\phi_k(t) \approx \eta_k'\mathbf{B}_{n1}(t)$, and $\rho_k(u) \approx \{\theta_k'\mathbf{B}_{n2}(u)\}^2$, where $\mathbf{B}_n(t,u) = \mathbf{B}_{n1}(t) \otimes \mathbf{B}_{n2}(u)$, $\otimes$ is the Kronecker product, and $\mathbf{B}_{n1}(\cdot) = \{b_{11}(\cdot), \ldots, b_{1,m_{n1}}(\cdot)\}'$ and $\mathbf{B}_{n2}(\cdot) = \{b_{21}(\cdot), \ldots, b_{2,m_{n2}}(\cdot)\}'$ are two sets of spline basis functions. We next define a sieve space as $\Pi_n = \mathcal{A} \times \Theta \times \prod_{k=1}^{K_n}\Theta_{1k} \times \prod_{k=1}^{K_n}\Theta_{2k}$, where $\mathcal{A} = \{(\beta', \alpha', \sigma^2) : (\beta', \alpha', \sigma^2) \in R^{p(1+K_n)+1}\}$, $\Theta = \{\gamma'\mathbf{B}_n(t,u) : t \in [0,T], u \in [-U,U]\}$, $\Theta_{1k} = \{\eta_k'\mathbf{B}_{n1}(t) : t \in [0,T]\}$, $\Theta_{2k} = \{\theta_k'\mathbf{B}_{n2}(u) : u \in [-U,U]\}$.

As eigenfunctions for individual i are selected by identifying whether or not $\rho_k(\mathbf{X}_i'\alpha_k)$ is zero, the estimator of the parameter π belonging to Π_n (or $\hat{\pi}_n$) can be obtained by maximizing

$$Q_n(\pi) = L_n(\pi) - \sum_{k=1}^{K_n}\sum_{i=1}^{n} p_\lambda(|\rho_k(\mathbf{X}_i'\alpha_k)|), \quad (6)$$

subject to the identifiable condition (IC), where the diagonal constraints on $\phi(t)$ are

$$\int_0^T \mathbf{B}_{n1}(t)^{\otimes 2}dt = \frac{1}{m_{n1}}\mathbf{I},\ \eta^{\otimes 2} = m_{n1}\mathbf{I}, \quad (7)$$

with $\eta = (\eta_1, \ldots, \eta_{K_n})'$, where $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}'$ for any vector $\mathbf{a}$. Given a set of spline basis functions $\widetilde{\mathbf{B}}_{n1}(t)$, its orthonormal version satisfying (7) is obtained as follows. Denote by $\mathbf{A} = \int_0^T \widetilde{\mathbf{B}}_{n1}(t)^{\otimes 2}dt$ a positive-definite matrix. It follows that $\int_0^T (\mathbf{A}^{-\frac{1}{2}}\widetilde{\mathbf{B}}_{n1}(t))^{\otimes 2}dt = \mathbf{I}/m_{n1}$; hence, $\mathbf{B}_{n1}(t) = \mathbf{A}^{-\frac{1}{2}}\widetilde{\mathbf{B}}_{n1}(t)$ satisfies (7).

For the choice of $p_\lambda(\cdot)$, the penalty function in (6), we take the smoothly clipped absolute deviation (SCAD) with a reasonable finite-sample performance, which satisfies $\dot{p}_\lambda(\beta) = \lambda I\{\beta \le \lambda\} + I\{\beta > \lambda\}(a\lambda - \beta)_+/(a-1)$ for some $a > 0$ and $\beta > 0$, with $\dot{p}_\lambda(0) = 0$. We choose $a = 3.7$ by following Fan and Li (2001). The penalty is operated on the function $\rho_k(\cdot)$ and encourages it to be zero in some subregions; this is termed local sparsity. Locally sparse methods, including functional linear regression that is interpretable ("FLiRTI") (James et al., 2009), a two-stage method (Zhou et al., 2013) and smooth and locally sparse (SLoS) (Lin et al., 2017), typically identify the null subregions of a curve by splitting the domain into a large number of subintervals. However, adapting these methods to our case is not straightforward because, unlike in these existing methods, the argument of $\rho_k(\cdot)$ in our setting is $\mathbf{X}_i'\alpha_k$, which is individualized and involves an unknown α_k.

Instead, with a spline approximation, we propose to impose the penalty on $|\theta_k'\mathbf{B}_{n2}(\mathbf{X}_i'\alpha_k)|$ and select individual-specific principal components. Particularly, $|\theta_k'\mathbf{B}_{n2}(\mathbf{X}_i'\alpha_k)| = 0$, coupled with $E(\xi_{ik}) = 0$, would imply a zero ξ_{ik} almost surely, and hence, a reduction in the principal component $\phi_k(\cdot)$ for individual i. It is worth mentioning that the penalty on $|\theta_k'\mathbf{B}_{n2}(\mathbf{X}_i'\alpha_k)|$ cannot be replaced by a penalty on $|\theta_k|$. The latter may only identify whether component ϕ_k is important for all individuals and cannot address the question of whether component ϕ_k is important for individual i.

2.3 | Optimization via the ADMM

Since Equation (6) is a nonconcave function with a nonsmooth penalty function, we choose to use the ADMM (Boyd et al., 2011) for optimization. Specifically, let $\zeta_{ik} = \theta_k'\mathbf{B}_{n2}(\mathbf{X}_i'\alpha_k)$. It follows that maximizing (6) is equivalent to minimizing an augmented Lagrangian function:

$$\mathcal{L}_n(\pi, \zeta) = -L_n(\pi, \zeta) + \sum_{k=1}^{K_n}\sum_{i=1}^{n} p_\lambda(|\zeta_{ik}|) + \frac{\nu}{2}\sum_{k=1}^{K_n}\sum_{i=1}^{n}\left[\left\{\zeta_{ik} - \theta_k'\mathbf{B}_{n2}(\mathbf{X}_i'\alpha_k) + \frac{C_{ik}}{\nu}\right\}^2 - c_0\right], \quad (8)$$

where $\zeta = (\zeta_1, \ldots, \zeta_n)'$; $\zeta_i = (\zeta_{i1}, \ldots, \zeta_{i,K_n})'$; $L_n(\pi, \zeta)$ represents $L_n(\pi)$, as defined in (5), with Σ_i replaced by $\mathbf{B}_{n1}(\mathbf{t}_i)'\sum_{k=1}^{K_n}\eta_k\zeta_{ik}^2\eta_k'\mathbf{B}_{n1}(\mathbf{t}_i) + \sigma^2\mathbf{I}_{n_i}$; c_0 is a constant that is independent of π and ζ; and ν is a user-specified step length for updating C_{ik}.

Because the nonlinearity of $\mathbf{B}_{n2}(\cdot)$ and $\mathbf{B}_n(\cdot,\cdot)$ does not lead to closed-form expressions at each ADMM step, we apply the following Taylor expansions:

$$\mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\beta) \approx \mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\widetilde{\beta}) + \dot{\mathbf{B}}_n(\mathbf{t}_i, \mathbf{X}_i'\widetilde{\beta})\mathbf{X}_i'(\beta - \widetilde{\beta}),$$
$$\mathbf{B}_{n2}(\mathbf{X}_i'\alpha_k) \approx \mathbf{B}_{n2}(\mathbf{X}_i'\widetilde{\alpha}_k) + \dot{\mathbf{B}}_{n2}(\mathbf{X}_i'\widetilde{\alpha}_k)\mathbf{X}_i'(\alpha_k - \widetilde{\alpha}_k), \quad (9)$$

where $\dot{\mathbf{B}}_n(t,u) = \partial\mathbf{B}_n(t,u)/\partial u$, $\dot{\mathbf{B}}_{n2}(u) = \partial\mathbf{B}_{n2}(u)/\partial u$ and $\widetilde{\alpha}_k (k = 1, \ldots, K_n)$ and $\widetilde{\beta}$ are the estimators of

