Single-index Varying Coefficient Model for Functional Responses

Xinchao Luo[1,2], **Lixing Zhu**[3], and **Hongtu Zhu**[2]

[1]School of Finance and Statistics, East China Normal University, Shanghai, China

[2]Department of Biostatistics and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A

[3]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

Summary

Recently, massive functional data have been widely collected over space across a set of grid points in various imaging studies. It is interesting to correlate functional data with various clinical variables, such as age and gender, in order to address scientific questions of interest. The aim of this paper is to develop a single-index varying coefficient (SIVC) model for establishing a varying association between functional responses (e.g., image) and a set of covariates. It enjoys several unique features of both varying-coefficient and single-index models. An estimation procedure is developed to estimate varying coefficient functions, the index function, and the covariance function of individual functions. The optimal integration of information across different grid points are systematically delineated and the asymptotic properties (e.g., consistency and convergence rate) of all estimators are examined. Simulation studies are conducted to assess the finite-sample performance of the proposed estimation procedure. Furthermore, our real data analysis of a white matter tract dataset obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study confirms the advantage and accuracy of SIVC model over the popular varying coefficient model.

Keywords

Functional response; Image analysis; Single index; Uniform convergence; Varying coefficient

1. Introduction

As a semiparametric regression modelling strategy, single-index modelling has attracted much attention in the literature due to its balance between exibility and fidelity. A classical single-index model is often written as

---

$$Y = g(X^T \beta) + \varepsilon, \quad (1)$$

where Y is a response variable, $g(\cdot)$ is an unknown index function, X is a covariate vector, and ε is an error term such that $E(\varepsilon|X) = 0$. See Horowitz (2009) for a comprehensive review of various estimation methods for single-index models and references therein (Li, 1991; Cook and Weisberg, 1991; Zhu et al., 2010; Xia et al., 2002; Xia, 2007; Ma and Zhu, 2012, 2013). For instance, dimension reduction approaches, such as likelihood-based methods (Cook and Forzani, 2009), are also commonly used for estimation. However, the existing literature primarily considers univariate response observed from cross-sectional studies.

This paper is motivated by the analysis of a real diffusion weighted imaging (DWI) data set with $n = 213$ subjects collected from NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) study. For each subject, we calculated a Fractional Anisotropy (FA) curve at all the 83 grid points along the skeleton of the midsagittal corpus callosum as shown in Figure 1. We are interested in establishing an association between FA curves and several covariates of interest, such as age and gender. To establish such association, standard grid-wise methods are to fit a linear model to functional observations at each grid point as responses and clinical variables, such as age and gender, as covariates, and to generate a statistical map of test statistics or p-values across all grid points (Lazar, 2008; Worsley et al., 2004). These grid-wise methods have several major limitations. First, compared with model (1), the classical linear model used in the neuroimaging literature is often restrictive, since it assumes that the index function $g(\cdot)$ is an identity function. When $g(\cdot)$ is truly nonlinear, directly fitting a classical linear model can cause substantial efficiency loss and reduce prediction accuracy. Second, since the grid-wise methods treat all grid points as independent units, they ignore two key functional features of functional data including spatial smoothness and spatial correlation.

Some advanced methods have been developed to specifically incorporate these features by using function-on-scalar regression under the functional data analysis (FDA) framework (Zhu et al., 2012; Ramsay and Silverman, 2005; Staicu et al., 2010; Zhang and Chen, 2007; Reiss et al., 2010). Some important estimation methods for FDA include adaptive smoothing methods (Polzehl and Spokoiny, 2006; Li et al., 2011), the integration of FDA and adaptive smoothing methods (Zhu et al., 2014), and spatial priors within the Bayesian framework (Gossl et al., 2001; Penny et al., 2005; Bowman et al., 2008; Smith and Fahrmeir, 2007; Yue et al., 2010; Miranda et al., 2013), among others. See Morris (2015) and Wang et al. (2015) for a comprehensive review of various FDA models for functional responses. However, to the best of our knowledge, none of the references cited above address the two functional features and estimate the nonparametric index function simultaneously.

The aim of this paper is to develop a single-index varying coefficient (SIVC) model to establish a varying association between functional responses and a set of covariates. Specifically, we extend the single-index model (1) to SIVC for functional responses as follows:

$$Y(s) = g(X^T \beta(s)) + \varepsilon(s), \quad (2)$$

where $\{ Y(s) : s \in \mathscr{S} \}$ is an observed stochastic process on a compact set $\mathscr{S}$ and $\varepsilon(s)$ is a random function characterizing the within-subject correlations and measurement errors at different grid points such that $E\{\varepsilon(s)|X\} = 0$ for all $s \in \mathscr{S}$. The $\beta(s)$ allows us to characterize the dynamic association between covariate X and functional response, whereas $g(\cdot)$ is a nonparametric function, while being fixed across all $s \in \mathscr{S}$. Model (2) differs from the functional single index model in Jiang and Wang (2011), in which $g(\cdot)$ varies across s, but $\beta(s)$ is assumed to be stationary. When $g(x) = x$, model (2) reduces to standard voxel-wise methods based on linear model and the most popular FDA model considered in (Zhang and Chen, 2007; Ramsay and Silverman, 2005; Zhu et al., 2014). For notational simplicity, we set $\mathscr{S} = [0, 1]$. The results can be readily extended to more general cases with compact subset $\mathscr{S}$ of the Euclidean space.

Compared with the existing literature, we make several unique contributions. **(i)** We develop a new estimation procedure to estimate various parametric and non-parametric components of SIVC. **(ii)** Theoretically, we delineate the integration of information across all grid points by using an optimal weight function and then establish the asymptotic properties of various estimates for SIVC. **(iii)** Our analysis of the ADNI data confirms the advantage and accuracy of SIVC model over the popular varying coefficient model (Zhang and Chen, 2007; Ramsay and Silverman, 2005; Zhu et al., 2014).

