JOURNAL OF APPLIED STATISTICS
2024, VOL. 51, NO. 10, 2025–2038
https://doi.org/10.1080/02664763.2023.2253379



# A novel two-way functional linear model with applications in human mortality data analysis

Xingyu Yan[a], Jiaqian Yu[a], Weiyong Ding[a], Hao Wang[b] and Peng Zhao[a]

[a]School of Mathematics and Statistics and RIMS, Jiangsu Provincial Key Laboratory of Educational Big Data Science and Engineering, Jiangsu Normal University, Xuzhou, Jiangsu, People's Republic of China; [b]School of Mathematics and Statistics, Anhui Normal University, Wuhu, People's Republic of China

**ABSTRACT**

Recently, two-way or longitudinal functional data analysis has attracted much attention in many fields. However, little is known on how to appropriately characterize the association between two-way functional predictor and scalar response. Motivated by a mortality study, in this paper, we propose a novel two-way functional linear model, where the response is a scalar and functional predictor is two-way trajectory. The model is intuitive, interpretable and naturally captures relationship between each way of two-way functional predictor and scalar-type response. Further, we develop a new estimation method to estimate the regression functions in the framework of weak separability. The main technical tools for the construction of the regression functions are product functional principal component analysis and iterative least square procedure. The solid performance of our method is demonstrated in extensive simulation studies. We also analyze the mortality dataset to illustrate the usefulness of the proposed procedure.

**ARTICLE HISTORY**
Received 1 March 2023
Accepted 15 August 2023

**KEYWORDS**
Functional data; two-way functional data; two-way functional linear regression; product functional principal components analysis; weak separability; matrix variate

**MATHEMATIC SUBJECT CLASSIFICATION**
62-08

## 1. Introduction

Functional linear regression, due to its flexibility and ease of interpretation, is widely used in different fields of applied sciences to model all kinds of dynamic association when predictor is trajectory. There have been extensive researches on functional linear regression in the past decades. For example, see the monographs [10,14,15], as well as the articles by [3,4,13,19–22,33,34,36,38,40] and the references therein.

So far, most existing works focus on the case where the functional predictor is one-way trajectory. Due to advances in technology, functional data analysts increasingly encounter complex data types where functional data are no longer traditional one-way trajectory. Two-way or multi-way functional data are now commonly seen in many different fields of applied sciences including, for example, chemometrics, biomedical studies, and econometrics. Some progress has been made in developing methodologies for analyzing such functional data. For instance, [12] first considered the model and method for analysis of functional data observed at multiple time points, i.e. longitudinal observed functional data.

**CONTACT** Hao Wang hao.wang@ahnu.edu.cn School of Mathematics and Statistics, Anhui Normal University, Wuhu 241000, People's Republic of China

[5] developed a two-step functional principal components analysis (FPCA) for modeling longitudinal functional data, aiming at covering the case where the recordings of the curves are recorded either dense or sparse. [28] accounted for the longitudinal design of functional data, and proposed a parsimonious modeling approach to analyze the diffusion tensor imaging study of multiple sclerosis. [16] performed a two-way principal component analysis for analyzing the functional magnetic resonance imaging data. By multi-level FPCA approach, [35] proposed a methodology for statistical inference in the presence of three-level nested hierarchical functional data. Further, [32] proposed a multi-dimensional nonparametric covariance function estimation approach under the framework of reproducing kernel Hilbert spaces that can handle both sparse and dense multidimensional functional data.

The above works mainly focus on the methods of modeling two-way trajectories, without considering impact of the trajectories on some interested response. This limits the application of various developed estimation approaches in real regression problems. More recently, there have been some related researches about image data to address this issue. For example, [31] described a regularized Haar wavelet-based approach for the analysis of three-dimensional brain image data in the framework of functional data analysis. [11] developed scalar-on-image regression models by using a fast and scalable Bayes' inferential procedure to estimate the image coefficient. [30] proposed a parsimonious modeling framework to study longitudinal dynamic functional regression models that account for the association between time-varying functional covariate and time-varying responses. [22] developed a set of generalized functional partial linear varying-coefficient regression models for the association analysis of asynchronous functional covariates and functional response.

As far as we know, the aforementioned methods and theories are not applicable for separately analysing dynamic association between each way of two-way functional predictor and scalar response. An additional technical drawback is that existing regression procedure involves performing high dimensional non-parametric regression, with possible slow computing, curse of dimensionality and loss of efficiency. This motivates us to study a new model and method that is suitable for characterizing possibly asymmetric dynamic effects between each way of two-way functional predictor and response separately. The methodology is also motivated by and applied to a human mortality study designed to analyze the dynamic impact of mortality over past 40 calendar years and age over the 60 years old on the corresponding average population across different countries, respectively (we shall discuss in detail in Section 3.2).

In the current work, we propose a new and flexible two-way functional linear model, which separately depicts the dynamic impact of each way of two-way functional predictor on the scalar response. To perform estimation, given the mild weak separability assumption, we utilize the product functional principal component analysis (FPCA) to achieve the dimension reduction for two-way functional predictor. This procedure not only can reduce the computational burden, especially for dense trajectories, but also guarantee the orthogonality of eigenfunctions, which is crucial for subsequent procedure. Then, a (regularized) iterative least-squares estimation (LSE) approach is developed to estimate the coefficient after dimension reduction. Lastly, combining the above estimated coefficient and the product of eigenfunctions, we can easily obtain the estimated regression functions, respectively. To our knowledge, our work is the first one that attempts to tackle scalar on two-way functional regression via popular FPCA techniques, which is a well-established tool of functional data analysis.

The rest of this paper is organized as follows. In the Section 2, we first review some necessary preliminary concepts, such as, weak separability and product FPCA. Subsequently, we present the two-way functional linear model in Section 2.2 and the estimation procedure in Section 2.3. We illustrate the proposed methods by simulation studies in Section 3.1, and analyze our motivating dataset in Section 3.2. More discussion and future works are put in the last section.

## 2. Methodology

### 2.1. Preliminaries

Since the concepts of product FPCA and weak separability introduced by [6] and [27], respectively, are still relatively new, we briefly review them in this section.

Suppose that the two-way stochastic process is $X(s,t)$ with mean and covariance functions, denoted by $E\{X(s,t)\} = \mu(s,t)$ and $G(s,t;u,v) = \mathrm{cov}\{X(s,t), X(u,v)\}$, respectively, where $s \in \mathcal{S}$ and $t \in \mathcal{T}$. Here, $\mathcal{S}$ and $\mathcal{T}$ denote closed interval of the real line in this paper. Aiming to avoid difficulties of estimation of multidimensional covariance function $G(s,t;u,v)$, such as, curse of dimensionality, slow computing, by simplified two-way Karhunen-Loève (K-L) representation, [6] proposed the product FPCA as follows:

$$X(s,t) = \mu(s,t) + \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \xi_{jk}\psi_j(s)\phi_k(t), \tag{1}$$

where $\xi_{jk} = \int_{\mathcal{T}}\int_{\mathcal{S}}\{X(s,t) - \mu(s,t)\}\psi_j(s)\phi_k(t)\,\mathrm{d}s\,\mathrm{d}t, j,k = 1,2\ldots$, are viewed as marginal projection scores. The functions $\{\psi_j(s)\}_{j=1,2,\ldots}$ and $\{\phi_k(t)\}_{k=1,2,\ldots}$ are the eigenbasis of the marginal covariance functions, respectively, which are defined as follows

$$G_{\mathcal{S}}(s,u) = \int_{\mathcal{T}} G(s,t;u,t)\,\mathrm{d}t, \quad G_{\mathcal{T}}(t,v) = \int_{\mathcal{S}} G(s,t;s,v)\,\mathrm{d}s. \tag{2}$$

