MULTIVARIATE SPLINE ESTIMATION AND INFERENCE FOR IMAGE-ON-SCALAR REGRESSION

Shan Yu[1], Guannan Wang[2], Li Wang[1] and Lijian Yang[3]

[1]*Iowa State University, Ames, IA 50011, USA.*

[2]*College of William & Mary, Williamsburg, VA 23187, USA.*

[3]*Tsinghua University, Beijing 100084, China.*

Abstract: Motivated by recent work of analyzing data in the biomedical imaging studies, we consider a class of image-on-scalar regression models for imaging responses and scalar predictors. We propose to use flexible multivariate splines over triangulations to handle the irregular domain of the objects of interest on the images and other characteristics of images. The proposed estimators of the coefficient functions are proved to be root-n consistent and asymptotically normal under some regularity conditions. We also provide a consistent and computationally efficient estimator of the covariance function. Asymptotic pointwise confidence intervals (PCIs) and data-driven simultaneous confidence corridors (SCCs) for the coefficient functions are constructed. Our method can simultaneously estimate and make inferences of the coefficient functions while incorporating the spatial heterogeneity and spatial correlation. A highly efficient and scalable estimation algorithm is developed. Monte Carlo simulation studies are conducted to examine the finite-sample performance of the proposed method. The proposed method is applied to the spatially normalized Positron Emission Tomography (PET) data of Alzheimer's Disease Neuroimaging Initiative (ADNI).

Key words and phrases: Multivariate splines; Coefficient maps; Confidence corridors; Image Analysis; Triangulation.

1. Introduction

In recent years, there has been explosive growth in the amount of imaging data collected from medical and public health studies, such as functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET) imaging, Computed Tomography (CT) and ultrasonic imaging. A large fraction of these data can be characterized as functional data. Comparing with traditional one-dimensional (1D) functional data, these imaging data are increasingly complex, high-dimensional, and structured, which poses exceptional challenges to traditional statistical methods.

In this paper, we propose a unifying approach to characterize the varying association between imaging responses with a set of explanatory variables. To investigate such kind of association, there are mainly three types of statistical methods. The first category is the univariate approaches and pixel/voxel-based methods (Worsley et al., 2004; Stein et al., 2010; Hibar et al., 2015), which take each pixel/voxel as a basic analytic unit. Since all the pixels/voxels are treated to be independent, a major drawback of the pixel/voxel-wise methods is that the correlation among the pixels/voxels is totally ignored. The second category is the tensor regression. This approach considers an image as a multi-dimensional array (Zhou et al., 2013; Li and Zhang, 2017), which is turned into a vector to perform regression. However, naively doing so yields an ultra-high dimensionality, and it requires a novel dimension reduction technique and highly scalable algorithms (Li and Zhang, 2017). The third category is the functional data analysis (FDA) approach, in which an image is viewed as the realizations of a function

defined on a given domain (Zhu et al., 2012, 2014; Reiss et al., 2017). Using the FDA, we are able to combine the information both across and within functions.

We take the FDA approach in this paper. Functional linear models (FLMs) are widely used to model the regression relationship between the response and some set of predictors from multiple subjects. In the literature (Ramsay and Silverman, 2005; Müller, 2005; Morris, 2015; Wang et al., 2016), FLMs are often categorized by whether the outcome, the predictor, or both are functional into three categories: (i) functional predictor regression (scalar-on-function) (Cardot et al., 1999, 2003; Hall and Horowitz, 2007); (ii) functional response regression (function-on-scalar) (Morris and Carroll, 2006; Reiss et al., 2010; Staicu et al., 2010; Zhu et al., 2014; Zhang and Wang, 2015; Chen et al., 2017); and (iii) function-on-function regression (Ramsay and Dalzell, 1991; Yao et al., 2005; Sentürk and Müller, 2010; Wu and Müller, 2011).

Motivated by the structure of brain imaging data, we propose a novel image-on-scalar regression model with spatially varying coefficients, which is able to capture the varying association between imaging phenotypes with a set of explanatory variables. Figure 1 shows a schematic diagram of the proposed modeling approach. To be more specific, let Ω be a two-dimensional bounded domain, and $\boldsymbol{z} = (z_1, z_2)$ be the location point on Ω. For the ith subject, $i = 1, \ldots, n$, let $Y_i(\boldsymbol{z})$ be the imaging measurement at location $\boldsymbol{z} \in \Omega$, and let $X_{i\ell}$, $\ell = 0, 1, \ldots, p$, with $X_{i0} \equiv 1$, be scalar predictors, for example, clinic variables (such as age and sex) and genetic factors. The spatially varying coefficient regression characterizes the association between imaging measures

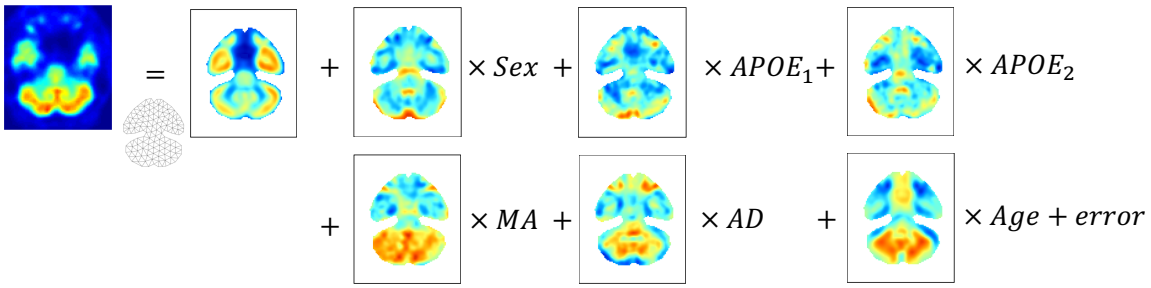

Figure 1: A schematic diagram of proposed modeling approach.

and covariates and is given by the following model:

$$Y_i(\boldsymbol{z}) = \widetilde{\mathbf{X}}_i^\top \boldsymbol{\beta}^o(\boldsymbol{z}) + \eta_i(\boldsymbol{z}) + \sigma(\boldsymbol{z})\varepsilon_i(\boldsymbol{z}),\ i = 1, \ldots, n,\ \boldsymbol{z} \in \Omega,$$

where $\widetilde{\mathbf{X}}_i = (X_{i0}, X_{i1}, \ldots, X_{ip})^\top$, $\boldsymbol{\beta}^o = (\beta_0^o, \beta_1^o, \ldots, \beta_p^o)^\top$ is a vector of some unknown bivariate functions, $\eta_i(\boldsymbol{z})$ characterizes the individual image variations, $\varepsilon_i(\boldsymbol{z})$ represents additional measurement errors, and $\sigma(\boldsymbol{z})$ is a positive deterministic function. In the following, we assume that $\eta_i(\boldsymbol{z})$ and $\varepsilon_i(\boldsymbol{z})$ are mutually independent. Moreover, we assume that $\eta_i(\boldsymbol{z})$, $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) copies of an L_2 stochastic process with mean zero and covariance function $G_\eta(\boldsymbol{z}, \boldsymbol{z}') = \text{cov}\{\eta_i(\boldsymbol{z}), \eta_i(\boldsymbol{z}')\}$, and $\varepsilon_i(\boldsymbol{z})$, $i = 1, \ldots, n$, are i.i.d. copies of a stochastic process with zero mean and covariance function $G_\varepsilon(\boldsymbol{z}, \boldsymbol{z}') = \text{cov}\{\varepsilon_i(\boldsymbol{z}), \varepsilon_i(\boldsymbol{z}')\} = I(\boldsymbol{z} = \boldsymbol{z}')$.

For 1D function-on-scalar regression, Chapter 13 of Ramsay and Silverman (2005) provides a common model-fitting strategy, in which the coefficient functions are expanded using some sets of basis functions and basis coefficients are estimated using ordinary least squares. However, it is not trivial to extend it to image-on-scalar regression, particularly with biomedical imaging responses. For biomedical images, the

objects (e.g. organs) on the images are usually irregularly shaped. For instance, breast tumors often have an irregular shape. Another example is the brain image, as shown in Figure 1, especially slices from the bottom and the top of the brain. Even though some images seem to be of a rectangle shape, the true signal only comes from the domain of an object, and the image only contains noises outside the boundary of the object. Many smoothing methods, for example, tensor product smoothing (Reiss et al., 2017; Chen et al., 2017), kernel smoothing (Zhu et al., 2014) and wavelet smoothing (Morris and Carroll, 2006), tend to provide poor estimation over difficult regions by smoothing inappropriately across boundary features, which refers to the "leakage" problem in the smoothing literature; see the discussions in Ramsay (2002) and Sangalli et al. (2013). Next, due to technical reasons, imaging data often have different visual qualities. General characteristics of medical images are determined and limited by the technology for each specific modality. There is a great interest in the development of a flexible method of various smoothness to adaptively smooth biomedical imaging data.

In this paper, we tackle the above challenges by using the bivariate splines on triangulations (Lai and Wang, 2013) to effectively model the spatially nonstationary relationship and preserve important features (shape, smoothness) of imaging data. Triangulation can effectively represent any two-dimensional (2D) geometric domain as any polygon can be decomposed into triangles. We study the asymptotic properties of the bivariate spline estimators of the coefficient functions, and specifically, we show that our spline estimators are root-n consistent and asymptotically normal. The asymptotic

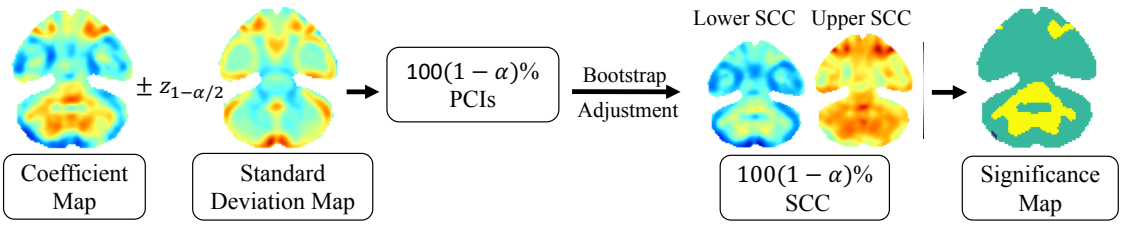

Figure 2: A schematic diagram of proposed inferential approach.

results are used as a guideline to construct pointwise confidence intervals (PCIs) and simultaneous confidence corridors (SCCs; also referred to as "simultaneous confidence band/regions") for the true coefficient functions. See Figure 2 for an illustration of our proposed inferential approach. Compared with the tensor regression (Li and Zhang, 2017) and the three-stage estimation (Zhu et al., 2014), our method is statistically more efficient, as it is able to accommodate complex domains of arbitrary shape and adjust the individual smoothing needs desired by different coefficient functions through the use of multiple smoothing parameters. In addition, our method does not rely on estimating the spatial similarity and adaptive weights repeatedly as in Zhu et al. (2014), thus it is much simpler.