The rest of this paper is organized as follows. Section 2 introduces the estimation procedure to estimate varying coefficient functions, index function, and the covariance function of individual functions. Section 3 systematically investigates the asymptotic properties of all estimators. A simulation study and a real data analysis of Alzheimer's Disease Neuroimaging Initiative (ADNI) study are presented in Section 4 to demonstrate the finite sample performance of SIVC. Section 5 concludes with some discussions.

2. Methods

2.1 Single-index Varying Coefficient Model

We formally introduce the single-index varying coefficient model as follows. Consider $\{(\{ Y_i(s) : s \in \mathscr{S} \}, X_i) : i = 1, \ldots, n\}$ from n independent subjects, where $\mathscr{S}$ is a compact set that characterizes the range of all possible grid points. Our single-index varying coefficient (SIVC) model is written as

$$Y_i(s) = g(X_i^T \beta(s)) + \eta_i(s) + \varepsilon_i(s) \text{ for all } s \in \mathscr{S} \text{ and } i = 1, \ldots, n, \quad (3)$$

where X_i is a $p \times 1$ covariate vector, $\beta(s) = (\beta_1(s), \cdots, \beta_p(s))^T$ is a $p \times 1$ vector of varying coefficient functions, $g(\cdot)$ is an unknown index function, $\eta_i(s)$ characterizes individual curve variations, and $\varepsilon_i(s)$ is a random function of measurement errors. The process $\{\eta_i(s) : s \in \mathscr{S}\}$ is assumed to be a Gaussian process with zero mean and a covariance function $\Sigma_\eta(s, t) =$

Cov$\{\eta(s), \eta(t)\}$. The error terms $\varepsilon_i(s)$ follow a Gaussian process with zero mean and a diagonal covariance function Cov$\{\varepsilon(s), \varepsilon(t)\}$. That is, $\varepsilon_i(s)$ and $\varepsilon_i(t)$ are assumed to be independent for $s \neq t$, and Cov$\{\varepsilon(s), \varepsilon(t)\}$ takes the form of $\sigma_\varepsilon^2(s)\mathbf{1}(s=t)$, where $\mathbf{1}(\cdot)$ is an indicator function. Moreover, $Y_i(s)$ are usually measured at the same set of locations for all subjects and exhibit both the within-curve and between-curve dependence structures. Thus, without loss of generality, it is assumed that $Y_i(s)$ are observed on M grid points $\mathscr{S}_M = \{s_m : 0 = s_1 \leq \cdots \leq s_M = 1\}$ for all subjects.

For single index model, $\beta(s)$ is not identifiable since $g(X^T\beta(s))$ and $g(\beta_0(s) + \delta X^T\beta(s))$ are not distinguishable, where δ is any nonzero scalar. A simple solution in the literature (Zhu et al., 2010; Xia et al., 2002) is to impose some constraints on $\beta_0(s)$ and $\beta(s)$, such as $\beta_0(s) = 0$ and $\beta(s)^T\beta(s) = 1$. Therefore, throughout the paper, it is assumed that X does not contain the intercept, $\beta(s)^T\beta(s) = 1$ holds for all $s \in \mathscr{S}$ and the first entry of $\beta(s)$ is positive at each s.

2.2 Estimation Procedure

Our estimation procedure consists of three steps for estimating the varying coefficient functions $\beta(\cdot)$, the index function $g(\cdot)$, and the covariance function $\Sigma_\eta(s, s')$.

Step I. Estimating varying coefficient $\beta(s)$—We first consider the estimation of $\beta(s_m)$ at each given grid point s_m. Model (3) reduces to a classical single-index model given by

$$Y_i(s_m) = g(X_i^T\beta(s_m)) + \varepsilon_i^*(s_m), \quad (4)$$

where $\varepsilon_i^*(s_m) = \eta_i(s_m) + \varepsilon_i(s_m)$ such that $E\{\varepsilon_i^*(s_m)|X_i\} = 0$ and $\mathrm{Var}\{\varepsilon^*(s_m)|X_i\} = \sigma_{\varepsilon*}^2(s_m) = \Sigma_\eta(s_m, s_m) + \sigma_\varepsilon^2(s_m)$. The likelihood function of a random observation $(X, Y(s_m))$ in model (4) is given by

$$f_1(X) \times f_2(Y(s_m) - g(X^T\beta(s_m)); X),$$

where f_1 is the probability density function of X, and f_2 is the conditional probability density function of $\varepsilon^*(s_m) = Y(s_m) - g(X^T\beta(s_m))$ given X.

Based on the Gaussian assumption of $\varepsilon_i^*(s)$, $f_2(\cdot)$ is the normal density with mean zero and variance $\sigma_{\varepsilon*}^2(s_m)$. Thus, the score function for $\beta(s_m)$, denoted as $S(\beta(s_m))$, is given by

$$\begin{aligned}
S(\beta(s_m)) &= -X\varepsilon^*(s_m)\dot{g}(X^T\beta(s_m))/\sigma_{\varepsilon*}^2(s_m) = (\mathrm{I}) + (\mathrm{II})\\
&= -E\{X|X^T\beta(s_m)\}\varepsilon^*(s_m)\dot{g}(X^T\beta(s_m))/\sigma_{\varepsilon*}^2(s_m)\\
&\quad - [X - E\{X|X^T\beta(s_m)\}]\varepsilon^*(s_m)\dot{g}(X^T\beta(s_m))/\sigma_{\varepsilon*}^2(s_m), \quad (5)
\end{aligned}$$

where $\dot{g}(t) = dg(t)/dt$. Following the reasoning in Ma and Zhu (2014), (I) and (II) belong to the tangent space of model (4) with respect to $\beta(s_m)$, denoted by $\Lambda_g(s_m)$, and its orthogonal

component, denoted by $\Lambda_g(s_m)^\perp$, respectively. For any $s \in \mathscr{S}$, $\Lambda_g(s)$ and $\Lambda_g(s)^\perp$ are, respectively, given by

$$\Lambda_g(s) = \{-\varepsilon^*(s)h(X^T\beta(s))/\sigma^2_{\varepsilon*}(s) : \forall h(\cdot)\},$$

$$\Lambda_g(s)^\perp = \{-\varepsilon^*(s)[\alpha(X) - E\{\alpha(X)|X^T\beta(s)\}] : \forall \alpha(\cdot)\}.$$