$\boldsymbol{\alpha}_k (k = 1, \ldots, K_n)$ and $\boldsymbol{\beta}$, respectively, from the previous step. By substituting $\mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\boldsymbol{\beta})$ and $\mathbf{B}_{n2}(\mathbf{X}_i'\boldsymbol{\alpha}_k)$ in (8) with their Taylor expansions in (9), differentiating (8) with respect to $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, θ_k, and $\boldsymbol{\alpha}_k$, and setting the derivatives to zero, we obtain the following closed-form estimates at each step:

$$\boldsymbol{\beta} = \left\{ \sum_{i=1}^{n} \mathbf{X}_i \boldsymbol{\gamma}' \dot{\mathbf{B}}_n(\mathbf{t}_i, \mathbf{X}_i'\widetilde{\boldsymbol{\beta}}) \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{B}}_n(\mathbf{t}_i, \mathbf{X}_i'\widetilde{\boldsymbol{\beta}})' \boldsymbol{\gamma} \mathbf{X}_i' \right\}^{-1} \times \left[\sum_{i=1}^{n} \mathbf{X}_i \boldsymbol{\gamma}' \dot{\mathbf{B}}_n(\mathbf{t}_i, \mathbf{X}_i'\widetilde{\boldsymbol{\beta}}) \boldsymbol{\Sigma}_i^{-1} \left\{ \mathbf{Y}_i - \mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\widetilde{\boldsymbol{\beta}})' \boldsymbol{\gamma} + \dot{\mathbf{B}}_n(\mathbf{t}_i, \mathbf{X}_i'\widetilde{\boldsymbol{\beta}})' \boldsymbol{\gamma} \mathbf{X}_i' \widetilde{\boldsymbol{\beta}} \right\} \right], \tag{10}$$

$$\boldsymbol{\gamma} = \left\{ \sum_{i=1}^{n} \mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1} \mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\boldsymbol{\beta})' \right\}^{-1} \sum_{i=1}^{n} \left\{ \mathbf{B}_n(\mathbf{t}_i, \mathbf{X}_i'\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1} \mathbf{Y}_i \right\}, \tag{11}$$

$$\theta_k = \left\{ \sum_{i=1}^{n} \mathbf{B}_{n2}(\mathbf{X}_i'\boldsymbol{\alpha}_k) \mathbf{B}_{n2}(\mathbf{X}_i'\boldsymbol{\alpha}_k)' \right\}^{-1} \sum_{i=1}^{n} \left\{ (\widetilde{\zeta}_{ik} + \frac{C_{ik}}{\nu}) \mathbf{B}_{n2}(\mathbf{X}_i'\boldsymbol{\alpha}_k) \right\}, \tag{12}$$

$$\boldsymbol{\alpha}_k = \left[\sum_{i=1}^{n} \mathbf{X}_i \left\{ \theta_k' \dot{\mathbf{B}}_{n2}(\mathbf{X}_i'\widetilde{\boldsymbol{\alpha}}_k) \right\}^2 \mathbf{X}_i' \right]^{-1} \times \left[\sum_{i=1}^{n} \mathbf{X}_i \theta_k' \dot{\mathbf{B}}_{n2}(\mathbf{X}_i'\widetilde{\boldsymbol{\alpha}}_k) \left\{ \widetilde{\zeta}_{ik} + \frac{C_{ik}}{\nu} - \theta_k' \mathbf{B}_{n2}(\mathbf{X}_i'\widetilde{\boldsymbol{\alpha}}_k) + \theta_k' \dot{\mathbf{B}}_{n2}(\mathbf{X}_i'\widetilde{\boldsymbol{\alpha}}_k) \mathbf{X}_i'\widetilde{\boldsymbol{\alpha}}_k \right\} \right], \tag{13}$$

for $k = 1, \ldots, K_n$.

We next estimate $\boldsymbol{\zeta}$. Denoting $\mathbf{C} = (\mathbf{C}_1, \ldots, \mathbf{C}_n)'$, $\mathbf{C}_i = (C_{i1}, \ldots, C_{i,K_n})'$ and $\mathbf{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_n)'$, $\mathbf{W}_i = \{\theta_k' \mathbf{B}_{n2}(\mathbf{X}_i'\boldsymbol{\alpha}_k), k = 1, \ldots, K_n\}'$, with (8), we estimate $\boldsymbol{\zeta}$ by minimizing $-L_n(\boldsymbol{\pi}, \boldsymbol{\zeta}) + \frac{\nu}{2}\|\boldsymbol{\zeta} - \mathbf{W} + \mathbf{C}/\nu\|_F^2 + p_\lambda(|\boldsymbol{\zeta}|)$, where $p_\lambda(|\boldsymbol{\zeta}|) = \sum_{k=1}^{K_n} \sum_{i=1}^{n} p_\lambda(|\zeta_{ik}|)$ and $\|\cdot\|_F$ is the Frobenius norm of a matrix. Denote by $\widetilde{\boldsymbol{\zeta}}$ the estimate of $\boldsymbol{\zeta}$ at the previous step; additionally, $H(\boldsymbol{\zeta}; \boldsymbol{\pi}) = -L_n(\boldsymbol{\pi}, \boldsymbol{\zeta}) + \frac{\nu}{2}\|\boldsymbol{\zeta} - \mathbf{W} + \mathbf{C}/\nu\|_F^2$. We have $H(\boldsymbol{\zeta}; \boldsymbol{\pi}) \le H(\widetilde{\boldsymbol{\zeta}}; \boldsymbol{\pi}) + \dot{H}(\widetilde{\boldsymbol{\zeta}}; \boldsymbol{\pi})'(\boldsymbol{\zeta} - \widetilde{\boldsymbol{\zeta}}) + \frac{1}{2h}(\boldsymbol{\zeta} - \widetilde{\boldsymbol{\zeta}})'(\boldsymbol{\zeta} - \widetilde{\boldsymbol{\zeta}})$, where h is sufficiently small such that the quadratic term dominates the Hessian of $H(\boldsymbol{\zeta}; \boldsymbol{\pi})$. Then, we update $\boldsymbol{\zeta}$ by

$$\boldsymbol{\zeta} = \arg\min_{\boldsymbol{\zeta}} \frac{1}{2h}\|\boldsymbol{\zeta} - (\widetilde{\boldsymbol{\zeta}} - h\dot{H}(\widetilde{\boldsymbol{\zeta}}; \boldsymbol{\pi}))\|_F^2 + p_\lambda(|\boldsymbol{\zeta}|), \tag{14}$$

which can be completed with the *ncvreg* function in R. Finally, we use the gradient descent method to update σ^2 and $\boldsymbol{\eta}_k$. Detailed steps for estimating $\boldsymbol{\pi}$ and $\boldsymbol{\zeta}$ are summarized in Algorithm 1, and the selection of the initial values and tuning parameters is given in the Supporting Information.

In practice, K_n is often chosen to be 2–4, as in Yao et al. (2005), so the computation mainly focuses on inverting $\boldsymbol{\Sigma}_i$, whose dimensionality is the number of time points. It takes less than 1 s to calculate the inverse of a 1000-dimensional matrix with a single two-core personal laptop with 8 GB of RAM. Hence, our algorithm is feasible for functional data with a few hundred time points for each individual, which actually satisfies the requirements for real data or is enough to contain sufficient information for functional data analysis with the smoothing assumption. Particularly, in the case with $n = 100, n_i = 100$, and $K_n = 4$, it takes only 7 s to obtain the estimation via the proposed FRIS.

3 | THEORETICAL PROPERTIES

We are positioned to establish the estimation and selection consistency and the asymptotic normality of the proposed estimator. For the notational ease, we write $m_{n1} = m_{n2} = m_n$, where $m_n = O(n^\upsilon)$. The true value of $\boldsymbol{\pi}$ is denoted as $\boldsymbol{\pi}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\alpha}_0', \sigma_0^2, \mu_0, \boldsymbol{\phi}_0', \boldsymbol{\rho}_0')'$. We specify the following sufficient conditions.

(C1) The covariates $\mathbf{X}$ are bounded.

(C2) $(\boldsymbol{\beta}_0', \boldsymbol{\alpha}_0', \sigma_0^2)' \in \mathcal{A}$, which is a bounded closed set, and the true functions $(\mu_0, \boldsymbol{\phi}_0', \boldsymbol{\rho}_0')' \in \mathcal{H}_{r,2} \times \prod_{k=1}^{K_n} \mathcal{H}_{r,1} \times \prod_{k=1}^{K_n} \mathcal{H}_{r,1}$ with $r > 1$, where

$$\mathcal{H}_{r,d} = \left\{ f(\cdot) : \left| \frac{\partial^l f}{\partial x_1^{a_1} \ldots \partial x_d^{a_d}}(x) - \frac{\partial^l f}{\partial y_1^{a_1} \ldots \partial y_d^{a_d}}(y) \right| \le c\|x - y\|^s, \text{ for any } x, y \in R^d \right\}$$

for $l \in \mathbb{N}_+, s \in (0, 1]$ with $r = l + s$, for any $a = (a_1, \ldots, a_d) \in \mathbb{N}_+^d$ with $\sum_{j=1}^{d} a_j = l$, and for a $c > 0$.

(C3) Denote by $\triangle_1 = \max_{l+1 \le j \le k_n + l + 1} |t_j - t_{j-1}|$ and $\triangle_2 = \min_{l+1 \le j \le k_n + l + 1} |t_j - t_{j-1}|$ the maximum and the minimum spacing of knots, respectively. We assume that $\triangle_1 = O(n^{-\upsilon})$ with $\upsilon \in (0, 0.5)$, and $\triangle_1 / \triangle_2$ is bounded.

(C4) The penalty function $p_\lambda(t)$ is nondecreasing and concave on $[0, \infty)$. There exists a constant b such that $p_\lambda(t)$ is a constant for all $t \ge b\lambda$. Additionally, $\dot{p}_\lambda(0+) = O(\lambda)$.

(C5) $K_n = n^\tau$ with $\tau \le \min(1 - \upsilon, 2\upsilon r)$.