We organize our paper as follows. Section 2 describes spline estimators for the coefficient functions, and establish their asymptotic properties. Section 3 describes a bootstrap method for constructing the SCC and how to estimate the unknown variance functions involved in the SCC. Section 4 contains the details of the implementation of the proposed estimation and inference. Section 5 reports our findings from two simulation studies. In Section 6, we illustrate the proposed method through the PET

data of Alzheimer's Disease Neuroimaging Initiative (ADNI). Section 7 contains some concluding remarks. The technical proofs of the theoretical results and additional numerical results are deferred in the Supplementary Material.

2. Models and Estimation Method

2.1 Image-on-scalar regression model

Let $\boldsymbol{z}_j \in \Omega$ be the center point of the jth pixel in the domain Ω, and let Y_{ij} be the imaging response of subject i at location j, and the actual data set consists of $\{(Y_{ij}, \widetilde{\mathbf{X}}_i, \boldsymbol{z}_j), i = 1, \ldots, n, j = 1, \ldots, N\}$, which can be modeled via the following:

$$Y_{ij} = \sum_{\ell=0}^{p} X_{i\ell}\beta_\ell^o(\boldsymbol{z}_j) + \eta_i(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_{ij}. \tag{2.1}$$

Denote the eigenvalues and eigenfunctions of the covariance operator $G_\eta(\boldsymbol{z}, \boldsymbol{z}')$ as $\{\lambda_k\}_{k=1}^{\infty}$, $\{\psi_k(\boldsymbol{z})\}_{k=1}^{\infty}$, in which $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, $\sum_{k=1}^{\infty}\lambda_k < \infty$, and $\{\psi_k\}_{k=1}^{\infty}$ form an orthonormal basis of $L^2(\Omega)$. It follows from spectral theory that $G_\eta(\boldsymbol{z}, \boldsymbol{z}') = \sum_{k=1}^{\infty}\lambda_k\psi_k(\boldsymbol{z})\psi_k(\boldsymbol{z}')$. The ith trajectory $\{\eta_i(\boldsymbol{z}), \boldsymbol{z} \in \Omega\}$ allows the Karhunen-Loéve L^2 representation (Li and Hsing, 2010; Sang and Huang, 2012): $\eta_i(\boldsymbol{z}) = \sum_{k=1}^{\infty}\lambda_k^{1/2}\xi_{ik}\psi_k(\boldsymbol{z})$, $\lambda_k^{1/2}\xi_{ik} = \int_{\boldsymbol{z}\in\Omega}\eta_i(\boldsymbol{z})\psi_k(\boldsymbol{z})d\boldsymbol{z}$, where the random coefficients ξ_{ik}'s are uncorrelated random variables with mean 0 and $E(\xi_{ik}\xi_{ik'}) = I(k = k')$, referred to as the kth functional principal component score (FPCA) of the ith subject. Thus, the response measurements in (2.1) can be represented as follows:

$$Y_{ij} = \sum_{\ell=0}^{p}\beta_\ell^o(\boldsymbol{z}_j)X_{i\ell} + \sum_{k=1}^{\infty}\lambda_k^{1/2}\xi_{ik}\psi_k(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_{ij}. \tag{2.2}$$

2.2 Spline approximation over triangulations and penalized regression

Note that the object of interest on many biomedical images are often distributed over an irregular domain Ω, and triangulation is an effective strategy to handle such type of data. For example, the spatial smoothing problem over difficult regions in Ramsay (2002) and Sangalli et al. (2013) was solved by the finite element method (FEM) on triangulations, which is mainly developed to solve partial differential equations. In this paper, we approximate each coefficient function in (2.2) by bivariate splines over triangulations (Lai and Schumaker, 2007). The idea is to approximate each function $\beta_\ell(\cdot)$ by Bernstein basis polynomials that are piecewise polynomial functions over a 2D triangulated domain. Compared with FEM, our proposed approach is appealing in the sense that we can use spline functions of more flexible degrees and various smoothness so that we are able to better approximate the coefficient functions. In this section, we briefly introduce the techniques of triangulations and describe the bivariate penalized spline smoothing (BPST) method for approximating the spatial data.

Triangulation is an effective tool to deal with data distributed over difficult regions with complex boundaries and/or interior holes. In the following, we use T to denote a triangle which is a convex hull of three points not located in one line. A collection $\triangle = \{T_1, \ldots, T_H\}$ of H triangles is called a triangulation of $\Omega = \cup_{h=1}^{H} T_h$ provided that any nonempty intersection between a pair of triangles in $\triangle$ is either a shared vertex or a shared edge. Given a triangle $T \in \triangle$, let $|T|$ be its longest edge length, and ϱ_T be the radius of the largest disk inscribed in T. Define the shape parameter of T as

the ratio $\pi_T = |T|/\varrho_T$. When π_T is small, the triangles are relatively uniform in the sense that all angles of triangles in $\triangle$ are relatively the same. Denote the size of $\triangle$ by $|\triangle| = \max\{|T|, T \in \triangle\}$, i.e., the length of the longest edge of $\triangle$. For an integer $r \geq 0$, let $\mathcal{C}^r(\Omega)$ be the collection of all r-th continuously differentiable functions over Ω. Given $\triangle$, let $\mathcal{S}_d^r(\triangle) = \{s \in \mathcal{C}^r(\Omega) : s|_T \in \mathbb{P}_d(T), T \in \triangle\}$ be a spline space of degree d and smoothness r over $\triangle$, where $s|_T$ is the polynomial piece of spline s restricted on triangle T, and $\mathbb{P}_d$ is the space of all polynomials of degree less than or equal to d. Note that the major difference between the FEM and the BPST is the flexibility of smoothness, r, and degree of polynomials, d. Specifically, FEM in Sangalli et al. (2013) requires that $r = 0$ and $d = 1$ or 2, while BPST allows the smoothness $r \geq 0$ and various degrees of polynomials.

We use Bernstein basis polynomials to represent the bivariate splines. For any $\ell = 0, 1, \ldots, p$, denote by $\triangle_\ell$ the triangulation of the ℓth component. Define

$$\mathcal{G}^{(p+1)} \equiv \mathcal{G}^{(p+1)}(\triangle_0 \times \cdots \times \triangle_p) = \left\{\boldsymbol{g} = (g_0, \ldots, g_p)^\top, g_\ell \in \mathcal{S}_d^r(\triangle_\ell), \ell = 0, \ldots, p\right\},$$

and let $\{B_{\ell m}\}_{m \in \mathcal{M}_\ell}$ be the set of degree-d bivariate Bernstein basis polynomials for $\mathcal{S}_d^r(\triangle_\ell)$, where $\mathcal{M}_\ell$ stands for an index set of Bernstein basis polynomials. Denote by $\mathbf{B}_\ell$ the evaluation matrix of the Bernstein basis polynomials for the ℓth component, and the jth row of $\mathbf{B}_\ell$ is given by $\mathbf{B}_\ell^\top(\boldsymbol{z}_j) = \{B_{\ell m}(\boldsymbol{z}_j), m \in \mathcal{M}_\ell\}$. For each $\beta_\ell(\cdot)$, we approximate it by $\beta_\ell(\boldsymbol{z}_j) \approx \mathbf{B}_\ell^\top(\boldsymbol{z}_j)\boldsymbol{\gamma}_\ell$, $\ell = 0, 1, \ldots, p$, where $\boldsymbol{\gamma}_\ell^\top = (\gamma_{\ell m}, m \in \mathcal{M}_\ell)$ is the spline coefficient vector.

Penalized spline smoothing has gained a lot of popularity over the last two decades;

see Hall and Opsomer (2005); Claeskens et al. (2009); Schwarz and Krivobokova (2016). To define the penalized spline method, for any direction z_q, $q = 1, 2$, let $\nabla^v_{z_q} s(\boldsymbol{z})$ denote the vth order derivative in the direction z_q at the point $\boldsymbol{z}$. We consider the following penalized least squares problem:

$$\min_{(\beta_0,\ldots,\beta_p)^\top \in \mathcal{G}^{(p+1)}} \sum_{i=1}^{n}\sum_{j=1}^{N}\left\{Y_{ij} - \sum_{\ell=0}^{p} X_{i\ell}\beta_\ell(\boldsymbol{z}_j)\right\}^2 + \sum_{\ell=0}^{p}\rho_{n,\ell}\mathcal{E}(\beta_\ell),$$

where $\mathcal{E}(s) = \sum_{T\in\triangle}\int_T \sum_{i+j=2}\binom{2}{i}(\nabla^i_{z_1}\nabla^j_{z_2}s)^2 dz_1 dz_2$ is the roughness penalty, and $\rho_{n,\ell}$ is the penalty parameter for the ℓth function.

To satisfy the smoothness condition of the splines, we need to impose some linear constraints on the spline coefficients $\boldsymbol{\gamma}_\ell$: $\mathbf{H}_\ell\boldsymbol{\gamma}_\ell = \mathbf{0}$, for $\ell = 0, 1, \ldots, p$. Thus, we have to minimize the following constrained least squares:

$$\sum_{i=1}^{n}\sum_{j=1}^{N}\left\{Y_{ij} - \sum_{\ell=0}^{p} X_{i\ell}\mathbf{B}_\ell^\top(\boldsymbol{z}_j)\boldsymbol{\gamma}_\ell\right\}^2 + \sum_{\ell=0}^{p}\rho_{n,\ell}\boldsymbol{\gamma}_\ell^\top\mathbf{P}_\ell\boldsymbol{\gamma}_\ell, \text{ subject to } \mathbf{H}_\ell\boldsymbol{\gamma}_\ell = 0,$$

where $\mathbf{P}_\ell$ is the block diagonal penalty matrix satisfying that $\boldsymbol{\gamma}_\ell^\top\mathbf{P}_\ell\boldsymbol{\gamma}_\ell = \mathcal{E}(\mathbf{B}_\ell^\top\boldsymbol{\gamma}_\ell)$.

We first remove the constraint via QR decomposition of the transpose of the constraint matrix $\mathbf{H}_\ell$. Applying QR decomposition on $\mathbf{H}_\ell^\top$, one has $\mathbf{H}_\ell^\top = \mathbf{Q}_\ell\mathbf{R}_\ell = (\mathbf{Q}_{\ell,1}\ \mathbf{Q}_{\ell,2})\binom{\mathbf{R}_{\ell,1}}{\mathbf{R}_{\ell,2}}$, where $\mathbf{Q}_\ell$ is an orthogonal matrix and $\mathbf{R}_\ell$ is an upper triangular matrix, the submatrix $\mathbf{Q}_{\ell,1}$ is the first r columns of $\mathbf{Q}_\ell$, where r is the rank of matrix $\mathbf{H}_\ell$, and $\mathbf{R}_{\ell,2}$ is a matrix of zeros. We reparametrize using $\boldsymbol{\gamma}_\ell = \mathbf{Q}_{\ell,2}\boldsymbol{\theta}_\ell$ for some $\boldsymbol{\theta}_\ell$, then it is guaranteed that $\mathbf{H}_\ell\boldsymbol{\gamma}_\ell = \mathbf{0}$. Then the minimization problem is now converted to a conventional penalized regression problem without restrictions:

$$\sum_{i=1}^{n}\sum_{j=1}^{N}\left\{Y_{ij} - \sum_{\ell=0}^{p} X_{i\ell}\mathbf{B}_\ell^\top(\boldsymbol{z}_j)\mathbf{Q}_{\ell,2}\boldsymbol{\theta}_\ell\right\}^2 + \sum_{\ell=0}^{p}\rho_{n,\ell}\boldsymbol{\theta}_\ell^\top\mathbf{D}_\ell\boldsymbol{\theta}_\ell, \tag{2.3}$$

where $\mathbf{D}_\ell = \mathbf{Q}_{\ell,2}^\top \mathbf{P}_\ell \mathbf{Q}_{\ell,2}$.