Therefore, the efficient score function for $\beta(s_m)$ is given by

$$S_{\mathrm{eff}}(\beta(s_m); X, Y(s_m)) = \varepsilon^*(s_m)[X - E\{X|X^T\beta(s_m)\}]\dot{g}(X^T\beta(s_m)). \quad (6)$$

To calculate an efficient estimator of $\beta(s_m)$, denoted as $\tilde{\beta}(s_m)$, we can solve

$$\sum_{i=1}^{n} S_{\mathrm{eff}}(\tilde{\beta}(s_m); X_i, Y_i(s_m)) = \mathbf{0}. \quad (7)$$

Since S_{eff} depends on three unknown quantities $E\{X|X^T\beta(s_m)\}$, $g(X^T\beta(s_m))$ and $\dot{g}(X^T\beta(s_m))$, we construct their nonparametric estimators as follows (Ma and Zhu, 2013). The Nadaraya-Watson kernel estimator of $E\{X|X^T\beta(s_m)\}$ is given by

$$\hat{E}\{X|X^T\beta(s_m)\} = \frac{\sum_{i=1}^{n} X_i K_{h_x}(X_i^T\beta(s_m) - X^T\beta(s_m))}{\sum_{i=1}^{n} K_{h_x}(X_i^T\beta(s_m) - X^T\beta(s_m))}, \quad (8)$$

where $K_h(\cdot) = K(\cdot/h)/h$ is a kernel function and h_x is a given bandwidth. We can calculate the estimates of $g(X^T\beta(s_m))$ and $\dot{g}(X^T\beta(s_m))$ at $X^T\beta(s_m) = X^T\beta_0(s_m)$, denoted by $\hat{g}(X^T\beta_0(s_m))$ and $\hat{\dot{g}}(X^T\beta_0(s_m))$, by minimizing

$$\sum_{i=1}^{n} [Y_i(s_m) - g(X^T\beta_0(s_m)) - \dot{g}(X^T\beta_0(s_m))(X_i^T - X^T)\beta_0(s_m)]^2 K_{h_y}((X_i^T - X^T)\beta_0(s_m)).$$

$$(9)$$

As suggested by Ma and Zhu (2014), we set $h_x = cn^{-1/3}$ and $h_y = cn^{-1/5}$, where c is the average standard deviation of X. By plugging $\hat{E}\{X|X^T\beta(s_m)\}$, $\hat{g}(X^T\beta_0(s_m))$ and $\hat{\dot{g}}(X^T\beta_0(s_m))$ into (7), we get an estimate of $S_{\mathrm{eff}}(\beta(s_m); X_i, Y_i(s_m))$, denoted by $\hat{S}_{\mathrm{eff}}(\beta(s_m); X_i, Y_i(s_m))$ and then calculate $\tilde{\beta}(s_m)$ by solving

$$\sum_{i=1}^{n} \hat{S}_{\text{eff}}(\tilde{\beta}(s_m); X_i, Y_i(s_m)) = \mathbf{0}. \tag{10}$$

To estimate $\beta(s)$ at any $s \in \mathscr{S}$, we need to construct a weighted estimating equation to borrow information across all grid points by using the functional features of imaging data (Wang et al., 2004; Zhu et al., 2012; Polzehl and Spokoiny, 2006; Li et al., 2011). Specifically, the weighted estimating equation is given by

$$\hat{S}_{nM}(\beta(s); w) = \sum_{i=1}^{n} \sum_{m=1}^{M} w(s_m, s) \hat{S}_{\text{eff}}(\beta(s); X_i, Y_i(s_m)), \tag{11}$$

where $w(s_m, s)$ is a weight function of (s_m, s) and may depend a few parameters, such as bandwidth. Then, we calculate an estimate of $\beta(s)$, denoted by $\hat{\beta}(s)$, by solving the following equation:

$$\hat{S}_{nM}(\hat{\beta}(s); w) = \mathbf{0}. \tag{12}$$

A critical issue in (12) is how to select an optimal weight function $w(s_m, s)$. Theoretically, one may choose $w(s_m, s)$ by minimizing the mean integrated squares error (MISE) of $\hat{\beta}(\cdot)$ for all $s \in \mathscr{S}$, but it can be challenging due to the lack of precise information about $\beta(s)$. Without such information, a simple solution is to set $w(s_m, s) = K((s_m - s)/h)/h$ and then select the bandwidth h by using some criteria, such as cross-validation method, based on MISE. Furthermore, when $\beta(s)$ is a piecewise constant function, we will show how to optimally select $w(s_m, s)$ in Section 3.

Step II. Estimating the unknown index function $g(\cdot)$—We use the local linear technique to estimate $g(X^T\beta(s))$. We define $Z_{i,m}(s) = (1, \{X_i^T\beta(s_m) - X^T\beta(s)\}/h_1)^T$ and

$$\Sigma(X^T\beta(s), h_1) = \sum_{i=1}^{n} \sum_{m=1}^{M} K_{h_1}(X_i^T\beta(s_m) - X^T\beta(s)) Z_{i,m}(s)^{\otimes 2},$$

where $Z^{\otimes 2} = ZZ^T$ for any vector Z. By replacing $\beta(s_m)$ by $\hat{\beta}(s_m)$, we get $\hat{Z}_{i,m}(s)$ and $\hat{\Sigma}(X^T\beta(s), h_1)$. Denote $G(X^T\beta(s)) = (g(X^T\beta(s)), h_1\dot{g}(X^T\beta(s)))^T$, we directly minimize a weighted least square function given by

$$\sum_{i=1}^{n} \sum_{m=1}^{M} \{Y_i(s_m) - \hat{Z}_{i,m}(s)^T G(X^T\beta(s))\}^2 K_{h_1}(X_i^T\hat{\beta}(s_m) - X^T\beta(s)). \tag{13}$$

Thus, we have $\hat{g}(X^T\beta(s)) = [1\ 0]\hat{G}(X^T\beta(s))$, where $\hat{G}(X^T\beta(s))$ is given by

$$\hat{\Sigma}(X^T\beta(s), h_1)^{-1} \sum_{i=1}^{n} \sum_{m=1}^{M} K_{h_1}(X_i^T\hat{\beta}(s_m) - X^T\beta(s))\hat{Z}_{i,m}(s)Y_i(s_m).$$

The bandwidth h_1 is chosen by using the cross-validation method.