Here $G_{\mathcal{S}}(s,u)$ and $G_{\mathcal{T}}(t,v)$ satisfy the spectral decomposition as ordinary covariance function, that is, $G_{\mathcal{S}}(s,u) = \sum_{j=1}^{\infty}\lambda_j\psi_j(s)\psi_j(u)$ and $G_{\mathcal{T}}(t,v) = \sum_{k=1}^{\infty}\kappa_k\phi_k(t)\phi_k(v)$, where $\lambda_1 \geq \lambda_2 \geq \cdots$ and $\kappa_1 \geq \kappa_2 \geq \cdots$ are the marginal eigenvalues, respectively. Although the simplified two-way K-L representation in (1) has a nice expression, without additional conditions, the scores $\xi_{jk}$ in general will not be uncorrelated. This restricts the application of product FPCA.

When aforementioned two-way (or multi-way) functional data are encountered, structural assumptions can often simplify matters greatly, both in terms of interpretation and computation. In this case, there is some literature on the structural restriction of multidimensional covariance function of $X(\cdot,\cdot)$ in the context of functional data analysis, see, such as, [1,2,7,8,25,37] and references therein. Noticeably, to alleviate the difficulties associated with modeling of the two-way functional data, [27] introduced following definition of weak separability.

**Definition 2.1:** $X(s,t)$ is weakly separable if there exist orthonormal bases $\{f_j(s)\}_{j=1}^{\infty}$ and $\{g_k(t)\}_{k=1}^{\infty}$, such that the scores array $\{\xi_{jk}, j = 1,2,\ldots; k = 1,2,\ldots\}$ are uncorrelated to each other. That is, $\mathrm{cov}(\xi_{jk}, \xi_{j'k'}) = 0$ for $j \neq j'$ or $k \neq k'$.

If weak separability holds, as established by [27], it need not be necessary that the covariance function $G(s,t;u,v)$ be factored as a product of two marginal covariance function, so that $\{\psi_j(s)\phi_k(t)\}_{j,k=1,\ldots,}$ are not necessarily eigenfunctions of $G(s,t;u,v)$. But following lemma can be established.

**Lemma 2.2:** *If $X(s,t)$ is weakly separable, the pair of bases $\{f_j(s)\}_{j=1}^{\infty}$ and $\{g_k(t)\}_{k=1}^{\infty}$ that satisfies weak separability is unique, i.e. for any $j,k \geq 1$, $f_j(s) \equiv \psi_j(s)$ and $g_k(t) \equiv \phi_k(t)$, where $\psi_j(s)$ and $\phi_k(t)$ are the eigenfunctions of the marginal covariance functions $G_{\mathcal{S}}(s,u)$ and $G_{\mathcal{T}}(t,v)$, respectively.*

Therefore, the key point of weak separability is that the K-L representation is the same as the product representation in (1). This means that we only need to calculate the marginal covariances $G_{\mathcal{S}}(s,u)$ and $G_{\mathcal{T}}(t,v)$ instead of the full covariance function $G(s,t;u,v)$. Hence, this provides us with a convenient vehicle for modeling two-way functional data and based on the assumption we will devise a new model and method for analysing two-way functional linear regression in next section.

### 2.2. Model and eigenbasis representation

Assume that the two-way functional predictor $X(s,t)$ is weakly separable, which has mean function $E\{X(s,t)\} = \mu(s,t)$ and covariance function $G(s,t;u,v) = \mathrm{cov}\{X(s,t), X(u,v)\}$. The response $Y$ is a real-valued random variable. Without loss of generality, we assume that the response variable and the two-way trajectory are all centered so that no intercept term is included hereafter.

Now assume that we collect pair data $(X_1, Y_1), \ldots, (X_n, Y_n)$, which are independent and identically distributed (i.i.d.) copies of $(X, Y)$, then the novel two-way functional linear model proposed is introduced as follows

$$Y_i = \int_{\mathcal{S}} \int_{\mathcal{T}} \beta_1(s) X_i(s,t) \beta_2(t)\, \mathrm{d}t\, \mathrm{d}s + \varepsilon_i, \quad i = 1, \ldots, n, \tag{3}$$

where $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are both square-integrable regression functions. The main object of our interest in this paper is to estimate the two regression functions $\beta_1(\cdot)$ and $\beta_2(\cdot)$. The random error term $\varepsilon_i$'s are i.i.d. copies of $\varepsilon$ with $E(\varepsilon) = 0$ and $\mathrm{var}(\varepsilon) = \sigma^2$. To avoid identification problems, we assume that $\int_{\mathcal{S}} \beta_1^2(s)\, \mathrm{d}s = 1$ and $\beta_2(0) = 1$. Similar constraints can be found in [18,26]. Different constraints mainly change the scale of the parameter coefficients and have no effect on model fitting or prediction. Note that model (3) allows the asymmetry of the roles played by $t$ and $s$. This makes it easy to separate and visualize the regression effects of $s$ and $t$.

To estimate the regression functions $\beta_1(s)$ and $\beta_2(t)$, we first utilize the product FPCA to obtain a data driven basis function, that is, $X_i(s,t)$ admits being decomposed as $X_i(s,t) = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \xi_{jk} \psi_j(s) \phi_k(t)$, where $\{\xi_{ijk}\}_{j=1,2,\ldots;k=1,2,\ldots}$ are uncorrelated with each other by the weak separability from Definition 2.1. Moreover, the product marginal eigenfunction $\{\psi_j(s)\phi_k(t)\}_{j=1,2,\ldots;k=1,2,\ldots}$ forms an orthonormal basis by Lemma 2.2, where $\psi_j(s)$ and $\phi_k(t)$ are marginal eigenfunctions of $G_{\mathcal{S}}(s,u)$ and $G_{\mathcal{T}}(t,v)$, respectively, which are defined in (2). Based on above results, we then expand $\beta_1(s)$ and $\beta_2(t)$ in terms of $\psi_j(s)$ and $\phi_k(t)$, respectively, such that

$$\beta_1(s) = \sum_{j=1}^{\infty} a_j \psi_j(s) \quad \text{and} \quad \beta_2(t) = \sum_{k=1}^{\infty} b_k \phi_k(t). \tag{4}$$

Then, it is easy to see that model (3) can be represented as follows

$$\begin{aligned}
Y_i &= \int_{\mathcal{S}} \int_{\mathcal{T}} \beta_1(s) X(s,t) \beta_2(t)\, \mathrm{d}s\, \mathrm{d}t \\
&= \int_{\mathcal{S}} \int_{\mathcal{T}} \left\{ \sum_{j=1}^{\infty} a_j \psi_j(s) \right\} \left\{ \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \xi_{ijk} \psi_j(s) \phi_k(t) \right\} \left\{ \sum_{k=1}^{\infty} b_k \phi_k(t) \right\} \mathrm{d}s\, \mathrm{d}t \\
&= \int_{\mathcal{S}} \int_{\mathcal{S}} \left\{ \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \xi_{ijk} \psi_j(s) \phi_k(t) \right\} \left\{ \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} a_j b_k \psi_j(s) \phi_k(t) \right\} \mathrm{d}s\, \mathrm{d}t \\
&= \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} a_j \xi_{ijk} b_k, \quad i = 1, \ldots, n.
\end{aligned} \tag{5}$$

