Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

A Dynamic Interaction Semiparametric Function-on-Scalar Model

Hua Liu, Jinhong You & Jiguo Cao

To cite this article: Hua Liu, Jinhong You & Jiguo Cao (2021): A Dynamic Interaction Semiparametric Function-on-Scalar Model, Journal of the American Statistical Association, DOI: 10.1080/01621459.2021.1933496

To link to this article: https://doi.org/10.1080/01621459.2021.1933496

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 00, NO. 0, 1–14: Theory and Methods
https://doi.org/10.1080/01621459.2021.1933496

A Dynamic Interaction Semiparametric Function-on-Scalar Model

Hua Liu[a,c], Jinhong You[b,c], and Jiguo Cao[d]

[a]School of Economics and Finance, Xi'an Jiaotong University, Xi'an, Shaanxi, China; [b]Institute of Data Science and Interdisciplinary Studies, Shanghai Lixin University of Accounting and Finance, Shanghai, China; [c]School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China; [d]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia, Canada

ABSTRACT

Motivated by recent work studying massive functional data, such as the COVID-19 data, we propose a new dynamic interaction semiparametric function-on-scalar (DISeF) model. The proposed model is useful to explore the dynamic interaction among a set of covariates and their effects on the functional response. The proposed model includes many important models investigated recently as special cases. By tensor product B-spline approximating the unknown bivariate coefficient functions, a three-step efficient estimation procedure is developed to iteratively estimate bivariate varying-coefficient functions, the vector of index parameters, and the covariance functions of random effects. We also establish the asymptotic properties of the estimators including the convergence rate and their asymptotic distributions. In addition, we develop a test statistic to check whether the dynamic interaction varies with time/spatial locations, and we prove the asymptotic normality of the test statistic. The finite sample performance of our proposed method and of the test statistic are investigated with several simulation studies. Our proposed DISeF model is also used to analyze the COVID-19 data and the ADNI data. In both applications, hypothesis testing shows that the bivariate varying-coefficient functions significantly vary with the index and the time/spatial locations. For instance, we find that the interaction effect of the population aging and the socio-economic covariates, such as the number of hospital beds, physicians, nurses per 1000 people and GDP per capita, on the COVID-19 mortality rate varies in different periods of the COVID-19 pandemic. The healthcare infrastructure index related to the COVID-19 mortality rate is also obtained for 141 countries estimated based on the proposed DISeF model.

ARTICLE HISTORY
Received May 2020
Accepted May 2021

KEYWORDS
Dynamic effect; Functional data analysis; Hypothesis testing; Profile least squares; Tensor product B-spline

1. Introduction

When a variable is measured or observed at multiple times, the variable can be treated as a function of time. The variable is therefore called a functional variable, and the data for the variable are called functional data (Ramsay and Silverman 2005; Ferraty and Vieu 2006). The functional data are often functions of time, but may also be functions of spatial locations, wavelengths, etc. Functional regression models the relationship among functional and scalar variables, and is widely used in functional data analysis (Morris 2015). The existing literature about functional regression can be divided into three categories depending on whether the responses or covariates are functional or scalar data (i) functional responses with functional covariates (Jiang and Wang 2011; Cai, Xue, and Cao 2021); (ii) scalar responses with functional covariates (Ferré and Yao 2003; Ainsworth, Routledge, and Cao 2011; Liu, Wang, and Cao 2017; Lin et al. 2017; Guan, Lin, and Cao 2020; Jiang et al. 2020); and (iii) functional responses with scalar covariates (Zhu, Li, and Kong 2012; Luo, Zhu, and Zhu 2016; Li, Huang, and Zhu 2017).

The functional regression with functional responses and scalar covariates is also called the function-on-scalar model. Several extensions of the basic function-on-scalar model have been proposed due to the complexity of the real data. For instance, a varying-coefficient model, proposed by Hastie and Tibshirani (1993), allows the regression coefficients to vary over some predictors of interest and is extended to functional data by Zhu, Li, and Kong (2012). Luo, Zhu, and Zhu (2016) proposed a single-index varying-coefficient (SIVC) model to establish a varying association between the functional response (e.g., images) and a set of covariates, especially various clinical variables, such as age and gender. In the SIVC model, the single-index vector is allowed to change with the observed times or spatial locations. This allows one to characterize the dynamic association between the covariates and the functional response. Li, Huang, and Zhu (2017) proposed a novel functional varying coefficient single-index model (FVCSIM) to allow that the relationship between some covariates, denoted as $\boldsymbol{X}$, and the functional response changes with the location/time, and the effect of the remaining covariates, denoted as $\boldsymbol{Z}$, on the functional response is characterized by an index function. However, these researches fail to capture the interaction effects between covariates $\boldsymbol{X}$ and $\boldsymbol{Z}$, besides, it ignores the dynamic effect of the single index of the covariates $\boldsymbol{Z}$.

CONTACT Jiguo Cao *jiguo_cao@sfu.ca* Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A1S6

Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.

Table 1. The number of COVID-19 death cases in various age groups for eight countries including Australia, Brazil, France, Germany, Italy, Japan, Norway, and Sweden.

Country	Australia	Brazil	France	Germany	Italy	Japan	Norway	Sweden
Date of report[a]	12.16	12.16	12.10	12.08	12.02	12.06	12.10	12.07
Age 0-9	0	288	4	7	7	0	0	0
Age 10–19	0	1167	5	3	5	0	0	0
Age 20-29	1	1709	35	24	26	2	0	0
Age 30-39	2	5217	143	46	118	6	3	0
Age 40-49	2	11219	420	153	493	23	6	76
Age 50-59	15	21975	1446	595	1893	67	12	181
Age 60-69	38	39136	4279	1635	5452	194	35	455
Age 70-79	157	45764	8690	4047	14142	556	83	1442
Age 80+	693	50399	24089	12804	33679	1285	243	4881
Total deaths	908	176874	39111	19314	55815	2133	382	7035

[a]For each country, we extracted information from the most up-to-date situational reports as of December 16, 2020.

The dynamic interaction effects between covariates $\boldsymbol{X}$ and $\boldsymbol{Z}$ are found when we analyze the Coronavirus disease 2019 data. Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). According to the official website of the World Health Organization (WHO) (*https://covid19.who.int*), as of December 16, 2020, there have been 72,196,732 confirmed cases of COVID-19 across about 271 countries and territories, resulting in 1,630,521 deaths. It should be pointed out that due to differences in the testing capabilities and policy reasons of various countries, there may be some biases in the published number of reported cases and deaths of COVID-19. The COVID-19 pandemic is impacting the global population in drastic ways. To roughly show the distribution of age among COVID-19 deaths, we choose eight countries with at least 100 reported COVID-19 deaths, including Australia, Brazil, France, Germany, Italy, Japan, Norway, and Sweden. These eight countries also have a breakdown of mortality by age available. Table 1 provides the number of COVID-19 death cases in various age groups for the eight countries. See Table S.3 in the supplementary document for a complete listing of the data source. Figure 1 displays the distribution of COVID-19 deaths by age groups for the eight countries. From Table 1 and Figure 1, we can see that in many countries, although all age groups are at risk of dying from COVID-19, older people face the greatest risk of death due to underlying health conditions (Carroll et al. 2020).

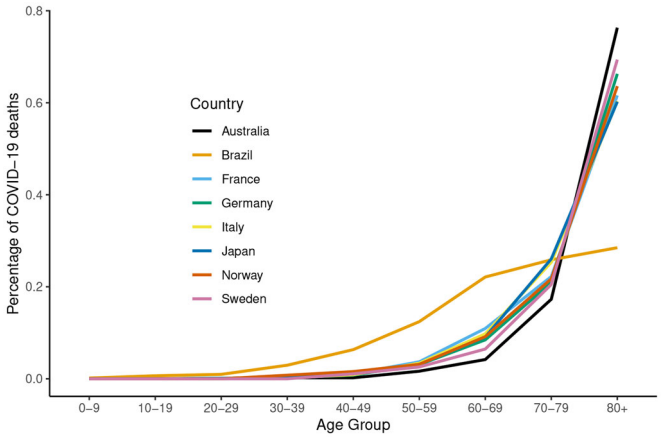


Figure 1. Distribution of COVID-19 deaths by age groups for eight countries including Australia, Brazil, France, Germany, Italy, Japan, Norway, and Sweden.

In this article, we focus on studying how the population aging (age $\geq$ 65) affects the mortality rate of COVID-19. The country-level data on COVID-19 death and confirmed number of cases is obtained from the repository of John Hopkins University (*https://github.com/CSSEGISandData/COVID-19*) and covers up the period from January 22, 2020, when Wuhan City was locked down, to November 30, 2020 (exactly 120 days after 173 countries reach 100 confirmed cases). The COVID-19 mortality rate can be calculated by the equation: death number of cases/total population. We also collect the socio-economic data from the World Bank (*https://data.worldbank.org/indicator*), including the latest total population, the percent of the population with aged 65 and above, the GDP per capita in the U. S. Dollar, the number of physicians per 1000 people, the number of nurses per 1000 people, and the number of hospital beds per 1000 people.

Note that the relationship between the COVID-19 mortality rate and the population aging may not only depend on these socio-economic variables but also changes with time. Due to the different outbreak times in different countries, we choose the time when a country reports the first 100 confirmed cases as the original time of the analysis (Lee et al. 2021; Liu, Moon, and Schorfheide 2021) and focus on the 120 days thereafter. Note that the original time between neighboring countries are very close. For example, the original time in Japan and South Korea are very close, and the original time in the United States and Canada are also very close. After preprocessing and matching, we have data of 141 countries for both the mortality rate and socio-economic covariates variable. To assess how this dependence varies with time, we divide the time into four equal periods: Period I (the 1st–30th day since 100 confirmed cases), Period II (the 31st–60th day since 100 confirmed cases), Period III (the 61st–90th day since 100 confirmed cases), and Period IV (the 91st–120th day since 100 confirmed cases). In each period, we fit an SIVC model $Y(t) = \alpha_1(\boldsymbol{Z}^T\boldsymbol{\beta}) + \texttt{aging} \times \alpha_2(\boldsymbol{Z}^T\boldsymbol{\beta}) + \eta(t)$, where $Y(t)$ is the COVID-19 mortality rate of each country at time t, `aging` is the percentage of the population with aged 65 and above, and $\boldsymbol{Z}$ is the vector of four socio-economic variables including the GDP per capita, the number of hospital beds per 1000 people, the number of physicians per 1000 people, and the number of nurses per 1000 people for each country. Figure 2 displays the two estimated varying coefficients $\alpha_1(\boldsymbol{Z}^T\boldsymbol{\beta})$ and $\alpha_2(\boldsymbol{Z}^T\boldsymbol{\beta})$ from the data in the four periods. These estimated varying coefficients show different patterns for different periods. Such preliminary evidence suggests the dynamic effect of the

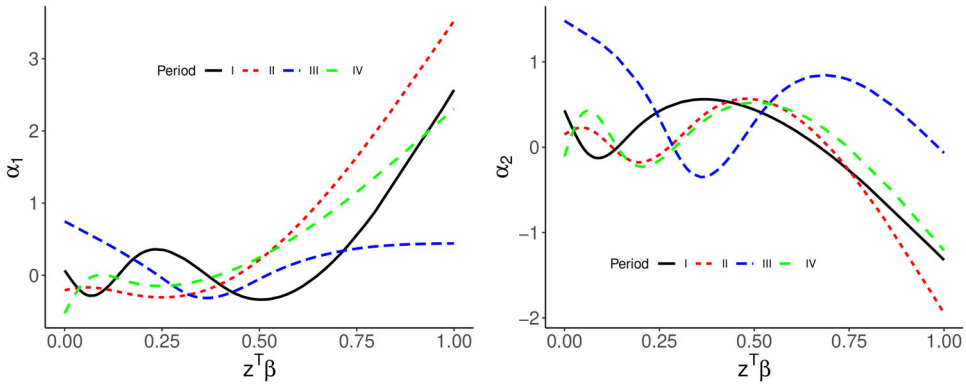


Figure 2. The estimated varying coefficients $\alpha_1(\boldsymbol{Z}^T\boldsymbol{\beta})$ and $\alpha_2(\boldsymbol{Z}^T\boldsymbol{\beta})$ for the SIVC model $Y(t) = \alpha_1(\boldsymbol{Z}^T\boldsymbol{\beta}) + \texttt{aging} \times \alpha_2(\boldsymbol{Z}^T\boldsymbol{\beta}) + \eta(t)$ in the four time periods: Period I (the 1st–30th day since 1000 confirmed cases), Period II (the 31st–60th day since 100 confirmed cases), and Period III (the 61st–90th day since 100 confirmed cases), and Period IV (the 91st–120th day since 100 confirmed cases).

population aging on the COVID-19 mortality rate, which will be confirmed with the formal hypothesis test result in Section 6.

To capture the dynamic interaction effects between the population aging and the socio-economic variables on the COVID-19 mortality rate, we propose a novel semiparametric varying-coefficient model

$$Y(t) = \boldsymbol{X}^T\boldsymbol{\alpha}(t, \boldsymbol{Z}^T\boldsymbol{\beta}) + \eta(t), \tag{1}$$

where $Y(t)$ is a functional response process for $t \in [a, b]$ and superscript T denotes the transpose of a vector or matrix. The effect of the covariate vector $\boldsymbol{X} = (X_1, \ldots, X_p)^T$, quantified by the bivariate varying-coefficient functions $\boldsymbol{\alpha}(t, \boldsymbol{Z}^T\boldsymbol{\beta})$, not only depends on the covariate $\boldsymbol{Z} = (Z_1, \ldots, Z_q)^T$ through a linear index $\boldsymbol{Z}^T\boldsymbol{\beta}$, but also changes with t. Therefore, $\boldsymbol{\alpha}(t, \boldsymbol{Z}^T\boldsymbol{\beta})$ represents the dynamic interaction between $\boldsymbol{X}$ and $\boldsymbol{Z}$. We will estimate $\boldsymbol{\alpha}(\cdot, \cdot)$ with no parametric assumption, while keep the interpretability of the interaction by using the linear index $\boldsymbol{Z}^T\boldsymbol{\beta}$, so model (1) is semiparametric. We call (1) the dynamic interaction semiparametric function-on-scalar (DISeF) model in this article. The random function $\eta(t)$ characterizes individual curve variations and is assumed to be a stochastic process with mean zero and covariance function $R(s, t) = \text{cov}\{\eta(s), \eta(t)\}$.

In real applications, the process $Y(t)$ is not observable, but can be measured at any given time/spatial point s_m with a random error. We sample n subjects $\{(y_i(s_m), \boldsymbol{X}_i, \boldsymbol{Z}_i)|i = 1, \ldots, n, m = 1, \ldots, M\}$, and the sample version of DISeF (1) is written as

$$y_i(s_m) = \boldsymbol{X}_i^T\boldsymbol{\alpha}(s_m, \boldsymbol{Z}_i^T\boldsymbol{\beta}) + \eta_i(s_m) + \epsilon_i(s_m), \tag{2}$$

where $y_i(s_m)$ is the observation for the ith subject at s_m, $\eta_i(s_m)$ is a realization of the subject-specific random function $\eta_i(t)$ at s_m, and $\epsilon_i(s_m) \triangleq \epsilon_{i,m}$ are independent measurement errors with mean zero and variances $\sigma^2(s_m) = \text{Var}(\epsilon_i(s_m))$. In addition, $\eta_i(s_m)$ and $\epsilon_i(s_m)$ are independent.