Step III. Estimating the covariance function $\Sigma_\eta(s, t)$—Let $d_i(s) = (\eta_i(s), h_2\dot{\eta}_i(s))^T$, $W_{m,s} = (1, (s_m-s)/h_2)^T$, and $Y_i^*(s_m)=Y_i(s_m) - \hat{g}(X_i^T\hat{\beta}(s_m))$. We minimize the following function

$$\sum_{m=1}^{M} \{Y_i^*(s_m) - W_{m,s}^T d_i(s)\}^2 K_{h_2}(s_m - s) \tag{14}$$

to obtain

$$\hat{d}_i(s) = \left(\sum_{m=1}^{M} K_{h_2}(s_m - s)W_{m,s}^{\otimes 2}\right)^{-1} \sum_{m=1}^{M} K_{h_2}(s_m - s)W_{m,s}Y_i^*(s_m).$$

Then, $\eta(s)$ can be estimated by

$$\hat{\eta}_i(s) = [1\,0]\hat{d}_i(s) = \sum_{m=1}^{M} \tilde{K}_{h_2}(s_m - s)Y_i^*(s_m), \tag{15}$$

where $\tilde{K}_{h_2}(s_m - s) = [1\,0](\sum_{m=1}^{M} K_{h_2}(s_m - s)W_{m,s}^{\otimes 2})^{-1} K_{h_2}(s_m - s)W_{m,s}$, which is the empirical equivalent kernel. The bandwidth h_2 is also chosen by using the cross-validation method.

We consider the spectral decomposition of $\Sigma_\eta(s, t)$ and its approximation. Suppose $\Sigma_\eta(s, t)$ is continuous on $\mathscr{S}^2$, then according to Mercer's theorem, it can be decomposed as

$$\Sigma_\eta(s, t) = \sum_{k=1}^{\infty} \lambda_k \psi_k(s)\psi_k(t),$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are ordered eigenvalues and $\psi_k(s)$ are the corresponding orthonormal eigenfunctions. Furthermore, the eigenfunctions form an orthonormal system on the space of square-intergrable function on $\mathscr{S}$, and $\eta_i(s)$ admits the Karhunen-Loeve expansion as $\eta_i(s) = \sum_{k=1}^{\infty} \xi_{ik}\psi_k(s)$, where $\xi_{ik} = \int_0^1 \eta_i(s)\psi_k(s)ds$ is the k-th functional principal component scores of the i-th subject. For a fixed i, ξ_{ik} are uncorrelated random variables with mean zero and variance λ_k.

By following Rice and Silverman (1991), the covariance matrix $\Sigma_\eta(s,\,t)$ and its spectral decomposition can be estimated by

$$\hat{\Sigma}_\eta(s,t)=(n-p)^{-1}\sum_{i=1}^{n}\hat{\eta}_i(s)\hat{\eta}_i(t)=\sum_{k=1}^{\infty}\hat{\lambda}_k\hat{\psi}_k(s)\hat{\psi}_k(t), \qquad (16)$$

where $\hat{\lambda}_1 \quad \hat{\lambda}_2 \quad \cdots \quad 0$ are estimated eigenvalues and $\hat{\psi}_k(s)$ are the corresponding estimated eigenfunctions. Moreover, the k-th functional principal component scores can be computed by $\hat{\xi}_{ik}=\sum_{m=1}^{M}\hat{\eta}_i(s_m)\hat{\psi}_k(s_m)(s_m-s_{m-1})$ for $i=1,\ldots,n$.

2.3 Simultaneous Confidence Bands

Given a Confidence level α, we construct a simultaneous Confidence band for each $\beta_l(s)$ such that $P(\hat{\beta}_l^{L,\alpha}(s)<\beta_l(s)<\hat{\beta}_l^{U,\alpha}(s) \text{ for all } s\in\mathscr{S})=1-\alpha$ where $\hat{\beta}_l^{L,\alpha}(s)$ and $\hat{\beta}_l^{U,\alpha}(s)$ are the lower and upper limits of simultaneous Confidence band, respectively. Specifically, we set

$$\hat{\beta}_l^{L,\alpha}(s)=\hat{\beta}_l(s)-\text{bias}(\hat{\beta}_l(s))-C_{\beta_l}(\alpha) \text{ and } \hat{\beta}_l^{U,\alpha}(s)=\hat{\beta}_l(s)-\text{bias}(\hat{\beta}_l(s))+C_{\beta_l}(\alpha), \qquad (17)$$

where $\text{bias}(\hat{\beta}_l(s))$ is the bias of $\hat{\beta}_l(s)$ at $s\in\mathscr{S}$ and $C_{\beta_l}(\alpha)$ is a scalar. By following the arguments in Zhu et al. (2012), we use the local polynomial technique to estimate $\text{bias}(\hat{\beta}_l(s))$ for each l and then approximate $C_{\beta_l}(\alpha)$ by using the wild bootstrap as follows:

- Step 1: We calculate $\hat{r}_i(s_m)=Y_i(s_m)-\hat{g}(X_i^T\hat{\beta}(s_m))$ for all i and m.

- Step 2: For $q=1,\ldots,Q$, we independently generate $\{\tau_i^{(q)}: i=1,\cdots,n\}$ from $N(0, 1)$ and construct $\hat{Y}_i(s_m)^{(q)}=\hat{g}(X_i^T\hat{\beta}(s_m))+\tau_i^{(q)}\hat{r}_i(s_m)$. Then, based on $\{Y_i(s_m)^{(q)}\}$, we recalculate $\hat{\beta}(s)^{(q)}$, and obtain a stochastic process $G_{\beta_l}(s)^{(q)}=|\hat{\beta}_l(s)(q)-\hat{\beta}_l(s)^{(q)}|$ for each l.