ALGORITHM 1

1: Give initial values $\boldsymbol{\beta}^{(0)\prime}, \boldsymbol{\gamma}^{(0)\prime}, \boldsymbol{\eta}_k^{(0)\prime}, \boldsymbol{\alpha}_k^{(0)\prime}, \boldsymbol{\theta}_k^{(0)\prime}, \sigma^{2(0)}$, and $\zeta_{ik}^{(0)} = \boldsymbol{\theta}_k^{(0)\prime}\mathbf{B}_{n2}(\mathbf{X}_i'\boldsymbol{\alpha}_k^{(0)}), i = 1,\ldots,n; k = 1,\ldots,K_n$, $\mathbf{C}^{(0)} = 0$.

2: Set the step length κ, h, ν and the tuning parameters λ and K_n.

3: **repeat**

4: At the $(t+1)$-th iteration, update $\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\theta}_k, \boldsymbol{\zeta}$ by (10)-(14), where $\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\alpha}}, \boldsymbol{\gamma}, \boldsymbol{\theta}_k, \widetilde{\boldsymbol{\zeta}}$ and $\mathbf{C}$ on the right sides of (10)-(14) are replaced by the estimators from the $t-$th iteration.

5: $\boldsymbol{\eta}_k^{(t+1)} = \boldsymbol{\eta}_k^{(t)} - \kappa\partial\mathcal{L}_n(\boldsymbol{\pi}^{(t)}, \boldsymbol{\zeta}^{(t)})/\partial\boldsymbol{\eta}_k, \quad \text{for} \quad k = 1,\ldots,K_n,$

6: $\sigma^{2(t+1)} = \sigma^{2(t)} - \kappa\partial\mathcal{L}_n(\boldsymbol{\pi}^{(t)}, \boldsymbol{\zeta}^{(t)})/\partial\sigma^2,$

7: $\mathbf{C}^{(t+1)} = \mathbf{C}^{(t)} + \nu(\boldsymbol{\zeta}^{(t+1)} - \mathbf{W}^{(t+1)}).$

8: **until** meeting the convergence criterion

Conditions (C1) and (C2) are commonly assumed in the nonparametric literature (Chen & Tong, 2010). Condition (C1) is a technical condition for controlling the tail of random variables. Condition (C2) ensures the smoothness of the true functions. Condition (C3) facilitates spline analysis with uniformly distributed knots. Condition (C4) can be satisfied by some generally used penalties, including the SCAD penalty, we use for our proposal. Condition (C5) restricts the number of components to ensure the estimability of the model. In the FPCA literature, K_n is often chosen to be finite, automatically guaranteeing that this condition holds (see, e.g., Hall and Mohammad, 2006; Yao et al., 2005).

We denote $\rho_{ik} = \rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)$, $\rho_{ik0} = \rho_{k0}(\mathbf{X}_i'\boldsymbol{\alpha}_{k0})$, and $\mathcal{O} = \{(i,k) : \rho_{ik0} \neq 0\}$. We define a mapping $T : \rho_{ik} \to \rho_{ik}^{or}$, where $\rho_{ik}^{or} = T(\rho_{ik}) = \rho_{ik}$ if $(i,k) \in \mathcal{O}$ and 0 otherwise. Then, with the notation of $\rho_{ik}^{or} = \rho_k^{or}(\mathbf{X}_i'\boldsymbol{\alpha}_k^{or})$, we define $\mathcal{A}^{or} = \{(\boldsymbol{\beta}', \boldsymbol{\alpha}^{or\prime}, \sigma^2) : (\boldsymbol{\beta}', \boldsymbol{\alpha}^{or\prime}, \sigma^2) \in R^{p(1+K_n)+1}, \boldsymbol{\rho}_{\mathcal{O}^c} = 0\}$, $\mathcal{H}_{r,1}^{or} = \{\boldsymbol{\rho} : \boldsymbol{\rho} \in \prod_{k=1}^{K_n}\mathcal{H}_{r,1}, \boldsymbol{\rho}_{\mathcal{O}^c} = 0\}$, and an oracle parameter space $\boldsymbol{\Theta}_{2n}^{or} = \{\boldsymbol{\rho} : \boldsymbol{\rho} \in \prod_{k=1}^{K_n}\boldsymbol{\Theta}_{2k}, \boldsymbol{\rho}_{\mathcal{O}^c} = 0\}, \boldsymbol{\Pi}_n^{or} = \mathcal{A}^{or} \times \boldsymbol{\Theta} \times \prod_{k=1}^{K_n}\boldsymbol{\Theta}_{1k} \times \boldsymbol{\Theta}_{2n}^{or}$.

To establish the asymptotic properties, we first consider an oracle estimator (defined as $\widehat{\boldsymbol{\pi}}_n^{or} = \arg\max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_n^{or}} L_n(\boldsymbol{\pi})$), the estimator of parameter $\boldsymbol{\pi}$ in $\boldsymbol{\Pi}_n^{or}$. Since we focus only on the asymptotic normality of $\boldsymbol{\vartheta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')$, we rewrite $\widehat{\boldsymbol{\pi}}_n^{or}$ as $(\widehat{\boldsymbol{\vartheta}}_n^{or\prime}, \widehat{\boldsymbol{\psi}}_n^{or\prime})'$, where $\widehat{\boldsymbol{\vartheta}}_n^{or} = (\widehat{\boldsymbol{\beta}}_n^{or\prime}, \widehat{\boldsymbol{\alpha}}_n^{or\prime})'$ and $\widehat{\boldsymbol{\psi}}_n^{or} = (\widehat{\sigma}_n^{or2}, \widehat{\mu}_n^{or}, \widehat{\boldsymbol{\phi}}_n^{or\prime}, \widehat{\boldsymbol{\rho}}_n^{or\prime})'$. We use $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ to denote the largest and the smallest eigenvalues of a matrix $\mathbf{A}$, respectively. For any vector $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_n)' \in R^n$, a function $f(t)$, and $\boldsymbol{\pi}$ and $\widetilde{\boldsymbol{\pi}}$, we define $\|\boldsymbol{\zeta}\| = \sqrt{\sum_{i=1}^n \zeta_i^2}$, $\|f\| = (\int_{t\in\mathcal{T}} f(t)^2 dt)^{1/2}$, and $\|\boldsymbol{\pi} - \widetilde{\boldsymbol{\pi}}\| = \{\|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\|^2 + (\sigma^2 - \widetilde{\sigma}^2)^2 + \|\mu - \widetilde{\mu}\|^2 + \sum_{k=1}^{K_n}(\|\boldsymbol{\alpha}_k - \widetilde{\boldsymbol{\alpha}}_k\|^2 + \|\phi_k - \widetilde{\phi}_k\|^2 + \|\rho_k - \widetilde{\rho}_k\|^2)\}^{1/2}$.

Theorem 1 (Consistency and convergence rate of oracle estimators). *Under conditions (C1)–(C5), denoting* $\delta_n = n^{-(1-2\upsilon)/2} + \sqrt{K_n}n^{-(1-\upsilon)/2} + \sqrt{K_n}n^{-\upsilon r}$, *we have*

$$\|\widehat{\boldsymbol{\pi}}_n^{or} - \boldsymbol{\pi}_0\| = O_p(\delta_n).$$

The first and second terms in δ_n, corresponding to the estimation error for $\mu(t,u)$ and $2K_n$ univariate functions $(\phi_k, \rho_k), k = 1,\ldots,K_n$, respectively, are related to the spline order n^υ and the structural parameter $K_n = n^\tau$ of the model (3); the last term in δ_n is the approximation error for using spline functions to approximate nonparametric functions. When K_n does not vary with n, that is, $\tau = 0$, Theorem 1 implies that $\|\widehat{\boldsymbol{\pi}}_n^{or} - \boldsymbol{\pi}_0\| = O_p(n^{-r/(2r+2)})$ with $\upsilon = 1/(2r+2)$, in which case, the convergence rate δ_n is indeed the optimal rate for approximating a nonparametric function (Stone, 1980).

Theorem 2 (Asymptotic normality of oracle estimators). *Denote by* $I(\vartheta_0) = \mathbb{P}\{l^*(\pi_0)\}^{\otimes 2}$ *and* $\Lambda = \lambda_{\min}\{I(\vartheta_0)\}$, *where* $l^*(\pi_0)$ *is defined as in the Appendix. Under conditions (C1)–(C5), if* $0 < \upsilon < 1/4, \tau < \min\{1/2 - \upsilon, 2\upsilon(r-1), \upsilon(2r-1)/2\}$ *and* $n^{\tau-1/2}/\Lambda = o_p(1)$ *for* $r > 1$, *for any vector* $\mathbf{u}$ *with* $||\mathbf{u}|| = 1$, *as* $n \to \infty$, *we have*

$$\sqrt{n}\mathbf{u}' I(\vartheta_0)^{1/2}(\widehat{\vartheta}_n^{or} - \vartheta_0) \xrightarrow{d} N(0,1).$$

Theorem 2 leads to the selection consistency and the asymptotic normality of $\widehat{\vartheta}_n$, as summarized by the following.