Note that under mild weak separability condition, the utilization of data-driven marginal basis naturally attends dimension reduction so that the model (3) is transformed to a matrix-valued regression model (5), instead of estimating the full covariance function of two-way trajectories. Consequently, the objective of estimating the regression functions $\beta_1(s)$ and $\beta_2(t)$ is transformed into the one of estimating the coefficients $\{a_j\}_{j=1,2,\ldots}$ and $\{b_k\}_{k=1,2,\ldots}$ as well as the principal component (PC) scores $\{\xi_{jk}\}_{j=1,2,\ldots;k=1,2,\ldots}$. To address the difficulty caused by the infinite dimensionality of the predictors, we assume that all useful information in the functional predictor is contained in the first $J$ and $K$ principal components, respectively. With this truncation, model (5) is approximately expressed in the following matrix regression form:

$$Y_i = \mathbf{a}^{\mathrm{T}} \mathbf{Z}_i \mathbf{b} + \varepsilon_i, \tag{6}$$

where $\mathbf{a} = (a_1, \ldots, a_J)^{\mathrm{T}} \in \mathbb{R}^{J \times 1}$, $\mathbf{b} = (b_1, \ldots, b_K)^{\mathrm{T}} \in \mathbb{R}^{K \times 1}$ and $\mathbf{Z}_i = (\xi_{ijk}) \in \mathbb{R}^{J \times K}$. For identifiability issues of (6), similar to [9], we further set the first element of $\mathbf{b}$ to be positive and note that $\|\mathbf{a}\|_2 = 1$ derived by $\int_{\mathcal{S}} \beta_1^2(s)\, \mathrm{d}s = 1$.

Intuitively, model (6) is the form of matrix-variate regression [17], and by the fact $\mathbf{a}^{\mathrm{T}} \mathbf{Z}_i \mathbf{b} = (\mathbf{b}^{\mathrm{T}} \otimes \mathbf{a}^{\mathrm{T}}) \times \mathrm{vec}(\mathbf{Z}_i)$, we can simply vectorize the matrix-valued PC scores $\mathbf{Z}_i$, such that the (6) is equaviently written as

$$Y_i = (\mathbf{b}^{\mathrm{T}} \otimes \mathbf{a}^{\mathrm{T}}) \times \mathrm{vec}(\mathbf{Z}_i) + \varepsilon_i, \tag{7}$$

Thus, the two-way functional regression problem is transformed into a standard vector variate regression with a Kronecker product structured on regression coefficient, so one can adopt the LSE techniques to obtain estimators. However, different from conventional matrix-variate regression, this naive approach would lead to at least two potential difficulties in our two-way functional regression setting. First, the vectorization inevitably loses meaningful information in the scores array structure as in (6). This may lead to inefficient and unstable estiamtion. Second, the number of parameters for the coefficient in model (7) is $J \times K$ whereas it is $J + K$ for the coefficients $\mathbf{a}$ and $\mathbf{b}$ in model (6). So, model (6) significantly reduces the dimension of the problem, especially for diverging dimensionality principal components involved. Moreover, the regression coefficient constraint $\mathbf{b}^{\mathrm{T}} \otimes \mathbf{a}^{\mathrm{T}}$ may lead to any estimator making interpretation less natural. It is hard to identify the regression effect along $s$ and $t$, respectively.

Therefore, it is necessary to provide a new approach that takes advantage of the structural information and makes the estimation more accurate and parsimonious. We describe the detailed approach in the next section.

### 2.3. Estimation

In this section, we turn to the estimation procedure. Different from conventional regression problems, our $\{\mathbf{Z}_i, i = 1, \ldots, n\}$ in (6) is matrix variate and unknown, and thus needs to be estimated from the data.

In this paper we focus on the case where trajectories $\{X_i(\cdot,\cdot), i = 1, \ldots, n\}$ are recorded on arbitrarily dense and equally spaced grid points. In practice, if the two-way functional data are not fully observed but rather, for each subject are recorded at series of different time points, then one can use two-dimensional local polynomial smoothing to smooth the trajectories before next step. Here, we adopt the moment estimation for brevity. Specifically, use the predictor $\{X_i(s,t), i = 1, \ldots, n\}$ to obtain estimates of the marginal covariance $G_{\mathcal{S}}(s,u)$ and then change the roles of the two arguments to obtain estimates of $G_{\mathcal{T}}(t,v)$ where

$$\widehat{G}_{\mathcal{S}}(s,u) = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{T}} X_i(s,t)X_i(u,t)\,\mathrm{d}t, \quad \widehat{G}_{\mathcal{T}}(t,v) = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{S}} X_i(s,t)X_i(s,v)\,\mathrm{d}s.$$

Then, we employ the spectral decomposition for $\widehat{G}_{\mathcal{S}}(s,u)$ and $\widehat{G}_{\mathcal{T}}(t,v)$ to obtain $\{\widehat{\psi}_j(s)\}_{j=1,\ldots,J}$ and $\{\widehat{\phi}_k(t)\}_{k=1,\ldots,K}$, respectively, where $J$ and $K$ are chosen by the percentage of variation explained throughout this paper [23]. This yields that the estimated PC scores $\widehat{\mathbf{Z}}_i = (\widehat{\xi}_{ijk})_{j=1,\ldots,J;k=1,\ldots,K}$ for $i = 1, \ldots, n$, where $\widehat{\xi}_{ijk} = \int_{\mathcal{S}}\int_{\mathcal{T}} \widehat{X}_i(s,t)\widehat{\psi}_j(s)\widehat{\phi}_k(t)\,\mathrm{d}s\,\mathrm{d}t$. Thus, the complex two-way functional predictor is transformed to PC scores array. This achieves the dimension reduction. Next, we take advantage of the array structural information to propose an iterative LSE estimaiton procedure, which can be regarded as matrix version of block relaxation algorithm [39]. For easy illustration, a four-step algorithm is summarized as follows.

Step 1: Take the initial estimator $\widehat{\mathbf{a}}^{(1)}$ generated from $U(0,1)$, and standardize $\widehat{\mathbf{a}}^{(1)}$, so that $\|\widehat{\mathbf{a}}^{(1)}\| = 1$.

Step 2: At the $m$-th step, fix $\widehat{\mathbf{a}}^{(m)}$ and apply the LSE algorithm with $(\widehat{\mathbf{Z}}_i^{\mathrm{T}}\widehat{\mathbf{a}}^{(m)}, Y_i)$ as the $i$th observation to estimate $\mathbf{b}$ and obtain $\widehat{\mathbf{b}}^{(m)}$. Adjust $\widehat{\mathbf{b}}^{(m)}$ such that $\mathrm{sign}(\widehat{\mathbf{b}}^{(m)}_{(1)}) = 1$.

Then fix $\widehat{\mathbf{b}}^{(m)}$ and again apply the LSE algorithm with $(\widehat{\mathbf{Z}}_i^{\mathrm{T}}\widehat{\mathbf{b}}^{(m+1)}, Y_i)$ as the $i$th observation to obtain $\widehat{\boldsymbol{a}}^{(m+1)}$, then scale the estimator such that $\|\widehat{\boldsymbol{a}}^{(m+1)}\| = 1$.