Various existing models are special cases of the DISeF model (2). For example, when the varying coefficient function $\boldsymbol{\alpha}(s_m, \boldsymbol{Z}_i^T\boldsymbol{\beta})$ in Equation (2) is only a univariate function of s_m, the DISeF model can be reduced to the functional varying coefficient model proposed by Zhu, Li, and Kong (2012). When the covariates vector $\boldsymbol{X}_i$ is 1, the DSIeF model can turn into the bivariate single-index model proposed by Jiang and Wang (2011). Besides, when the varying coefficient function is specified as $\boldsymbol{\alpha}(s_m, \boldsymbol{Z}_i^T\boldsymbol{\beta}) = \alpha_0(\boldsymbol{Z}_i^T\boldsymbol{\beta}) + \tilde{\boldsymbol{\alpha}}(s_m)$, the DSIeF model becomes the functional varying coefficient single-index model proposed by Li, Huang, and Zhu (2017).

We propose a three-step estimation procedure to estimate the bivariate varying-coefficient function $\boldsymbol{\alpha}(\cdot, \cdot)$, the index parameter vector $\boldsymbol{\beta}$, and the covariance functions $R(s, t)$ and $\sigma^2(s_m)$ in the DISeF model (2). At the first step, we develop a profile least-square (PLS) estimation approach for estimating the index parameter vector $\boldsymbol{\beta}$ and the bivariate function $\boldsymbol{\alpha}(\cdot, \cdot)$, in which $\boldsymbol{\alpha}(\cdot, \cdot)$ is approximated by tensor product B-spline basis functions (de Boor 1978). At the second step, we estimate the covariance functions $R(s, t)$ and $\sigma^2(s, s)$ with the help of the PLS estimators. At the final step, we propose a weighted profile least-square (WPLS) method to improve the efficiency of the PLS estimators by borrowing information from the dependence and heteroscedasticity among the time/spatial point t. The asymptotic properties of the proposed estimators are also established. The proposed estimator for the index parameter vector $\boldsymbol{\beta}$ is shown to be $\sqrt{n}$-consistent and asymptotically normally distributed. We also show that the proposed estimator for the bivariate varying-coefficient function $\boldsymbol{\alpha}(\cdot, \cdot)$ is asymptotically consistent and normally distributed.

The rest of this article is organized as follows. In Section 2, an efficient estimation procedure is introduced to iteratively estimate all unknown parameters and functions in the DISeF model. In addition, we extend the estimation procedure to the DISeF model with multi-index. Section 3 systematically investigates the asymptotic properties of the proposed estimators. Section 4 describes the hypothesis testing procedure for the varying coefficient component. The finite-sample performance of the proposed model is evaluated with several simulation studies in Section 5. We also demonstrate the proposed model by analyzing the COVID-19 data in Section 6. Some conclusions and discussion are provided in Section 7. Appendix gives the conditions used in the asymptotic properties.

2. The Estimation Procedure

In this section, we propose to estimate the DISeF model (2) in three steps, which is represented by the flowchart in Figure 3.

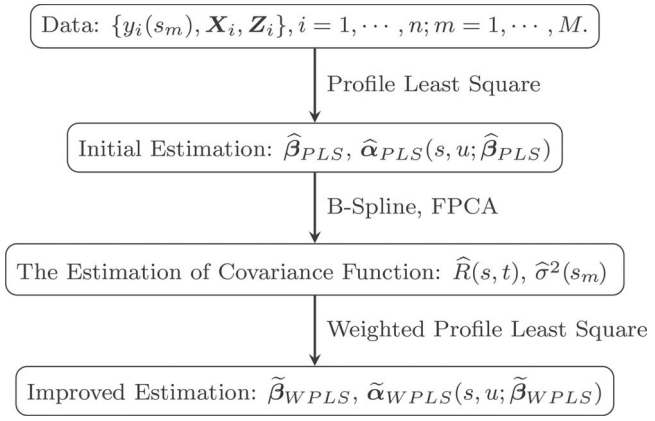


Figure 3. A flowchart for estimating the dynamic interaction DISeF model (2) in three steps.

We will introduce the estimation details for each step in the next three subsections.

A single-index model is not identifiable in the absence of constraints on its structure. For the univariate semiparametric single-index model, Xia et al. (2002) and Lin and Kulasekera (2007) proposed a standard model identification method by assuming that the parameter vector has a positive first component and the norm of the parameter vector equals to one, which means $\boldsymbol{\beta}$ belongs to the parameter space $\boldsymbol{\Theta} = \{\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^T : ||\boldsymbol{\beta}||_2 = 1, \beta_1 > 0\}$ with $\|\cdot\|_2$ being the Euclidean norm. In this paper, to ensure identifiability, we consider using the same identification method for the DISeF model (2). Denote an index by $U(\boldsymbol{\beta}) = \boldsymbol{Z}^T\boldsymbol{\beta}$, which is assumed to be confined in a compact set S_ω, and without loss of generality, set $S_\omega = [0, 1]$.

Proposition 1. Assume that $\boldsymbol{\beta} \in \boldsymbol{\Theta}$ and $\boldsymbol{\alpha}(s, u)$ is continuous and nonconstant over $s \in [0, 1]$ and $u \in [0, 1]$. Then the model (2) is identifiable.

Denote $\boldsymbol{\beta}_{-1} = (\beta_2, \ldots, \beta_q)$, then $\boldsymbol{\beta}_{-1}$ belongs to the space $\boldsymbol{\Theta}_{-1} = \{\boldsymbol{\beta}_{-1} = (\beta_2, \ldots, \beta_q)^T : ||\boldsymbol{\beta}_{-1}||_2^2 < 1\}$, and the original index parameter $\boldsymbol{\beta}$ can be rewritten as

$$\boldsymbol{\beta} = (\sqrt{1 - ||\boldsymbol{\beta}_{-1}||_2^2}, \boldsymbol{\beta}_{-1}^T)^T. \tag{3}$$

The Jacobian matrix is given by $\boldsymbol{J} = \partial\boldsymbol{\beta}/\partial\boldsymbol{\beta}_{-1} = \left(-\boldsymbol{\beta}_{-1}/\sqrt{1 - ||\boldsymbol{\beta}_{-1}||_2^2}, \boldsymbol{I}_{q-1}\right)^T$, where $\boldsymbol{I}_{q-1}$ is an identity matrix.

2.1. Profile Least-Squares Estimation

Due to its desirable numerical stability and computational efficiency in practice (Ma, Song, and Wang 2013; Wang and Yang 2009), tensor product B-splines (de Boor 1978) are commonly used to approximate unknown bivariate nonparametric functions. Let $\mathcal{B}(s) = (B_1(s), \ldots, B_{N_1}(s))^T$ and $\mathcal{B}(u) = (B_1(u), \ldots, B_{N_2}(u))^T$ be two sets of B-spline basis functions of the order κ with L_1 and L_2 internal knots, where the order $\kappa \geq 2$. Let $N_1 = L_1 + \kappa$, $N_2 = L_2 + \kappa$. When $\kappa = 4$, denote the vector of the tensor product cubic B-spline basis functions as $\mathcal{B}(s, u) = \mathcal{B}(s) \otimes \mathcal{B}(u) = (B_l(s)B_k(u) : 1 \leq l \leq N_1, 1 \leq k \leq N_2)^T$. We approximate the bivariate varying-coefficient function $\alpha_d(\cdot, \cdot)$ by a linear combination of tensor product B-spline basis functions: $\alpha_d(\cdot, \cdot) \approx \mathcal{B}(s, u)^T\boldsymbol{\theta}_d$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_p^T)^T$ is the vector of spline coefficients with $\boldsymbol{\theta}_d = (\theta_{lk,d} : 1 \leq l \leq N_1, 1 \leq k \leq N_2)^T$ for $d = 1, \ldots, p$.

We can obtain the estimators of the spline coefficients $\boldsymbol{\theta}$ and the index parameter vector $\boldsymbol{\beta}$ by minimizing

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{m=1}^{M}\sum_{i=1}^{n}\left\{y_i(s_m) - \sum_{d=1}^{p} X_{id}\mathcal{B}\left(s_m, U_i(\boldsymbol{\beta})\right)^T\boldsymbol{\theta}_d\right\}^2. \tag{4}$$

Minimizing (4) is difficult to compute, which requires constrained nonlinear programming. Therefore, we consider an iterative procedure to estimate the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. The detailed estimation steps are given below.

Step 0. Start with an initial value $\widehat{\boldsymbol{\beta}}^{(0)}$ such that $||\widehat{\boldsymbol{\beta}}^{(0)}||_2 = 1$ and $\|\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}\| = O_p(1/\sqrt{n})$.

Step 1. For any given $\boldsymbol{\beta}$, then $\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})$ can be obtained by

$$\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}) = \underset{\boldsymbol{\theta}(\boldsymbol{\beta}) \in \mathbb{R}^{p \times N_1 \times N_2}}{\arg\min} \; L(\boldsymbol{\theta}, \boldsymbol{\beta}). \tag{5}$$

Denote $\mathbf{B}_{mi}(\boldsymbol{\beta}) = (X_{i1}\mathcal{B}\left(s_m, U_i(\boldsymbol{\beta})\right)^T, \ldots, X_{ip}\mathcal{B}(s_m, U_i(\boldsymbol{\beta}))^T)^T = \boldsymbol{X}_i \otimes \mathcal{B}(s_m, U_i(\boldsymbol{\beta}))$. The solution to (5) is

$$\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}) = \left\{\sum_{m=1}^{M}\sum_{i=1}^{n}\mathbf{B}_{mi}(\boldsymbol{\beta})\mathbf{B}_{mi}(\boldsymbol{\beta})^T\right\}^{-1}\sum_{m=1}^{M}\sum_{i=1}^{n}\mathbf{B}_{mi}(\boldsymbol{\beta})y_i(s_m). \tag{6}$$

Let $\boldsymbol{B}_i(\boldsymbol{\beta}) = (\boldsymbol{B}_{1i}(\boldsymbol{\beta}), \ldots, \boldsymbol{B}_{Mi}(\boldsymbol{\beta}))^T$ and $\boldsymbol{y}_i = (y_i(s_1), \ldots, y_i(s_M))^T$. Then Equation (6) can be expressed as $\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta}) = \left\{\sum_{i=1}^{n}\boldsymbol{B}_i(\boldsymbol{\beta})^T\boldsymbol{B}_i(\boldsymbol{\beta})\right\}^{-1}\sum_{i=1}^{n}\boldsymbol{B}_i(\boldsymbol{\beta})^T\boldsymbol{y}_i$. Thus, the estimator of the bivariate varying coefficient function is $\widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(s, u) = (\mathcal{B}(s, u)^T\widehat{\boldsymbol{\theta}}_1(\boldsymbol{\beta}), \ldots, \mathcal{B}(s, u)^T\widehat{\boldsymbol{\theta}}_p(\boldsymbol{\beta}))^T = (\boldsymbol{I}_p \otimes \mathcal{B}(s, u)^T)\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})$. Then the partial derivative $\boldsymbol{\alpha}^{(2)}(s, u) = \partial\boldsymbol{\alpha}(s, u)/\partial u$ is given by $\widehat{\boldsymbol{\alpha}}^{(2)}_{\mathrm{PLS}}(s, u) = (\boldsymbol{I}_p \otimes (\mathcal{B}(s) \otimes \dot{\mathcal{B}}(u))^T)\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})$, where $\dot{\mathcal{B}}(u)$ is the first-order derivative of $\mathcal{B}(u)$.

Step 2. Denote $\widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(\mathcal{S}, U_i(\boldsymbol{\beta})) = (\widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(s_1, U_i(\boldsymbol{\beta})), \ldots, \widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(s_M, U_i(\boldsymbol{\beta})))^T$, and $\widehat{\boldsymbol{\alpha}}^{(2)}_{\mathrm{PLS}}(\mathcal{S}, U_i(\boldsymbol{\beta})) = (\widehat{\boldsymbol{\alpha}}^{(2)}_{\mathrm{PLS}}(s_1, U_i(\boldsymbol{\beta})), \ldots, \widehat{\boldsymbol{\alpha}}^{(2)}_{\mathrm{PLS}}(s_M, U_i(\boldsymbol{\beta})))^T$. We can use the current estimate $\widehat{\boldsymbol{\beta}}^{(k)}$ to update the estimator of $\boldsymbol{\beta}_{-1}$ with the following Fisher scoring algorithm:

$$\widehat{\boldsymbol{\beta}}_{-1}^{(k+1)} = \widehat{\boldsymbol{\beta}}_{-1}^{(k)} - \left\{\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{-1}\partial\boldsymbol{\beta}_{-1}^T}\right\}^{-1}\frac{\partial L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{-1}}|_{\boldsymbol{\beta}_{-1}=\widehat{\boldsymbol{\beta}}_{-1}^{(k)}}, \tag{7}$$

where $\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{-1}\partial\boldsymbol{\beta}_{-1}^T} = \frac{\partial(\partial L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}_{-1})}{\partial\boldsymbol{\beta}_{-1}^T}$, and $\frac{\partial L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{-1}} = -\sum_{i=1}^{n}\{\widehat{\boldsymbol{\alpha}}^{(2)}_{\mathrm{PLS}}(\mathcal{S}, U_i(\boldsymbol{\beta}))\boldsymbol{X}_i\boldsymbol{Z}_i^T\boldsymbol{J} + \boldsymbol{B}_i(\boldsymbol{\beta})\frac{\partial\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{-1}}\}^T\left\{\boldsymbol{y}_i - \widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(\mathcal{S}, U_i(\boldsymbol{\beta}))\boldsymbol{X}_i\right\}$.

Step 3. Repeat Steps 1 and 2 until convergence to obtain the final estimator $\widehat{\boldsymbol{\beta}}_{-1,\mathrm{PLS}}$. We then apply formula (3) to get the estimators $\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}$. Then the tensor product cubic B-spline estimators of the bivariate nonparametric functions $\boldsymbol{\alpha}(s, u)$ and $\boldsymbol{\alpha}^{(2)}(s, u)$ are $\widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(s, u; \widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}) = (\boldsymbol{I}_p \otimes \mathcal{B}(s, u)^T)\widehat{\boldsymbol{\theta}}(\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}})$ and $\widehat{\boldsymbol{\alpha}}^{(2)}_{\mathrm{PLS}}(s, u; \widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}) = (\boldsymbol{I}_p \otimes (\mathcal{B}(s) \otimes \dot{\mathcal{B}}(u))^T)\widehat{\boldsymbol{\theta}}(\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}})$, respectively.