Theorem 3 (Oracle properties). *Under conditions (C1)–(C5), if* $\lambda_{max}\{\mathbb{P}\partial^2 L_n(\pi_0)/\partial\rho\partial\rho'\}$ *is finite,* $\inf_{(i,k)\in\mathcal{O}} |\rho_{ik0}| \geq b\lambda$ *and* $\lambda \gg \delta_n$ *for some constant* $b > 0$, *we have*

(1) $P(\widehat{\pi}_n = \widehat{\pi}_n^{or}) \to 1$;
(2) $\|\widehat{\pi}_n - \pi_0\| = O_p(\delta_n)$, *where* δ_n *is defined as in Theorem 2;*
(3) Under the conditions in Theorem 2, we have

$$\sqrt{n}\mathbf{u}' I(\vartheta_0)^{1/2}(\widehat{\vartheta}_n - \vartheta_0) \xrightarrow{d} N(0,1)$$

for any vector $\mathbf{u}$ *with* $\|\mathbf{u}\| = 1$.

The asymptotic normality of $\widehat{\vartheta}_n$ given by Theorem 3 lays the groundwork for inference; however, it is not feasible to draw an inference by directly using the theoretical covariance, which involves infinite-dimensional unknown parameters. We use bootstrap resampling to estimate the covariance of $\widehat{\vartheta}_n$. Denote by $\widehat{\vartheta}_n^* = (\widehat{\boldsymbol{\beta}}_n^{*\prime}, \widehat{\boldsymbol{\alpha}}_n^{*\prime})'$ the bootstrap estimator of ϑ. We show that $\widehat{\vartheta}_n^*$ has the same asymptotic distribution as the proposed estimator.

Theorem 4 (Distribution consistency of bootstrap estimators). *Under conditions (C1)–(C5), if* $\tau < 1/2 - \upsilon$, *we have that for any k,*

$$\sup_{x\in R^p} |P\left\{\sqrt{n}(\widehat{\boldsymbol{\beta}}_n^* - \widehat{\boldsymbol{\beta}}_n) \leq x\right\} - P(\sqrt{n}\left\{\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \leq x\right\}| = o_p(1),$$

$$\sup_{x\in R^p} |P\left\{\sqrt{n}(\widehat{\alpha}_{nk}^* - \widehat{\alpha}_{nk}) \leq x\right\} - P\left\{\sqrt{n}(\widehat{\alpha}_{nk} - \alpha_{k0}) \leq x\right\}| = o_p(1),$$

where the inequalities are taken in a componentwise manner.

4 | SIMULATION

We assess the finite-sample performance of the proposed FRIS by comparing it with two competing methods, namely, the FSREM (Chiou et al., 2003b) and the method with a covariate-dependent mean structure and a covariate-independent covariance structure, that is, the same score variance (SSV) across individuals. We vary $n =$ 100 to 500 and $n_i = 10$ to 20. For each $1 \leq i \leq n$, observation time points $\mathbf{t}_i$ are randomly sampled from $U(0,1)$, and the covariates $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3})'$ are independently generated from the uniform distribution $U(-1,1)$. For a fair comparison, we consider three scenarios, each following the assumptions of the FRIS, SSV, and FSREM. We compare the methods by using the bias, standard deviation (sd), and root-mean-square error (RMSE) criteria, which is defined as

$$bias = \left[\frac{1}{n_{grid}}\sum_{i=1}^{n_{grid}}\{E\widehat{f}(x_i) - f(x_i)\}^2\right]^{1/2}, sd$$

$$= \left[\frac{1}{n_{grid}}\sum_{i=1}^{n_{grid}} E\{\widehat{f}(x_i) - E\widehat{f}(x_i)\}^2\right]^{1/2},$$

$RMSE = (bias^2 + sd^2)^{1/2}$, where $x_i (i = 1, \ldots, n_{grid})$ are the grid points at which $f(\cdot)$ is estimated, $E\widehat{f}(x_i)$ is approximated by the sample mean, and $n_{grid} = 100$. The predictive performance is assessed by the normalized mean squared error (NMSE) of the out-of-sample period, which is defined as $NMSE = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}(\widehat{Y}_i(t_{ij}) - Y_i(t_{ij}))^2}{\sum_{i=1}^{n}\sum_{j=1}^{n_i}(\overline{Y} - Y_i(t_{ij}))^2}$, where $\overline{Y} = \frac{1}{\sum_{i=1}^{n} n_i}\sum_{i=1}^{n}\sum_{j=1}^{n_i} Y_i(t_{ij})$ and $\{\mathbf{X}_i, Z_i(\cdot), Y_i\}, i = 1, \ldots, n$ are independent of the samples used to predict Y_i. The NMSE is a unitless error measure, where a smaller NMSE indicates better performance. In addition, the nominal Type-1 error rates and power for the $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ of the proposed FRIS can also be assessed sbecause inference is required to identify the subsets of covariates that are significant for the mean and covariance, respectively. For each simulation configuration, a total of $N = 500$ datasets are independently generated.

Example 1. We generate $\mathbf{Y}_i = \{Y_i(t_{i1}), \ldots, Y_i(t_{i,n_i})\}'$ from a model satisfying the assumptions of the FRIS: $E(\mathbf{Y}_i|\mathbf{X}_i) = \boldsymbol{\mu}_i(\mathbf{X}_i) = \mu(\mathbf{t}_i, \mathbf{X}_i'\boldsymbol{\beta}), \mathrm{cov}(\mathbf{Y}_i|\mathbf{X}_i) = \boldsymbol{\Sigma}_i$, where $\mu(\mathbf{t}_i, u) = (\mu(t_{i1}, u), \ldots, \mu(t_{i,n_i}, u))'$, $\boldsymbol{\Sigma}_i = \sum_{k=1}^{3} \phi_k(\mathbf{t}_i)\rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)\phi_k(\mathbf{t}_i)' + \sigma^2\mathbf{I}_{n_i}$ with $\sigma^2 = 1$, $\mu(t,u) = 10 \times \{u \cdot \cos(t) + (1-u)\cdot\sin(t)\}$, $\phi_k(\mathbf{t}_i) = \{\phi_k(t_{i1}), \ldots, \phi_k(t_{i,n_i})\}'$ with $\phi_1(t) = \sqrt{2}\cos(\pi t), \phi_2(t) = \sqrt{2}\sin(\pi t)$, $\phi_3(t) = \sqrt{2}\cos(3\pi t)$, and $\rho_k(u) = 10^{2-k}u^2 I(u < 0), k = 1, 2, 3$. We set $\boldsymbol{\beta} = (0.2, 0.8, 0.6)'$, $\boldsymbol{\alpha}_1 = (0.8, 0.1, 0.4)'$, $\boldsymbol{\alpha}_2 = (0.2, 0.6, 0.8)'$, and $\boldsymbol{\alpha}_3 = (0.5, 0.8, 0.3)'$. In this example, we consider two kinds of distributions for $\mathbf{Y}_i$: (1) Normal: $\mathbf{Y}_i \sim N\{\boldsymbol{\mu}_i(\mathbf{X}_i), \boldsymbol{\Sigma}_i\}$; (2) Mixture Normal: $\boldsymbol{\Sigma}_i^{-1/2}\{\mathbf{Y}_i - \boldsymbol{\mu}_i(\mathbf{X}_i)\} \sim \frac{1}{2}N(-\mathbf{1}/2, \mathbf{I}) + \frac{1}{2}N(\mathbf{1}/2, \mathbf{I})$;

TABLE 1 Comparisons of FRIS and SSV under Example 1; presented are bias (sd)