Step 3: Go back to Step 2 until the convergence is achieved. Then, the coefficient estimator are finally given by $\widehat{\mathbf{a}} = (\widehat{a}_1, \ldots, \widehat{a}_J)^{\mathrm{T}}$ and $\widehat{\mathbf{b}} = (\widehat{b}_1, \ldots, \widehat{b}_K)^{\mathrm{T}}$.

Step 4: Construct the estimates $\widehat{\beta}_1(s) = \sum_{j=1}^{J} \widehat{a}_j \widehat{\psi}_j(s)$ and $\widehat{\beta}_2(t) = \sum_{k=1}^{K} \widehat{b}_k \widehat{\phi}_k(t)$.

The similar choice of initial estimator in our Step 1 is suggested by [9]. Further, it is noted that the truncation $J$ and $K$ were chosen to explain the majority of the total marginal variation in $X(s,t)$, respectively. The effect of PC scores on the response may does not necessarily coincide with their magnitudes specified by the marginal variation alone, i.e. some higher order PC scores may contribute to the regression significantly more than the leading PC scores, especially when the number of principal components is diverging.

Therefore, to simultaneously estimate the coefficient and select related significant principal components, we also propose an adjusted estimation procedure by using the idea of variable selection. Specifically, regularized estimation for (6) incurs slight changes in Step 2. That is, in Step 2, when updating $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{b}}$ at each step, we fit a penalized LSE regression, such as, Lasso, instead of ordinary LSE. Other Steps in the algorithm remain the same. For similar regularized consideration for classical one-way functional linear model, one can refer to [20,21,24,29,33] and references therein.

## 3. Numerical studies

### 3.1. Simulation studies

In this section we present a simulation study to evaluate the numerical performance of the proposed method. The scalar responses $\{Y_i, i = 1, \ldots, n\}$ are generated as follows

$$\begin{aligned} Y_i &= \int_0^1 \int_0^1 \beta_1(s) X_i(s,t) \beta_2(t) \, \mathrm{d}t \, \mathrm{d}s + \varepsilon_i, \\ &= \mathbf{a}^{\mathrm{T}} \mathbf{Z}_i \mathbf{b} + \varepsilon_i, \quad i = 1, \ldots, n. \end{aligned} \tag{8}$$

Let $s$ and $t$ take values from 0 to 1 on an equally-spaced grid of 200 points. The centered two-way predictor trajectories are generated as $X_i(s,t) = \sum_{j=1}^{p_1} \sum_{k=1}^{p_2} \xi_{ijk} \psi_j(s) \phi_k(t)$, where the scores $\xi_{ijk}$'s are i.i.d. as $\mathcal{N}(0, \varsigma_{jk})$ for $i = 1, \ldots, n$, where $\varsigma_{jk} = 45.25 \times 0.64^{(j+k)}$. The marginal eigenfunction $\{\psi_j(t), j = 1, \ldots, p_1\}$ is derived from the Fourier basis $\psi_{2\ell-1}(t) = \sqrt{2}\cos\{(2\ell-1)\pi t\}$ and $\psi_{2\ell}(t) = \sqrt{2}\sin\{(2\ell-1)\pi t\}$ $(\ell = 1, \ldots, p_1/2)$, and define $\phi_k(t) = \sqrt{2}\cos\{(k-1)\pi t\}$ for $k = 1, \ldots, p_2$. The random errors $\varepsilon_i \sim \mathcal{N}(0, 0.03)$. The regression functions are $\beta_1(s) = \sum_{j=1}^{p_1} a_j \psi_j(s)$ and $\beta_2(t) = \sum_{k=1}^{p_2} b_k \phi_k(t)$ linear combinations of the marginal eigenfunctions, respectively, where $a_j$ and $b_k$ are the $j$th and $k$th element of $\mathbf{a}$ and $\mathbf{b}$, respectively. We consider the following two scenarios.

**Example 3.1 (Fixed principal components):** The true principal components $(p_1, p_2)$ is set as $(2,2)$, $(5,5)$, and $(4,6)$, respectively, and the corresponding coefficients are set as follows: $\mathbf{a} = (1, 0.5)^{\mathrm{T}}$ and $\mathbf{b} = (0.9, 0.6)^{\mathrm{T}}$; $\mathbf{a} = (1, 0.9, 0.6, 0.4, 0.2)^{\mathrm{T}}$ and $\mathbf{b} = (0.9, 0.8, 0.6, 0.5, 0.3)^{\mathrm{T}}$; $\mathbf{a} = (1, 0.9, 0.6, 0.4)^{\mathrm{T}}$ and $\mathbf{b} = (0.9, 0.8, 0.6, 0.4, 0.3, 0.2)^{\mathrm{T}}$.

**Table 1.** Simulation results are displayed for the mean squared error (MSE), the integrated mean squared error (IMSE) and the relative prediction error (RPE) based on 200 replication.

| # PC | Methods | $n$ | MSE($\widehat{\mathbf{a}}$) | MSE($\widehat{\mathbf{b}}$) | IMSE($\widehat{\beta}_1$) | IMSE($\widehat{\beta}_2$) | RPE($\widehat{Y}_i$) |
|---|---|---|---|---|---|---|---|
| (2,2) | Proposed | 100 | 2.18 (4.38) | 1.64 (1.84) | 1.71 (5.86) | 1.07 (2.43) | 0.46 (1.27) |
| | | 200 | 1.08 (1.65) | 0.82 (0.91) | 0.55 (1.06) | 0.42 (0.62) | 0.09 (0.23) |
| | | 300 | 0.57 (0.93) | 0.55 (0.59) | 0.27 (0.59) | 0.22 (0.38) | 0.03 (0.06) |
| | Naive | 100 | 5.43 (8.07) | 3.25 (3.00) | 17.18 (23.07) | 10.40 (12.21) | 12.75 (15.74) |
| | | 200 | 3.73 (5.03) | 2.26 (2.08) | 9.19 (11.52) | 6.10 (6.79) | 7.23 (9.40) |
| | | 300 | 3.23 (3.12) | 1.94 (1.33) | 7.83 (7.98) | 5.54 (5.30) | 6.55 (7.07) |
| (5,5) | Proposed | 100 | 1.70 (2.16) | 1.33 (0.90) | 3.50 (6.88) | 2.44 (3.30) | 0.27 (0.34) |
| | | 200 | 0.80 (0.89) | 0.71 (0.47) | 1.42 (2.46) | 1.08 (1.47) | 0.06 (0.08) |
| | | 300 | 0.62 (0.74) | 0.53 (0.38) | 1.18 (1.96) | 0.89 (1.23) | 0.03 (0.04) |
| | Naive | 100 | 3.90 (3.22) | 2.57 (1.29) | 27.36 (18.43) | 17.84 (9.99) | 9.65 (7.15) |
| | | 200 | 2.59 (1.55) | 1.76 (0.77) | 16.85 (9.53) | 11.92 (5.92) | 6.08 (4.37) |
| | | 300 | 2.37 (1.37) | 1.63 (0.61) | 15.01 (7.99) | 10.34 (4.33) | 5.08 (2.98) |
| (4,6) | Proposed | 100 | 1.65 (1.67) | 1.06 (0.66) | 2.44 (3.58) | 2.07 (2.37) | 0.23 (0.24) |
| | | 200 | 0.77 (0.79) | 0.61 (0.37) | 1.07 (1.48) | 0.93 (1.06) | 0.06 (0.08) |
| | | 300 | 0.60 (0.67) | 0.49 (0.33) | 0.91 (1.43) | 0.74 (0.96) | 0.03 (0.05) |
| | Naive | 100 | 5.58 (3.58) | 9.98 (3.94) | 28.72 (20.32) | 61.93 (24.26) | 16.39 (8.89) |
| | | 200 | 4.76 (2.37) | 9.45 (2.91) | 21.86 (10.99) | 56.53 (18.22) | 13.02 (4.72) |
| | | 300 | 4.37 (2.22) | 9.35 (2.84) | 20.22 (9.69) | 55.15 (16.91) | 12.21 (3.97) |