Let $\widetilde{\mathbf{Y}}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iN})^\top$, $\mathbf{B}_\ell(\boldsymbol{z}) = \{B_{\ell m}(\boldsymbol{z}), m \in \mathcal{M}_\ell\}^\top$, $\mathbb{Y} = (\widetilde{\mathbf{Y}}_1^\top, \ldots, \widetilde{\mathbf{Y}}_n^\top)^\top$, and $\mathbb{U} = (\mathbf{U}_{11}, \mathbf{U}_{12}, \ldots, \mathbf{U}_{nN})^\top$, where

$$\mathbf{U}_{ij} = \{X_{i0}\mathbf{B}_0(\boldsymbol{z}_j)^\top \mathbf{Q}_{0,2}, X_{i1}\mathbf{B}_1(\boldsymbol{z}_j)^\top \mathbf{Q}_{1,2}, \cdots, X_{ip}\mathbf{B}_p(\boldsymbol{z}_j)^\top \mathbf{Q}_{p,2}\}^\top. \tag{2.4}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_0^\top, \boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_p^\top)^\top$ and $\mathbb{D}(\rho_{n,0}, \ldots, \rho_{n,p}) = \mathrm{diag}\{\rho_{n,0}\mathbf{D}_0, \ldots, \rho_{n,p}\mathbf{D}_p\}$. Minimizing (2.3) is then equivalent to minimizing $\|\mathbb{Y} - \mathbb{U}\boldsymbol{\theta}\|^2 + \boldsymbol{\theta}^\top \mathbb{D}(\rho_{n,0}, \ldots, \rho_{n,p})\boldsymbol{\theta}$, hence,

$$\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_0^\top, \widehat{\boldsymbol{\theta}}_1^\top, \ldots, \widehat{\boldsymbol{\theta}}_p^\top)^\top = \{\mathbb{U}^\top\mathbb{U} + \mathbb{D}(\rho_{n,0}, \ldots, \rho_{n,p})\}^{-1}\mathbb{U}^\top\mathbb{Y}.$$

Thus, the estimators of $\boldsymbol{\gamma}_\ell$ and $\beta_\ell(\cdot)$ are:

$$\widehat{\boldsymbol{\gamma}}_\ell = \mathbf{Q}_{\ell,2}\widehat{\boldsymbol{\theta}}_\ell, \quad \widehat{\beta}_\ell(\boldsymbol{z}) = \mathbf{B}_\ell(\boldsymbol{z})^\top \widehat{\boldsymbol{\gamma}}_\ell. \tag{2.5}$$

2.3 Asymptotic Properties of the BPST Estimators

This section studies the asymptotics for the proposed estimators. Given random variables U_n for $n \geq 1$, we write $U_n = O_P(b_n)$ if $\lim_{c\to\infty} \limsup_n P(|U_n| \geq cb_n) = 0$. Similarly, we write $U_n = o_P(b_n)$ if $\lim_n P(|U_n| \geq cb_n) = 0$, for any constant $c > 0$. Next, to facilitate discussion, we introduce some notation of norms. For any function g over the closure of domain Ω, denote $\|g\|_{L^2(\Omega)}^2 = \int_\Omega g^2(\boldsymbol{z})d\boldsymbol{z}$ the regular L_2 norm of g, and $\|g\|_{\infty,\Omega} = \sup_{\boldsymbol{z}\in\Omega} |g(\boldsymbol{z})|$ the supremum norm of g. Let $|g|_{v,\infty,\Omega} = \max_{i+j=v} \|\nabla_{z_1}^i \nabla_{z_2}^j g\|_{\infty,\Omega}$ be the maximum norms of all the vth order derivatives of g over Ω. Let $\mathcal{W}^{d,\infty}(\Omega) = \{g : |g|_{k,\infty,\Omega} < \infty, 0 \leq k \leq d\}$ be the standard Sobolev space. Next, we introduce some technical conditions.

(A1) For any $\ell = 0, \ldots, p$, $\beta_\ell^o(\cdot) \in \mathcal{W}^{d+1,\infty}(\Omega)$ for an integer $d \geq 1$.

(A2) For any $i = 1, \ldots, n$, $j = 1, \ldots, N$, ε_{ij}'s are independent with mean 0, variance 1, and for any $k \geq 1$, ξ_{ik}'s are uncorrelated random variables with mean 0 and variance 1.

(A3) For any $\ell = 0, 1, \ldots, p$, there exists a positive constant C_ℓ such that $E|X_\ell|^8 \leq C_\ell$. The eigenvalues of $\boldsymbol{\Sigma}_X = E(\boldsymbol{X}\boldsymbol{X}^\top)$ are bounded away from 0 and infinity.

(A4) The function $\sigma(\boldsymbol{z}) \in \mathcal{C}^{(1)}(\Omega)$ with $0 < c_\sigma \leq \sigma(\boldsymbol{z}) \leq C_\sigma \leq \infty$, for any $\boldsymbol{z} \in \Omega$; for any k, $\psi_k(\boldsymbol{z}) \in \mathcal{C}^{(1)}(\Omega)$ and $0 < c_G \leq G_\eta(\boldsymbol{z}, \boldsymbol{z}) \leq C_G \leq \infty$, for any $\boldsymbol{z} \in \Omega$.

(A5) Let $|\underline{\triangle}| = \min_{0 \leq \ell \leq p} |\triangle_\ell|$, and $|\overline{\triangle}| = \max_{0 \leq \ell \leq p} |\triangle_\ell|$. Triangulations $\triangle_\ell$'s satisify that $\limsup_n (|\overline{\triangle}|/|\underline{\triangle}|) < \infty$. The triangulations are π-quasi-uniform, that is, there exists a positive constant π such that $\max_{0 \leq \ell \leq p} \{(\min_{T \in \triangle_\ell} \varrho_T)^{-1} |\triangle_\ell|\} \leq \pi$.

(A6) As $N \to \infty$, $n \to \infty$, for some $0 < \kappa < 1$, $N^{-1} n^{1/(d+1)+\kappa} \to 0$, $n^{1/2} |\overline{\triangle}|^{d+1} \to 0$, $N^{1/2} |\underline{\triangle}| \to \infty$, and the smoothing parameters satisfy that $n^{-1/2} N^{-1} |\underline{\triangle}|^{-3} \rho_n \to 0$, where $\rho_n = \max_{0 \leq \ell \leq p} \rho_{n,\ell}$.

The above assumptions are mild conditions that can be satisfied in many practical situations. Assumption (A1) describes the requirement on the coefficient functions as usually used in the literature of nonparametric estimation. Assumption (A1) can be relaxed to Assumption (A1′) in Section 2.4 which only requires $\beta_\ell^o(\cdot) \in \mathcal{C}^{(0)}(\Omega)$ when dealing with the imaging data with sharp edges; see Section 2.4. Assumptions (A1) and (A2) are similar as Assumptions (A1) and (A2) in Gu et al. (2014) and Assumptions (A1)–(A3) in Huang et al. (2004). Assumption (A3) is analog to Assumption (A5) in Gu et al. (2014), ensuring that the $X_{i\ell}$'s are not multicollinear. Assumption

(A5) requires that the $\triangle_\ell$'s the of similar sizes, and it also suggests the use of more uniform triangulations with smaller shape parameters. Assumption (A6) implies that the number of pixels for each image N diverges to infinity, and the sample size n grows as $N \to \infty$, a well-developed asymptotic scenario for dense functional data (Li and Hsing, 2010). Assumption (A6) also describes the requirement of the growth rate of the dimension of the spline spaces relative to the sample size and the image resolution. This assumption is easily satisfied since images measured using new technology are usually of much higher resolution than the previous generation.

The following theorem provides the L_2 convergence rate of $\widehat{\beta}_\ell(\cdot)$, $\ell = 0, 1, \ldots, p$. Its detailed proofs are given in Appendix 1 in the Supplementary Material.

Theorem 1. *Suppose Assumptions (A1)–(A5) hold,* $N^{1/2}|\underline{\triangle}| \to \infty$ *as* $N \to \infty$*, then for any* $\ell = 0, 1, \ldots, p$*, the BPST estimator* $\widehat{\beta}_\ell(\cdot)$ *is consistent and satisfies that* $\|\widehat{\beta}_\ell - \beta_\ell^o\|_{L^2(\Omega)} = O_P\left\{\frac{\rho_n}{nN|\underline{\triangle}|^3}\|\boldsymbol{\beta}^o\|_{2,\infty} + \left(1 + \frac{\rho_n}{nN|\underline{\triangle}|^5}\right)|\overline{\triangle}|^{d+1}\|\boldsymbol{\beta}^o\|_{d+1,\infty} + n^{-1/2}\right\}$.

Theorem 2 below states the asymptotic normality of the $\widehat{\beta}_\ell$ at any given point $\boldsymbol{z} \in \Omega$, $\ell = 0, 1, \ldots, p$. See Appendix 1 in the Supplemental Material for detailed proofs. Denote

$$\boldsymbol{\Xi}_n(\boldsymbol{z}) = \widetilde{\mathbb{B}}(\boldsymbol{z})^\top E\left\{\boldsymbol{\Gamma}_{n,\rho}^{-1}\frac{1}{n^2N^2}\sum_{i=1}^{n}\sum_{j,j'=1}^{N}\mathbf{U}_{ij}\mathbf{U}_{ij'}^\top G_\eta(\boldsymbol{z}_j, \boldsymbol{z}_{j'})\boldsymbol{\Gamma}_{n,\rho}^{-1}\right\}\widetilde{\mathbb{B}}(\boldsymbol{z}), \tag{2.6}$$

where $\mathbf{U}_{ij}$ and $\boldsymbol{\Gamma}_{n,\rho}$ are given in (2.4) and (S17) in Appendix 1, respectively, $\widetilde{\mathbb{B}}(\boldsymbol{z}) = \text{diag}\{\widetilde{\mathbf{B}}_0(\boldsymbol{z}), \cdots, \widetilde{\mathbf{B}}_p(\boldsymbol{z})\}$, and for $\ell = 0, \ldots, p$, $\widetilde{\mathbf{B}}_\ell(\boldsymbol{z}) = \mathbf{Q}_{2,\ell}^\top\mathbf{B}_\ell(\boldsymbol{z})$.