2.2. The Estimation of Covariance Functions

The covariance matrix for the response observed at M grid points in the sample, denoted by $\boldsymbol{\Phi}$, is given by $\boldsymbol{\Phi} = \mathrm{cov}(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i) = (R(s_l, s_k) + \sigma^2(s_l)I(s_l = s_k))_{k,l=1,\ldots,M}$, where $I(\cdot)$ is an indicator function. We want to adapt the covariance matrix to obtain more efficient estimators for $\boldsymbol{\alpha}(\cdot,\cdot)$ and $\boldsymbol{\beta}$. However, $\boldsymbol{\Phi}$ is generally unknown in practical applications, so we need to estimate it to gain the efficient estimators of $\boldsymbol{\alpha}(\cdot,\cdot)$ and $\boldsymbol{\beta}$. With the estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}$ and $\widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(\cdot,\cdot)$ from Section 2.1, we can get

$$y_i(s_m) \approx \boldsymbol{X}_i^T \widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(s_m, U_i(\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}})) + \eta_i(s_m) + \epsilon_i(s_m). \quad (8)$$

Let $y_i^\star(s_m) = y_i(s_m) - \boldsymbol{X}_i^T \widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(s_m, U_i(\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}))$, then Equation (8) can be expressed as $y_i^\star(s_m) = \eta_i(s_m) + \epsilon_i(s_m)$.

Let $\boldsymbol{\mathcal{B}}_\eta(s) = (B_1(s), \ldots, B_{K_\eta}(s))^T$ and $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{iK_\eta})^T$, where K_η is the number of basis functions. Then the individual functions $\eta_i(s_m), i = 1, \ldots, n, m = 1, \ldots, M$, can be approximated by the cubic spline functions: $\eta_i(s_m) \approx \boldsymbol{\mathcal{B}}_\eta(s_m)^T \boldsymbol{b}_i = \sum_{k=1}^{K_\eta} B_k(s_m) b_{ik}$, in which $\boldsymbol{b}_i$ is estimated by

$$\widehat{\boldsymbol{b}}_i = \underset{\boldsymbol{b}_i \in \mathbb{R}^{K_\eta}}{\arg\min} \sum_{m=1}^{M} \left\{ y_i^\star(s_m) - \boldsymbol{\mathcal{B}}_\eta(s_m)^T \boldsymbol{b}_i \right\}^2. \quad (9)$$

Denote $\boldsymbol{y}_i^\star = (y_i^\star(s_1), \ldots, y_i^\star(s_M))^T$ and $\boldsymbol{B}_\eta = (\boldsymbol{\mathcal{B}}_\eta(s_1), \ldots, \boldsymbol{\mathcal{B}}_\eta(s_M))^T$. The solution to Equation (9) is expressed as $\widehat{\boldsymbol{b}}_i = (\boldsymbol{B}_\eta^T \boldsymbol{B}_\eta)^{-1} \boldsymbol{B}_\eta^T \boldsymbol{y}_i^\star$. Thus, the cubic spline estimator of $\eta_i(s_m)$ is $\widehat{\eta}_i(s_m) = \boldsymbol{\mathcal{B}}_\eta(s_m)^T \widehat{\boldsymbol{b}}_i$.

The covariance function $R(s,t)$ can thus be estimated by an empirical covariance estimator $\widehat{R}(s,t) = 1/n \sum_{i=1}^n \widehat{\eta}_i(s)\widehat{\eta}_i(t)$. Subsequently, we can calculate the spectral decomposition (Li and Hsing 2010; Hall and Hosseini-Nasab 2006) of $\widehat{R}(s,t)$ as follows: $\widehat{R}(s,t) = \sum_{l=1}^{\infty} \widehat{\lambda}_l \widehat{\psi}_l(s)\widehat{\psi}_l(t)$, where $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq 0$ are estimated eigenvalues and the $\widehat{\psi}_l(s), l = 1, \ldots,$ are the corresponding estimated eigenfunctions.

The variance function $\sigma^2(s)$ measures the variation of $\epsilon(s)$. It can be estimated based on the residuals $\widehat{\epsilon}_i(s) = y_i^\star(s) - \widehat{\eta}_i(s)$. Define the spline estimator for $\sigma^2(s) = \mathbb{E}(\epsilon_i^2(s))$ as $\widehat{\sigma}^2(s) = \sum_{k=1}^{K_{\sigma^2}} B_k(s)\widehat{c}_k$, where K_{σ^2} is the number of basis functions, and $\widehat{\boldsymbol{c}} = (\widehat{c}_1, \ldots, \widehat{c}_{K_{\sigma^2}})^T$ is given by $\widehat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c} \in \mathbb{R}^{K_{\sigma^2}}} \sum_{i=1}^n \sum_{m=1}^M \left\{ \widehat{\epsilon}_i^2(s_m) - \sum_{k=1}^{K_{\sigma^2}} B_k(s_m) c_k \right\}^2$. Denote $\widehat{\boldsymbol{\epsilon}}_i^2 = (\widehat{\epsilon}_i^2(s_1), \ldots, \widehat{\epsilon}_i^2(s_M))^T$, $\boldsymbol{\mathcal{B}}_\epsilon(s_m) = (B_1(s_m), \ldots, B_{K_{\sigma^2}}(s_m))^T$, and $\boldsymbol{B}_\epsilon = (\boldsymbol{\mathcal{B}}_\epsilon(s_1), \ldots, \boldsymbol{\mathcal{B}}_\epsilon(s_M))^T$, then $\widehat{\boldsymbol{c}}$ can be expressed as $\widehat{\boldsymbol{c}} = (\boldsymbol{B}_\epsilon^T \boldsymbol{B}_\epsilon)^{-1} (\boldsymbol{B}_\epsilon^T \sum_{i=1}^n \widehat{\boldsymbol{\epsilon}}_i^2 / n)$. So, the spline estimator of $\sigma^2(s)$ is given by $\widehat{\sigma}^2(s) = \boldsymbol{\mathcal{B}}_\epsilon(s)^T \widehat{\boldsymbol{c}}$. Thus, $\boldsymbol{\Phi}$ can be estimated as $\widehat{\boldsymbol{\Phi}} = (\widehat{R}(s_l, s_k) + \widehat{\sigma}^2(s_l) I(s_l = s_k))_{k,l=1,\ldots,M}$.

2.3. WPLS Estimation

In this section, we adapt the covariance matrix to obtain more efficient estimators of the spline coefficients $\boldsymbol{\theta}$ and the index parameter vector $\boldsymbol{\beta}$ by minimizing

$$\widetilde{L}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \boldsymbol{y}_i - \boldsymbol{B}_i(\boldsymbol{\beta})\widetilde{\boldsymbol{\theta}}(\boldsymbol{\beta}) \right\}^T \boldsymbol{\Phi}^{-1} \left\{ \boldsymbol{y}_i - \boldsymbol{B}_i(\boldsymbol{\beta})\widetilde{\boldsymbol{\theta}}(\boldsymbol{\beta}) \right\}. \quad (10)$$

Similarly, it is difficult to directly minimize (10). Therefore, we consider an iterative estimation procedure as follows.

Step 0. Get an initial value $\widehat{\boldsymbol{\beta}}^{(0)}$ such that $\|\widehat{\boldsymbol{\beta}}^{(0)}\|_2 = 1$ and $\|\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}\| = O_p(1/\sqrt{n})$.

Step 1. For any given $\boldsymbol{\beta}$, we can obtain

$$\widetilde{\boldsymbol{\theta}}(\boldsymbol{\beta}) = \left\{ \sum_{i=1}^{n} \boldsymbol{B}_i(\boldsymbol{\beta})^T \boldsymbol{\Phi}^{-1} \boldsymbol{B}_i(\boldsymbol{\beta}) \right\}^{-1} \left\{ \sum_{i=1}^{n} \boldsymbol{B}_i(\boldsymbol{\beta})^T \boldsymbol{\Phi}^{-1} \boldsymbol{y}_i \right\}, \quad (11)$$

and $\widetilde{\boldsymbol{\alpha}}(s,u) = (\boldsymbol{I}_p \otimes \mathcal{B}(s,u)^T)\widetilde{\boldsymbol{\theta}}(\boldsymbol{\beta})$.

Step 2. We can update the estimator of $\boldsymbol{\beta}_{-1}$ using the following iterative procedure:

$$\widehat{\boldsymbol{\beta}}_{-1}^{(k+1)} = \widehat{\boldsymbol{\beta}}_{-1}^{(k)} - \left\{ \frac{\partial^2 \widetilde{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{-1} \partial \boldsymbol{\beta}_{-1}^T} \right\}^{-1} \frac{\partial \widetilde{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{-1}} \Big|_{\boldsymbol{\beta}_{-1} = \widehat{\boldsymbol{\beta}}_{-1}^{(k)}}, \quad (12)$$

where

$$\partial \widetilde{L}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}_{-1} = -\sum_{i=1}^{n} \left\{ \widetilde{\boldsymbol{\alpha}}^{(2)}(\mathcal{S}, U_i(\boldsymbol{\beta})) \boldsymbol{X}_i \boldsymbol{Z}_i^T \boldsymbol{J} + \boldsymbol{B}_i(\boldsymbol{\beta}) \frac{\partial \widetilde{\boldsymbol{\theta}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{-1}} \right\}^T \boldsymbol{\Phi}^{-1} \left\{ \boldsymbol{y}_i - \widetilde{\boldsymbol{\alpha}}(\mathcal{S}, U_i(\boldsymbol{\beta})) \boldsymbol{X}_i \right\}.$$

Step 3. Repeat Steps 1 and 2 until convergence to obtain the final estimator $\widehat{\boldsymbol{\beta}}_{-1,\mathrm{WPLS}}$. Then, we can get the estimators $\widehat{\boldsymbol{\beta}}_{\mathrm{WPLS}} = (\sqrt{1 - ||\widehat{\boldsymbol{\beta}}_{-1,\mathrm{WPLS}}||_2^2}, \widehat{\boldsymbol{\beta}}_{-1,\mathrm{WPLS}}^T)^T$, $\widehat{\boldsymbol{\alpha}}_{\mathrm{WPLS}}(s,u; \widehat{\boldsymbol{\beta}}_{\mathrm{WPLS}}) = (\boldsymbol{I}_p \otimes \mathcal{B}(s,u)^T)\widetilde{\boldsymbol{\theta}}(\widehat{\boldsymbol{\beta}}_{\mathrm{WPLS}})$ and $\widehat{\boldsymbol{\alpha}}^{(2)}_{\mathrm{WPLS}}(s,u; \widehat{\boldsymbol{\beta}}_{\mathrm{WPLS}}) = (\boldsymbol{I}_p \otimes (\mathcal{B}(s) \otimes \dot{\mathcal{B}}(u))^T)\widetilde{\boldsymbol{\theta}}(\widehat{\boldsymbol{\beta}}_{\mathrm{WPLS}})$.

When we replace the unknown covariance matrix $\boldsymbol{\Phi}$ in Equations (10) and (11) by its estimator gained from Section 2.2, we can obtain the new objective function

$$\widehat{L}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \boldsymbol{y}_i - \boldsymbol{B}_i(\boldsymbol{\beta})\widetilde{\boldsymbol{\theta}}(\boldsymbol{\beta}) \right\}^T \widehat{\boldsymbol{\Phi}}^{-1} \left\{ \boldsymbol{y}_i - \boldsymbol{B}_i(\boldsymbol{\beta})\widetilde{\boldsymbol{\theta}}(\boldsymbol{\beta}) \right\}. \quad (13)$$

For any given $\boldsymbol{\beta}$, we minimize $\widehat{L}(\boldsymbol{\theta}, \boldsymbol{\beta})$ and obtain the estimator $\widetilde{\boldsymbol{\theta}}^*(\boldsymbol{\beta}) = \left\{ \sum_{i=1}^n \boldsymbol{B}_i(\boldsymbol{\beta})^T \widehat{\boldsymbol{\Phi}}^{-1} \boldsymbol{B}_i(\boldsymbol{\beta}) \right\}^{-1} \sum_{i=1}^n \boldsymbol{B}_i(\boldsymbol{\beta})^T \widehat{\boldsymbol{\Phi}}^{-1} \boldsymbol{y}_i$. The other details in the iterative procedure of minimizing (13) is similar with those of minimizing (10) and will not be repeated here. Finally, we obtain the feasible refined estimator $\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}}$ for the parameter vector $\boldsymbol{\beta}$. Thus, the feasible refined tensor product cubic B-spline estimator of the bivariate nonparametric function $\boldsymbol{\alpha}(s,u)$ and $\boldsymbol{\alpha}^{(2)}(s,u)$ can be expressed as $\widetilde{\boldsymbol{\alpha}}_{\mathrm{WPLS}}(s,u) = (\boldsymbol{I}_p \otimes \mathcal{B}(s,u)^T)\widetilde{\boldsymbol{\theta}}^*(\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}})$ and $\widetilde{\boldsymbol{\alpha}}^{(2)}_{\mathrm{WPLS}}(s,u) = (\boldsymbol{I}_p \otimes (\mathcal{B}(s) \otimes \dot{\mathcal{B}}(u))^T)\widetilde{\boldsymbol{\theta}}^*(\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}})$, respectively.

2.4. Tuning Parameter Selection

In Sections 2.1 and 2.3, there are four tuning parameters to be chosen. In the PLS and WPLS steps, we need to choose the number of interior knots L_1 and L_2. As a common practice in the spline literature, we select the number of interior knots via a data-driven method and position them at equal intervals on the sample quantiles. According to the Bayes information

criterion (BIC), we choose the optimal number of interior knots by minimizing the following BIC function:

$$\mathrm{BIC}(L_1,L_2)=\log\left\{\frac{1}{nM}\sum_{i=1}^{n}\sum_{m=1}^{M}\left(y_i(s_m)-\boldsymbol{X}_i^T\widehat{\boldsymbol{\alpha}}(s_m,U_i(\widehat{\boldsymbol{\beta}}))\right)^2\right\}+\frac{\log(nM)}{2nM}(L_1+\kappa)(L_2+\kappa).$$

If $\widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(\cdot,\cdot)$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}$ are used in the BIC function, we can choose the optimal number of interior knots for the profile least squares estimation at the first step. Similarly, we use $\widetilde{\boldsymbol{\alpha}}_{\mathrm{WPLS}}(\cdot,\cdot)$ and $\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}}$ in the BIC function to choose the optimal number of interior knots for the weighted profile least squares estimation at the third step. At both steps, BIC is minimized on the range $[(nM)^{\frac{1}{6}}]\le L_1\le[2(nM)^{\frac{1}{6}}]+1$ and $[(nM)^{\frac{1}{6}}]\le L_2\le[2(nM)^{\frac{1}{6}}]+1$, where $[a]$ denotes the closest integer to a.

In Section 2.2, we need to choose the number of basis functions K_η and K_{σ^2}, which are selected by generalized cross-validation (GCV): $\mathrm{GCV}(K_\eta)=1/(nM)\sum_{i=1}^{n}\sum_{m=1}^{M}(y_i^*(s_m)-\hat{\eta}_i(s_m))^2\times\left(1-K_\eta/(nM)\right)^{-2}$, and $\mathrm{GCV}(K_{\sigma^2})$ has the same form as $\mathrm{GCV}(K_\eta)$ except that $y_i^*(s_m)$ and $\hat{\eta}_i(s_m)$ are replaced by $\widehat{\epsilon}_i^2(s_m)$ and $\hat{\sigma}^2(s_m)$, respectively.