Notes: The true number of principal components (# PC) is set as $(2, 2)$, $(5, 5)$ and $(4, 6)$, respectively. We compare the proposed approach (Proposed) with the naive vectoring approach (Naive) with sample size taken as 100, 200 and 300. (The reported results are original results multiplied by $10^2$).

**Example 3.2 (Diverging principal components):** The simulated data is also generated by (8), except that the true number of principal components is diverging. Specifically, $(p_1, p_2)$ is set as $(10, 10)$, $(10, 15)$ and $(15, 15)$, respectively, where $\mathbf{a} = (1, 0.8, 0.6, 0.5, 0.4, 0.3, 8 \times (j-2)^{-4})^{\mathrm{T}}$ and $\mathbf{b} = (0.4, 0.3, 0.2, 3 \times (k-2)^{-4})^{\mathrm{T}}$ with $j \geq 6$ and $k \geq 4$.

To evaluate the accuracy of estimation and prediction, we simulate $M = 200$ Monte Carlo runs, each run consisting of a collection of $n = 100$, 200, 300 predictor trajectories $X_i$ and associated scalar response $Y_i$ that serves as the training sample for estimation, respectively. In addition, for each run, we further generate another 100 pairs of $(X_i, Y_i)$ that constitute the validation sample, which will be used towards the end of this section for assessing the predictive power. To measure the estimation accuracy, we consider the integrated mean squared error (IMSE) for the functional coefficients, which is defined as $\int_0^1 \{\widehat{\beta}_j(t) - \beta_j(t)\}^2 \, \mathrm{d}t$, when estimating $\beta_j(t)$ for $j = 1, 2$. We also provide the mean squared error (MSE) for the corresponding vector coefficients, which is defined as $E\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2$, where $\boldsymbol{\theta} = \mathbf{a}$ and $\mathbf{b}$. Lastly, to inspect the predictive ability of the proposal, relative prediction error (RPE) is also reported based on validation samples, which is defined as $\mathrm{RPE} = \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2 / (\sum_{i=1}^n Y_i^2)$, where $Y_i$ is the response of the $i$th new subject in the validation sample and $\widehat{Y}_i$ is its predicted value.

We summarize the simulation results for Example 3.1 and 3.2 in Tables 1 and 2, respectively, where the true number of principal components varies from being fixed to diverging. For diverging example, we use the regularized estimation. We compare the estimation accuracy of the proposed method with that of the naive vectoring approach as in (7) with sample size $n$ changing. For each case, we include the empirical means of IMSE($\widehat{\beta}_j$) for $j = 1, 2$, MSE($\widehat{\mathbf{a}}$) and MSE($\widehat{\mathbf{b}}$), as well as RPE with their standard deviation in the parentheses. As we can see, Tables 1 and 2 show that our proposed estimation methods always

**Table 2.** Simulation results are displayed for the mean squared error (MSE), the integrated mean squared error (IMSE) and the relative prediction error (RPE) based on 200 replication.

| # PC | Methods | $n$ | MSE($\widehat{\mathbf{a}}$) | MSE($\widehat{\mathbf{b}}$) | IMSE($\widehat{\beta}_1$) | IMSE($\widehat{\beta}_2$) | RPE($\widehat{Y}_i$) |
|---|---|---|---|---|---|---|---|
| (10,10) | Proposed | 100 | 2.19 (1.52) | 0.21 (0.19) | 14.61 (11.51) | 1.40 (1.52) | 3.88 (3.92) |
| | | 200 | 1.18 (0.78) | 0.11 (0.09) | 8.01 (6.58) | 0.68 (0.67) | 1.82 (1.77) |
| | | 300 | 0.77 (0.53) | 0.07 (0.04) | 4.87 (4.11) | 0.45 (0.36) | 1.09 (1.00) |
| | Naive | 100 | 3.00 (0.98) | 0.34 (0.13) | 34.51 (12.63) | 3.69 (1.79) | 11.15 (7.35) |
| | | 200 | 2.40 (0.55) | 0.30 (0.07) | 25.24 (6.72) | 2.87 (1.06) | 7.03 (3.85) |
| | | 300 | 2.17 (0.42) | 0.29 (0.07) | 22.33 (5.53) | 2.58 (0.84) | 5.81 (2.80) |
| (10,15) | Proposed | 100 | 2.63 (1.74) | 0.21 (0.19) | 17.60 (12.14) | 1.71 (1.80) | 4.72 (4.02) |
| | | 200 | 1.22 (0.80) | 0.09 (0.07) | 8.48 (6.16) | 0.73 (0.67) | 2.01 (1.68) |
| | | 300 | 0.88 (0.64) | 0.06 (0.05) | 5.96 (4.95) | 0.55 (0.45) | 1.40 (1.23) |
| | Naive | 100 | 7.49 (2.09) | 1.37 (0.39) | 76.62 (23.60) | 15.04 (4.69) | 23.08 (12.46) |
| | | 200 | 6.54 (1.42) | 1.46 (0.33) | 61.43 (15.19) | 15.80 (4.00) | 17.42 (7.19) |
| | | 300 | 6.57 (1.29) | 1.50 (0.27) | 60.51 (12.79) | 16.11 (3.32) | 16.86 (6.10) |
| (15,15) | Proposed | 100 | 2.19 (1.23) | 0.20 (0.17) | 18.46 (11.98) | 1.66 (1.55) | 4.97 (4.26) |
| | | 200 | 1.10 (0.75) | 0.10 (0.08) | 9.22 (7.14) | 0.87 (0.79) | 2.27 (2.03) |
| | | 300 | 0.86 (0.64) | 0.07 (0.05) | 7.34 (6.09) | 0.59 (0.52) | 1.70 (1.53) |
| | Naive | 100 | 5.36 (2.66) | 0.40 (0.32) | 63.02 (32.16) | 4.72 (3.80) | 16.68 (10.55) |
| | | 200 | 3.70 (2.52) | 0.35 (0.26) | 42.03 (28.07) | 3.86 (3.03) | 11.01 (8.62) |
| | | 300 | 3.26 (2.19) | 0.32 (0.26) | 37.59 (23.32) | 3.43 (2.91) | 9.11 (6.46) |

Notes: The true number of principal components (# PC) is set as (10, 10), (10, 15) and (15, 15), respectively. We compare the proposed approach (Proposed) with the naive vectoring approach (Naive) with sample size taken as 100, 200 and 300. (The reported results are original results multiplied by $10^2$).

perform much better than the naive approach in all different settings. As expected, the performance improves as the number of subjects increases. Fewer number of principal components lead to better performance of estimation. This is because there are fewer parameters to be estimated under fewer principal components. Moreover, the developed method outperforms the naive methods in terms of relative prediction error, and the larger the number of principal components is, the worse the performance of naive gets. These results also indicate the advantages of the proposed methods.