Theorem 2. *Suppose Assumptions (A1)–(A6) hold. If for any* $\ell = 0, 1, \ldots, p$, $|X_{i\ell}| \leq C_\ell < \infty$, $\boldsymbol{\Xi}_n^{-1/2}(\boldsymbol{z})\{\widehat{\boldsymbol{\beta}}(\boldsymbol{z}) - \boldsymbol{\beta}^o(\boldsymbol{z})\} \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \mathbf{I}_{(p+1)\times(p+1)}\right)$, *as* $N \to \infty$, $n \to \infty$, *where* $\boldsymbol{\Xi}_n(\boldsymbol{z})$ *is given in (2.6). Furthermore, there exist positive constants* $c_V < C_V < +\infty$ *such that* $c_V n^{-1}\left(1 + \frac{\rho_n}{nN|\underline{\triangle}|^4}\right)^{-2} \leq \mathrm{Var}\{\widehat{\beta}_\ell(\boldsymbol{z})\} \leq C_V n^{-1}$, *for any* $\ell = 0, 1, \ldots, p$.

2.4 Piecewise Constant Spline over Triangulation Smoothing

Many imaging data can usually be regarded as a noisy version of piecewise-smooth function of $\boldsymbol{z} \in \Omega$ with sharp edges, which often reflect the functional or structural changes. The penalized bivariate spline smoothing method introduced in Section 2.2 assumes some degrees of smoothness over the entire image. To relax this assumption, and preserve the features of sharp edges, we make the following less stringent assumption on the smoothness of the coefficient functions:

(A1′) For any $\ell = 0, \ldots, p$, the bivariate function $\beta_\ell^o(\cdot) \in \mathcal{C}^{(0)}(\Omega)$.

For the estimation, we consider the piecewise constant spline over triangulation (PCST) method in this section. For any $\ell = 1, \ldots, p$, denote by $\mathcal{PC}(\triangle_\ell)$ the space of piecewise constant functions over each T_m, $m \in \mathcal{M}_\ell$. The bivariate spline basis functions of $\mathcal{PC}(\triangle_\ell)$ are denoted as $\{B_{\ell m}(\boldsymbol{z})\}_{m\in\mathcal{M}_\ell}$, which are simply indicator functions over triangle T_m, $B_{\ell m}(\boldsymbol{z}) = I(\boldsymbol{z} \in T_m)$, $m \in \mathcal{M}_\ell$. Assumption (A1′) controls the bias of the piecewise constant spline estimator for β_ℓ^o and leads to the estimation consistency.

When using the constant bivariate spline basis functions, one has $\mathcal{E}(s) = 0$ for all $s \in \mathcal{PC}(\triangle)$, and for any $\boldsymbol{z} \in \Omega$, $\mathbf{B}_\ell(\boldsymbol{z})\mathbf{B}_\ell(\boldsymbol{z})^\top = \mathrm{diag}\{B_{\ell m}^2(\boldsymbol{z}), m \in \mathcal{M}_\ell\}$. Then

$\widehat{\boldsymbol{\gamma}}_m = (\widehat{\gamma}_{0m}, \widehat{\gamma}_{1m}, \ldots, \widehat{\gamma}_{pm})^\top = \widehat{\mathbf{V}}_m^{-1} \left\{ (nN)^{-1} \sum_{i=1}^n \sum_{j=1}^N B_{\ell m}(\boldsymbol{z}_j) X_{i\ell} Y_{ij} \right\}_{\ell=0}^p$, where

$$\widehat{\mathbf{V}}_m = \frac{1}{nN} \sum_{j=1}^N B_{\ell m}^2(\boldsymbol{z}_j) \sum_{i=1}^n \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^\top = \left\{ \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N B_{\ell m}^2(\boldsymbol{z}_j) X_{i\ell} X_{i\ell'} \right\}_{\ell,\ell'=0}^p . \tag{2.7}$$

By simple linear algebra, for any $\ell = 0, \ldots, p$, the PCST estimator

$$\widehat{\beta}_\ell^{\mathrm{c}}(\boldsymbol{z}) = \sum_{m \in \mathcal{M}_\ell} \widehat{\gamma}_{\ell m} B_{\ell m}(\boldsymbol{z}). \tag{2.8}$$

For any $\boldsymbol{z} \in \Omega$, define the index of the triangle containing $\boldsymbol{z}$ as $m(\boldsymbol{z})$, i.e., $m(\boldsymbol{z}) = m$, if $\boldsymbol{z} \in T_m$. Then $\widehat{\beta}_\ell(\boldsymbol{z}) = \widehat{\boldsymbol{\gamma}}_{\ell m(\boldsymbol{z})}$, and $\widehat{\boldsymbol{\beta}}^{\mathrm{c}}(\boldsymbol{z}) = (\widehat{\beta}_0^{\mathrm{c}}(\boldsymbol{z}), \ldots, \widehat{\beta}_p^{\mathrm{c}}(\boldsymbol{z}))^\top = (\widehat{\gamma}_{0m(\boldsymbol{z})}, \ldots, \widehat{\gamma}_{pm(\boldsymbol{z})})^\top = \widehat{\boldsymbol{\gamma}}_{m(\boldsymbol{z})}$. For any $\boldsymbol{z} \in \Omega$, denote

$$\boldsymbol{\Sigma}_n(\boldsymbol{z}) = n^{-1} \boldsymbol{\Sigma}_X^{-1} G_\eta(\boldsymbol{z}, \boldsymbol{z}). \tag{2.9}$$

Theorem 3 below shows the asymptotic normality of the piecewise constant estimators $\widehat{\boldsymbol{\beta}}(\boldsymbol{z})$. See the Supplementary Material for detailed proofs. To obtain the asymptotic variance-covariance function, we also need the following assumption:

(C1) The variables ξ_{ik}'s and ε_{ij}'s are independent and satisfy that $E|\xi_{ik}|^{4+\delta_1} < +\infty$ for some $\delta_1 > 0$, and $E|\varepsilon_{ij}|^{4+\delta_2} < \infty$ for some $\delta_2 > 0$.

Theorem 3. *Under Assumptions (A1′), (A2)–(A5), and (C1), as $N \to \infty$, $n \to \infty$, if for some $0 < \kappa < 1$, $N^{-1} n^{1+\kappa} \to 0$, $N^{-1/2} \ll |\underline{\triangle}| \leq |\overline{\triangle}| \ll n^{1/4} N^{-1/2}$, and $\| \sum_{k=1}^\infty \lambda_k^{1/2} \psi_k \|_\infty < \infty$, then for any $\boldsymbol{z} \in \Omega$, $\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{z}) \{ \widehat{\boldsymbol{\beta}}^{\mathrm{c}}(\boldsymbol{z}) - \boldsymbol{\beta}^o(\boldsymbol{z}) \} \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \mathbf{I}_{(p+1)\times(p+1)}\right)$, where $\boldsymbol{\Sigma}_n(\boldsymbol{z})$ is (2.9);* $\mathrm{pr}\left\{ (\sigma_{n,\ell\ell}^{\mathrm{c}})^{-1}(\boldsymbol{z}) \left| \widehat{\beta}_\ell(\boldsymbol{z}) - \beta_\ell^o(\boldsymbol{z}) \right| \leq Z_{1-\alpha/2} \right\} \to 1 - \alpha$, *for any $\alpha \in (0,1)$, as $N \to \infty$, $n \to \infty$, where $\sigma_{n,\ell\ell}^{\mathrm{c}}(\boldsymbol{z})$ is the square root of the (ℓ,ℓ)th entry of the matrix $\boldsymbol{\Sigma}_n(\boldsymbol{z})$, and $Z_{1-\alpha/2}$ is the $100\,(1-\alpha/2)^{th}$ percentile of the standard normal distribution.*

3. Variance Function Estimation and Simultaneous Confidence Corridors

3.1 Estimation of the Variance Function

Define the estimated residual $\widehat{R}_{ij} = Y_{ij} - \sum_{\ell=0}^{p} X_{i\ell}\widehat{\beta}_\ell(\boldsymbol{z}_j)$ or $Y_{ij} - \sum_{\ell=0}^{p} X_{i\ell}\widehat{\beta}^{\mathrm{c}}_\ell(\boldsymbol{z}_j)$, for any $i = 1, \ldots, n$, $j = 1, \ldots, N$. We employ the bivariate spline smoothing method to $\{(\widehat{R}_{ij}, \boldsymbol{z}_j)\}_{j=1}^{N}$. To be more specific, we define

$$\widehat{\eta}_i(\boldsymbol{z}) = \underset{g_i \in \mathcal{S}_d^r(\triangle_\eta)}{\arg\min} \sum_{j=1}^{N} \left\{ \widehat{R}_{ij} - g_i(\boldsymbol{z}_j) \right\}^2, \ i = 1, \ldots, n, \tag{3.1}$$

as the spline estimator of $\eta_i(\boldsymbol{z})$, where the triangulation $\triangle_\eta$ may be different from the one introduced in Section 2 when estimating $\beta^o_\ell(\boldsymbol{z})$'s. Next, let $\widehat{\epsilon}_{ij} = \widehat{R}_{ij} - \widehat{\eta}_i(\boldsymbol{z}_j)$. Define the estimator of $G_\eta(\boldsymbol{z}, \boldsymbol{z}')$ and $\sigma^2(\boldsymbol{z}_j)$ as:

$$\widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') = n^{-1} \sum_{i=1}^{n} \widehat{\eta}_i(\boldsymbol{z}) \widehat{\eta}_i(\boldsymbol{z}'), \quad \widehat{\sigma}^2(\boldsymbol{z}_j) = n^{-1} \sum_{i=1}^{n} \widehat{\epsilon}_{ij} \widehat{\epsilon}_{ij}. \tag{3.2}$$

In general, for spline estimators ($d \geq 0$), denote $\widehat{\boldsymbol{\Xi}}_n(\boldsymbol{z}) = \left\{ \widehat{\sigma}^2_{n,\ell\ell'}(\boldsymbol{z}) \right\}_{\ell,\ell'=0}^{p}$, where

$$\widehat{\boldsymbol{\Xi}}_n(\boldsymbol{z}) = \frac{1}{n^2 N^2} \widetilde{\mathbb{B}}(\boldsymbol{z})^\top \sum_{i=1}^{n} \left\{ \sum_{j,j'=1}^{N} \boldsymbol{\Gamma}_{n,\rho}^{-1} \mathbf{U}_{ij} \mathbf{U}_{ij'}^\top \widehat{G}_\eta(\boldsymbol{z}_j, \boldsymbol{z}_{j'}) \boldsymbol{\Gamma}_{n,\rho}^{-1} + \sum_{j=1}^{N} \mathbf{U}_{ij} \mathbf{U}_{ij}^\top \widehat{\sigma}^2(\boldsymbol{z}_j) \right\} \widetilde{\mathbb{B}}(\boldsymbol{z}). \tag{3.3}$$

It is also worth mentioning that the estimation can be much simplified if PCST smoothing is applied. In this case, the variance-covariance matrix $\boldsymbol{\Sigma}_n(\boldsymbol{z})$ can be simply estimated by

$$\widehat{\boldsymbol{\Sigma}}_n(\boldsymbol{z}) = \left\{ (\widehat{\sigma}^{\mathrm{c}}_{n,\ell\ell'})^2(\boldsymbol{z}) \right\}_{\ell,\ell'=0}^{p} = \frac{1}{n} \left(n^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^\top \right)^{-1} \left\{ \widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}) + \frac{\widehat{\sigma}^2(\boldsymbol{z})}{N A_{m(\boldsymbol{z})}} \right\},$$

where $A_{m(\boldsymbol{z})}$ is the area of triangle $T_{m(\boldsymbol{z})}$ divided by the area of the domain. The following conditions (C2)–(C3) are required for the bivariate spline approximation in

the covariance estimation and establishing the estimation consistency. Proofs of results in this section are in the Supplementary Material.