2.5. Extension to Multiple-index Models

In some applications, the index parameters $\boldsymbol{\beta}$ could be different for each covariate $X_d, d=1,\dots,p$. By some minor modifications, our method could be extended to tackle this situation. Especially, we consider the following dynamic interaction DISeF model which allows the index parameters to vary with different X_d:

$$y_i(s_m)=\sum_{d=1}^{p}X_{id}\alpha_d(s_m,\boldsymbol{Z}_i^T\boldsymbol{\beta}_d)+\eta_i(s_m)+\epsilon_i(s_m).$$

Correspondingly, the identification condition is $\boldsymbol{\beta}\in\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}=\{\boldsymbol{\beta}=(\boldsymbol{\beta}_1^T,\dots,\boldsymbol{\beta}_p^T)^T:\|\boldsymbol{\beta}_d\|_2=1,\beta_{d1}>0,d=1,\dots,p\}$. For each d, let $\boldsymbol{\beta}_{d,-1}=(\beta_{d2},\dots,\beta_{dq})^T$, $U_i(\boldsymbol{\beta}_d)=\boldsymbol{Z}_i^T\boldsymbol{\beta}_d$ and the Jacobian matrix $\boldsymbol{J}_d=\left(-\boldsymbol{\beta}_{d,-1}/\sqrt{1-\|\boldsymbol{\beta}_{d,-1}\|_2^2},\boldsymbol{I}_{q-1}\right)^T$. The computational steps described in Sections 2.1 and 2.3 are modified as follows to accommodate estimating the unknown index parameters $\boldsymbol{\beta}_d$ and bivariate functions $\alpha_d(\cdot,\cdot),d=1,\dots,p$. Replacing $\boldsymbol{B}_{mi}(\boldsymbol{\beta})$ in these sections by $\mathbf{B}_{mi}(\boldsymbol{\beta})=(X_{i1}\mathcal{B}\left(s_m,U_i(\boldsymbol{\beta}_1)\right)^T,\dots,X_{ip}\mathcal{B}\left(s_m,U_i(\boldsymbol{\beta}_p)\right)^T)^T$. Then $\partial L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}_{-1}$ in Equation (7) and $\partial\widetilde{L}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}_{-1}$ in Equation (12) are replaced, respectively, by $\partial L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}_{-1}=-\sum_{i=1}^{n}\{\widehat{\boldsymbol{\mathcal{A}}}_{i1},\dots,\widehat{\boldsymbol{\mathcal{A}}}_{ip}\}^T\left\{\boldsymbol{y}_i-\sum_{d=1}^{p}\widehat{\boldsymbol{\alpha}}_{d,\mathrm{PLS}}(\mathcal{S},U_i(\boldsymbol{\beta}_d))X_{id}\right\}$ and $\partial\widetilde{L}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}_{-1}=-\sum_{i=1}^{n}\left\{\widetilde{\boldsymbol{\mathcal{A}}}_{i1},\dots,\widetilde{\boldsymbol{\mathcal{A}}}_{ip}\right\}^T\boldsymbol{\Phi}^{-1}\left\{\boldsymbol{y}_i-\sum_{d=1}^{p}\widetilde{\boldsymbol{\alpha}}_d\left(\mathcal{S},U_i(\boldsymbol{\beta}_d)\right)X_{id}\right\}$, where $\widehat{\boldsymbol{\mathcal{A}}}_{id}=\widehat{\boldsymbol{\alpha}}_{d,\mathrm{PLS}}^{(2)}(\mathcal{S},U_i(\boldsymbol{\beta}_d))X_{id}\boldsymbol{Z}_i^T\boldsymbol{J}_d+\boldsymbol{B}_i(\boldsymbol{\beta})\frac{\partial\widehat{\boldsymbol{\theta}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{d,-1}}$, $\widetilde{\boldsymbol{\mathcal{A}}}_{id}=\widetilde{\boldsymbol{\alpha}}_d^{(2)}(\mathcal{S},U_i(\boldsymbol{\beta}_d))X_{id}\boldsymbol{Z}_i^T\boldsymbol{J}_d+\boldsymbol{B}_i(\boldsymbol{\beta})\frac{\partial\widetilde{\boldsymbol{\theta}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_{d,-1}}$, $\widehat{\boldsymbol{\alpha}}_{d,\mathrm{PLS}}(\mathcal{S},U_i(\boldsymbol{\beta}_d))=(\widehat{\alpha}_{d,\mathrm{PLS}}(s_1,U_i(\boldsymbol{\beta}_d)),\dots,\widehat{\alpha}_{d,\mathrm{PLS}}(s_M,U_i(\boldsymbol{\beta}_d)))^T$ and $\widehat{\boldsymbol{\alpha}}_{d,\mathrm{PLS}}^{(2)}(\mathcal{S},U_i(\boldsymbol{\beta}_d))=(\widehat{\alpha}_{d,\mathrm{PLS}}^{(2)}(s_1,U_i(\boldsymbol{\beta}_d)),\dots,\widehat{\alpha}_{d,\mathrm{PLS}}^{(2)}(s_M,U_i(\boldsymbol{\beta}_d)))^T$. After these modifications, the iterative algorithm described in Sections 2.1 and 2.3 can be carried out to estimate $\boldsymbol{\beta}_d,d=1,\dots,p$ simultaneously.

3. Asymptotic Properties

We first define some notations. The true values of $\boldsymbol{\beta}$ are denoted by $\boldsymbol{\beta}_0$. For positive number sequences a_n and b_n, let $a_n\lesssim b_n$ mean that a_n/b_n is bounded, $a_n\asymp b_n$mean that $a_n\lesssim b_n$ and $b_n\lesssim a_n$. Besides, $a_n<<b_n$ means $\lim_n a_n/b_n=0$. Let $|\mathcal{S}|$ be the cardinality of the set $\mathcal{S}$ and $\mathcal{S}_{(i)}$ be the ith element of $\mathcal{S}$ from small to large. For any matrix $\boldsymbol{A}$, let $\boldsymbol{A}^{\otimes2}=\boldsymbol{A}\boldsymbol{A}^T$, $\mathrm{tr}(A)$ be the trace of $\boldsymbol{A}$, and for two positive semi-definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{A}\ge\boldsymbol{B}$ means that $\boldsymbol{A}-\boldsymbol{B}$ is positive semi-definite. Denote the space of r-order smooth functions defined on $[a,b]$ as $\mathcal{C}^{(r)}[a,b]=\{m|m^r\in\mathcal{C}[a,b]\}$, where $\mathcal{C}[a,b]$ is the collection of real-valued functions that are bounded and continuous in $[a,b]$. Similar with Ma and Song (2015), define the space $\mathcal{M}=\{\boldsymbol{g}(\boldsymbol{S},U(\boldsymbol{\beta}_0))=(\boldsymbol{g}(s_1,U(\boldsymbol{\beta}_0)),\dots,\boldsymbol{g}(s_M,U(\boldsymbol{\beta}_0)))^T,\boldsymbol{g}(s_m,U(\boldsymbol{\beta}_0))=(g_1(s_m,U(\boldsymbol{\beta}_0)),\dots,g_p(s_m,U(\boldsymbol{\beta}_0)))^T,\mathbb{E}(g_d(s_m,U(\beta_0))^2)<\infty,d=1,\dots,p;m=1,\dots,M\}$. For $1\le k\le q$, denote $\boldsymbol{g}_k(\boldsymbol{S},U(\boldsymbol{\beta}_0))=\arg\min_{\boldsymbol{g}_k^0\in\mathcal{M}}\mathbb{E}\{(\mathbf{1}_M\otimes Z_k-\boldsymbol{g}_k^0(\boldsymbol{S},U(\boldsymbol{\beta}_0))\boldsymbol{X})^T(\mathbf{1}_M\otimes Z_k-\boldsymbol{g}_k^0(\boldsymbol{S},U(\boldsymbol{\beta}_0))\boldsymbol{X})\}$,

$$\mathbb{P}(Z_k)=\frac{1}{M}\mathbf{1}_M^T\boldsymbol{g}_k(\boldsymbol{S},U(\boldsymbol{\beta}_0))\boldsymbol{X}=\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{g}_k(s_m,U(\boldsymbol{\beta}_0))^T\boldsymbol{X},\tag{14}$$

$\mathbb{P}(\boldsymbol{Z})=\{\mathbb{P}(Z_1),\dots,\mathbb{P}(Z_q)\}^T$, and $\widetilde{\boldsymbol{Z}}=\boldsymbol{Z}-\mathbb{P}(\boldsymbol{Z})$, where $\mathbf{1}_M$ is an $M\times1$ vector of 1s. Define $\mathbb{P}_n(Z_{ik})=M^{-1}\mathbf{1}_M^T\boldsymbol{B}_i(\boldsymbol{\beta}_0)\widehat{\boldsymbol{\theta}}_{Z_k}(\boldsymbol{\beta}_0)$, where $\widehat{\boldsymbol{\theta}}_{Z_k}(\boldsymbol{\beta}_0)=\left\{\sum_{i=1}^{n}\boldsymbol{B}_i(\boldsymbol{\beta}_0)^T\boldsymbol{B}_i(\boldsymbol{\beta}_0)\right\}^{-1}\sum_{i=1}^{n}\boldsymbol{B}_i(\boldsymbol{\beta}_0)^T(\mathbf{1}_M\otimes Z_{ik})$. Besides, define $\boldsymbol{g}_k^*$ by $\boldsymbol{g}_k^*(\boldsymbol{S},U(\boldsymbol{\beta}_0))=\arg\min_{\boldsymbol{g}_k^{*,0}\in\mathcal{M}}\mathbb{E}\{(\mathbf{1}_M\otimes Z_k-\boldsymbol{g}_k^{*,0}(\boldsymbol{S},U(\boldsymbol{\beta}_0))\boldsymbol{X})^T\boldsymbol{\Phi}^{-1}(\mathbf{1}_M\otimes Z_k-\boldsymbol{g}_k^{*,0}(\boldsymbol{S},U(\boldsymbol{\beta}_0))\boldsymbol{X})\}$,

$$\mathbb{Q}(Z_k)=\frac{1}{M}\mathbf{1}_M^T\boldsymbol{g}_k^*(\boldsymbol{S},U(\boldsymbol{\beta}_0))\boldsymbol{X}=\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{g}_k^*(s_m,U(\boldsymbol{\beta}_0))^T\boldsymbol{X},\tag{15}$$

$\mathbb{Q}(\boldsymbol{Z})=\{\mathbb{Q}(Z_1),\dots,\mathbb{Q}(Z_q)\}^T$, and $\overline{\boldsymbol{Z}}=\boldsymbol{Z}-\mathbb{Q}(\boldsymbol{Z})$. Denote $\mathbb{Q}_n(Z_{ik})=M^{-1}\mathbf{1}_M^T\boldsymbol{B}_i(\boldsymbol{\beta}_0)\widetilde{\boldsymbol{\theta}}_{Z_k}(\boldsymbol{\beta}_0)$, where $\widetilde{\boldsymbol{\theta}}_{Z_k}(\boldsymbol{\beta}_0)=\left\{\sum_{i=1}^{n}\boldsymbol{B}_i(\boldsymbol{\beta}_0)^T\boldsymbol{\Phi}^{-1}\ \boldsymbol{B}_i(\boldsymbol{\beta}_0)\right\}^{-1}\sum_{i=1}^{n}\boldsymbol{B}_i(\boldsymbol{\beta}_0)^T\boldsymbol{\Phi}^{-1}(\mathbf{1}_M\otimes Z_{ik})$. Here, $\mathbb{P}(Z_k)$ and $\mathbb{Q}(Z_k)$ are both averaged projections of Z_k.

We state the following theorems, whose detailed conditions and proofs can be found in the appendix and supplementary document, respectively. The first theorem establishes the weak convergence and asymptotic result of $\widehat{\boldsymbol{\beta}}_{-1,\mathrm{PLS}}$.

Theorem 1. Under Conditions (C1)-(C5), $nM/(N_1N_2^3)\to\infty$ and $nM^2(N_1N_2)^{-2r}\to0$ as $n,M\to\infty$, we have (i) (consistency) $||\widehat{\boldsymbol{\beta}}_{-1,\mathrm{PLS}}-\boldsymbol{\beta}_{-1,0}||_2=O_p(n^{-1/2})$; (ii) (asymptotic normality) $\sqrt{n}\boldsymbol{\Sigma}^{-1/2}(\widehat{\boldsymbol{\beta}}_{-1,\mathrm{PLS}}-\boldsymbol{\beta}_{-1,0})\xrightarrow{D}\mathbb{N}(\mathbf{0},\boldsymbol{I}_{q-1})$, where

$$\boldsymbol{\Sigma}=\left\{\mathbb{E}(\boldsymbol{\mathcal{H}}_i^{\otimes2})\right\}^{-1}\mathbb{E}(\boldsymbol{\mathcal{H}}_i\boldsymbol{\Phi}\boldsymbol{\mathcal{H}}_i^T)\left\{\mathbb{E}(\boldsymbol{\mathcal{H}}_i^{\otimes2})\right\}^{-1},\tag{16}$$

$\boldsymbol{\mathcal{H}}_i^T=\boldsymbol{\alpha}^{(2)}(\mathcal{S},U_i(\boldsymbol{\beta}_0))\boldsymbol{X}_i\widetilde{\boldsymbol{Z}}_i^T\boldsymbol{J}$ and $\boldsymbol{\alpha}^{(2)}(\mathcal{S},U_i(\boldsymbol{\beta}_0))=(\boldsymbol{\alpha}^{(2)}(s_1,U_i(\boldsymbol{\beta}_0)),\dots,\boldsymbol{\alpha}^{(2)}(s_M,U_i(\boldsymbol{\beta}_0)))^T$.

Remark 1. Due to formula (3) and $\partial\boldsymbol{\beta}/\partial\boldsymbol{\beta}_{-1} = \boldsymbol{J}$, by Theorem 1 with an application of the multivariate delta method, we can get (i) $||\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}} - \boldsymbol{\beta}_0||_2 = O_p(n^{-1/2})$ and (ii) $\sqrt{n}(\boldsymbol{J}\boldsymbol{\Sigma}\boldsymbol{J}^T)^{-1/2}(\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}} - \boldsymbol{\beta}_0) \xrightarrow{\mathrm{D}} \mathbb{N}(\boldsymbol{0}, \boldsymbol{I}_{\mathrm{q}})$.

The following theorem provides the convergence rate and asymptotic distribution of $\widehat{\alpha}_d(s,u;\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}})$, $d = 1,\ldots,p$.

Theorem 2. Under Conditions (C1)–(C5), $nM/(N_1N_2^3) \to \infty$ and $nM^2(N_1N_2)^{-2r} \to 0$ as $n, M \to \infty$, for any s,u on $[0,1]^2$, we have (i) (consistency) for each $1 \le d \le p$, $|\widehat{\alpha}_{d,\mathrm{PLS}}(s,u;\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}) - \alpha_d(s,u)| = O_p(N_1^{-r} + N_2^{-r} + \sqrt{\frac{N_1N_2}{nM}})$; (ii) (asymptotic normality) under $(N_1N_2)^{2r+1}/(nM) \to \infty$ as $n \to \infty$, $\sqrt{Mn/(N_1N_2)}\boldsymbol{\Sigma}_\alpha(s,u)^{-1/2}(\widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(s,u;\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}) - \boldsymbol{\alpha}(s,u)) \xrightarrow{D} \mathbb{N}(\boldsymbol{0}, \boldsymbol{I}_p)$, where

$$\boldsymbol{\Sigma}_\alpha(s,u) = \frac{M}{N_1N_2}\boldsymbol{L}_i(s,u)\,\mathbb{E}(\boldsymbol{B}_i(\boldsymbol{\beta_0})^T\boldsymbol{\Phi}\boldsymbol{B}_i(\boldsymbol{\beta_0}))\boldsymbol{L}_i(s,u)^T, \quad (17)$$

with $\boldsymbol{L}_i(s,u) = (\boldsymbol{I}_p \otimes \mathcal{B}(s,u))^T\left\{\mathbb{E}(\boldsymbol{B}_i(\boldsymbol{\beta_0})^T\boldsymbol{B}_i(\boldsymbol{\beta_0}))\right\}^{-1}$.