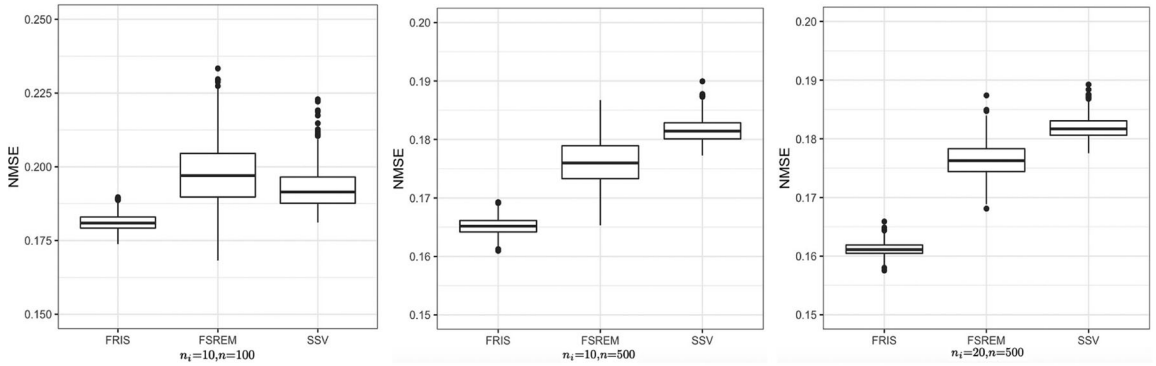
			Normal		Mixture normal	
			FRIS	SSV	FRIS	SSV
$n_i = 10$	$n = 100$	β_1	0.0010(0.0117)	0.0023(0.0163)	0.0003(0.0147)	0.0036(0.0244)
		β_2	0.0004(0.0069)	0.0008(0.0102)	0.0005(0.0084)	0.0022(0.0152)
		β_3	0.0007(0.0076)	0.0007(0.0118)	0.0009(0.0093)	0.0009(0.0167)
		$\mu(\cdot,\cdot)$	0.0152(0.2844)	0.0247(0.4859)	0.0164(0.4131)	0.0369(0.5704)
		$\rho_1(\cdot)$	0.1435(0.4244)	4.3897(2.2415)	0.1673(0.4384)	4.4381(2.5859)
		$\rho_2(\cdot)$	0.0281(0.1374)	0.4939(0.2575)	0.0346(0.1369)	0.5228(0.2598)
		$\rho_3(\cdot)$	0.0054(0.0212)	0.0442(0.0444)	0.0071(0.0248)	0.0483(0.0481)
	$n = 500$	β_1	0.0010(0.0058)	0.0016(0.0113)	0.0009(0.0066)	0.0011(0.0129)
		β_2	0.0004(0.0031)	0.0005(0.0080)	0.0003(0.0039)	0.0007(0.0086)
		β_3	0.0002(0.0032)	0.0007(0.0085)	0.0002(0.0043)	0.0003(0.0094)
		$\mu(\cdot,\cdot)$	0.0056(0.1204)	0.0163(0.2458)	0.0065(0.1337)	0.0395(0.4699)
		$\rho_1(\cdot)$	0.1409(0.4134)	4.3890(2.2071)	0.1518(0.4239)	3.9318(2.3274)
		$\rho_2(\cdot)$	0.0230(0.1245)	0.4714(0.2311)	0.0231(0.1349)	0.5212(0.2451)
		$\rho_3(\cdot)$	0.0051(0.0207)	0.0436(0.0440)	0.0064(0.0215)	0.0521(0.0495)
$n_i = 20$	$n = 100$	β_1	0.0008(0.0109)	0.0016(0.0128)	0.0002(0.0149)	0.0017(0.0248)
		β_2	0.0001(0.0054)	0.0004(0.0083)	0.0005(0.0082)	0.0016(0.0142)
		β_3	0.0002(0.0056)	0.0006(0.0099)	0.0010(0.0084)	0.0006(0.0156)
		$\mu(\cdot,\cdot)$	0.0136(0.3247)	0.0215(0.4428)	0.0157(0.3521)	0.0299(0.4783)
		$\rho_1(\cdot)$	0.1429(0.4241)	4.4168(2.2785)	0.1753(0.4607)	4.4328(2.5274)
		$\rho_2(\cdot)$	0.0306(0.1405)	0.4759(0.2023)	0.0257(0.1357)	0.5312(0.2351)
		$\rho_3(\cdot)$	0.0056(0.0215)	0.0466(0.0433)	0.0069(0.0229)	0.0520(0.0495)
	$n = 500$	β_1	0.0006(0.0050)	0.0005(0.0061)	0.0009(0.0070)	0.0009(0.0100)
		β_2	0.0002(0.0029)	0.0003(0.0054)	0.0006(0.0039)	0.0007(0.0068)
		β_3	0.0001(0.0030)	0.0002(0.0069)	0.0005(0.0042)	0.0009(0.0077)
		$\mu(\cdot,\cdot)$	0.0059(0.0958)	0.0133(0.2534)	0.0062(0.1125)	0.0152(0.2814)
		$\rho_1(\cdot)$	0.1408(0.4126)	4.1304(1.9150)	0.1511(0.4431)	4.3713(2.2832)
		$\rho_2(\cdot)$	0.0264(0.1319)	0.4604(0.1676)	0.0232(0.1355)	0.5265(0.2219)
		$\rho_3(\cdot)$	0.0048(0.0205)	0.0438(0.0428)	0.0058(0.0219)	0.0533(0.0481)

Let $m_{n1} = m_{n2} = 6$ and $K_n = 4$; Tables 1 and S1 show that the proposed estimator presents smaller biases and variances than the SSV for both kinds of distributions; the SSV is seriously biased for $\rho_k(\cdot)$ because it ignores its heteroskedasticity. The performance of the proposed approach improves as n or n_i increases. Figures 1(A) and 1(B), which display the NMSEs of the FRIS, SSV, and FSREM, further reveal that the proposed method achieves the best predictive performance in all of the examined cases. Of note, the FSREM performs better than the SSV when $n = 500$, but worse than the SSV when $n = 100$, possibly because the FSREM requires a larger sample size due to the increased model complexity.

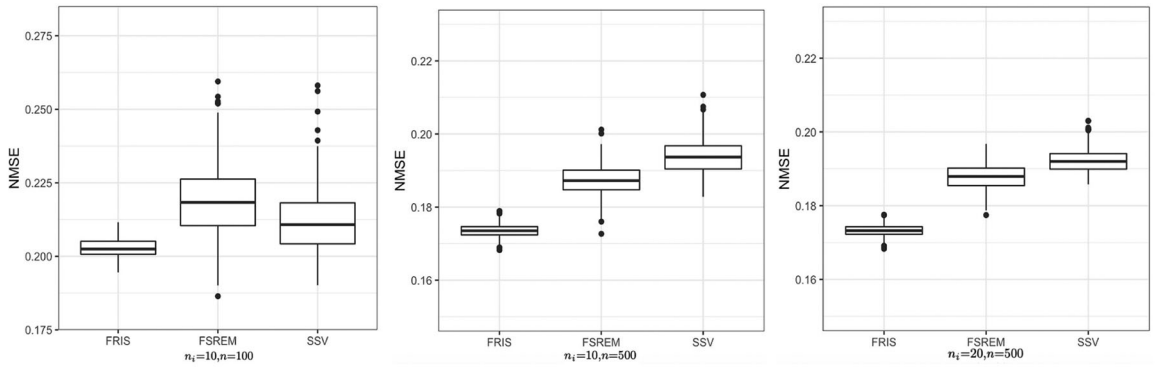
Table 2 summarizes the selection results of the eigenfunctions under Example 1, where the false positive rate (FPR) records the fraction of the individuals that are incorrectly selected, the false negative rate (FNR) corresponds to the fraction of the individuals that are incorrectly excluded for an eigenfunction, and the accuracy (ACC) reflects the fraction of the individuals whose eigenfunctions are correctly identified. With small FPRs, FNRs and large ACCs for $\phi_k, k = 1, 2, 3$, as in Table 2, it appears that the FRIS can correctly identify the eigenfunction contributions for each individual with high probability under both normal and mixed normal distributions.

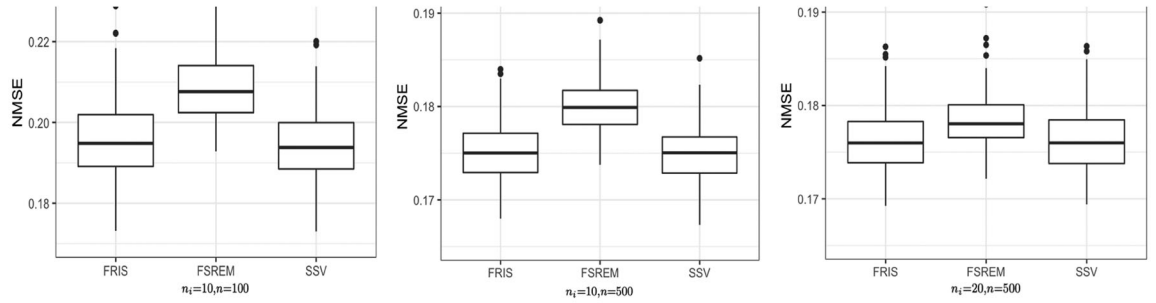
In addition, Figures S1–S3 in the Supporting Information display the estimates of the eigenfunctions $\phi_k(t)$, the score variances $\rho_k(u), k = 1, 2, 3$, and the mean function $\mu(t, u)$ to evaluate the performance of the nonparametric estimators. The proposed method performs well in that the estimated functions are close to the true functions, and its confidence bands completely cover the true functions.