(C2) For any $k \geq 1$, $\psi_k(\boldsymbol{z}) \in \mathcal{W}^{s+1,\infty}$ for an integer $s \geq 0$, and for a sequence $\{K_n\}_{n=1}^{\infty}$ of increasing positive integers with $\lim_n K_n \to \infty$, $|\triangle_\eta|^{s+1} \sum_{k=1}^{K_n} \lambda_k^{1/2} \|\psi_k\|_{s+1,\infty} \to 0$ as $N \to \infty$, $n \to \infty$.

(C3) As $N \to \infty$, $n \to \infty$, for some $0 < \kappa < 1$, $N^{-1} n^{1/(d+1)+\kappa} \to 0$, $N|\triangle_\eta|^2 \to \infty$ and $n|\triangle_\eta|^2/(\log n)^{1/2} \to \infty$.

Assumption (C2) concerns the bounded smoothness of principal components for bounding the bias terms in the spline covariance estimator.

Theorem 4. *Under Assumptions (A1)–(A6), (C1)–(C3),* $\widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}')$ *uniformly converges to* $G_\eta(\boldsymbol{z}, \boldsymbol{z}')$ *in probability, i.e.,* $\sup_{(\boldsymbol{z},\boldsymbol{z}') \in \Omega^2} |\widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') - G_\eta(\boldsymbol{z}, \boldsymbol{z}')| = o_P(1)$.

Corollary 1. *Under Assumptions (A1)–(A6), (C1)–(C3), the estimator of* $\widehat{\boldsymbol{\Sigma}}_n(\boldsymbol{z})$ *uniformly converges to to* $\boldsymbol{\Sigma}_n(\boldsymbol{z})$ *in probability, i.e.,* $\sup_{\boldsymbol{z} \in \Omega} |\widehat{\boldsymbol{\Sigma}}_n(\boldsymbol{z}) - \boldsymbol{\Sigma}_n(\boldsymbol{z})| = o_P(1)$.

Denote

$$\widehat{\sigma}^{\mathrm{c}}_{n,\ell\ell}(\boldsymbol{z}) = n^{-1/2} \left[\mathbf{e}_\ell^\top \left(n^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^\top \right)^{-1} \mathbf{e}_\ell \left\{ \widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}) + \frac{\widehat{\sigma}^2(\boldsymbol{z})}{N A_{m(\boldsymbol{z})}} \right\} \right]^{1/2}. \tag{3.4}$$

As a result of Corollary 1, $\widehat{\sigma}^{\mathrm{c}}_{n,\ell\ell}(\boldsymbol{z})$ is a consistent estimator of $\sigma^{\mathrm{c}}_{n,\ell\ell}(\boldsymbol{z})$ in (2.9).

3.2 Bootstrap Simultaneous Confidence Corridors (SCCs)

Following from Theorems 2, 3 and the Slutzky's Theorem, we can have the following asymptotic PCIs.

Corollary 2. *(a) For the BPST estimators, under Assumptions (A1)–(A6), for any* $\ell = 0, \ldots, p$, $\alpha \in (0,1)$, *as* $N \to \infty$, $n \to \infty$, *an asymptotic* $100(1-\alpha)\%$ *PCI for* $\beta_\ell^o(\boldsymbol{z})$, *is* $\widehat{\beta}_\ell(\boldsymbol{z}) \pm \sigma_{n,\ell\ell}(\boldsymbol{z}) Z_{1-\alpha/2}$, *for any* $\boldsymbol{z} \in \Omega$, *where* $\sigma^2_{n,\ell\ell}(\boldsymbol{z})$ *is the* (ℓ,ℓ)*th entry of the matrix* $\boldsymbol{\Xi}_n^{-1/2}(\boldsymbol{z})$, *and* $Z_{1-\alpha/2}$ *is the* $100\,(1-\alpha/2)^{th}$ *percentile of the standard normal distribution.*

(b) For the PCST estimators, under Assumptions (A1′), (A2)–(A6), if for some $0 < \kappa < 1$, $N^{-1}n^{1+\kappa} \to 0$, *an asymptotic* $100(1-\alpha)\%$ *PCI for* $\beta_\ell^o(\boldsymbol{z})$, *is* $\widehat{\beta}_\ell^{\mathrm{c}}(\boldsymbol{z}) \pm \sigma^{\mathrm{c}}_{n,\ell\ell}(\boldsymbol{z}) Z_{1-\alpha/2}$, *for any* $\boldsymbol{z} \in \Omega$, *where* $\sigma^c_{n,\ell\ell}(\boldsymbol{z})$ *is the standard deviation function of* $\widehat{\beta}_\ell^{\mathrm{c}}(\boldsymbol{z})$ *in Theorem 3.*

Next, we introduce a simple bootstrap approach to extend the PCIs to the SCCs. Our approach is based on the nonparametric bootstrap method used in Hall and Horowitz (2013). We triangulate the domain Ω with quasi-uniform triangles, and obtain a set of approximate $100(1-\alpha)\%$ piecewise confidence intervals (PCIs). In the following α_0 denote the nominal confidence level of the desired SCCs. We recalibrate the PCIs using the following bootstrap method.

Step 1. Based on $\left\{(\widetilde{\mathbf{X}}_i, Y_{ij})\right\}_{j=1,i=1}^{N,n}$, obtain the coefficient functions $\beta_\ell^o(\boldsymbol{z})$ via the BPST estimators $\widehat{\beta}_\ell(\boldsymbol{z})$ in (2.5) or the PCST estimators $\widehat{\beta}_\ell^{\mathrm{c}}(\boldsymbol{z})$ in (2.8), for $\ell = 0, \ldots, p$. Let $\widehat{\mu}(\boldsymbol{z}) = \sum_{\ell=0}^p X_{i\ell}\widehat{\beta}_\ell(\boldsymbol{z})$ or $\sum_{\ell=0}^p X_{i\ell}\widehat{\beta}_\ell^{\mathrm{c}}(\boldsymbol{z})$.

Step 2. Obtain $\widehat{\eta}_i(\boldsymbol{z})$ and $\widehat{\varepsilon}_{ij}$ presented in (3.1)–(3.2), and estimate $G_\eta(\boldsymbol{z},\boldsymbol{z})$, $\sigma^2(\boldsymbol{z})$ and $\sigma^2_{n,\ell\ell}(\boldsymbol{z})$ by $\widehat{G}_\eta(\boldsymbol{z},\boldsymbol{z})$ and $\widehat{\sigma}^2(\boldsymbol{z})$ in (3.2) and $\widehat{\sigma}^2_{n,\ell\ell}(\boldsymbol{z})$ in (3.3) or (3.4), respectively.

Step 3. Obtain adjusted nominal confidence level $\widehat{\alpha}_\ell(\alpha_0)$.

(i) Generate an independent random sample $\delta_i^{(b)}$ and $\delta_{ij}^{(b)}$ from $\{-1, 1\}$ with probability 0.5 each, and define $Y_{ij}^{*(b)} = \widehat{\mu}(\boldsymbol{z}_j) + \delta_i^{(b)}\widehat{\eta}_i(\boldsymbol{z}_j) + \delta_{ij}^{(b)}\widehat{\varepsilon}_{ij}$.

(ii) Based on $\left\{(\widetilde{\mathbf{X}}_i, Y_{ij}^{*(b)})\right\}_{j=1,i=1}^{N,n}$, obtain $\widehat{\beta}_\ell^{*(b)}(\boldsymbol{z})$ using (2.5) or (2.8), and calculate $\widehat{\sigma}_{n,\ell\ell}^{*(b)}$ using (3.3) or (3.4).

(iii) Construct SCCs, for resampled data $\left\{(\widetilde{\mathbf{X}}_i, Y_{ij}^{*(b)})\right\}_{j=1,i=1}^{N,n}$: $\mathcal{B}_{(b)}^*(\alpha)$, $b = 1, \cdots, B$,

$$\mathcal{B}_{(b)}^*(\alpha) = \{(\boldsymbol{z}, y) : \boldsymbol{z} \in \Omega, \widehat{\beta}_\ell^{*(b)}(\boldsymbol{z}) - \widehat{\sigma}_{n,\ell\ell}^{*(b)}(\boldsymbol{z}) Z_{1-\alpha/2} \le y \le \widehat{\beta}_\ell^{*(b)}(\boldsymbol{z}) + \widehat{\sigma}_{n,\ell\ell}^{*(b)}(\boldsymbol{z}) Z_{1-\alpha/2}\}.$$

(iv) Estimate coverage rate $\tau_\ell(\boldsymbol{z}_j, \alpha) = P\{(\boldsymbol{z}_j, \widehat{\beta}_\ell(\boldsymbol{z}_j)) \in \mathcal{B}^*(\alpha)|\mathbb{X}\}$ using $\widehat{\tau}_\ell(\boldsymbol{z}_j, \alpha) = \frac{1}{B}\sum_{b=1}^B I\{(\boldsymbol{z}_j, \widehat{\beta}_\ell(\boldsymbol{z}_j)) \in \mathcal{B}_{(b)}^*(\alpha)\}$.

(v) Find the root to the equation $\widehat{\tau}_\ell(\boldsymbol{z}_j, \alpha) = 1 - \alpha_0$, $j = 1, \ldots, N$, and denote them as $\{\widehat{\alpha}_\ell(\boldsymbol{z}_j, \alpha_0)\}_{j=1}^N$. The root can be found using grid method by repeating the last two steps with respect to different values of α.

(vi) Take the minimum of $\{\widehat{\alpha}_\ell(\boldsymbol{z}_j, \alpha_0)\}_{j=1}^N$ and denote it as $\widehat{\alpha}_\ell \equiv \widehat{\alpha}_\ell(\alpha_0)$.