- Step 3: For all q, we calculate the $1-\alpha$ empirical percentile of $G_{\beta_l}(s)^{(q)}$, denoted by $C_{\beta_l}(s,\alpha)$, at each $s\in\mathscr{S}$. Finally, an estimate of $C_{\beta_l}(\alpha)$ is $\sup_s |C_{\beta_l}(s,\alpha)|$.

Similarly, for a given α, we construct a simultaneous Confidence band for $g(\cdot)$ as follows:

$$P(\hat{g}^{L,\alpha}(u)<g(u)<\hat{g}^{U,\alpha}(u) \text{ for all } u\in\mathscr{U})=1-\alpha,$$

where $\mathscr{U}$ is a compact set in R and $\hat{g}^{L,\alpha}(u)$ and $\hat{g}^{U,\alpha}(u)$ are the lower and upper limits of simultaneous Confidence band, respectively. Specifically, we set

$$\hat{g}^{L,\alpha}(u)=\hat{g}(u)-\text{bias}(\hat{g}(u))-C_g(\alpha) \text{ and } \hat{g}^{U,\alpha}(u)=\hat{g}(u)-\text{bias}(\hat{g}(u))+C_g(\alpha). \qquad (18)$$

Subsequently, we use the local polynomial technique to estimate $\mathrm{bias}(\hat{g}(u))$ and then use the wild bootstrap to approximate $C_g(\alpha)$ as follows:

- Step 1: We calculate $\hat{r}_i(s_m) = Y_i(s_m) - \hat{g}(X_i^T \hat{\beta}(s_m))$ for all i and m.

- Step 2: For $q = 1, \ldots, Q$, we independently generate $\{\tau_i^{(q)} : i = 1, \cdots, n\}$ from $N(0, 1)$ and construct $\hat{Y}_i(s_m)^{(q)} = \hat{g}(X_i^T \hat{\beta}(s_m)) + \tau_i^{(q)} \hat{r}_i(s_m)$. Then, based on $\{Y_i(s_m)^{(q)}\}$, we recalculate $\hat{g}(\cdot)^{(q)}$, and obtain a stochastic process $G_g(u)^{(q)} = |\hat{g}(u) - \hat{g}(u)^{(q)}|$.

- Step 3: For all ℓ, we calculate the $1 - \alpha$ empirical percentile of $G_g(u)^{(q)}$ denoted by $C_g(u, \alpha)$ at every $u \in \mathscr{U}$ and then approximate $C_g(\alpha)$ by using $\sup_{u \in \mathscr{U}} |C_g(u, \alpha)|$.

3. Theoretical Results

3.1 Optimal weight Functions

We consider a challenging issue of optimally selecting the weight function $w(s_m, s)$ in order to gain efficiency, since $\hat{S}_{nM}(\beta(s); w)$ for a given weight function $w(\cdot, \cdot)$ may not be an efficient estimating equation. Specifically, $\Lambda_g(s)$ and $\Lambda_g(s')$ may interact with each other for $s \neq s'$ by noting that

$$E[\varepsilon^*(s)h(X^T\beta(s))\varepsilon^*(s')h'(X^T\beta(s'))/\{\sigma_{\varepsilon*}^2(s)\sigma_{\varepsilon*}^2(s')\}] = \Sigma_\eta(s,s')E\{h(X^T\beta(s))h'(X^T\beta(s'))\}/\{\sigma_{\varepsilon*}^2(s)\sigma_{\varepsilon*}^2(s')\}.$$

$$(19)$$

Therefore, $\Lambda_g(s)$ and $\Lambda_g(s')$ are orthogonal with each other for $s \neq s'$ if and only if $\Sigma_\eta(s, s') = 0$ for $s \neq s'$. It also holds for $\{\Lambda_g(s)^\perp : s \in \mathscr{S}\}$ due to $\Lambda_g(s) \perp \Lambda_g(s)^\perp$.

First, we consider how to choose the weight function when $\beta(s) = \beta_0$ does not vary across $s \in \mathscr{S}$. With some calculations, we can show that the covariance matrix of $\sqrt{n}\hat{\beta}(s)$ can be approximated by

$$\Sigma_1(\mathbf{w}(s)) = D(\mathbf{w}(s))E\{[X - E\{X|X^T\beta_0\}]^{\otimes 2}\hat{g}(X^T\beta_0)^2\}^{-1}, \quad (20)$$

where $\mathbf{w}(s) = (w(s_1, s), \cdots, w(s_M, s))^T$ and $D(\mathbf{w}(s))$ is given by

$$\frac{\sum_{m,m'=1}^M w(s_m, s)w(s_{m'}, s)\{\Sigma_\eta(s_m, s_{m'}) + \sigma_\varepsilon^2(s_m)\mathbf{1}(s_m = s_{m'})\}}{\{\sum_{m=1}^M w(s_m, s)\}^2}. \quad (21)$$

We can obtain an optimal weight vector, denoted by $\mathbf{w}_*$, by minimizing $\Sigma_1(\mathbf{w}(s))$ such that $\mathbf{w}_* = \mathrm{argmin}_{\mathbf{w}(s)}\Sigma_1(\mathbf{w}(s))$. As shown in Theorem 1 (i) below, $\mathbf{w}_*$ is associated with the

eigenvalue-eigenvector pairs of $\Lambda_{\varepsilon*,M}^{-1/2}\Sigma_{\eta*,M}\Lambda_{\varepsilon*,M}^{-1/2}$, denoted by $\{(\lambda_{m*,M}, \psi_{m*,M}) : m = 1, \ldots, M\}$, where $\Sigma_{\eta*,M} = (\Sigma_\eta(s_m, s_m'))$ and $\Lambda_{\varepsilon*,M} = (\sigma_\varepsilon^2(s_m)\mathbf{1}(s_m = s_{m'}))$ are two $M \times M$ matrices.