We also provide graphical illustration for estimated functional regression coefficients in Figures 1 and 2, where we randomly choose the case $(p_1, p_2) = (2, 2)$ and $(10, 10)$ corresponding to fixed and diverging case, respectively, with sample size $n = 100$. We compare in each panel the 2.5% and 97.5% pointwise percentiles of our coefficient estimator with the truth. It is remarkable that the newly proposed estimator performs very well. In Figure 1, the estimated coefficients are close to the true values, and the true coefficients are always nicely covered by the 95% confidence bands. Similarly, it can be seen from Figure 2 that the curves based on the proposed method are closer to the true ones, which also indicates that our method can fit the functional regression coefficients very well.

### 3.2. *Application to mortality data*

The analysis of human mortality is important to assess the future demographic prospects of societies, quantify differences between countries with regard to this overall public health measure, and to appraise biological limits of longevity. We now use the developed methods to analyze the human mortality dataset, which is downloaded from www.mortality.org. The human mortality database contains data in the form of yearly life-tables differentiated by country and age. We select 28 life-tables made up of data from 28 countries, one table for one country. Each row consists of the rates of mortality available for each of the calendar

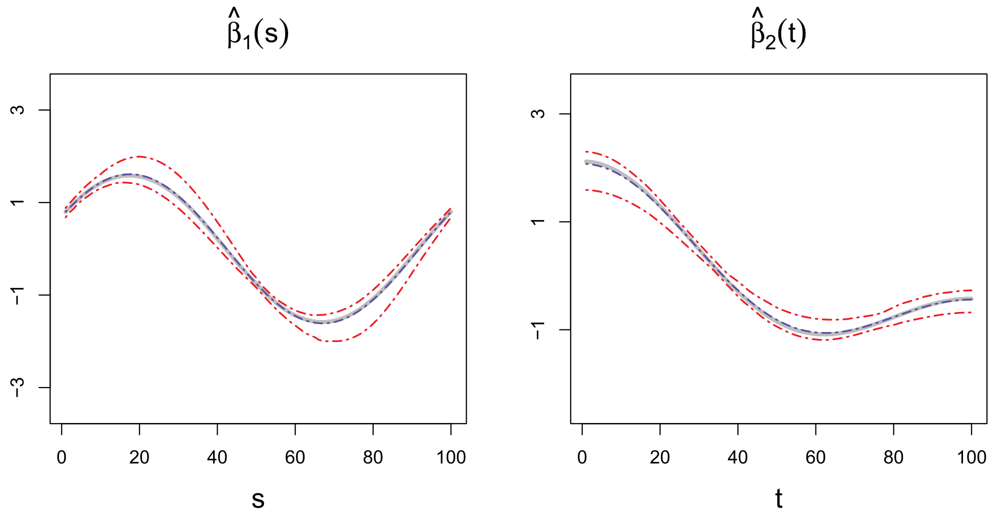


**Figure 1.** The functional coefficients estimation $\widehat{\beta}_1(s)$ and $\widehat{\beta}_2(t)$ with sample size $n = 100$. The solid lines are the true values, the dashed lines are the average estimated values, and the dotted lines are the pointwise 2.5% and 97.5% percentiles of the estimators based on 200 replications.

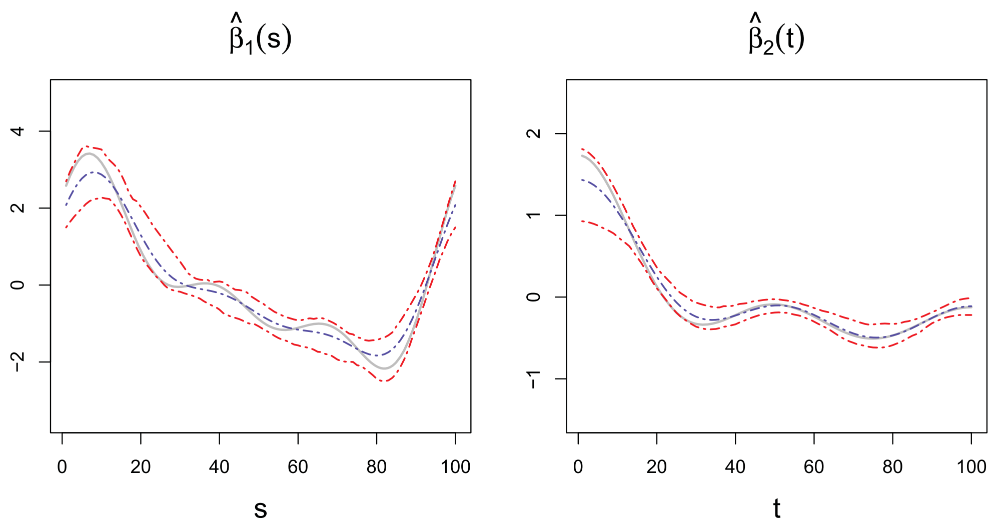


**Figure 2.** The functional coefficients estimation $\widehat{\beta}_1(s)$ and $\widehat{\beta}_2(t)$ with sample size $n = 100$. The solid lines are the true values, the dashed lines are the average estimated values, and the dotted lines are the pointwise 2.5% and 97.5% percentiles of the estimators based on 200 replications.

years from 1960 to 2006; each column denotes mortality of people from 60 to 100 years of age. Due to the high correlation and density of the life sheet mortality singal, we naturally view each life sheet mortality as a two-way functional trajectory. We randomly choose a country (i.e. UK) in Figure 3 to visualize the complex two-way data structure. In addition, we also collect the corresponding total population for each country.

What we are interested in is to investigate the dynamic impact of mortality of those over the age of 60 in the past 40 calendar years on the corresponding average population

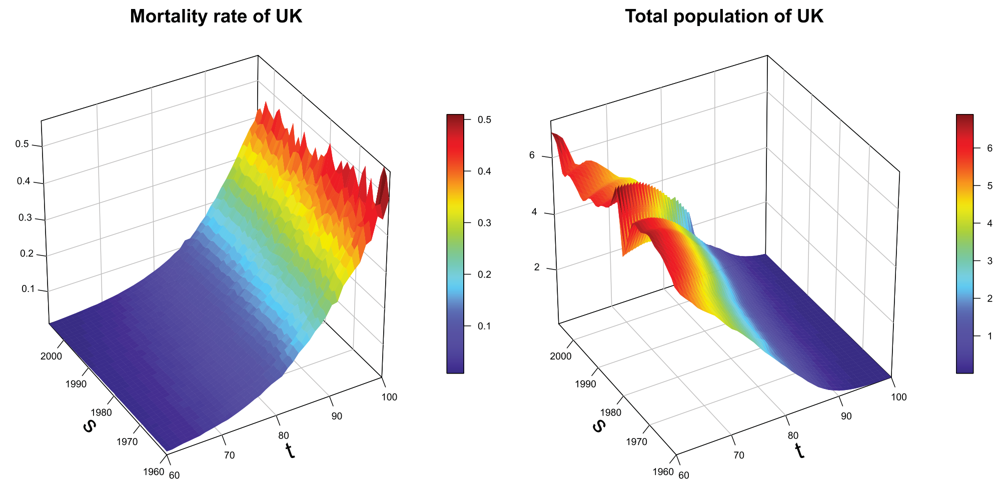


**Figure 3.** Left panel: the mortality rate of UK, where $60 \leq t \leq 100$ and $1960 \leq s \leq 2006$. Right panel: the total population of UK, where $60 \leq t \leq 100$ and $1960 \leq s \leq 2006$.

across different countries, respectively. Following the notation introduced in Section 2, for $i = 1, \ldots, 28$, $X_i(s,t)$ denotes the mortality rate for the $i$th country for subjects at age $t$ for calendar year $s$, where $60 \leq t \leq 100$, i.e. focusing on the death rates of older individuals, and on a recent block of 47 years, $1960 \leq s \leq 2006$. We use $Y_i$ to denote the average population across the same age and year in $i$th country, which is our interested scalar response.