Step 4. Construct the final SCCs: $\mathcal{B}(\widehat{\alpha}_\ell) = \{(\boldsymbol{z}, y) : \boldsymbol{z} \in \Omega, \widehat{\beta}_\ell(\boldsymbol{z}) - \widehat{\sigma}_{n,\ell\ell}(\boldsymbol{z}) Z_{1-\widehat{\alpha}_\ell/2} \le y \le \widehat{\beta}_\ell(\boldsymbol{z}) + \widehat{\sigma}_{n,\ell\ell}(\boldsymbol{z}) Z_{1-\widehat{\alpha}_\ell/2}\}$.

4. Implementation

The proposed procedure can be easily implemented using our R package "FDAimage" (Yu et al., 2019), in which the bivariate spline basis are generated via the R package

"BPST" (Wang et al., 2019). When the response imaging seems to be a realization from some smooth function, we suggest using smooth parameter $r = 1$, and degree $d \geq 5$, which achieves full estimation power asymptotically (Lai and Schumaker, 2007). In contrast, if there are sharp edges on the images, we suggest considering PCST presented in Section 2.4.

Selecting suitable values of smoothing parameters is important to good model fitting. To select $\rho_{n,\ell}$, $\ell = 0, \ldots, p$, we used K-fold cross validation (CV). The individuals are randomly partitioned into K groups, and one groups is retained as test set and the remaining $K - 1$ groups are used as training set. The cross-validation process is then repeated K times (the folds), with each of the K groups used exactly once as the validation data. Then the K-fold CV score is $\mathrm{CV}(\rho_{n,0}, \ldots, \rho_{n,p}) = K^{-1} \sum_{k=1}^{K} (|\mathcal{V}_k|N)^{-1} \sum_{i \in \mathcal{V}_k} \sum_{j=1}^{N} \{Y_{ij} - \widetilde{\mathbf{X}}_i^\top \widehat{\boldsymbol{\beta}}_{-k}(\boldsymbol{z}_j)\}^2$, where $\mathcal{V}_k$ is the kth testing set for $k = 1, \ldots, K$ and $\widehat{\boldsymbol{\beta}}_{-k}$ is the corresponding estimator after removing the kth testing set. We have used $K = 5$ in the numerical examples below.

To determine an optimal triangulation, the criterion usually considers the shape, size or number of triangles. In terms of shape, a "good" triangulation usually refers to those with well-shaped triangles without small angles or/and obtuse angles. Therefore, for a given number of triangles, Lai and Schumaker (2007) and Lindgren et al. (2011) recommended selecting the triangulation according to "max-min" criterion, which maximizes the minimum angle of all the angles of the triangles in the triangulation. With respect to the number of triangles, our numerical studies show that a lower limit of

the number of triangles is necessary to capture the features of the images, but once this minimum number has been reached, further refining the triangulation usually has little effect on the fitting process. In practice, when using higher order BPST smoothing, we suggest taking the number of triangles as: $H_n = \min\{\lfloor c_1 n^{1/(2d+2)} N^{1/2} \rfloor, N/10\}$, where c_1 is a tuning parameter and we find that $c_1 \in [0.3, 2.0]$ works very well in our numerical studies, while using PCST, we suggest taking the number of triangles as: $H_n = \min\{\lfloor c_2 n^{-1/4} N \rfloor, N/2\}$ with $c_2 \in [0.3, 2.0]$. Once H_n is chosen, one can build the triangulation using typical triangulation construction methods such as Delaunay Triangulation and Distmesh (Persson and Strang, 2004).

5. Simulation Studies

In this section, we conduct two Monte Carlo simulation studies using our R package "FDAimage" (Yu et al., 2019) to examine the finite sample performance of the proposed methodology. The triangulations used below can be found from the dataset in the "FDAimage" package. To illustrate the performance of our estimation method, we compare the proposed spline method with the Kernel method proposed by Zhu et al. (2014) (Kernel), and the tensor regression method in Li and Zhang (2017) (Tensor). To implement the Kernel method, we use R Package *SVCM*, which is publicly available at `https://github.com/BIG-S2/SVCM`, and for Tensor method, the accompanying matlab code at `https://ani.stat.fsu.edu/~henry/TensorEnvelopes_html.html` is used. We compare the proposed method with the tensor regression approach in Li and

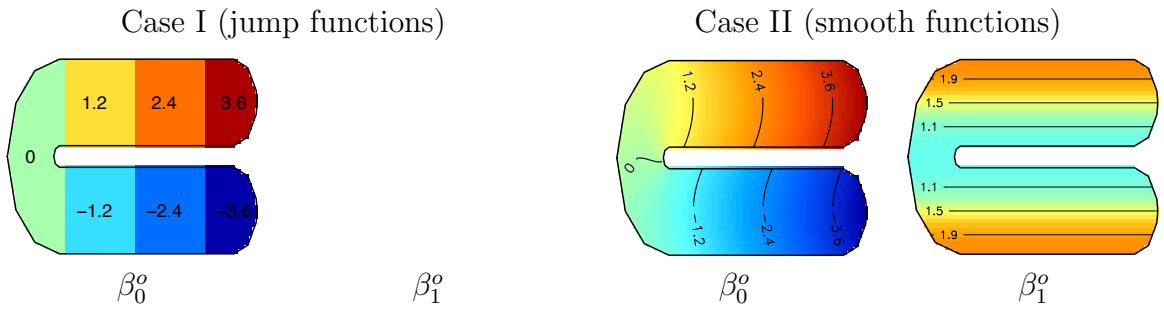

Figure 3: The true coefficient functions in the Simulation Example 1.

Zhang (2017) and the three-stage FDA approach in Zhu et al. (2014).

5.1 Example 1

To illustrate the advantage of the proposed method over complex domain, we study the horseshoe domain in Sangalli et al. (2013). The response images are generated from the following model: $Y_{ij} = \beta_0^o(\boldsymbol{z}_j) + X_i\beta_1^o(\boldsymbol{z}_j) + \eta_i(\boldsymbol{z}_j) + \sigma\varepsilon_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, N$, and $\boldsymbol{z}_j \in \Omega$. To understand the advantages and disadvantages of different methods, we consider two types of coefficient functions in the above image-on-scalar regression model: (I) functions with jumps; and (II) smooth functions. The true coefficient functions are demonstrated in Figure 3.

For each image, we set the resolution as 100×50 (pixels). The true signal falls only within the horseshoe domain (3182 pixels), while outside the domain, there are pure noises. We generate scalar covariate $X_i \sim N(0, 1)$ and then truncate it by $[-3, +3]$. We set $\eta_i(\boldsymbol{z}) = \sum_{k=1}^{2} \lambda_k^{1/2} \xi_{ik} \psi_k(\boldsymbol{z})$, where $(\lambda_1, \lambda_2) = (0.1, 0.02)$ or $(0.2, 0.05)$, ξ_{i1} and $\xi_{i2} \sim N(0, 1)$, $\psi_1(\boldsymbol{z}) = c_1 \sin(2\pi z_1)$ and $\psi_2(\boldsymbol{z}) = c_2 \cos(2\pi z_2)$. Let $c_1 = 0.56$, $c_2 = 0.61$,

Table 1: Estimation errors of the coefficient estimators, $\sigma = 2.0$.

Function Type	n	Method	$\lambda_1 = 0.03,\ \lambda_2 = 0.006$		$\lambda_1 = 0.2,\ \lambda_2 = 0.05$	
			β_0	β_1	β_0	β_1
Jump	50	BPST	0.0139	0.0182	0.0145	0.0189
		PCST	0.0088	0.0090	0.0094	0.0097
		Kernel	0.0801	0.0819	0.0807	0.0826
		Tensor	0.0799	0.0248	0.0799	0.0254
	100	BPST	0.0090	0.0118	0.0093	0.0122
		PCST	0.0044	0.0044	0.0047	0.0047
		Kernel	0.0400	0.0405	0.0403	0.0409
		Tensor	0.0395	0.0166	0.0399	0.0171
Smooth	50	BPST	0.0026	0.0032	0.0032	0.0041
		PCST	0.0088	0.0090	0.0119	0.0139
		Kernel	0.0801	0.0819	0.0807	0.0826
		Tensor	0.0799	0.0256	0.0806	0.0271
	100	BPST	0.0016	0.0019	0.0019	0.0022
		PCST	0.0070	0.0086	0.0073	0.0090
		Kernel	0.0400	0.0405	0.0403	0.0409
		Tensor	0.0399	0.0168	0.0402	0.0179

so that ψ_1 and ψ_2 are orthonormal functions on Ω. The measurement error ε_{ij} is independently generated from $N(0, 1)$ and $\sigma = 1.0,\ 2.0$.

To fit the model, we consider the BPST and PCST methods presented in Section 2. To obtain the BPST estimators, we set $d = 5$ and $r = 0$ when generating the bivariate spline basis functions. Figure S1 in the Supplementary Material illustrates the triangulations used for the BPST and PCST. The triangulation used for BPST ($\triangle_1$) contains 90 triangles (73 vertices), and the triangulation used for PCST ($\triangle_2$) contains 346 triangles (226 vertices).

We quantify the estimation accuracy of the coefficient functions using the mean

squared error (MSE). Table 1 provides the average MSE (across 500 Monte Carlo experiments) for two types of coefficient functions. To save the space, we only present results for $\sigma = 2.0$ here, and the results for $\sigma = 1.0$ are presented in Table S1 in the Supplementary Material. As expected, the estimation accuracy for all the methods improves as the sample size increases or noise level decreases. In both scenarios, the BPST and PCST outperform the other two competitors, reflecting the advantage of our method over a complex domain. When the true coefficient functions are smooth, BPST provides the best estimation, followed by PCST. On the other hand, when the true coefficient function contains jumps, PCST provides a better result. For Tensor regression, the estimator of $\beta_1^o(\cdot)$ is much more accurate than the estimator of $\beta_0^o(\cdot)$ due to the design of the coefficient function. Based on Figure 3, one sees that, in contrast to the intercept function of $\beta_0^o(\cdot)$, the true slope function of $\beta_1^o(\cdot)$ is still smooth across the complex boundary. Moreover, even when the coefficient function is smooth across the boundary, the estimation accuracy is also affected by the domain of the true signal. The performance of the Kernel method is not affected by the design of the coefficient functions, instead, it heavily depends on the noise level due to the three-stage structure.