Second, we set $\omega(s_m, s)$ as $K_h(s_m - s)$, which is a kernel function of $(s_m - s)/h$, when $\beta(s)$ may vary across $s \in \mathscr{S}$. If $h \to 0$, then it can be shown that the covariance matrix of $\sqrt{n}\hat\beta(s)$ can be approximated by $\Sigma_1(\mathbf{w}(s))$ in (20). We will show in Theorem 1 (ii) that the use of the kernel function can lead to substantial efficiency gain even under this general scenario.

Theorem 1—*We have the following results.*

    i.   *Suppose that $\beta(s) = \beta_0$ does not vary across $s \in \mathscr{S}$. The optimal $\mathbf{w}_*$ is given by*

$$\mathbf{w}_* = \sum\nolimits_{\varepsilon*,M}^{-1} \mathbf{1}_M / \left\| \sum\nolimits_{\varepsilon*,M}^{-1} \mathbf{1}_M \right\|_2, \quad (22)$$

        where $\|\cdot\|_2$ is the Euclidean norm of a vector, $\Sigma_{e*,M} = \Sigma_{\eta*,M} + \Lambda_{e*,M}$ is an $M \times M$ matrix, and $\mathbf{1}_M$ is an $M \times 1$ vector of ones. Thus, the optimal $D(\mathbf{w})$ is given by $D(\mathbf{w}_*) = (\mathbf{1}_M^T \sum\nolimits_{\varepsilon*,M}^{-1} \mathbf{1}_M)^{-1}$ and is independent of s. The $\Sigma_{\varepsilon*,M}^{-1}\mathbf{1}_M$ can be written as

$$\{\mathbf{I}_M - \Lambda_{\varepsilon*,M}^{-1/2} \sum_{m=1}^M \frac{\lambda_{m*,M}}{1+\lambda_{m*,M}} \psi_{m*,M}\psi_{m*,M}^T \Lambda_{\varepsilon*,M}^{-1/2}\}\Lambda_{\varepsilon*,M}^{-1/2}\mathbf{1}_M. \quad (23)$$

    ii.   *Suppose that $\beta(s)$ may vary across $s \in \mathscr{S}$. Under Assumptions (C6) and (C7), if $w(s_m, s) = K_h(s_m - s)$, $h \to 0$, and $Mh \to \infty$, then $D(\mathbf{w}(s))$ can be approximated by $\Sigma_\eta(s, s)$.*

Theorem 1 has several interesting implications. If $\Sigma_\eta(s, s') = 0$ for any $s \neq s'$, then $\mathbf{w}_*$ is proportional to $\Lambda_{\varepsilon*,M}^{-1}\mathbf{1}_M = (\sigma_\varepsilon^{-2}(s_1), \cdots, \sigma_\varepsilon^{-2}(s_M))^T$. In this case, we can set $w(s_m, s) = \sigma_\varepsilon^{-2}(s_m)$ for all m, and then the optimal weighted estimating equation is given by

$$\hat S_{nM}(\beta(s); \mathbf{w}_*) = \sum_{m=1}^M \sum_{i=1}^n \varepsilon_i^*(s_m)[X_i - E\{X_i | X_i^T\beta(s_m)\}]\dot g(X_i^T\beta(s_m))/\sigma_{\varepsilon*}^2(s_m).$$

Theorem 1 (i) also implies that in general cases, $w(s_m, s)$ is given by

$$\sigma_\varepsilon^{-2}(s_m) - \sigma_\varepsilon(s_m)^{-1} \sum_{m'=1}^M \frac{\lambda_{m'*,M}}{1+\lambda_{m'*,M}} \mathbf{e}_m^T \psi_{m'*,M}\psi_{m'*,M}^T \Lambda_{\varepsilon*,M}^{-1/2}\mathbf{1}_M,$$

where $\mathbf{e}_m$ is an $M \times 1$ vector with the m-th element one and zero otherwise. For functional data, it is common to assume that $\lambda_{m*,M} = 0$ for all $m > K$, where K is a positive integer. Under some additional conditions, it can be shown that $\lambda_{m*,M}$ and $\psi_{m*,M}$ converge to the m–th eigenvalue and its corresponding eigenfunction of the covariance function $\Sigma_\eta(s, s')/\{\sigma_\varepsilon(s)\sigma_\varepsilon(s')\}$, respectively.

Another implication of Theorem 1 (i) is the lower bound of $D(\mathbf{w}_*)$. Specifically, $D(\mathbf{w}_*)$ is given by

$$(\mathbf{1}_M^T \Lambda_{\varepsilon*,M}^{-1} \mathbf{1}_M - \mathbf{1}_M^T \Lambda_{\varepsilon*,M}^{-1/2} \sum_{m=1}^{M} \frac{\lambda_{m*,M}}{1+\lambda_{m*,M}} \psi_{m*,M} \psi_{m*,M}^T \Lambda_{\varepsilon*,M}^{-1/2} \mathbf{1}_M)^{-1}, \tag{24}$$

which is greater than $(\mathbf{1}_M^T \Lambda_{\varepsilon*,M}^{-1} \mathbf{1}_M)^{-1} = 1/\{\sum_{m=1}^{M} \sigma_\varepsilon^{-2}(s_m)\}$. In general, the presence of spatial correlation increases the covariance matrix of $\hat{\beta}(s)$. Such lower bound is achievable only when $\Sigma_\eta(s, s') = 0$ holds for any $s \quad s'$. Furthermore, such lower bound is asymptotically achievable when $\lambda_{1*,M} = o_p(1)$, since $x/(1 + x)$ is a monotone function of x.