Remark 2. The order assumptions regarding N_1 and N_2, in Theorem 2 (i) imply that $N_1 \asymp (nM)^{1/(2r+2)}$ and $N_2 \asymp (nM)^{1/(2r+2)}$, which is the optimal order for the number of interior knots needed to estimate the bivariate nonparametric functions. The convergence rate of estimator $\widehat{\alpha}_{d,\mathrm{PLS}}(s,u;\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}})$ is $O_p((nM)^{-r/(2r+2)})$. In Theorem 2 (ii), the required condition $(N_1N_2)^{r+1}/(nM) \to \infty$ for making the bias asymptotically negligible simply means that one need use larger number of knots than what are needed for achieving the optimal rate of convergence.

We next study the convergence rate of the estimators $\widehat{\eta}_i(s)$ and $\widehat{\sigma}^2(s)$ for the individual random function $\eta_i(s)$ and the covariance function of error term.

Theorem 3. (i) Under Conditions of Theorem 1-2, (C6), (C8), and $K_\eta \to \infty, K_\eta M^{-1} \to 0$, as $M, n \to \infty$, one has $|\widehat{\eta}_i(s) - \eta_i(s)| = O_p(K_\eta^{-r_1} + \sqrt{K_\eta M^{-1}})$ for any $s \in [0,1]$.

(ii) Under Conditions of Theorem 1-2, (C8) and $K_{\sigma^2} \to \infty$, $K_{\sigma^2}(nM)^{-1} \to 0$, as $M, n \to \infty$, one has $|\widehat{\sigma}^2(s) - \sigma^2(s)| = O_p(K_\eta^{-2r_1} + K_\eta M^{-1} + K_{\sigma^2}^{-r_2})$ for any $s \in [0,1]$.

The next theorem provides the asymptotic properties of the estimated covariance function of the individual random function $\{\eta_i(s) : s \in [0,1]\}$ and its spectrum decomposition.

Theorem 4. Under Conditions (C1), (C2)-(i),(ii),(iv), (C3)–(C6) and (C8), and $N_1, N_2, K_\eta \to \infty$, $nM/(N_1N_2^3) \to \infty$, $nM^2(N_1N_2)^{-2r} \to 0$, $K_\eta M^{-1} \to 0$ as $n, M \to \infty$, we have

(i) $|\widehat{R}(s,t) - R(s,t)| = O_p(K_\eta^{-r_1} + \sqrt{\frac{K_\eta}{M}})$ for any $(s,t) \in [0,1]^2$;

(ii) under Condition (C7), for $l \ge 1$, (a) $\int_0^1\{\widehat{\psi}_l(s) - \psi_l(s)\}^2 ds = O_p(K_\eta^{-r_1} + \sqrt{K_\eta/M})$ and (b) $|\widehat{\lambda}_l - \lambda_l| = O_p(K_\eta^{-r_1} + \sqrt{K_\eta/M})$.

We then present the asymptotic results for the feasible refined estimators which fully acknowledge the dependence among different time/spatial points in the error process.

Theorem 5. Under Conditions (C1)–(C8), $nM/(N_1N_2^3) \to \infty$, $nM^2(N_1N_2)^{-2r} \to 0$, $K_\eta M^{-1} \to 0$ as $n, M \to \infty$,

(i) for feasible refined estimator $\widetilde{\boldsymbol{\beta}}_{-1,\mathrm{WPLS}}$, we have: (a) (consistency) $||\widetilde{\boldsymbol{\beta}}_{-1,\mathrm{WPLS}} - \boldsymbol{\beta}_{-1,0}||_2 = O_p(n^{-1/2})$; (b) (asymptotic normality) $\sqrt{n}\boldsymbol{\Omega}^{-1/2}(\widetilde{\boldsymbol{\beta}}_{-1,\mathrm{WPLS}} - \boldsymbol{\beta}_{-1,0}) \xrightarrow{D} \mathbb{N}(\boldsymbol{0}, \boldsymbol{I}_{q-1})$, where

$$\boldsymbol{\Omega} = \left\{\mathbb{E}(\overline{\boldsymbol{\mathcal{H}}}_i\boldsymbol{\Phi}^{-1}\overline{\boldsymbol{\mathcal{H}}}_i^T)\right\}^{-1}, \quad \text{and} \quad \overline{\boldsymbol{\mathcal{H}}}_i^T = \boldsymbol{\alpha}^{(2)}(\mathcal{S}, U_i(\boldsymbol{\beta}_0))\boldsymbol{X}_i\overline{\boldsymbol{Z}}_i^T\boldsymbol{J}. \quad (18)$$

(ii) for any s,u on $[0,1]^2$ and the feasible refined estimator of nonparametric function, we have: (a)(consistency) for each $1 \le d \le p$, $|\widetilde{\alpha}_{d,\mathrm{WPLS}}(s,u;\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}}) - \alpha_d(s,u)| = O_p(\sqrt{\frac{N_1N_2}{nM}})$; (b) (asymptotic normality) $\sqrt{\frac{nM}{N_1N_2}}\boldsymbol{\Omega}_\alpha(s,u)^{-1/2}$ $(\widetilde{\boldsymbol{\alpha}}_{\mathrm{WPLS}}(s,u;\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}}) - \boldsymbol{\alpha}(s,u)) \xrightarrow{D} \mathbb{N}(\boldsymbol{0}, \boldsymbol{I}_p)$, where

$$\boldsymbol{\Omega}_\alpha(s,u) = \frac{M}{N_1N_2}(\boldsymbol{I}_p \otimes \mathcal{B}(s,u))^T\left\{\mathbb{E}(\boldsymbol{B}_i(\boldsymbol{\beta_0})^T\boldsymbol{\Phi}^{-1}\boldsymbol{B}_i \times (\boldsymbol{\beta_0}))\right\}^{-1}(\boldsymbol{I}_p \otimes \mathcal{B}(s,u)). \quad (19)$$

Remark 3. Similarly, by Theorem 5-(i) with an application of the multivariate delta method, we can get (i) $||\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}} - \boldsymbol{\beta}_0||_2 = O_p(n^{-1/2})$, and (ii) $\sqrt{n}(\boldsymbol{J}\boldsymbol{\Omega}\boldsymbol{J}^T)^{-1/2}(\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}} - \boldsymbol{\beta}_0) \xrightarrow{D} \mathbb{N}(\boldsymbol{0}, \boldsymbol{I}_q)$. By Lemma 5.1 in Jiang (2010), it is easy to prove that the covariance matrices of $\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}}$ and $\widetilde{\boldsymbol{\alpha}}_{\mathrm{WPLS}}(s,u;\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}})$ are smaller than those of $\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}$ and $\widehat{\boldsymbol{\alpha}}_{\mathrm{PLS}}(s,u;\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}})$, respectively. So, the improved estimators are more asymptotically efficient.

Remark 4. In Theorem 1–5, all asymptotic covariance matrices (16)–(19) are unknown and needed to estimate. There, $\widetilde{\boldsymbol{Z}}$ can be estimated by $\widehat{\widetilde{\boldsymbol{Z}}} = \boldsymbol{Z} - \mathbb{P}_n(\boldsymbol{Z};\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}})$, $\overline{\boldsymbol{Z}}$ can be estimated by $\widehat{\overline{\boldsymbol{Z}}} = \boldsymbol{Z} - \mathbb{Q}_n(\boldsymbol{Z};\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}},\widehat{\boldsymbol{\Phi}})$, $\widehat{\boldsymbol{\mathcal{H}}}_i^T = \widehat{\boldsymbol{\alpha}}^{(2)}_{\mathrm{PLS}}(\mathcal{S}, U_i(\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}))\boldsymbol{X}_i\widehat{\widetilde{\boldsymbol{Z}}}_i^T\boldsymbol{J}$ and $\widehat{\overline{\boldsymbol{\mathcal{H}}}}_i^T = \widetilde{\boldsymbol{\alpha}}^{(2)}_{\mathrm{WPLS}}(\mathcal{S}, U_i(\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}}))\boldsymbol{X}_i\widehat{\overline{\boldsymbol{Z}}}_i^T\boldsymbol{J}$. Thus we can estimate the covariance matrix (16) by $\widehat{\boldsymbol{\Sigma}} = \left\{\sum_{i=1}^n \widehat{\boldsymbol{\mathcal{H}}}_i^{\otimes 2}\right\}^{-1}\sum_{i=1}^n \widehat{\boldsymbol{\mathcal{H}}}_i\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\mathcal{H}}}_i^T \left\{\sum_{i=1}^n \widehat{\boldsymbol{\mathcal{H}}}_i^{\otimes 2}\right\}^{-1}$. For $\boldsymbol{\Sigma}_\alpha$ in (17), it can be estimated by $\widehat{\boldsymbol{\Sigma}}_\alpha(s,u) = nM/(N_1N_2)\widehat{\boldsymbol{L}}_i(s,u)\sum_{i=1}^n(\boldsymbol{B}_i(\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}})^T\widehat{\boldsymbol{\Phi}}\boldsymbol{B}_i(\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}))\widehat{\boldsymbol{L}}_i(s,u)^T$, where $\widehat{\boldsymbol{L}}_i(s,u) = (\boldsymbol{I}_p \otimes \mathcal{B}(s,u))^T\{\sum_{i=1}^n(\boldsymbol{B}_i(\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}})^T\boldsymbol{B}_i(\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}))\}^{-1}$. Similarly, the estimator of $\boldsymbol{\Omega}$ is $\widehat{\boldsymbol{\Omega}} = \{\sum_{i=1}^n \widehat{\overline{\boldsymbol{\mathcal{H}}}}_i\widehat{\boldsymbol{\Phi}}^{-1}\widehat{\overline{\boldsymbol{\mathcal{H}}}}_i^T\}^{-1}$. And, for $\boldsymbol{\Omega}_\alpha$, it can be estimated by $\widehat{\boldsymbol{\Omega}}_\alpha(s,u) = nM/(N_1N_2)(\boldsymbol{I}_p \otimes \mathcal{B}(s,u))^T\{\sum_{i=1}^n \boldsymbol{B}_i(\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}})^T\widehat{\boldsymbol{\Phi}}^{-1}\boldsymbol{B}_i(\widetilde{\boldsymbol{\beta}}_{\mathrm{WPLS}})\}^{-1}(\boldsymbol{I}_p \otimes \mathcal{B}(s,u))$.

4. Hypothesis Test

We are interested in testing whether some bivariate varying-coefficient functions in the DISeF model (2), $\alpha_d(s_m, U_i(\boldsymbol{\beta}))$, $d = 1,\ldots,p$, do not change with the index $U_i(\boldsymbol{\beta})$. In other words, the null hypothesis can be expressed as follows:

$$H_0 : \alpha_d(t,u) = \alpha_d(t), \qquad \text{a.s. on } [0,1]^2,$$

where $d \in \mathcal{S}^\star \subseteq \{1,\ldots,p\}$, and $\mathcal{S}^\star$=$\{d$: we want to test whether $\alpha_d(t,u)$ changes with the index $U_i(\boldsymbol{\beta})\}$. Note that the

DISeF model (2) reduces to a varying coefficient model under the null hypothesis and $\mathcal{S}^\star = \{1, \ldots, p\}$. Let $\boldsymbol{\alpha}_{\mathcal{S}^\star}(t,u)$ be the $|\mathcal{S}^\star| \times 1$ vector of all $\alpha_d(t,u)$, $d \in \mathcal{S}^\star$, and $\boldsymbol{\alpha}(t)$ be an $|\mathcal{S}^\star| \times 1$ vector of univariate functions $\alpha_d(t)$, $d \in \mathcal{S}^\star$. There always exists a matrix $\boldsymbol{A}$ such that $\boldsymbol{\alpha}_{\mathcal{S}^\star}(t,u) = \boldsymbol{A}\boldsymbol{\alpha}(t,u)$, where $\boldsymbol{A}$ is a $|\mathcal{S}^\star| \times p$ matrix with the elements $a_{ij} = 1$ if $j \in \mathcal{S}^\star$ and $a_{ij} = 0$ otherwise. Then the null hypothesis can be expressed equivalently as

$$H_0 : \int_0^1 \int_0^1 ||\boldsymbol{A}\boldsymbol{\alpha}(t,u) - \boldsymbol{\alpha}(t)||^2 dtdu = 0. \tag{20}$$

A natural way to come up with a test statistic for Equation (20) is to replace the unknown functions $\boldsymbol{\alpha}(t,u)$, and $\boldsymbol{\alpha}(t)$ by their nonparametric spline estimators. We can then define the L_2-distance test statistic

$$\mathcal{T} = \frac{nM}{N_1 N_2} \int_0^1 \int_0^1 ||\boldsymbol{A}\widetilde{\boldsymbol{\alpha}}_{\text{WPLS}}(t,u;\widetilde{\boldsymbol{\beta}}_{\text{WPLS}}) - \widetilde{\boldsymbol{\alpha}}(t;\widetilde{\boldsymbol{\beta}}^0_{\text{WPLS}})||^2 dtdu, \tag{21}$$

where $\widetilde{\boldsymbol{\alpha}}(t;\widetilde{\boldsymbol{\beta}}^0_{\text{WPLS}})$ and $\widetilde{\boldsymbol{\beta}}^0_{\text{WPLS}}$ are the improved estimators taking the correlation structure into account under the null hypothesis. The estimation procedure under the null hypothesis is put in Section S1 of the supplementary document.

In the next theorem, we present a central limit theorem for the test statistic $\mathcal{T}$ under the null hypothesis.

Theorem 6. Suppose that Conditions (C1)–(C8) hold. If $nM/(N_1N_2^3) \to \infty$, $nM^2(N_1N_2)^{-2r} \to 0$, $K_\eta M^{-1} \to 0$ as $n, M \to \infty$, then under H_0, it follows that $\mathcal{T} - B \xrightarrow{D} \mathbb{N}(0, V)$, where $B = M/(N_1N_2)\int_0^1\int_0^1 \text{tr}(\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\Omega}_\alpha(t,u))dtdu$, and $V = 8(M/(N_1N_2))^2 \int_0^1\int_0^1 \text{tr}(\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\Omega}_\alpha^{\otimes 2}(t,u))dtdu$.

To investigate the power of the test, we define the local alternatives converging to the null hypothesis as the sample size grows; that is $H_{1,\alpha} : \boldsymbol{A}\boldsymbol{\alpha}(t,u) = \boldsymbol{\alpha}(t) + c_{n,M}\boldsymbol{\delta}(t,u)$, where $c_{n,M} \to 0$ as $n, M \to \infty$ and $\boldsymbol{\delta}(\cdot,\cdot)$ is a continuous function vector.