5.2 Example 2

In this example, we consider the domain of the 5th and the 35th slices of the brain images illustrated in Section 6 as the domain Ω to simulate data. We generate response images based on a set of smooth coefficient functions from the following model: $Y_{ij} =$

$\sum_{\ell=0}^{2} X_{i\ell}\beta_\ell^o(\boldsymbol{z}_j) + \eta_i(\boldsymbol{z}_j) + \sigma\varepsilon_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, N$, and $\boldsymbol{z}_j \in \Omega$, where $\beta_0^o(\boldsymbol{z}) = 5\{(z_1-0.5)^2+(z_2-0.5)^2\}$, $\beta_1^o(\boldsymbol{z}) = -1.5z_1^3+1.5z_2^3$, and $\beta_2^o(\boldsymbol{z}) = 2-2\exp[-8\{(z_1-0.5)^2+(z_2-0.5)^2\}]$, and the true coefficient images are shown in the first column of Figures S5 and S6 in the Supplementary Material for the 5th and 35th slices, respectively. For each image, we simulate the data at all 79×95 pixels. To mimic the real brain images, the true signals are only generated on the pixels/voxels (3476 or 5203 pixels in total) within the brain domain, while outside the boundary of the brain, the image only contains noises. We set $X_{i0} = 1$ and generate $\widetilde{\mathbf{X}}_i = (X_{i1}, X_{i2})^\top \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \left(\begin{smallmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{smallmatrix}\right)$, and the $X_{i\ell}$'s truncated by $[-3, +3]$. For the error terms, we set $\eta_i(\boldsymbol{z}) = \sum_{k=1}^{2} \lambda_k^{1/2}\xi_{ik}\psi_k(\boldsymbol{z})$, where ξ_{i1} and $\xi_{i2} \sim N(0,1)$, $\psi_1(\boldsymbol{z}) = 1.488\{\sin(\pi z_1) - 1.5\}$, $\psi_2(\boldsymbol{z}) = 1.939\cos(2\pi z_2)$, $(\lambda_1, \lambda_2) = (0.1, 0.02)$ or $(0.2, 0.05)$. The measurement error ε_{ij} is independently generated from $N(0,1)$ and $\sigma = 0.5,\ 1.0$. For the sake of space, we only demonstrate the results based on the domain of the 5th slice for $\sigma = 1.0$ in the main article. The results based on $\sigma = 0.5$ and the results based on the domain of the 35th slice are shown in Section S2 in the Supplementary Material.

Since the functions in this example are smooth, for the bivariate spline approach, we only consider the BPST method. To further study the effect of different triangulations, we consider $\triangle_3$ and $\triangle_4$; see Figure S4 in the Supplementary Material. Similar to Section 5.1, we summarize the MSE for different coefficient functions based on 500 Monte Carlo experiments in Table 2. Columns 2–5 in Figure S5 in the Supplementary Material display the image of estimated coefficient functions using the Kernel, Tensor

Table 2: Estimation errors of the coefficient function estimators, $\sigma = 1.0$.

n	Method	$\lambda_1 = 0.1,\ \lambda_2 = 0.02$			$\lambda_1 = 0.2,\ \lambda_2 = 0.05$		
		β_0	β_1	β_2	β_0	β_3	β_2
50	BPST($\triangle_3$)	0.003	0.005	0.005	0.007	0.011	0.010
	BPST($\triangle_4$)	0.003	0.005	0.005	0.007	0.010	0.009
	Kernel	0.023	0.032	0.032	0.026	0.037	0.037
	Tensor	0.023	0.013	0.019	0.026	0.017	0.024
100	BPST($\triangle_3$)	0.002	0.002	0.002	0.003	0.005	0.005
	BPST($\triangle_4$)	0.002	0.002	0.002	0.003	0.004	0.004
	Kernel	0.011	0.015	0.015	0.013	0.018	0.018
	Tensor	0.011	0.007	0.011	0.013	0.009	0.013

and BPST methods, respectively. Based on Table 2 and Figure S5 in the Supplementary Material, one sees that the estimation accuracy for all methods improves as the sample size increases or the noise level decreases. In all the settings, the BPST method has the smallest MSE compared with the Kernel and Tensor methods, reflecting the advantage of our method in estimating the coefficient functions and hence the regression function. Since both the Tensor and Kernel methods are designed for a rectangle domain, the estimation accuracy can be affected by the noises outside the domain. One also observes that the MSE is invariable across two triangulations, which shows that $\triangle_3$ might be sufficient to capture the feature in the dataset. It also implies that, when this minimum number of triangles is reached, further refining the triangulation will have little effect on the fitting process, but makes the computational burden unnecessarily heavy.

Finally, we illustrate the finite sample performance of the proposed SCCs for the coefficient functions described in Section 3. In particular, we report empirical coverage probabilities of nominal 95% SCCs using triangulation $\triangle_3$. We evaluate the coverage

Table 3: The coverage rate of the 95% SCCs for the coefficient functions.

n	λ	σ	Coverage			Width		
			β_0	β_1	β_2	β_0	β_1	β_2
50	(0.1,0.02)	0.5	0.976	0.928	0.938	0.332	0.362	0.377
		1.0	0.976	0.940	0.952	0.358	0.392	0.413
	(0.2,0.05)	0.5	0.962	0.918	0.932	0.445	0.497	0.513
		1.0	0.970	0.930	0.940	0.478	0.527	0.544
100	(0.1,0.02)	0.5	0.970	0.956	0.956	0.234	0.250	0.267
		1.0	0.978	0.968	0.978	0.262	0.285	0.297
	(0.2,0.05)	0.5	0.956	0.958	0.936	0.313	0.348	0.357
		1.0	0.966	0.964	0.954	0.344	0.378	0.389

of the proposed SCCs over all pixels on the interior of Ω and test whether the true functions are entirely covered by the SCCs at these pixels. Table 3 summarizes the empirical coverage rate (ECR) among 500 Monte Carlo experiments of the 95% SCCs and the average width of the SCCs. The results clearly show the ECRs of SCCs are well approximated to 95%. In particular, when the sample size increases, the ECRs are more approximate to 95%. Table 3 also reveals that the SCCs tend to be narrower when the sample size becomes larger or the noise level gets smaller.

6. ADNI Data Analysis

To illustrate the proposed method, we consider the spatially normalized FDG (fludeoxyglucose) PET data of Alzheimer's Disease Neuroimaging Initiative (ADNI). As pointed out in Marcus et al. (2014), FDG-PET image has been proven to be a promising modality for detecting functional brain changes in Alzheimer's Disease (AD). The data can be obtained from the ADNI database at `http://adni.loni.usc.edu/`. In the database,

there are spatially normalized PET images of 447 subjects. Among those 447 subjects, 112 have normal cognitive functions which are considered to be the control group, 213 are diagnosed as mild cognitive impairment (MCI) and 122 are diagnosed as AD. Table S5 in the Supplementary Material summarizes the distribution of patients by diagnosis status and sex.

In this study, we examine several patient-level features including (i) demographical features such as age (Age) and sex (Sex); (ii) dummy variable of the abnormal diagnosis status named as "MA" (1 = either "AD" or "MCI", 0, otherwise); (iii) dummy variable of AD (1 = "AD", 0, otherwise); and (iv) dummy variables of APOE genotype which has already been discovered as the strongest genetic risk factor for "AD" from previous study; see Corder et al. (1993). We code APOE_1 as a dummy variable for subjects with one epsilon 4 allele, and APOE_2 as subjects who have two alleles.

Noting that the PET images are 3D, we select the 5th, 8th, 15th, 35th, 55th, 62nd, and 65th horizontal slices (bottom to up) of the brain from a total of 68 slices to illustrate our method. Each slice of the image contains 79×95 pixels, but the domain of brain for different slices are quite different. Specifically, the domain boundary for the bottom slices and upper slices are much more complex than the slices in the middle, more examples can be found in Figure S7 in the Supplementary Material. For each slice, we consider the following image-on-scalar regression:

$$\begin{aligned} Y_i(\boldsymbol{z}_j) =& \beta_0(\boldsymbol{z}_j) + \beta_1(\boldsymbol{z}_j)\text{MA}_i + \beta_2(\boldsymbol{z}_j)\text{AD}_i + \beta_3(\boldsymbol{z}_j)\text{Age}_i + \beta_4(\boldsymbol{z}_j)\text{Sex}_i \\ &+ \beta_5(\boldsymbol{z}_j)\text{APOE}_{1i} + \beta_6(\boldsymbol{z}_j)\text{APOE}_{2i} + \eta_i(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_i(\boldsymbol{z}_j),\ i = 1, \ldots, n. \end{aligned}$$

Table 4: 10-fold CV results for the ADNI dataset. ($\times 10^{-2}$)

Method	Slice 5	Slice 8	Slice 15	Slice 35	Slice 55	Slice 62	Slice 65
BPST	1.4508	1.4809	1.5013	1.5633	2.0693	2.3020	2.6239
Kernel	1.4533	1.4828	1.5021	1.5638	2.0715	2.3060	2.6303
Tensor	1.5010	1.5260	1.5400	1.5900	2.1000	2.3340	2.6400

We fit the above model using the BPST method for each slice; see Figure S7 in the Supplementary Material for the set of triangulations used for the BPST method. The image maps in Figures 4, S8 and S9 in the Supplementary Material present the estimated coefficient functions using the BPST ($d = 5$, $r = 1$) method. To evaluate the predictive performance, Table 4 reports the 10-fold cross-validation (10-fold CV; part of images are left out as training sets) MSPE results for the BPST method, Kernel method in Zhu et al. (2014) and the tensor regression method in Li and Zhang (2017). Based on the table, one can see that the MSPEs of the BPST method are uniformly smaller than that of the Kernel method and the tensor regression method.

Next, we construct the 95% SCCs to check if the covariates are significant. The yellow and blue color on the "significance" map in Figure 4 indicate the regions that zero is below the lower SCC or above the upper SCC, respectively. From these estimated coefficient functions and 95% SCCs, one can assess the impact of covariates on the response images. Taking the 5th slice as an example, the main impact of "AD" on PET images is an increase of activity in the cerebellum compared with the normal individual. The cerebellum obtains information from the sensory systems, the spinal cord, and other parts of the brain and then regulates motor movements, resulting in