Example 2. To assess the Type-1 error rates and power of the FRIS for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, we generate data in the same way as in Example 1(1) except that $\boldsymbol{\beta} = (0.8, 0, 0.6)$,

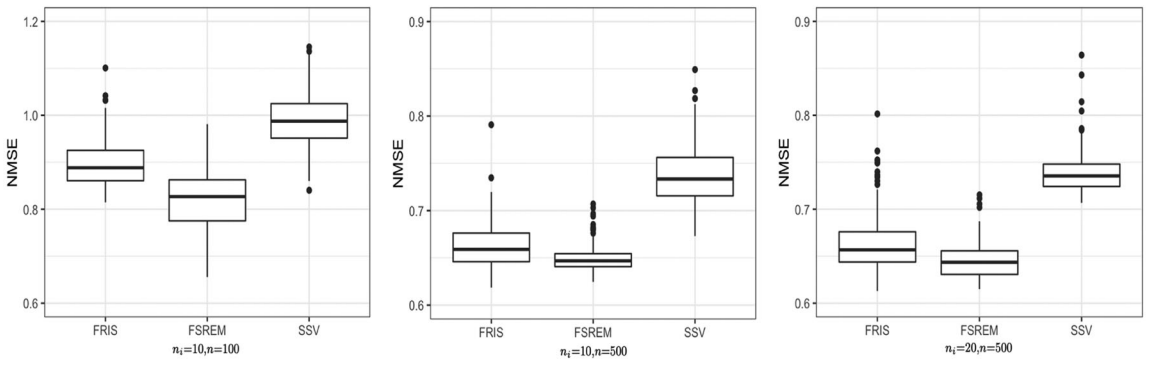


(A) Example 1(1): Normal FRIS data





(C) Example 3: SSV data



(D) Example 4: FSREM data

FIGURE 1 NMSE of three methods: FRIS, SSV, and FSREM for Examples 1, 3, and 4

TABLE 2 The selection results of the eigenfunctions under Example 1; presented are mean (sd)

		Normal				Mixture normal			
		$n_i = 10$		$n_i = 20$		$n_i = 10$		$n_i = 20$	
		$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$
ϕ_1	FPR	0.0398(0.0837)	0.0502(0.1010)	0.0549(0.0929)	0.0523(0.0963)	0.0475(0.0950)	0.0773(0.1628)	0.0626(0.1030)	0.0630(0.1074)
	FNR	0.0291(0.0291)	0.0398(0.0426)	0.0277(0.0302)	0.0369(0.0408)	0.0282(0.0290)	0.0421(0.0434)	0.0281(0.0297)	0.0413(0.0443)
ϕ_2	FPR	0.1067(0.1635)	0.0959(0.1583)	0.1024(0.1664)	0.1021(0.1643)	0.0920(0.1573)	0.1018(0.1576)	0.0958(0.1591)	0.0987(0.1542)
	FNR	0.0572(0.0635)	0.0641(0.0724)	0.0520(0.0604)	0.0559(0.0647)	0.0634(0.0657)	0.0578(0.0664)	0.0517(0.0598)	0.0572(0.0677)
ϕ_3	FPR	0.0663(0.1425)	0.0708(0.1460)	0.0805(0.1538)	0.0746(0.1471)	0.0834(0.1551)	0.0649(0.1389)	0.0829(0.1574)	0.0656(0.1375)
	FNR	0.0832(0.0881)	0.0885(0.0928)	0.0804(0.0894)	0.0783(0.0907)	0.0797(0.0923)	0.0807(0.0927)	0.0874(0.0991)	0.0806(0.0943)
ACC		0.8755(0.0532)	0.8503(0.0543)	0.8722(0.0564)	0.8577(0.0551)	0.8680(0.0561)	0.8603(0.0531)	0.8686(0.0551)	0.8565(0.0547)

$\boldsymbol{\alpha}_k = (0, 0.9, 0.4), \rho_k(u) = 10^{-k}u^2 I(u < 0), k = 1, 2, 3$ and $\mu(t, u) = 2 \times \{u \cdot \cos(t) + (1 - u) \cdot \sin(t)\}$. Table S2 in the Supporting Information shows the calculated Type-1 error rates and power levels obtained under a significance level of 0.05, suggesting that the Type-1 error rates of all parameters are controlled at approximately 0.05, and the power tends to 1 as the sample size increases. It should be mentioned that some powers remain at 1 due to the identification condition that β_1, α_{12}, α_{22}, and α_{32} are positive.

Example 3. We generate data with a common score variance, satisfying the assumption of the SSV. Specifically, we generate data in the same way as in Example 1(1), except that $\rho_k(u) = \rho_k$ for $k = 1, 2, 3$ and $\rho_1 = 5, \rho_2 = 1, \rho_3 = 0.5$. Table S3 presents the biases and empirical standard deviations produced for the parametric and nonparametric estimates obtained by the FRIS and SSV, whereas Figure 1(C) shows the NMSEs obtained for the FRIS, SSV, and FSREM estimators. Table S3 and Figure 1(C) indicate that, in general, the FRIS performs as well as the SSV and is better than the FSREM in terms of the NMSE metric. This result is reasonable because the SSV correctly uses the information of constant score variances, whereas the FRIS estimates score variances as nonparametric functions of covariates.

Example 4. We generate data following Chiou et al. (2003b). That is, given covariates $\mathbf{X}_i$, $Y_i(t)$ follows a normal distribution, $Y_i(t) = \mu(t) + \sum_{k=1}^{3} A_{ik}\phi_k(t)$, and we assume that the observed random curves are conditional on the covariates,

$$E\{Y_i(t)|\mathbf{X}_i\} = \mu(t) + \sum_{k=1}^{3} E(A_{ik}|\mathbf{X}_i)\phi_k(t), \ \operatorname{cov}\{Y_i(s), Y_i(t)|\mathbf{X}_i\} = \sum_{k=1}^{3} \operatorname{var}(A_{ik}|\mathbf{X}_i)\phi_k(s)\phi_k(t),$$

where $E(A_{ik}|\mathbf{X}_i) = \mu_k(\mathbf{X}_i'\boldsymbol{\beta}_k)$, $\operatorname{var}(A_{ik}|\mathbf{X}_i) = \rho_k(\mathbf{X}_i'\boldsymbol{\alpha}_k)$, and $\mu(t) = t^2 + 1$; $\mu_1(u) = 1 - \cos(u \cdot \pi), \mu_2(u) = \{1 - \cos(u \cdot \pi)\}/5, \mu_3(u) = \{1 - \cos(u \cdot \pi)\}/10$, $\rho_k(u) = \sqrt{\alpha_k(u)}, k = 1, 2, 3$, and $\phi_k(t)$ are the same as in Example 1, and $\boldsymbol{\beta}_k = (0.8, 0, 0.6)', \boldsymbol{\alpha}_k = (0, 1, 0)'$ for each k. These settings imply that X_{i1} and X_{i3} are associated with the mean, and the second covariate X_{i2} is related to the covariance of the functional response. Figure 1(D) shows the NMSEs obtained for the FRIS, SSV, and FSREM estimators under Example 4. Since the data are generated by following the assumptions of the FSREM, it is not surprising that the FSREM has the lowest NMSE among all three methods. However, the proposed FRIS is comparable to the FSREM and is better than the SSV for the FSREM

data. We also show the p-values obtained by the FRIS for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in Table S4 in the Supporting Information, which suggests that the mean significantly depends on X_{i1} and X_{i3} and that the covariance is associated with X_{i2} under a significance level of 0.05. This means that the Type-1 error can be effectively controlled by the FRIS even for the FSREM data.

In summary, by examining a variety of scenarios, the proposed method clearly has the best performance if the required assumptions are satisfied and still achieves a reasonable performance when such assumptions are violated, which might pave the way for its real-world application.

5 | ANALYSIS OF THE ALSPAC

The ALSPAC, known as Children of the 1990s, is a birth cohort study based in England. Between 1991 and 1992, 14,000 pregnant women were recruited; they, along with their children arising from the pregnancies and their partners, were followed up intensively over two decades, with the goal of investigating the environmental and genetic factors that affect a person's health and development. We apply our proposed method to analyze a dataset from the ALSPAC, and study a variety of pregnancy and maternal factors that may influence the growth trajectories of children. Specifically, the functional responses in our analysis are the body mass index (BMI) curves of children measured from 0 to 7–24 years, and the nine covariates include the birth weight, birth length, presence of maternal gestational diabetes, amniocentesis noted during pregnancy, number of children previously delivered by a mother, and method of delivery. Only one child of a single mother was included in this study. For the method of delivery, spontaneous delivery is set as the baseline, whereas other methods (e.g., assisted breech, caesarean section, and forceps delivery) are coded as distinct dummy variables. After conducting quality control and removing the subjects with missing values, we obtain 7313 individuals for the data analysis. We further center the birth weight and birth length values to better interpret their coefficients and the intercept.

We analyze the data by using our proposed FRIS. For comparisons, we also model the data by using the SSV and FSREM. We choose $m_{n1} = 4$, $m_{n2} = 5$, $K_n = 4$, $\kappa = 0.001$, $\nu = 0.1$, and $h = 0.01$ and select the tuning parameters using the generalized cross-validation (GCV) criterion . The GCV results obtained for the tuning parameter λ are shown in Figure 2(E), which suggests that $\lambda = 0.158$ might be the best choice. We draw inferences based on 200 bootstrap replications. Table 3 compares the FRIS, SSV, and FSREM in terms of their point estimates and p-values for $\boldsymbol{\beta}$, which accounts for covariate–mean relationships. Figure S4(a) and Figure S4(b) of the Supporting Information show the mean estimates of $\mu(t, u)$ for the FRIS and SSV, respectively. Table 3 shows that our proposed FRIS gives smaller p-values, suggesting that it may be more efficient than the SSV and the FSREM; that is consistent with the simulation findings. The FRIS identifies three additional significant covariates at the significance level of 0.05, which are not detected by either the SSV or the FSREM. Table 3 also shows that the birth weight, birth length, maternal gestational diabetes, amniocentesis noted during pregnancy, number of children previously delivered by a mother, and method of delivery covariates are all strongly associated with the mean childhood BMI.