Before presenting the estimated regression functions, we first show the mortality profile trajectories along calendar year $s$ at age of 80 for all countries in the left panel of Figure 4; as well as the mortality profile trajectories changing with age $t$ for all countries in 1980 as described in right panel of Figure 4. Then, apply the proposed methodology to the dataset. The plot of two estimated regression functions are shown in Figure 5. The two estimated curves are both nonlinear and one is smooth, the other is fluctuant. This is due to the different profile noise level of the predictor trajectories itself as shown in Figure 4. Here, the regression function $\widehat{\beta}_1(s)$ describes the dynamic relationship between mortality rate and average population size as calendar year $s$ changes, and $\widehat{\beta}_2(t)$ characterizes association between mortality rate and average population size along with age $t$ increasing. Observe that the shape of these two estimated regression functions is totally different, $\widehat{\beta}_1(s)$ is increasing as year changes, while $\widehat{\beta}_2(t)$ is decreasing as age increases. This shows that the closer it gets to the year 2006, the greater the impact of mortality on the total population is, while the older the age is, the smaller the impact on the total population, because the proportion of this group of people is smaller. This is consistent with our common knowledge.

## 4. Discussion

In this work, under the structure of weak separability, we provide a convenient model, which can separately investigate the dynamic relationship between each way of two-way functional predictor and scalar response. Aiming to simplify the modeling and with a view

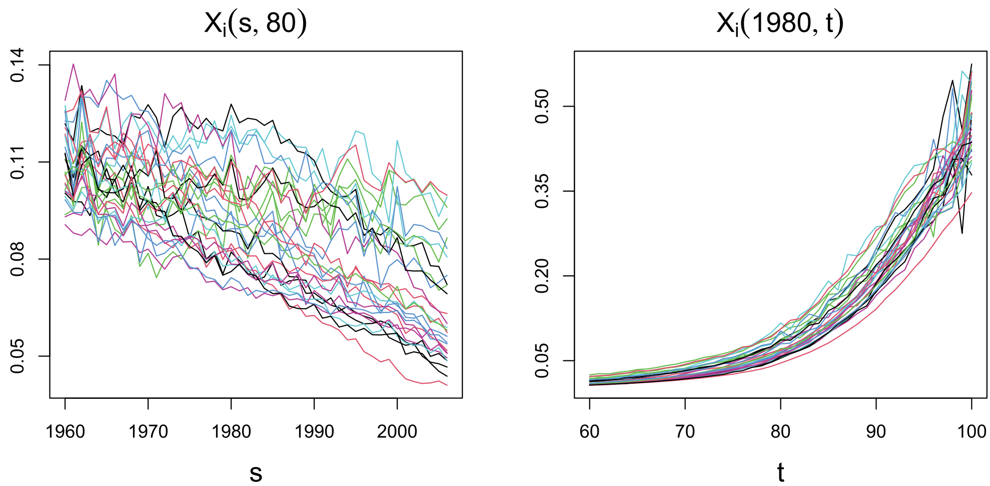


**Figure 4.** Left panel: the mortality profile trajectories from 1960 to 2006 at age 80 from all 28 countries; Right panel: the mortality profile trajectories between age 60 and 100 from all 28 countries in year 1980.

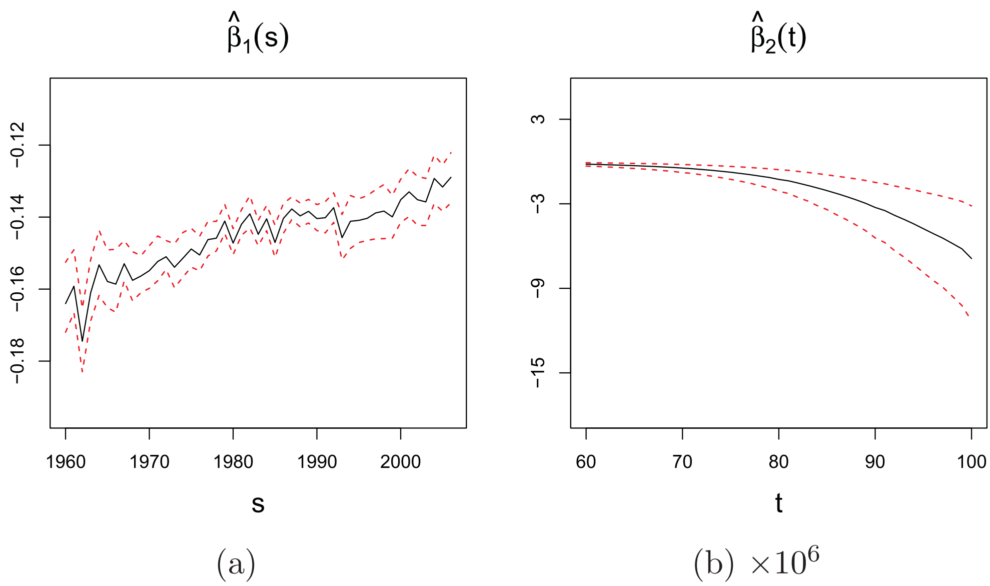


**Figure 5.** The estimated functional coefficients estimation $\widehat{\beta}_1(s)$ and $\widehat{\beta}_2(t)$. The solid black lines are estimated values and the dashed red lines are the pointwise 2.5% and 97.5% percentiles of the estimators based on 200 Bootstrap samples.

towards interpretability, we propose a flexible way to estimate the two time-varying regression coefficient function by combining ideas from two-way functional data analysis and regularized approach. Simulation studies demonstrate the satisfactory finite sample performance of developed procedure. The proposed method is motivated by and applied to the human mortality data study.

Many extensions are possible following our work, which can potentially generate interesting research topics for future studies. For example, a natural extension is to consider generalized two-way functional linear model to analyze the binary response for the purpose of classification. An immediate application is the EEG database data set (http://archive.ics.uci.edu/ml/datasets/EEG+Database). It is interesting to explore the dynamic relationship between alcoholism and the pattern of voltage over time and channel. In addition, the theoretical property of the proposed estimator is also very important and will be pursued in future work.

## Acknowledgments

The authors sincerely acknowledge the efforts of the Editor, the Associate Editor and two anonymous referees that have resulted in significant improvements of this paper.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 12101270, 12325109, 12301597], The Humanities and Social Sciences Project of the Ministry of Education [grant numbers 21YJC910001], and the Graduate Research and Innovation Program in Jiangsu Province [grant numbers KYCX22 2834].