Theorem 1 (ii) implies that the use of $w(s_m, s) = K_h(s_m - s)$ can lead to substantial efficiency gain if there are substantial measurement errors. Specifically, if $\mathbf{w}(s_m) = \mathbf{e}_m$, then we consider information at the m–th grid point. In this case, we have $D(\mathbf{e}(s_m)) = \Sigma_\eta(s_m, s_m) + \sigma_\varepsilon^2(s_m)$, which can be much larger than $\Sigma_\eta(s_m, s_m)$ when $\sigma_\varepsilon^2(m)$ is much larger than $\Sigma_\eta(s_m, s_m)$. The ratio of $\Sigma_1(\mathbf{e}(s_m))$ over $\Sigma_1((K_h(s_m-s)))$ is equal to $1 + \sigma_\varepsilon^2(s_m)/\Sigma_\eta(s_m, s_m)$. Therefore, the efficiency gain of using $w(s_m, s) = K_h(s_m - s)$ can be substantial if the value of $\sigma_\varepsilon^2(s_m)/\Sigma_\eta(s_m, s_m)$ is large.

3.2 Asymptotic Properties

Second, we investigate the asymptotic properties of $\hat{\beta}(\cdot)$, $\hat{g}(X^T \hat{\beta}(s))$ and $\hat{\Sigma}_\eta(s, t)$ when we set $w(s_m, s) = K_h(s_m - s)$. For any smooth function $f(s)$ and $g(s, t)$, define $\dot{f}(s) = df(s)/ds$, $\ddot{f}(s) = d^2 f(s)/ds^2$ and $g^{(a,b)}(s, t) = {}^{a+b}g(s, t)/ s^a t^b$, where a and b are any nonnegative integers. We state the following theorems regarding the weak convergence of $\{\hat{\beta}(s) : s \in \mathscr{S}\}$ and $\hat{g}(X^T \hat{\beta}(s))$, whose detailed assumptions and proofs can be found in Web Appendix. Moreover, we also include additional asymptotic properties and their proofs in the same Web Appendix.

Theorem 2—*Under Assumptions (C1)–(C11), as n, $M \to \infty$, we have the following results.*

    i.

$$\sqrt{n}(\hat{\beta}(s) - \beta_0(s) - 0.5h^2 A_n(s)^{-1}\{\frac{1}{n}\sum_{i=1}^{n} B_i(s)[\ddot{g}(X_i^T \beta(s)) + 2\dot{G}(X_i^T \beta(s))\dot{\pi}(s)/\pi(s)]\mu_2(K)\})$$

converges weakly to a Gaussian process with mean zero and covariance function, which is the limiting function of

$$A_n(s)^{-1}[n^{-1}\sum_{i=1}^{n} B_i(s)\Sigma_\eta(s,t)B_i(t)^T]A_n(t)^{-1}, \text{ where } A_n(\cdot) \text{ and } B_i(\cdot)$$
are defined in Web Appendix.

ii. $\sqrt{n}[\hat{g}(X^T\hat{\beta}(s)) - g(X^T\beta(s)) - 0.5h_1^2\mu_2(K)\ddot{g}(K^T\beta(s))]$ *converges weakly to a Gaussian process with mean zero and covariance function* $\Sigma_\eta(s, t)$.

Theorem 2 ensures that we can make formal statistical inference on $\beta(\cdot)$ and $g(\cdot)$. Based on Theorem 2 (i) and (ii), we develop a wild bootstrap method to construct the Confidence bands of $\hat{\beta}(\cdot)$ and $\hat{g}(\cdot)$ and include it in Section B of Web Appendix.

4. Numerical Studies

4.1 Simulation Results

We generated $Y(s_m)$ according to model (3) with $\varepsilon(s_m) \sim N(0, \sigma^2 = 0.3^2)$ and $X_i \sim N(\mathbf{0}_p, \Sigma)$, where Σ is a 4×4 matrix with elements $\rho^{|j'-j|}$ for $j, j' = 1, \dots, 4$. Moreover, the varying coefficients $\beta_j(s)$ are, respectively, given by

$$\beta_1(s) = 1 + s^2, \qquad \beta_2(s) = (1-s)^2,$$

$$\beta_3(s) = 4s(1-s), \qquad \beta_4(s) = -1 + 4(s-0.5)^2,$$

and then they were scaled as $\beta(s)/\|\beta(s)\|_2$. We set s_m as $M = 50$ equidistant grid points in [0,1] with $s_1 = 0$ and $s_M = 1$. We set $\eta_i(s) = \sum_{k=1}^{2} \xi_{ik}\psi_k(s)$ with $\xi_{i1} \sim N(0, \lambda_1^2 = 1), \xi_{i2} \sim N(0, \lambda_2^2 = 0.5^2), \psi_1(s) = \sqrt{2}\sin(2\pi s)$, and $\psi_2(s) = \sqrt{2}\cos(2\pi s)$. We consider two index functions including

$g_1(X_i^T\beta(s_m)) = \sin(2X_i^T\beta(s_m)) + 2\cos(2 + X_i^T\beta(s_m))$ and $g_2(X_i^T\beta(s_m)) = X_i^T\beta(s_m)$.

We conducted extensive simulation studies under different settings, but we only report some representative results for the sake of space. First, we set $n = 40$ and 200 and simulated data sets from model (3) for the first index function $g_1(X_i^T\beta(s_m))$ as described above. We fitted SIVC to each simulated data set and calculated all unknown quantities. Table 1 summarizes the mean absolute error (MAE) and root mean square error (RMSE) of all parameter estimates and the mean integrated absolute error (MIAE) and mean integrated squared error (MISE) of all estimated functions based on 200 simulations. The results in Table 1 indicate satisfactory performance of our estimators since all MAE, MSE, MIAE and MISE values are quite small. Typical estimated functions with mean performance are displayed in Figure 2 and Figure 1 in Web Appendix. The estimated curves (broken lines) closely resemble the corresponding true functions (solid lines) in these figures. As expected, all the errors increase as sample size decreases. Moreover, Table 2 includes the coverage probabilities of the simultaneous Confidence bands of $\beta_j(s)$s and $g_1(u)$ for $n = 40$ and 200 based on 500 simulations. These coverage probabilities get close to the specified Confidence level 95% as sample size increases, while the results for $g(\cdot)$ are slightly worse than those of $\beta_j(s)$.