We elaborate on the results obtained by the FRIS, which may have biological implications. For example, compared with spontaneous delivery, children delivered by assisted breech tend to have lower BMIs, whereas children delivered by caesarean section and vacuum extraction tend to have higher BMIs, confirming the impact of caesarean section on the risk of children becoming overweight and obese, as reported in Ireland (Masukume et al., 2018). Indeed, caesarean section is identified by the FRIS to be the most predominant risk factor for an increased BMI, whereas its effect is barely significant when using the SSV method. Furthermore, our FRIS detects that birth weight and maternal gestational diabetes are two significant risk factors with large effects on increased childhood BMI; this is consistent with the results of two recent studies (Wang et al., 2018). These two effects are also detected by the SSV but at less significant levels. Moreover, the FRIS identifies that amniocentesis during pregnancy is negatively correlated with childhood BMI. This makes sense biologically because amniocentesis, a prenatal test for diagnosing birth defects, carries high risks of procedure-related pregnancy loss, preterm labor, and respiratory distress (Seeds, 2004). Last, the FRIS identifies that birth order is related to childhood BMI, coinciding with Ochiai et al. (2012), who found that being the youngest child was associated with childhood obesity.

Table 4 reveals that our proposed FRIS gives much smaller p-values, again suggesting its efficiency compared to that of the FSREM. In particular, the FRIS identifies five additional significant covariates that are not detected by the FSREM at the significance level of 0.05 for covariance. Table 4 shows that all considered covariates are strongly associated with the covariance of childhood BMI.

Figures 2(A)–2(D) display the estimated eigenfunctions and score variance functions, as well as their 95% pointwise confidence bands. Figure 2(A) for the first group of eigenfunctions $\phi_1(\cdot)$ shows that the periodicity of BMI exhibits peaks and troughs at infancy, 5 years old, 12 years old, and 18 years old. Figure 2(B) for the second group of eigenfunctions $\phi_2(\cdot)$ implies that the BMI has large

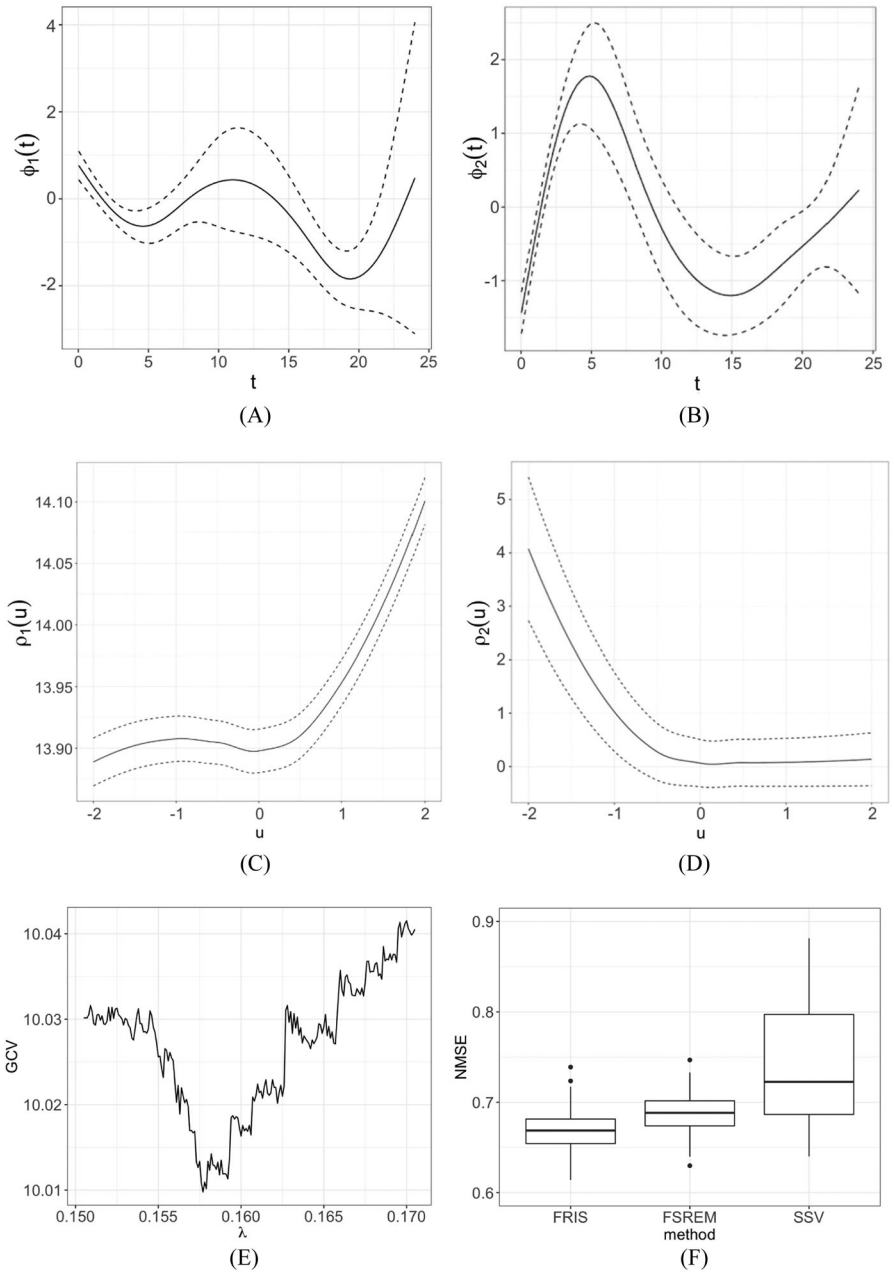


FIGURE 2 (A)(B) Estimates of the eigenfunctions for $\phi_1(t)$ and $\phi_2(t)$, respectively; (C)(D) Estimates of the score variance functions for $\rho_1(u)$ and $\rho_2(u)$, respectively (solid average of the estimated function; dashed-95% pointwise confident band); (E) GCV results of tuning parameters λ; (F) NMSEs of FRIS, SSV, and FSREM for Avon Longitudinal Study of Parents and Children

fluctuations at 5 years old and 18 years old. Figure 2(C) shows that $\rho_1(u)$ is nonzero for all individuals, whereas Figure 2(D) shows that some individuals may have zero values for $\rho_2(u)$, suggesting that $\phi_2(t)$ is not necessary for all individuals. Particularly, those satisfying $\mathbf{X}_i'\boldsymbol{\alpha}_2 \geqslant -0.3$ are expressed only by eigenfunctions $\phi_1(t)$, whereas the others are expressed by both $\phi_1(t)$ and $\phi_2(t)$, where $\boldsymbol{\alpha}_2$ is listed in Table 4. Hence, the individuals for which $\mathbf{X}_i'\boldsymbol{\alpha}_2 < -0.3$ have larger fluctuations at 5 and 18 years old.

We finally evaluate the three methods according to their prediction performances via the NMSE metric defined in Section 4; see Figure 2(F). The proposed FRIS performs the best, followed by the FSREM, suggesting the necessity of properly handling heterogeneous functional data.

6 | DISCUSSION

Under the framework of FPCA, we introduce a new strategy for functional regression by allowing FPC scores to be covariate-dependent and thus for the covariance structures of the functional responses to be individualized, which may lead to a more parsimonious and interpretable representation of heterogeneous functional data for each

TABLE 3 Comparisons of the estimates for $\boldsymbol{\beta}$, accounting for covariate-mean relationships and obtained by FRIS, SSV, and FSREM; presented are point estimates (Est.) and *p*-values for the ALSPAC study

		FRIS		SSV		FSREM (β_1)		FSREM (β_2)	
		Est.	*p*-value	Est.	*p*-value	Est.	*p*-Value	Est.	*p*-Value
Spontaneous	β_0	0.0827	0.7170	−0.0638	0.8208	0.0950	0.8181	0.1241	0.4301
Birth weight	β_1	0.3752	0.0000	0.3281	0.0283	0.5348	0.0000	0.7300	0.0000
Birth length	β_2	−0.1507	0.0000	−0.1389	0.0608	−0.4040	0.0002	−0.4297	0.0001
Diabetes	β_3	0.5470	0.0000	0.7021	0.0007	0.1534	0.1250	0.2731	0.0063
Amniocentesis	β_4	−0.3512	0.0002	−0.2938	0.0597	0.0187	0.8517	0.0558	0.5768
# of children	β_5	0.3859	0.0145	0.2068	0.3880	−0.0867	0.3859	−0.1153	0.2489
Assisted breech	β_6	−0.1928	0.0346	−0.2540	0.0742	0.1135	0.2564	0.1403	0.1606
Caesarean section	β_7	0.4306	0.0000	0.3955	0.0221	0.3550	0.0004	0.1154	0.2485
Forceps delivery	β_8	0.0276	0.4520	0.0519	0.1585	0.2785	0.4151	−0.1227	0.7417
Vacuum extraction	β_9	0.1864	0.0019	0.1523	0.0844	0.3665	0.0002	−0.0158	0.8745