Theorem 7 gives the asymptotic distribution of the test statistic under the local alternative hypothesis.

Theorem 7. Suppose the conditions of Theorem 6 hold and let $c_{n,M} = \sqrt{N_1N_2/(nM)}$. Then, under the local alternative hypothesis $H_{1,\alpha}$, we have $\mathcal{T} - B \xrightarrow{D} \mathbb{N}(\widetilde{\boldsymbol{\delta}}, V)$, with $\widetilde{\boldsymbol{\delta}} = \int_0^1\int_0^1 ||\boldsymbol{\delta}(t,u)||^2 dtdu$.

Although we can prove the asymptotic distribution of $\mathcal{T}$ under the null or alternative hypothesis, such a theoretical result does not produce a good approximation when the sample size is actively small. Thus, we propose the following wild bootstrap test procedure.

Step 1. Perform the proposed method to estimate $\widetilde{\boldsymbol{\beta}}_{\text{WPLS}}$, $\widetilde{\alpha}_d(s_m, U_i(\widetilde{\boldsymbol{\beta}}_{\text{WPLS}}))$, $\widehat{\eta}_i(s_m)$ and $\widehat{\epsilon}_i(s_m)$.

Step 2. Under H_0, we obtain the estimators $\widetilde{\alpha}_{d\notin\mathcal{S}^\star,\text{WPLS}}(s_m, U_i(\widetilde{\boldsymbol{\beta}}^0_{\text{WPLS}}))$, $\widetilde{\boldsymbol{\alpha}}(s_m;\widetilde{\boldsymbol{\beta}}^0_{\text{WPLS}})$, $\widetilde{\boldsymbol{\beta}}^0_{\text{WPLS}}$ and then compute the estimated $\mathcal{T}$ using Equation (21).

Step 3. Draw $\xi_i^{(b)}$ and $\xi_{im}^{(b)}$ for $i = 1, \ldots, n$ and $m = 1, \ldots, M$ independently from the standard normal distribution and the bootstrap sample is given by $\{y_i^{(b)}(s_m), \boldsymbol{X}_i, \boldsymbol{Z}_i\}$, where we define $y_i^{(b)}(s_m) = \sum_{d\notin\mathcal{S}^\star} \widetilde{\alpha}_{d,\text{WPLS}}(s_m, U_i(\widetilde{\boldsymbol{\beta}}^0_{\text{WPLS}}))X_{id} + (X_{id})^T_{d\in\mathcal{S}^\star}\widetilde{\boldsymbol{\alpha}}(s_m;\widetilde{\boldsymbol{\beta}}^0_{\text{WPLS}}) + \widehat{\eta}_i(s_m)\xi_i^{(b)} + \widehat{\epsilon}_i(s_m)\xi_{im}^{(b)}$.

Step 4. Given the bootstrap sample $\{y_i^{(b)}(s_m), \boldsymbol{X}_i, \boldsymbol{Z}_i\}$, we can then compute the corresponding bootstrap test statistics $T^{(b)}$

$$\mathcal{T}^{(b)} = \frac{nM}{N_1N_2}\int_0^1\int_0^1 ||\boldsymbol{A}\widetilde{\boldsymbol{\alpha}}^{(b)}_{\text{WPLS}}(t,u;\widetilde{\boldsymbol{\beta}}^{(b)}_{\text{WPLS}}) - \widetilde{\boldsymbol{\alpha}}^{(b)} \times (t;\widetilde{\boldsymbol{\beta}}^{0,(b)}_{\text{WPLS}})||^2 dtdu, \tag{22}$$

where $\widetilde{\boldsymbol{\alpha}}^{(b)}_{\text{WPLS}}(t,u;\widetilde{\boldsymbol{\beta}}^{(b)}_{\text{WPLS}})$, $\widetilde{\boldsymbol{\alpha}}^{(b)}(t;\widetilde{\boldsymbol{\beta}}^{0,(b)}_{\text{WPLS}})$, $\widetilde{\boldsymbol{\beta}}^{0,(b)}_{\text{WPLS}}$and $\widetilde{\boldsymbol{\beta}}^{(b)}_{\text{WPLS}}$ are the estimators for the bootstrap sample.

Step 5. Repeat Step 3 and 4 B times to obtain $\{\mathcal{T}^{(b)} : b = 1, \ldots, B\}$ and then estimate the test p-value as $B^{-1}\sum_{b=1}^B \boldsymbol{I}(\mathcal{T}^{(b)} \geq \mathcal{T})$. Reject H_0 when the p-value is lower than a prespecified significant level α.

The next theorem shows that the bootstrap test statistic $\mathcal{T}^{(b)}$ has the same asymptotic distribution as $\mathcal{T}$.

Theorem 8. Under the conditions of Theorem 6, the bootstrap statistic $\mathcal{T}^{(b)}$ defined in Equation (22) has the distribution $\mathcal{T}^{(b)} - B \xrightarrow{D} \mathbb{N}(0, V)$.

The test procedure about whether the bivariate varying coefficient functions are related to space or time is similar to the above test, which is omitted for simplicity.

5. Simulation Studies

In this section, we conduct two simulation studies to illustrate the finite performances of our proposed model, estimation procedure and hypothesis testing. In the supplementary document, we conduct an additional simulation study to compare the performance of the DISeF model with the SIVCM.

5.1. Simulation I

In this simulation study, the data are generated from the DISeF model (2), in which $s_m \sim U[0,1]$, $(X_{i1}, X_{i2})^T \overset{\text{iid}}{\sim} \mathbb{N}(\boldsymbol{0}, 2\boldsymbol{I}_2)$ and $\boldsymbol{X}_i = (1, X_{i1}, X_{i2})^T$ for all $i = 1, \ldots, n$ and $m = 1, \ldots, M$. For the random term, $\eta_i(s) = \xi_{i1}\psi_1(s) + \xi_{i2}\psi_2(s)$, where $\xi_{il} \overset{\text{iid}}{\sim} N(0, \lambda_l)$ for $l = 1, 2$. The measurement errors $\epsilon_i(s_m) = \epsilon_{i,m} \overset{\text{iid}}{\sim} N(0, \sigma^2(s_m))$. We generate $\boldsymbol{Z}$ from a three-dimensional standard normal distribution with the index parameter vector $\boldsymbol{\beta} = (1,2,3)^T/\sqrt{14}$. Besides, $s_m, X_{i1}, X_{i2}, \xi_{ijl}$ and $\epsilon_i(s_m)$ are independent random variables. We set $(\lambda_1, \lambda_2, \sigma^2(s_m)) = (1.2, 1.2, 0.2)$ and the eigenfunctions and varying coefficient functions as follows: $\psi_1(s) = \sqrt{2}\sin(2\pi s)$, $\psi_2(s) = \sqrt{2}\cos(2\pi s)$, $\alpha_1(s,u) = 5\sin(s)\cos(u)$, $\alpha_2(s,u) = 5\sin(u)\cos(s)$ and $\alpha_3 = 5u\cos(s)$.

We conduct extensive simulation studies under different settings and report the estimation results for $M = 30, 50$ and $n = 200, 300, 400$ in 500 simulation replicates. We apply the estimation procedure in Section 2 to each simulated dataset and calculate all unknown parameters and functions. Table 2 summarizes the biases for the estimator of individual parameters. It shows that all biases are close to 0, which confirms

Table 2. The biases ($\times 10^{-2}$), the averaged ASEs ($\times 10^{-2}$), the ESEs ($\times 10^{-2}$), the empirical CP of the 95% confidence intervals for the estimators of parameters β_1, β_2 and β_3; the means and SEs of the RASEs defined in Equation (23) of the estimators for the varying coefficient functions $\alpha_d(\cdot,\cdot), d = 1, 2, 3$, using the profile least-square (PLS) method and the WPLS method when the sample size $n = 200, 300, 400$ and the number of observations per subject $M = 30, 50$.

			β_1		β_2		β_3	
	n	M	30	50	30	50	30	50
Bias (CP)	200	PLS	0.140(0.97)	0.167(0.97)	0.129(0.97)	0.147(0.98)	0.087(0.96)	0.105(0.98)
		WPLS	**0.049(0.97)**	**0.039(0.96)**	**0.046(0.98)**	**0.025(0.98)**	**0.029(0.98)**	**0.020(0.98)**
	300	PLS	0.121(0.95)	0.138(0.96)	0.101(0.97)	0.115(0.97)	0.076(0.97)	0.084(0.97)
		WPLS	**0.035(0.95)**	**0.028(0.95)**	**0.033(0.97)**	**0.023(0.97)**	**0.025(0.97)**	**0.016(0.97)**
	400	PLS	0.107(0.95)	0.118(0.95)	0.093(0.95)	0.104(0.95)	0.070(0.95)	0.075(0.95)
		WPLS	**0.031(0.95)**	**0.025(0.95)**	**0.030(0.95)**	**0.021(0.95)**	**0.023(0.95)**	**0.014(0.95)**
ASE (ESE)	200	PLS	0.211(0.192)	0.226(0.212)	0.184(0.175)	0.210(0.193)	0.124(0.118)	0.147(0.133)
		WPLS	**0.059(0.060)**	**0.046(0.047)**	**0.064(0.065)**	**0.036(0.033)**	**0.040(0.041)**	**0.027(0.025)**
	300	PLS	0.175(0.163)	0.185(0.183)	0.132(0.130)	0.161(0.143)	0.101(0.095)	0.115(0.108)
		WPLS	**0.052(0.051)**	**0.035(0.033)**	**0.059(0.060)**	**0.030(0.029)**	**0.039(0.038)**	**0.021(0.020)**
	400	PLS	0.144(0.141)	0.177(0.164)	0.116(0.113)	0.124(0.122)	0.075(0.073)	0.085(0.083)
		WPLS	**0.042(0.041)**	**0.031(0.030)**	**0.056(0.055)**	**0.026(0.025)**	**0.033(0.032)**	**0.019(0.018)**
			α_1		α_2		α_3	
	n	M	30	50	30	50	30	50
Mean(SE)	200	PLS	0.847(1.383)	0.893(1.701)	0.752(0.545)	0.796(0.728)	0.837(0.627)	0.881(0.957)
		WPLS	**0.755(0.955)**	**0.624(0.770)**	**0.703(0.523)**	**0.666(0.503)**	**0.738(0.489)**	**0.643(0.423)**
	300	PLS	0.733(1.026)	0.787(1.415)	0.567(0.502)	0.603(0.613)	0.552(0.543)	0.575(0.594)
		WPLS	**0.642(0.887)**	**0.516(0.630)**	**0.526(0.486)**	**0.510(0.443)**	**0.541(0.403)**	**0.525(0.334)**
	400	PLS	0.590(0.926)	0.609(1.195)	0.495(0.484)	0.520(0.516)	0.513(0.325)	0.547(0.354)
		WPLS	**0.560(0.735)**	**0.435(0.594)**	**0.473(0.446)**	**0.456(0.390)**	**0.487(0.307)**	**0.448(0.278)**

Table 3. The means and SEs of the RASEs of the estimators for eigenfunctions $\psi_l(\cdot), l = 1, \ldots, 2$; the bias and ESE of the estimators for eigenvalues $\lambda_l, l = 1, \ldots, 2$ and error term variance σ^2 when the sample size $n = 200, 300, 400$ and the number of observations per subject $M = 30, 50$.

		$M = 30$			$M = 50$		
		$n = 200$	$n = 300$	$n = 400$	$n = 200$	$n = 300$	$n = 400$
Mean(SE)	ψ_1	0.158(0.176)	0.133(0.152)	0.111(0.131)	0.099(0.114)	0.098(0.092)	0.080(0.081)
	ψ_2	0.143(0.144)	0.126(0.121)	0.103(0.110)	0.102(0.102)	0.083(0.093)	0.075(0.083)
Bias(ESE)	λ_1	0.188(0.103)	0.168(0.092)	0.135(0.073)	0.160(0.097)	0.132(0.082)	0.108(0.066)
	λ_2	0.387(0.092)	0.356(0.080)	0.303(0.076)	0.330(0.081)	0.270(0.076)	0.236(0.067)
	σ^2	0.075(0.002)	0.072(0.003)	0.069(0.002)	0.046(0.003)	0.042(0.002)	0.041(0.001)

the consistency of these estimators. In addition, the WPLS estimators have much smaller biases than the PLS estimators. Table 2 also gives the average asymptotic standard errors (ASEs) calculated through Theorem 1 and Theorem 5 and the empirical standard errors (ESEs) among 500 simulation replicates. The ESE becomes smaller as n increases. The biases, ASEs and ESEs of the WPLS estimators also decrease as M increases. On the other hand, the PLS estimators do not have this trend due to the ignorance of spatial/temporal correlation. Moreover, the ASEs and the corresponding ESEs are very comparable in each case. Table 2 also shows the empirical coverage probabilities (CP) of the 95% confidence intervals for individual parameters $\beta_l, l = 1, \ldots, q$, where standard errors (SEs) are calculated according to the asymptotic formula. It is clear that all CP approach the 95% level as the sample size n increases. This result is confirmatory to the asymptotic normals of the index parameter estimators established in Theorems 1 and 5.

In the nonparametric part, to evaluate the performance of the estimator, $\widehat{\alpha}_d, d = 1, \ldots, p$, we consider the root-average-square-errors (RASEs):

$$\mathrm{RASE}(\widehat{\alpha}_d) = \sqrt{\frac{1}{nM}\sum_{m=1}^{M}\sum_{i=1}^{n}(\widehat{\alpha}_d(s_m, u_i) - \alpha_d(s_m, u_i))^2}. \quad (23)$$

Table 2 displays the means and SEs of the RASEs in 500 simulation replicates. It shows that the RASEs of $\widehat{\alpha}_d, d = 1, \ldots, p$, become smaller when n increases and $\widetilde{\alpha}_d$ always performs better than $\widehat{\alpha}_d$.

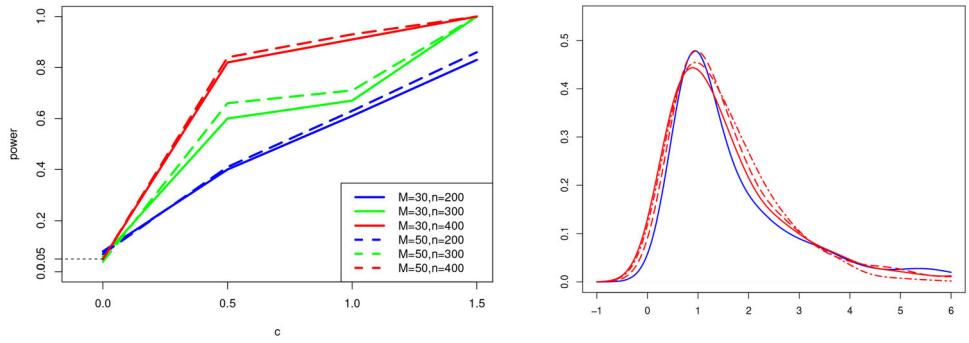
Now, we turn to the estimation of the covariance functions. In Table 3, we show the bias and ESEs of the estimators of eigenvalue $\lambda_l, l = 1, 2$ and variance σ^2. From this table, we can see that the bias and ESE of these estimators decrease as the sample size increases. About the estimated eigenfunctions $\widehat{\psi}_l, l = 1, 2$, we also consider the RASE. The result of $\widehat{\psi}_l$ is showed in Table 3. From this table, we can see that the RASE of $\psi_l(\cdot), l = 1, 2$ decreases as n, M increases.