## References

[1] J.A. Aston, D. Pigoli, and S. Tavakoli, *Tests for separability in nonparametric covariance operators of random surfaces*, Ann. Stat. 45 (2017), pp. 1431–1461.
[2] P. Bagchi and H. Dette, *A test for separability in covariance operators of random surfaces*, Ann. Stat. 48 (2020), pp. 2303–2322.
[3] T.T. Cai and H. Peter, *Prediction in functional linear regression*, Ann. Stat. 34 (2006), pp. 2159–2179.
[4] C. Chen, S.J. Guo, and X.H. Qiao, *Functional linear regression: Dependence and error contamination*, J. Bus. Econ. Stat. 40 (2022), pp. 444–457.
[5] K.H. Chen and H.-G. Müller, *Modeling repeated functional observations*, J. Am. Stat. Assoc. 107 (2012), pp. 1599–1609.
[6] K.H. Chen, D. Pedro, and H.-G. Müller, *Modelling function-valued stochastic processes, with applications to fertility dynamics*, J. R. Stat. Soc. Series. B. (Stat Methodol) 79 (2017), pp. 177–196.
[7] P. Constantinou, K. Piotr, and R. Matthew, *Testing separability of space-time functional processes*, Biometrika 104 (2017), pp. 425–437.
[8] H. Dette, G. Dierickx, and T. Kutta, Testing separability for continuous functional data, arXiv preprint arXiv:2301.04487
[9] J.H. Fang and Y.Y. Grace, *Matrix-variate logistic regression with measurement error*, Biometrika 108 (2021), pp. 83–97.
[10] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, Vol. 76, Springer, 2006.
[11] J. Goldsmith, L. Huang, and M.C. Ciprian, *Smooth scalar-on-image regression via spatial bayesian variable selection*, J. Comput. Graph. Stat. 23 (2014), pp. 46–64.
[12] S. Greven, C. Crainiceanu, B. Caffo, and D. Reich, *Longitudinal functional principal component analysis*, Electron. J. Stat. 4 (2011), pp. 1022–1054.
[13] P. Hall and L.H. Joel, *Methodology and convergence rates for functional linear regression*, Ann. Stat. 35 (2007), pp. 70–91.

[14] L. Horváth and P. Kokoszka, *Inference for Functional Data with Applications*, Vol. 200, Springer Science and Business Media, 2012.
[15] T. Hsing and R. Eubank, *Theoretical Foundations of Functional Data Analysis, with An Introduction to Linear Operators*, Vol. 997, John Wiley and Sons, 2015.
[16] L. Huang, P.T. Reiss, L. Xiao, V. Zipunnikov, M.A. Lindquist, and C.M. Crainiceanu, *Two-way principal component analysis for matrix-variate data, with an application to functional magnetic resonance imaging data*, Biostatistics 18 (2017), pp. 214–229.
[17] H. Hung and C.C. Wang, *Matrix variate logistic regression model with application to EEG data*, Biostatistics 14 (2013), pp. 189–202.
[18] F. Jiang, S. Baek, J.G. Cao, and Y.Y. Ma, *A functional single-index model*, Stat. Sin. 30 (2020), pp. 303–324.
[19] K. Kato, *Estimation in functional linear quantile regression*, Ann. Stat. 40 (2012), pp. 3108–3136.
[20] D. Kong, K. Xue, F. Yao, and H.H. Zhang, *Partially functional linear regression in high dimensions*, Biometrika 103 (2016), pp. 147–159.
[21] E.R. Lee and U.P. Byeong, *Sparse estimation in functional linear regression*, J. Multivar. Anal. 105 (2012), pp. 1–17.
[22] T. Li, T.F. Li, Z.Y. Zhu, and H.T. Zhu, *Regression analysis of asynchronous longitudinal functional and scalar data*, J. Am. Stat. Assoc. 117 (2022), pp. 1228–1242.
[23] Y.H. Li, N.Y. Wang, and R.J. Carroll, *Selecting the number of principal components in functional data*, J. Am. Stat. Assoc. 108 (2013), pp. 1284–1294.
[24] H. Lian, *Shrinkage estimation and selection for multiple functional regression*, Stat. Sin. 23 (2013), pp. 51–74.
[25] D. Liang, H. Huang, and Y. Guan and F. Yao, *Test of weak separability for spatially stationary functional field*, J. Am. Stat. Assoc. (2022), pp. 1–14. https://doi.org/10.1080/01621459.2021.2002156.
[26] X.C. Luo, L.P. Zhu, and H.H. Zhu, *Single index varying coefficient model for functional responses*, Biometrics 72 (2016), pp. 1275–1284.
[27] B. Lynch and K.H. Chen, *A test of weak separability for multi-way functional data, with application to brain connectivity studies*, Biometrika 105 (2018), pp. 815–831.
[28] S.Y. Park and A.-M. Staicu, *Longitudinal functional data analysis*, Stat 4 (2015), pp. 212–226.
[29] P.J. Sang, L.L. Wang, and J.G. Cao, *Estimation of sparse functional additive models with adaptive group LASSO*, Stat. Sin. 30 (2020), pp. 1191–1211.
[30] A.-M. Staicu, M.N. Islam, R. Dumitru, and E. Heugten, *Longitudinal dynamic functional regression*, J. R. Stat. Soc. Ser. C. Appl. Stat. 69 (2020), pp. 25–46.
[31] X.J. Wang, B. Nan, J. Zhu, and R. Koeppe, *Regularized 3D functional regression for brain image data via Haar wavelets*, Ann. Appl. Stat. 8 (2014), pp. 1045–1064.
[32] J.Y. Wang, R.K.W. Wong, and X.K. Zhang, *Low-rank covariance function estimation for multidimensional functional data*, J. Am. Stat. Assoc. 117 (2022), pp. 809–822.
[33] R.K. Wong, Y.H. Li, and Z.Y. Zhu, *Partially linear functional additive models for multivariate functional data*, J. Am. Stat. Assoc. 114 (2019), pp. 406–418.
[34] Y.C. Wu, J.Q. Fan, and H.-G. Müller, *Varying-coefficient functional linear regression*, Bernoulli 16 (2010), pp. 730–758.
[35] Y.H. Xu, Y.H. Li, and N. Dan, *Nested hierarchical functional data modeling and inference for the analysis of functional plant phenotypes*, J. Am. Stat. Assoc. 113 (2018), pp. 593–606.
[36] F. Yao, H.-G. Müller, and J.-L. Wang, *Functional linear regression analysis for longitudinal data*, Ann. Stat. 33 (2005), pp. 2873–2903.
[37] J. Zapata, S.-Y. Oh, and P. Alexander, *Partial separability and functional graphical models for multivariate Gaussian processes*, Biometrika 109 (2022), pp. 665–681.
[38] J.-T. Zhang and J.W. Chen, *Statistical inferences for functional data*, Ann. Stat. 35 (2007), pp. 1052–1079.
[39] H. Zhou, L.X. Li, and H.T. Zhu, *Tensor regression with applications in neuroimaging data analysis*, J. Am. Stat. Assoc. 108 (2013), pp. 540–552.
[40] X.N. Zhu, Z.R. Cai, and Y.Y. Ma, *Network functional varying coefficient model*, J. Am. Stat. Assoc. 117 (2021), pp. 1–12.