TABLE 4 Comparisons of the estimates of $\boldsymbol{\alpha}$, accounting for covariate-covariance relationships; presented are point estimates (Est.) and *p*-values for ALSPAC data

	FRIS				FSREM			
	α_1		α_2		α_1		α_2	
	Est.	*p*-Value	Est.	*p*-Value	Est.	*p*-Value	Est.	*p*-Value
Spontaneous	0.6297	0.0000	−0.1166	0.2823	0.3173	0.4020	−0.0982	0.8092
Birth weight	0.0588	0.3444	0.2781	0.0000	0.1918	0.0220	0.4033	0.0000
Birth length	−0.0014	0.9693	0.2771	0.0000	−0.0807	0.2799	−0.2047	0.0425
Diabetes	0.4916	0.0000	0.3391	0.0012	−0.1311	0.7338	−0.4010	0.2268
Amniocentesis	−0.3442	0.0006	0.4673	0.0001	0.4095	0.2559	−0.6868	0.0492
# of children	0.3822	0.0000	0.1369	0.1185	−0.5482	0.0400	0.0588	0.7893
Assisted breech	0.2492	0.0001	0.2662	0.0813	−0.0472	0.8417	0.0555	0.8450
Caesarean section	−0.1053	0.0248	0.3240	0.0001	0.3765	0.2805	−0.3448	0.3587
Forceps delivery	−0.0972	0.1556	0.4059	0.0003	0.4238	0.2242	0.1194	0.7331
Vacuum extraction	0.1051	0.0268	0.3737	0.0011	0.2167	0.5644	−0.1167	0.7834

individual. To tackle the nonconvexity and nonsmoothness of the penalized quasi-likelihood function when drawing inferences, we combine the ADMM and a locally linearized and majorized coordinate descendent algorithm to achieve computational readiness. The nonconvexity factor also requires reasonable initial values. In the Supporting Information, we provide initial values based on four steps, whose validity is confirmed by simulation studies. The proposed estimator exhibits selection and estimation consistency is asymptotically normally distributed, and possesses a reasonable finite sample performance.

Several directions are available for future research. First, it might be possible to extend our method to a more complicated mean model with time-varying coefficient, especially by utilizing a sufficiently long follow-up period. Second, by modeling the covariance matrix of the scores derived from the correlated functional data, it is possible to extend the proposed method to cases where functional profiles are correlated. Third, stemming from the established framework that accommodates a single-index structure for mean and covariance functions, we may be able to develop a more flexible model with multiple-index means and covariances; however, this work may require substantial efforts, especially for situations with unknown numbers of indices. Fourth, as our method requires the functional responses to be continuous, the accommodation of discrete responses is worthy of further investigation for addressing specific scientific questions. Finally, with the emergence of high-throughput predictors, how to extend the current modeling framework and inference procedures to allow high-dimensional covariates may be of substantial interest; we will conduct this work elsewhere.

ACKNOWLEDGMENTS

This research was partially supported by National Natural Science Foundation of China (Nos. 11931014, 11829101, 12171374).

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are available from the Avon Longitudinal Study of Parents and Children. Restrictions apply to the availability of these data, which were used under license in this paper. Data are available at https://www.bristol.ac.uk/alspac/ with the permission of the Avon Longitudinal Study of Parents and Children.

ORCID

Huazhen Lin https://orcid.org/0000-0002-4890-9550
Jin Liu https://orcid.org/0000-0002-5707-2078
Yi Li https://orcid.org/0000-0003-1720-2760

REFERENCES

Ash, R.B. & Gardner, M.F. (1978) *Topics in stochastic processes*. Academic, New York.

Backenroth, D., Goldsmith, J., Harran, M.D., Cortes, J.C., Krakauer, J.W. & Kitago, T. (2018) Modeling motor learning using heteroskedastic functional principal components analysis. *Journal of the American Statistical Association*, 113, 1003–1015.

Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trend in Machine learning*, 3(1), 1–122.

Cardot, H. (2006) Conditional functional principal components analysis. *Scandinavian Journal of Statistics*, 34, 317–335.

Chen, K. & Tong, X. (2010) Varying coefficient transformation models with censored data. *Biometrika*, 97, 969–976.

Chen, X., Li, H., Liang, H. & Lin, H. (2019) Functional response regression analysis. *Journal of Multivariate Analysis*, 169, 218–233.

Chiou, J., Müller, H. & Wang, J. (2003a) Functional response models. *Statistica Sinica*, 14, 675–693.

Chiou, J.M., Müller, H.G. & Wang, J.L. (2003b) Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society, Series B*, 65, 405–423.

Fan, J. & Li, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348–1360.

Hall, P. & Mohammad, H. (2006) On properties of functional principal components analysis. *Journal of the Royal Statistical Society, Series B*, 68, 109–126.

James, G.M., Wang, J. & Zhu, J. (2009) Functional linear regression that's interpretable. *The Annals of Statistics*, 37, 2083–2108.

Jiang, C. & Wang, J.L. (2011) Functional single index models for longitudinal data. *The Annals of Statistics*, 39, 362–388.

Li, G., Shen, H. & Huang, J. (2016) Supervised sparse and functional principal component analysis. *Journal of Computational and Graphical Statistics*, 26, 859–878.

Li, G., Yang, D., Nobel, A.B. & Shen, H. (2015) Supervised singular value decomposition and its asymptotic properties. *Journal of Multivariate Analysis*, 146, 7–17.

Li, J., Huang, C. & Zhu, H. (2017) A functional varying-coefficient single-index model for functional response data a functional varying-coefficient single-index model for functional response data. *Journal of the American Statistical Association*, 112(519), 1169–1181.

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

Li, Y., Wang, N. & Carroll, R.J. (2010) Generalized functional linear models with semiparametric single-index interactions. *Journal of the American Statistical Association*, 105, 621–633.

Lin, Z., Cao, J., Wang, L. & Wang, H. (2017) Locally sparse estimator for functional linear regression models. *Journal of Computational and Graphical Statistics*, 26, 306–318.

Lin, Z., Müller, H.G. & Yao, F. (2018) Mixture inner product spaces and their application to functional data analysis. *The Annals of Statistics*, 104, 545–560.

Masukume, G., O'Neill, S.M., Baker, P.N., Kenny, L.C., Morton, S.M. & Khashan, A.S. (2018) The impact of caesarean section on the risk of childhood overweight and obesity: new evidence from a contemporary cohort study. *Scientific Reports*, 8(1), 1–9.

Ochiai, H., Shirasawa, T., Ohtsu, T., Nishimura, R., Morimoto, A., Obuchi, R. et al. (2012) Number of siblings, birth order, and childhood overweight: a population-based cross-sectional study in japan. *BMC Public Health*, 12(1), 1–7.

Seeds, J.W. (2004) Diagnostic mid trimester amniocentesis: how safe? *American Journal of Obstetrics and Gynecology*, 191(2), 607–615.

Stone, C. (1980) Optimal rate of convergence for nonparametric estimators. *The Annals of Statistics*, 8, 1348–1360.

Wang, J., Wang, L., Liu, H., Zhang, S., Leng, J., Li, W. et al. (2018) Maternal gestational diabetes and different indicators of childhood obesity: a large study. *Endocrine Connections*, 7(12), 1464–1471.

Yao, F. & Müller, H.G. (2010) Functional quadratic regression. *Biometrika*, 97, 49–64.

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

Zhong, Q., Lin, H. & Li, Y. (2020) Cluster non-Gaussian functional data. *Biometrics*, 77(3), 852–865.

Zhou, J., Wang, N.Y. & Wang, N. (2013) Functional linear model with zero-value coefficient function at sub-regions. *Statistica Sinica*, 23, 25–50.

Zhou, L., Lin, H. & Liang, H. (2018) Efficient estimation of the nonparametric mean and covariance functions for longitudinal and sparse functional data. *Journal of the American Statistical Association*, 113, 1550–1564.

SUPPORTING INFORMATION

Tables and figures referenced in Sections 4–5 are available with this paper at the Biometrics website on Wiley Online Library. The R code for the proposed method is posted online with the article and is also available on GitHub https://github.com/LinhzLab/.

How to cite this article: Zhang, C., Lin, H., Liu, L., Liu, J., & Li, Y. (2023) Functional data analysis with covariate-dependent mean and covariance structures. *Biometrics*, 79, 2232–2245. https://doi.org/10.1111/biom.13744