5.2. Simulation II

In this simulation study, the simulation settings are analogous with those in Simulation I, except

$$\alpha_1(s,u) = \sin(s)(1 + c\cos(u)), \quad \alpha_2(s,u) = (1 + c\sin(u))s, \quad \text{and} \quad \alpha_3 = \cos(s)(1 + c2u), \quad (24)$$

which is designed to test $H_0 : \alpha_d(s,u) = \alpha_d(s), d = 1, \ldots, p$, where $c = 0, 0.5, 1$, and 1.5. When $c = 0$, the bivariate functions reduce to univariate functions about s, and a larger c indicates a stronger iteraction effect. We run the simulation



(a) (b)

Figure 4. (a) Estimated powers under different c in Equation (24) when the sample size $n = 200, 300, 400$ and the number of observations per subject $M = 30, 50$. (b) The simulated density of the L_2-distance test statistic $\mathcal{T}$ defined in Equation (21) (solid line) and three bootstrap approximations (dashed lines) when the sample size $n = 400$ and the number of observations per subject $M = 50$.

Table 4. Results for testing whether the bivariate varying coefficient functions in Equation (25) can reduce to univariate functions.

H_0	$\boldsymbol{\alpha}(s_m, U_i(\boldsymbol{\beta}))$ $= \boldsymbol{\alpha}(s_m)$	$\alpha_1(s_m, U_i(\boldsymbol{\beta}))$ $= \alpha_1(s_m)$	$\alpha_2(s_m, U_i(\boldsymbol{\beta}))$ $= \alpha_2(s_m)$	$\boldsymbol{\alpha}(s_m, U_i(\boldsymbol{\beta}))$ $= \boldsymbol{\alpha}^*(U_i(\boldsymbol{\beta}))$	$\alpha_1(s_m, U_i(\boldsymbol{\beta}))$ $= \alpha_1^*(U_i(\boldsymbol{\beta}))$	$\alpha_2(s_m, U_i(\boldsymbol{\beta}))$ $= \alpha_2^*(U_i(\boldsymbol{\beta}))$
p-value	0.039	0.006	0.038	0.032	0.027	0.032

Table 5. The estimated coefficients, SEs, the lower bound (LB), and the upper bound (UB) of 95% confidence intervals for index parameters in the dynamic interaction DISeF model (25) using the profile least square (PLS) method and the weighted profile least square (WPLS) method from the COVID-19 data.

	PLS				WPLS			
Covariates	Coefficients	SE	LB	UB	Coefficients	SE	LB	UB
GDP	0.983	0.045	0.894	1.072	0.870	0.003	0.865	0.876
Physician	0.011	0.223	-0.435	0.458	0.489	0.005	0.480	0.499
Nurse	-0.183	0.231	-0.644	0.279	0.059	0.006	0.048	0.070
Bed	0.006	0.190	-0.374	0.387	0.007	0.005	-0.003	0.018

500 times under different sample sizes. For each run, we draw 500 bootstrap resamples. The corresponding estimated powers are shown in Figure 4 (a). It shows that the actual size is close to the nominal size of 0.05 and the power goes to 1 rapidly as c increases. Moreover, the power increases as the sample size increases. We plot the density functions of the test statistic $\mathcal{T}$ based on the 500 simulation replications and its bootstrap approximation in Figure 4 (b) when the sample size $n = 400$ and the number of observations per subject $M = 50$. It shows that the bootstrap approximations are close to the asymptotic null distribution. For the sake of simplicity, we omit the plots for other c, which have similar results.

6. Real Data Analysis

The proposed DISeF model is demonstrated via the analysis of COVID-19 data and ADNI data. This section presents the analysis of COVID-19 data introduced in Section 1. The analysis results on ADNI data can be found in the supplementary document.

The functional response variable $y_i(s_m)$ is the mortality rate at the mth day since 100 confirmed cases for the ith country, where $m = 1, \ldots, 120$. The socio-economic covariates in the vector $\boldsymbol{Z} = (Z_1, Z_2, Z_3, Z_4)^T$ are the GDP per capita (Z_1), the number of physicians per 1000 people (Z_2), the number of nurses per 1000 people (Z_3), and the number of hospital beds per 1000 people (Z_4) of each country. Covariates of interest consist of the intercept ($X_1 = 1$), and the percentage of population with the age 65 and above (X_2). We standardized all covariates and the point-wise response variable to have mean 0 and variance 1.

We fit the following DISeF model:

$$y_i(s_m) = \alpha_1(s_m, \boldsymbol{Z}_i^T\boldsymbol{\beta}) + X_{i2}\alpha_2(s_m, \boldsymbol{Z}_i^T\boldsymbol{\beta}) + \eta_i(s_m) + \epsilon_i(s_m). \quad (25)$$

The above equation can be expressed as $\boldsymbol{y}(s_m) = (\boldsymbol{B}_{m1}(\boldsymbol{\beta}), \ldots, \boldsymbol{B}_{mn}(\boldsymbol{\beta}))^T\boldsymbol{\theta} + \boldsymbol{\eta}(s_m) + \boldsymbol{\epsilon}(s_m)$, where $\boldsymbol{y}(s_m) = (y_1(s_m), \ldots, y_n(s_m))^T$, $\boldsymbol{\eta}(s_m) = (\eta_1(s_m), \ldots, \eta_n(s_m))^T$, and $\boldsymbol{\epsilon}(s_m) = (\epsilon_1(s_m), \ldots, \epsilon_n(s_m))^T$.

In the COVID-19 data, there is obvious dependency among countries (Cao, Ayako, and Scott 2020). To address this problem, we first estimate the spatial dependence among countries by using the exponential correlation matrix $\boldsymbol{\Delta}$ (Minasny and McBratney 2005). The (i,j)th element of $\boldsymbol{\Delta}$ is $\Delta_{ij} = \exp\left(-d_{ij}/\rho\right)$, where d_{ij} is the distance between the centers of the two countries i and j based on their GPS coordinates, and ρ is the correlation parameter. We choose the Vincenty's formula (Vincenty 1975) to calculate the distance d_{ij}. Although the original time of each country has been reset, the maximum likelihood estimator (MLE) $\widehat{\rho}_{\text{MLE}} = 1.031$ of the parameter ρ (Bachoc 2018) in the spatial dependence is obtained based on

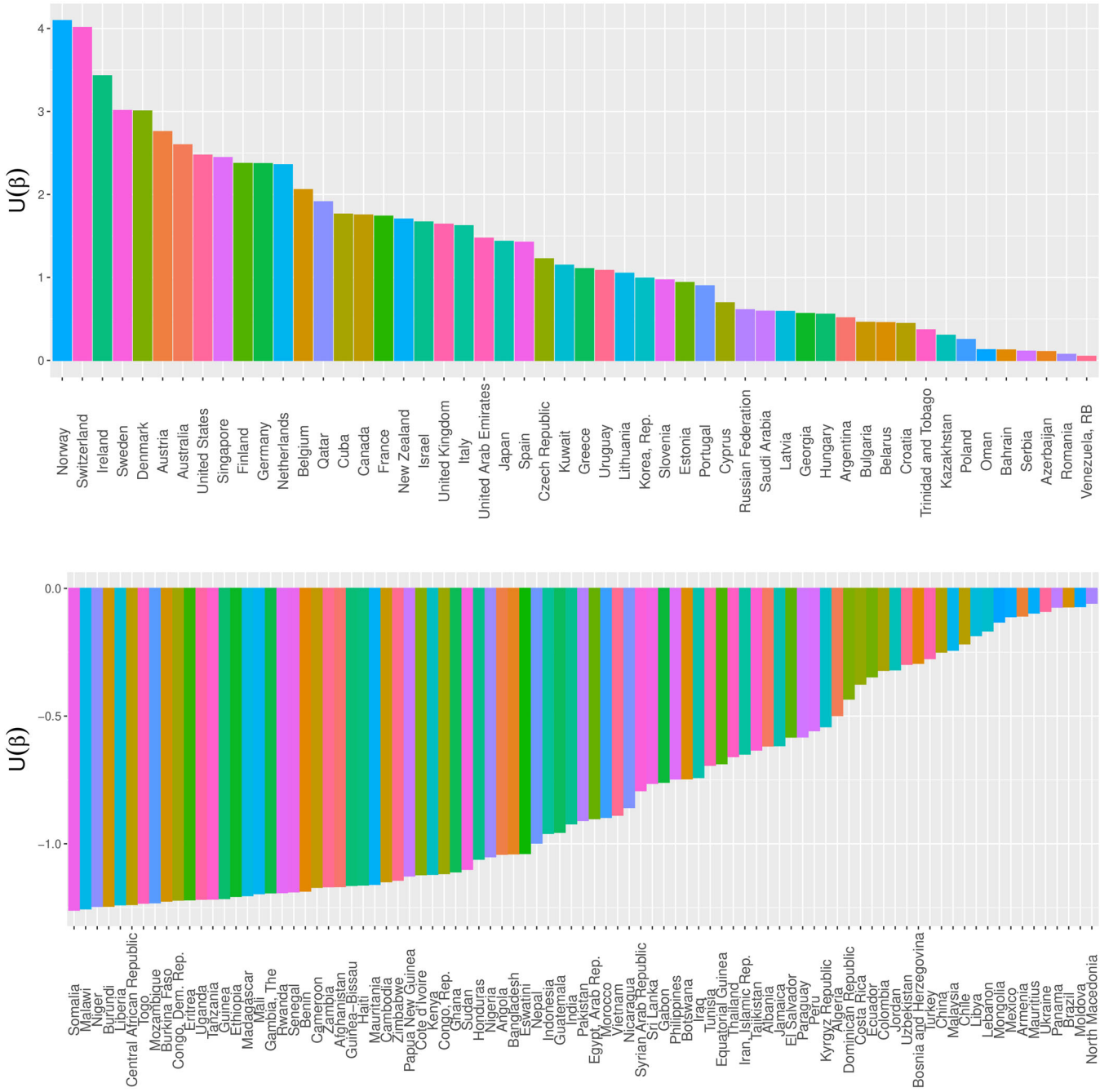


Figure 5. The barplots of the healthcare infrastructure index $U_i(\boldsymbol{\beta})$ related to the COVID-19 mortality rate for 141 countries estimated based on the dynamic interaction DISeF model (25) using the WPLS method. The top figure shows the indices of countries with positive healthcare infrastructure indices. The bottom figure shows the indices of countries with negative healthcare infrastructure indices.

the response $y_i(s_m)$. In the estimation procedure, for each point s_m, we first add a weight to the data using $\boldsymbol{\Delta}$: $\boldsymbol{\Delta}^{-1/2}\boldsymbol{y}(s_m) = \boldsymbol{\Delta}^{-1/2}\left(\boldsymbol{B}_{m1}(\boldsymbol{\beta}),\ldots,\boldsymbol{B}_{mn}(\boldsymbol{\beta})\right)^T\boldsymbol{\theta} + \boldsymbol{\Delta}^{-1/2}\boldsymbol{\eta}(s_m) + \boldsymbol{\Delta}^{-1/2}\boldsymbol{\epsilon}(s_m)$. We then estimate the dynamic interaction DISeF model using the weighted data $\boldsymbol{\Delta}^{-1/2}\boldsymbol{y}(s)$ as the functional response.

Next, we will test whether the bivariate varying coefficient functions in Equation (25) can reduce to univariate functions. The test results are displayed in Table 4. It shows that we can reject all null hypotheses at the 5% significant level, that is, all bivariate varying coefficient functions in Equation (25) vary with time and index.

Fitting the above model by the proposed PLS method and WPLS method, we obtain the estimators of $\beta_l, l = 1,\ldots,4$, the SEs, and the 95% confidence intervals (CI), which are displayed in Table 5. It shows that the SEs estimated by the WPLS method are smaller than those estimated with the PLS method without using the covariance matrix. The 95% confidence intervals for the socio-economic covariates including GDP, the number of physicians, and the number of nurses using the WPLS method do not contain zero, indicating that these three socio-economic covariates are significant. Table 5 also shows that the number of hospital beds is not significant, which may due to the fact that it is relatively easy for the governments to quickly increase the number of beds when the beds are insufficient.

All the index parameters are positive, therefore, the index $U_i(\boldsymbol{\beta}) = \boldsymbol{Z}_i^T\boldsymbol{\beta}$ can be interpreted as the level of healthcare

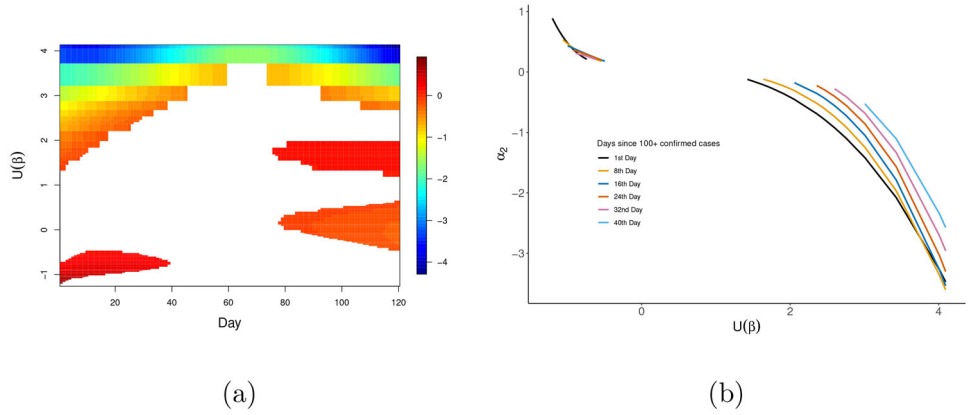


Figure 6. The results of the estimated bivariate varying-coefficient function for population aging $\alpha_2(s_m, U_i(\boldsymbol{\beta}))$ in the dynamic interaction DISeF model (25) using the WPLS method from the COVID-19 data, where $U_i(\boldsymbol{\beta}) = \boldsymbol{Z}_i^T\boldsymbol{\beta}$ is the healthcare infrastructure index related to the COVID-19 morality in the i-th country. (a). The heatmap of the estimated $\alpha_2(s_m, U_i(\boldsymbol{\beta}))$, and it only shows the area where $\alpha_2(s_m, U_i(\boldsymbol{\beta}))$ is significantly nonzero by the pointwise t-test. (b). The plots of the estimated $\alpha_2(s_m, U_i(\boldsymbol{\beta}))$ for the 1st, 8th, 16th, 24th, 32nd, and 40th days since 100+ confirmed cases.

infrastructure related to the COVID-19 morality rate in the ith country. The larger the $U_i(\boldsymbol{\beta})$, the better the healthcare infrastructure. Figure 5 displays the barplots of the healthcare infrastructure index $U_i(\boldsymbol{\beta})$ related to COVID-19 morality rate. The top subfigure in Figure 5 shows that the top five countries with relatively good healthcare infrastructure are all developed countries, namely Norway, Switzerland, Ireland, Sweden, and Denmark. The bottom subfigure in Figure 5 shows that the five countries with the lowest healthcare infrastructure index are Somalia, Malawi, Niger, Burundi, and Liberia. For the bivariate varying-coefficient function of population aging $\alpha_2(s_m, U_i(\boldsymbol{\beta}))$, we conduct a pointwise t-test of whether the function value $\alpha_2(s_m, U_i(\boldsymbol{\beta})) = 0$ according to the result of Theorem 5-(ii)-(b). Based on the above test, we only show the significant area of the heatmap of the estimated varying-coefficient function for population aging in Figure 6(a). We also plot $\alpha_2(s_m, U_i(\boldsymbol{\beta}))$ at fixed chosen days in Figure 6(b).