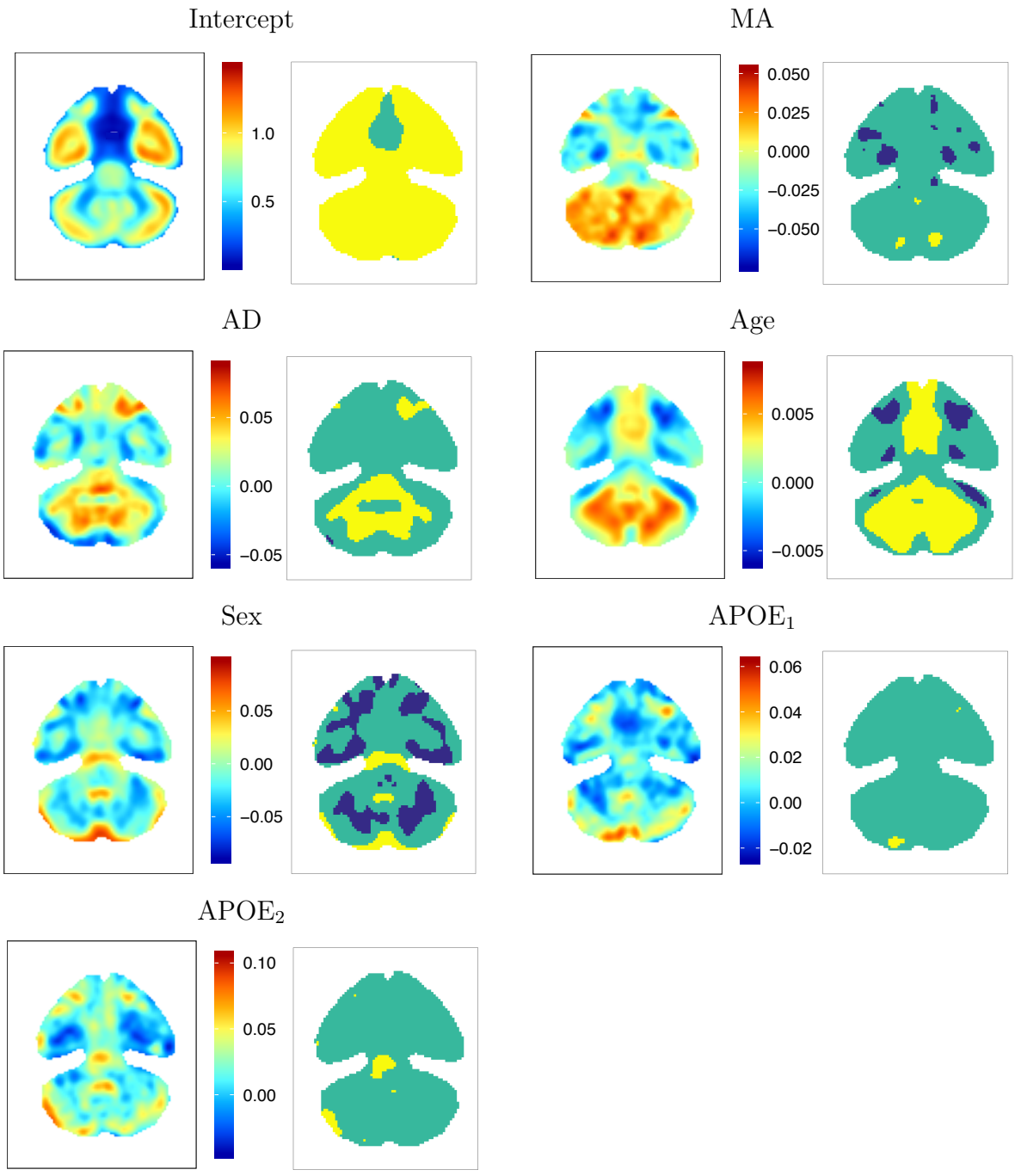

Figure 4: The BPST estimate and significance map of the coefficient functions for the 5th slice of the PET images. The yellow and blue color in the significance map indicates the regions that zero is below the lower SCC or above the upper SCC, respectively.

smooth and balanced muscular activities. The significance map of "Age" shows also an increase of activity in the cerebellum, and the "Sex" has a different impact on brain images between the female and male. The significance maps of the covariates for all other slices of PET image are shown in the Figures S10 – S11 in the Supplementary Materials. From these figures, one sees that the effect of the covariates on the brain activity level varies from one slice to another slice depending on the location of slice, and a more detailed description is given in the Supplementary Materials.

7. Conclusion

This article studies a class of image-on-scalar regression models to efficiently explore the spatial nonstationarity of a regression relationship between imaging responses and scalar predictors allowing regression coefficients to change with pixels. We have proposed an efficient estimation procedure to carry out statistical inference. We have developed a fast and accurate method for estimating the coefficient images, while consistently estimating their standard deviation images. Our method provides coefficient maps and significance maps that highlight and visualize the association of brain region and the potential risk factors adjusted for the other patient-level features and permits inference. In addition, it allows an easy implementation of piecewise polynomial representations of various degrees and various smoothness over an arbitrary triangulation, so it can handle irregular shaped 2D objects of different visual qualities. This provides enormous flexibility, accommodating various types of nonstationarity that are

commonly encountered in imaging data analysis. Our methodology is extendable to 3D images to fully realize its potential usefulness in biomedical imaging. Instead of using bivariate spline over triangulation, trivariate splines over tetrahedral partitions introduced in Lai and Schumaker (2007) could be well suited, and they have many properties in common with the bivariate splines over triangulation. However, it is a non-trivial task because computing is much more challenging for high-resolution 3D images compared to the 2D ones, which warrants further investigations.

Acknowledgement

The authors are truly grateful to the editor, the associate editor and two reviewers for their constructive comments and suggestions that led to significant improvement of the paper. Li Wang's research was partially supported by the National Science Foundation grants DMS-1542332 and DMS-1916204. Lijian Yang's research was supported in part by National Natural Science Foundation of China award 11771240. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`ADNI.loni.usc.edu`). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`.

Supplementary material

In the Supplementary Material, we provide the technical proofs for the main theorems, and additional results from simulation studies and ADNI data analysis.

References

Cardot, H., Ferraty, F., and Sarda, P. (1999), "Functional linear model," *Statistics & Probability Letters*, 45, 11–22.

— (2003), "Spline estimators for the functional linear model," *Statistica Sinica*, 13, 571–591.

Chen, K., Delicado, P., and Müller, H.-G. (2017), "Modelling function-valued stochastic processes, with applications to fertility dynamics," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 177–196.

Claeskens, G., Krivobokova, T., and Opsomer, J. D. (2009), "Asymptotic properties of penalized spline estimators," *Biometrika*, 96, 529–544.

Corder, E., Saunders, A., Strittmatter, W., Schmechel, D., Gaskell, P., Small, G., Roses, A., Haines, J., and Pericak-Vance, M. (1993), "Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families," *Science*, 261, 921–923.

Gu, L., Wang, L., Härdle, W. K., and Yang, L. (2014), "A simultaneous confidence corridor for varying coefficient regression with sparse functional data," *Test*, 23, 806–843.

Hall, P. and Horowitz, J. (2013), "A simple bootstrap method for constructing nonparametric confidence bands for functions," *The Annals of Statistics*, 41, 1892–1921.

Hall, P. and Horowitz, J. L. (2007), "Methodology and convergence rates for functional linear regression," *The Annals of Statistics*, 35, 70–91.

Hall, P. and Opsomer, J. D. (2005), "Theory for penalised spline regression," *Biometrika*, 92, 105–118.

Hibar, D. P., Stein, J. L., Renteria, M. E., Arias-Vasquez, A., Desrivières, S., Jahanshad, N., et al. (2015), "Common genetic variants influence human subcortical brain structures," *Nature*, 520, 224–229.

Huang, J. Z., Wu, C. O., and Zhou, L. (2004), "Polynomial spline estimation and inference for varying coefficient models with longitudinal data," *Statistica Sinica*, 14, 763–788.

Lai, M.-J. and Schumaker, L. L. (2007), *Spline functions on triangulations*, Cambridge University Press.

Lai, M. J. and Wang, L. (2013), "Bivariate penalized splines for regression." *Statistica Sinica*, 23, 1399–1417.

Li, L. and Zhang, X. (2017), "Parsimonious tensor response regression," *Journal of the American Statistical Association*, 112, 1131–1146.

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.

Lindgren, F., Rue, H., and Lindström, J. (2011), "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 423–498.

Marcus, C., Mena, E., and Subramaniam, R. M. (2014), "Brain PET in the diagnosis of Alzheimer's disease," *Clinical Nuclear Medicine*, 39, e413–e422.

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

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

Müller, H.-G. (2005), "Functional modelling and classification of longitudinal data," *Scandinavian Journal of Statistics*, 32, 223–240.

Persson, P. O. and Strang, G. (2004), "A simple mesh generator in MATLAB." *SIAM Review*, 46, 329–345.

Ramsay, J. O. and Dalzell, C. J. (1991), "Some tools for functional data analysis," *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, 539–572.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, New York, Springer.

Ramsay, T. (2002), "Spline smoothing over difficult regions." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 307–319.

Reiss, P. T., Goldsmith, J., Shang, H. L., and Ogden, R. T. (2017), "Methods for scalar-on-function regression," *International Statistical Review*, 85, 228–249.

Reiss, P. T., Huang, L., and Mennes, M. (2010), "Fast function-on-scalar regression with penalized basis expansions," *The International Journal of Biostatistics*, 6, Article 28.

Sang, H. and Huang, J. Z. (2012), "A full scale approximation of covariance functions for large spatial data sets," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74, 111–132.

Sangalli, L. M., Ramsay, J. O., and Ramsay, T. O. (2013), "Spatial spline regression models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75, 681–703.

Schwarz, K. and Krivobokova, T. (2016), "A unified framework for spline estimators," *Biometrika*, 103, 121–131.

Sentürk, D. and Müller, H.-G. (2010), "Functional varying coefficient models for longitudinal data," *Journal of the American Statistical Association*, 105, 1256–1264.

Staicu, A. M., Crainiceanu, C. M., and Carroll, R. J. (2010), "Fast methods for spatially correlated multilevel functional data," *Biostatistics*, 11, 177–194.

Stein, J. L., Hua, X., Lee, S., Ho, A. J., Leow, A. D., Toga, A. W., Saykin, A. J., Shen, L., Foroud, T., Pankratz, N., et al. (2010), "Voxelwise genome-wide association study (vGWAS)," *Neuroimage*, 53, 1160–1174.

Wang, G., Wang, L., and Lai, M. J. (2019), "Package 'BPST'," R package version 1.0. Available at "`https://github.com/funstatpackages/BPST`".

Wang, J.-L., Chiou, J.-M., and Muller, H.-G. (2016), "Functional data analysis," *Annal Review of Statistics and Its Applications*, 3, 257–295.

Worsley, K. J., Taylor, J. E., Tomaiuolo, F., and Lerch, J. (2004), "Unified univariate and multivariate random field theory," *NeuroImage*, 23, S189–S195.

Wu, S. and Müller, H.-G. (2011), "Response-Adaptive Regression for Longitudinal Data," *Biometrics*, 67, 852–860.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005), "Functional linear regression analysis for longitudinal data," *The Annals of Statistics*, 33, 2873–2903.

Yu, S., Wang, G., and Wang, L. (2019), "Package 'FDAimage'," R package version 1.0. Available at "`https://github.com/funstatpackages/FDAimage`".

Zhang, X. and Wang, J.-L. (2015), "Varying-coefficient additive models for functional data," *Biometrika*, 102, 15–32.

Zhou, H., Li, L., and Zhu, H. (2013), "Tensor regression with applications in neuroimaging data analysis," *Journal of the American Statistical Association*, 108, 540–552.

Zhu, H., Fan, J., and Kong, L. (2014), "Spatially varying coefficient model for neuroimaging data with jump discontinuities," *Journal of the American Statistical Association*, 109, 1084–1098.

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

Shan Yu

Department of Statistics and the Statistical Laboratory,

Iowa State University,

Ames, IA 50011

E-mail: (syu@iastate.edu)

Guannan Wang

Department of Mathematics,

College of William & Mary,

Williamsburg, VA 23187

E-mail: (gwang01@wm.edu)

Li Wang

Department of Statistics and the Statistical Laboratory,

Iowa State University,

Ames, IA 50011

E-mail: (lilywang@iastate.edu)

Lijian Yang

Center for Statistical Science and Department of Industrial Engineering,

Tsinghua University,

Beijing, 100084

E-mail: (yanglijian@tsinghua.edu.cn)