In the early stage (about from the first to the 40th day), Figure 6 shows that (i) $\alpha_2(s_m, U_i(\boldsymbol{\beta}))$ decreases with the healthcare infrastructure index when s_m is fixed. For countries with the healthcare infrastructure index $-1.260 \le U_i(\boldsymbol{\beta}) \le -0.433$, the population aging has a significant positive impact on the COVID-19 mortality rate; (ii) when the healthcare infrastructure index $-0.433 \le U_i(\boldsymbol{\beta}) \le 1.224$, the population aging has no significant impact on the COVID-19 mortality rate; and (iii) for countries with the healthcare infrastructure index $1.224 \le U_i(\boldsymbol{\beta}) \le 4.094$, the good healthcare infrastructure can suppress the impact of population aging on the COVID-19 mortality rate; the greater the value of $U_i(\boldsymbol{\beta})$, the greater the suppression impact.

During the early and middle stages, for countries with the healthcare infrastructure index $1.224 \le U_i(\boldsymbol{\beta}) \le 4.094$, the bigger $U_i(\boldsymbol{\beta})$ is, the later the significant suppression impact ends. When the healthcare infrastructure index belongs to the interval $[4.011, 4.094]$, such as Norway with $U(\boldsymbol{\beta}) = 4.094$, the suppression impact brought by the high economic and medical level can last at least 120 days after 100+ confirmed cases, although in the intermediate stage, the impact weakened. In Sweden, a developed Nordic country with $U(\boldsymbol{\beta}) = 3.011 \in [2.756, 4.011]$, the suppression impact vanishes during the middle stage. For countries with the healthcare infrastructure index $1.911 \le U_i(\boldsymbol{\beta}) \le 2.756$, for example, the United States ($U(\boldsymbol{\beta}) = 2.473$) and Singapore ($U(\boldsymbol{\beta}) = 2.443$), with the development of the epidemic, the available medical care becomes saturated, the suppression effect disappears in the mid-late stage and the population aging starts to have no significant impact on the COVID-19 mortality rate. Canada has the middle healthcare infrastructure index $U(\boldsymbol{\beta}) = 1.752 \in [1.224, 1.911]$. The function $\alpha_2(s_m, U_i(\boldsymbol{\beta}))$ is significantly negative only from March 11 to March 21. During the end of March to the end of May, the COVID-19 mortality rate increases rapidly, and suppression impact of good medical standards vanishes. Even starting June, due to the development of the epidemic, the mortality rate increases as population aging increases. During the 120 days, population aging has no significant effect on the mortality rate for those countries with $0.610 < U_i(\boldsymbol{\beta}) < 1.224$. For those countries with $0.610 < U_i(\boldsymbol{\beta}) < -0.433$ (e.g., $U(\boldsymbol{\beta}) = -0.072$ for Brazil), the healthcare infrastructure does not suppress the impact of population aging on the COVID-19 mortality rate until the later stage. In general, Figure 6 (a) shows that after about the 80th day since 100+ confirmed cases, the impact of population aging on the mortality rate is more complicated, which may be due to the policies, climate changes in various country and the mutation of coronavirus.

7. Concluding Remarks

We propose the DISeF model to characterize the dynamic interaction between covariates and their effects on the functional response. The proposed model includes many popular models as special cases. In this article, we develop a three-step estimation procedure to estimate the vector of index parameters, the bivariate varying-coefficient functions, the covariance function of random effects. And we also extend our estimation procedure to the DISeF model with multi-index. We also establish the asymptotic results when the sample size, n, and the number of observations per subject, M, tend to infinity but M increases relatively more slowly than n. Besides, we propose a hypothesis testing method to test whether the bivariate varying-coefficient

function can reduce to a univariate function. Our proposed DISeF model is also used to analyzing the COVID-19 data and the ADNI data. In both applications, the hypothesis testing shows that the bivariate varying-coefficient functions significantly vary with the index and the time/location. Therefore, the proposed DISeF model is more appropriate than some existing models. There are some challenging problems to be addressed in future research. First, we will extend this model with high dimensional covariates $\boldsymbol{Z}$. Second, we will consider a unified inference for longitudinal/functional data with sparse or dense observations for each subject. There are other methods to address the spatial dependence among different countries, such as the factor model or the copula model . We will consider these models in our future work.

Supplementary Materials

The supplementary document provides the estimation procedure under the null hypothesis of Section 4, the theoretical proofs of the asymptotic results, additional simulation studies and the analysis of ADNI data. The computing codes of the simulation studies and real data analysis are also provided.

Acknowledgments

The authors thank to the editor, the associate editor and two anonymous referees for many insightful comments. These comments are very helpful for us to improve our work.

Funding

Dr. Liu gratelfully acknowledges the financial support from China Scholarship Council. Dr. You's research is supported by grants from the National Natural Science Foundation of China (NSFC) (No. 11971291). Dr. Cao's research is supported by the Discovery grant (RGPIN-2018-06008) from the Natural Sciences and Engineering Research Council of Canada (NSERC).

ORCID

Orcid Hua Liu http://orcid.org/0000-0001-8994-9703
Jiguo Cao http://orcid.org/0000-0001-7417-6330

Appendix A: Conditions

The needed technical conditions are as follows:

(C1) (i) The density function $f_{U(\beta)}(\cdot)$ of random variable $U(\beta) = \boldsymbol{Z}^T\boldsymbol{\beta}$ is bounded away from 0 on S_ω and $f_{U(\beta)}(\cdot) \in \mathcal{C}^{0,1}(S_\omega)$ for $\boldsymbol{\beta}$ in the neighborhood of $\boldsymbol{\beta}_0$, where $S_\omega = \{\boldsymbol{Z}^T\boldsymbol{\beta}, \boldsymbol{Z} \in S\}$ and S is a compact support support set of $\boldsymbol{Z}$. Without loss of generality, we assume $S_\omega = [0,1]$. (ii) The density function $f_s(\cdot)$ of random variable s_m is bounded away from 0 on $[0,1]$ and $f_s(\cdot) \in \mathcal{C}^{0,1}[0,1]$; (iii) The joint density function $f(s,u)$ of random variable $(s_m, U_i(\beta))$ and its partial derivatives up to second order are continuous and $f(s,u)$ is bounded away from 0.

(C2) (i) For every $1 \le d \le p$, the nonparametric function $\alpha_d(\cdot,\cdot) \in \mathcal{C}^{(r)}[0,1]^2$; (ii) For all $1 \le i \le n$, the individual random function $\eta_i(\cdot) \in \mathcal{C}^{(r_1)}[0,1]$; (iii) The variance of error term $\sigma^2(\cdot) \in \mathcal{C}^{(r_2)}[0,1]$, where integers $r, r_1, r_2 \ge 2$ and the cubic spline order 4 satisfies $r, r_1, r_2 \le 4$; (iv) Besides, for every $1 \le d \le p$, $1 \le k \le q$, $g_{d,k}(\cdot,\cdot) \in \mathcal{C}^{(1)}[0,1]^2$ and $g^*_{d,k}(\cdot,\cdot) \in \mathcal{C}^{(1)}[0,1]^2$.

(C3) (i) The variance matrix of $\epsilon(s)$ at different time/spatial points is measurable and bounded. (ii) The matrix of the covariance function $R(s,t)$ at different time/spatial points is bounded. (iii) For the spatial covariance matrix $\boldsymbol{\Phi}$, its eigenvalues are bounded away from zero and infinity.

(C4) There exist constants $0 \le c_Q \le C_Q \le \infty$, such that $c_Q \le Q(z) = \mathbb{E}(\boldsymbol{X}\boldsymbol{X}^T|\boldsymbol{Z} = z) \le C_Q$ for all $z \in S$.

(C5) The eigenvalues of $V(\boldsymbol{\beta}_0) = \mathbb{E}(\boldsymbol{B}_i(\boldsymbol{\beta}_0)^T\boldsymbol{B}_i(\boldsymbol{\beta}_0))$ are bounded away from zero and infinity, which ensures $V(\boldsymbol{\beta}_0)$ is invertible.

(C6) The functional classes $\{\eta(s) : s \in [0,1]\}$, $\{\eta(s)\eta(t) : (s,t) \in [0,1]^2\}$ are Donsker.

(C7) All components of $R(s,t)$ have continuous second-order partial derivatives with respect to $(s,t) \in [0,1]^2$.

(C8) The relationship between M and n is $n^{1/5} << M \le Cn^{1/4}$ as both M and n converge to ∞.

Remark 5. Conditions (C1)-(i) and (C1)-(ii) are the same as Condition (C1) in Ma and Song (2015). (C1)-(iii) is the regularity condition about the joint density function of s_m and $U(\boldsymbol{\beta})$. Conditions (C2) and (C7) assume the degree of the smoothness of the varying coefficient functions, the covariance function $R(s,t)$, the covariance function of error term, the functions $\boldsymbol{g}_k$ defined in Equation (14) and the functions $\boldsymbol{g}^*_k$ defined in (15). Conditions (C3) and (C4) place some restrictions on the moments of covariates $\boldsymbol{X}_i$, error term and individual term, respectively. Condition (C6) is the same as Assumption 2 in Li, Huang, and Zhu (2017). Condition (C8) is used to prove the asymptotic properties of the estimators obtained from the WPLS method.

References

Ainsworth, L. M., Routledge, R., and Cao, J. (2011), "Functional Data Analysis in Ecosystem Research: The Decline of Oweekeno Lake Sockeye Salmon and Wannock River Flow," *Journal of Agricultural, Biological, and Environmental Statistics*, 16, 282–300. [1]

Bachoc, F. (2018), "Asymptotic Analysis of Covariance Parameter Estimation for Gaussian Processes in the Misspecified Case," *Bernoulli*, 24, 1531–1575. [10]

Cai, X., Xue, L., and Cao, J. (2021), "Variable Selection for Multiple Function-on-Function Linear Regression," *Statistica Sinica*. [1]

Cao, Y., Ayako, H., and Scott, M. (2020), "Covid-19 Case-Fatality Rate and Demographic and Socioeconomic Influencers: Worldwide Spatial Regression Analysis Based on Country-Level Data," *BMJ Open*, 10, e043560. [10]

Carroll, C., Bhattacharjee, S., Chen, Y., Dubey, P., Fan, J., Gajardo, A., Zhou, X., Mueller, H.-G., and Wang, J.-L. (2020), "Time Dynamics of Covid-19," *Scientific Reports*, 10, Article number 21040. [2]

de Boor, C. (1978), *A Practical Guide to Splines*, New York: Springer. [3,4]

Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis*, New York: Springer. [1]

Ferré, L., and Yao, A.-F. (2003), "Functional Sliced Inverse Regression Analysis," *Statistics*, 37, 475–488. [1]

Guan, T., Lin, Z., and Cao, J. (2020), "Estimating Truncated Functional Linear Models With a Nested Group Bridge Approach," *Journal of Computational and Graphical Statistics*, 29, 620–628. [1]

Hall, P., and Hosseini-Nasab, M. (2006), "On Properties of Functional Principal Components Analysis," *Journal of the Royal Statistical Society*, Series B, 68, 109–126. [5]

Hastie, T., and Tibshirani, R. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society*, Series B, 55, 757–779. [1]

Jiang, C. R., and Wang, J. L. (2011), "Functional Single Index Models for Longitudinal Data," *The Annals of Statistics*, 39, 362–388. [1,3]

Jiang, F., Baek, S., Cao, J., and Ma, Y. (2020), "A Functional Single Index Model," *Statistica Sinica*, 30, 303–324. [1]

Jiang, J. (2010), *Large Sample Techniques for Statistics*, New York: Springer. [7]

Lee, S., Liao, Y., Seo, M. H., and Shin, Y. (2021), "Sparse HP Filter: Finding Kinks in the Covid-19 Contact Rate," *Journal of Econometrics*, 220, 158–180. [2]

Li, Y., and Hsing, T. (2010), "Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data," *The Annals of Statistics*, 38, 3321–3351. [5]

Li, J., Huang, C., and Zhu, H. (2017), "A Functional Varying-Coefficient Single-Index Model for Functional Response Data," *Journal of the American Statistical Association*, 112, 1169–1181. [1,3,13]

Lin, W., and Kulasekera, K. (2007), "Identifiability of Single-Index Models and Additive-Index Models," *Biometrika*, 94, 496–501. [4]

Lin, Z., Cao, J., Wang, L., and Wang, H. (2017), "Locally Sparse Estimator for Functional Linear Regression Models," *Journal of Computational and Graphical Statistics*, 26, 306–318. [1]

Liu, L., Moon, H. R., and Schorfheide, F. (2021), "Panel Forecasts of Country-Level Covid-19 Infections," *Journal of Econometrics*, 220, 2–22. [2]

Liu, B., Wang, L., and Cao, J. (2017), "Estimating Functional Linear Mixed-Effects Regression Models," *Computational Statistics & Data Analysis*, 106, 153–164. [1]

Luo, X., Zhu, L., and Zhu, H. (2016), "Single-Index Varying Coefficient Model for Functional Responses," *Biometrics*, 72, 1275–1284. [1]

Ma, S., and Song, P. X. K. (2015), "Varying Index Coefficient Models," *Journal of the American Statistical Association*, 110, 341–356. [6,13]

Ma, S., Song, Q., and Wang, L. (2013), "Variable Selection and Estimation in Marginal Partially Linear Additive Models for Longitudinal Data," *Bernoulli*, 19, 252–274. [4]

Minasny, B., and McBratney, A. B. (2005), "The matérn Function as a General Model for Soil Variograms," *Geoderma*, 128, 192–207. [10]

Morris, J. S. (2015), "Functional Regression," *Annual Review of Statistics and Its Application*, 2, 321–359. [1]

Ramsay, J. O., and Silverman, B. W. (2005). *Functional Data Analysis* (2nd ed.), New York: Springer. [1]

Vincenty, T. (1975), "Direct and Inverse Solutions of Geodesics on the Ellipsoid With Application of Nested Equations," *Survey Review*, 23, 88–93. [10]

Wang, L., and Yang, L. (2009), "Spline Estimation of Single-Index Models," *Statistica Sinica*, 19, 765–783. [4]

Xia, Y., Tong, H., Li, W. K., and Zhu, L. (2002), "An Adaptive Estimation of Dimension Reduction Space," *Journal of the Royal Statistical Society*, Series B, 64, 363–410. [4]

Zhu, H., Li, R., and Kong, L. (2012), "Multivariate Varying Coefficient Model for Functional Responses," *The Annals of Statistics*, 40, 2634–2666. [1,3]