Second, we illustrate the superiority of SIVC over multivariate varying coefficient model (MVCM) in Zhu et al. (2012) in terms of prediction accuracy. We set $n = 200$ and then

simulated data sets from model (3) for both $g_1(X_i^T\beta(s_m))$ and $g_2(X_i^T\beta(s_m))$. For each simulated dataset, we randomly split it into a training set and a test set according to proportions π and $1 - \pi$, respectively. We used the training set to estimate all unknown parameters, and then predicted the responses in the test set. Finally, we calculated the prediction error for each simulated dataset. We consider three values of π including 0.3, 0.5, and 0.7. For each case, we repeated 200 times. Table 3 reports the MAE and RMSE of prediction errors for SIVC and MVCM. For the index function $g_1(X_i^T\beta(s_m))$, SIVC significantly outperforms MVCM with smaller MAE and RMSE. Even for $g_2(X_i^T\beta(s_m))$, SIVC is slightly better than MVCM. It may indicate that SIVC is a useful tool for modeling functional data.

4.2 Real Data Analysis

We applied model (3) to the DWI data set described in Section 1. One goal of NIH ADNI is to test whether genetic, structural and functional neuroimaging, and clinical data can be integrated to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). We downloaded the structural brain MRI data and corresponding clinical and genetic data from baseline and follow-up from the ADNI publicly available database (http://adni.loni.usc.edu/).

The DWI data were processed by using a FSL TBSS pipeline (Smith et al., 2006) to register DTIs from multiple subjects to create a mean image and a mean skeleton. We used FMRIB software library to compute maps of fractional anisotropy (FA) for all subjects from the DTI after eddy current correction and automatic brain extraction. Then, we fed FA maps into the TBSS tool, which is also part of FSL. In the TBSS analysis, we aligned the FA data of all the subjects into a common space by using non-linear registration and created and thinned the mean FA image to obtain a mean FA skeleton, which represents the centers of all white matter tracts common to the group. Subsequently, we projected each subject's aligned FA data onto this skeleton. Finally, we obtained the FA template curve measured at all the 83 grid points along the skeleton of the midsagittal corpus callosum as shown in Figure 2 in Appendix for all the subjects.

We are interested in establishing an association between FA and seven covariates including the gender variable (123 male and 91 female, coded by a dummy variable indicating for male), the age of the subject (years, ranges from 48.4 (years) to 90.4, mean 73.20), an indicator for handiness (193 right-hand and 20 left-hand, coded by a dummy variable indicating for left-hand), the education level (years, ranges from 9 to 20 (years), mean 15.91), an indicator for Alzheimer's disease (AD) status (19.6%), an indicator for mild cognitive impairment (MCI) status (55.1%) and Mini-Mental State Exam (MMSE) of ADNI. We standardized all variables to have mean zero and variance one. We fitted model (3) and applied the estimation procedure in Section 2 to the data set. Figure 3 presents the estimated varying coefficients corresponding to age, education, AD status, and MCI status. The results reveal that MMSE, age, education, MCI, and AD are the most important factors. Moreover, gender and handiness have little effects on FA.

We estimated the single index function $g(\cdot)$ and the covariance function $\Sigma_\eta(s, s')$ and its associated eigenvalues and eigenfunctions. See Figure 1 for details. The single index function shows a pattern with small values at the left and right ending points of the range $X_i^T \beta(s)$ and two modes in the middle of the range. The top five non-zero eigenvalues of $\hat{\Sigma}_\eta(s, s')$ are 0.8628, 0.2771, 0.0284, 0.0142, and 0.0102, respectively. The first two eigenvalues account for 94.8% of the total variability, while the remaining eigenvalues rapidly drop to zero. The first eigenfunction, with a dominant eigenvalue accounting for 71.8% of the total variation, is simple in structure and resembles a single cycle of a sine wave. The remaining eigenfunctions are also quite simple and roughly sinusoidal representing additional functional structure that cannot be captured by the mean structure of model (3).

Finally, we compared the prediction accuracy of SIVC with that of MVCM in Zhu et al. (2012). We randomly split the 213 subjects into a training set and a test set with corresponding proportions π and $1 - \pi$, respectively. We used the training set to estimate all the parameters, and then predicted the responses of the testing set. The 100 replications were used to calculate the prediction errors corresponding $\pi = 0.3$, 0.5 and 0.7. Table 4 reports the MAE and RMSE of the prediction errors and indicates that SIVC significantly outperforms MVCM in terms of both MAE and RMSE.

5. Conclusion

In this paper we have developed a single-index varying coefficient model for establishing a varying association between functional responses (e.g., image) and a set of covariates. We have developed an estimation procedure to estimate varying coefficient functions, link function, and the covariance function of individual functions. We have investigated a strategy to integrate the information across all grid points. We have used simulations and real data analysis to demonstrate that SIVC is a useful tool for modeling functional data.

Many important issues need to be addressed in future research. First, the computational burden associated with SIVC can be quite heavy making it infeasible for large-scale imaging data at this moment. We will develop more computationally efficient algorithms to address such challenge. Second, we need to develop an effective procedure to carry out statistical inference, such as hypothesis test. Third, it is scientifically interesting to extend SIVC to carry out regression analysis of longitudinal functional data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Bowman FD, Caffo B, Bassett SS, Kilts C. A bayesian hierarchical framework for spatial modeling of fmri data. NeuroImage. 2008; 39:146–156. [PubMed: 17936016]

Cook RD, Forzani L. Likelihood-based sufficient dimension reduction. Journal of the American Statistical Association. 2009; 104:197–208.

Cook RD, Weisberg S. Discussion of "sliced inverse regression for dimension reduction". Journal of the American Statistical Association. 1991; 86:328–332.

Gossl C, Auer DP, Fahrmeir L. Bayesian spatiotemporal inference in functional magnetic resonance imaging. Biometrics. 2001; 57:554–562. [PubMed: 11414583]

Horowitz, JL. Semiparametric and Nonparametric Methods in Econometrics. Springer; 2009.

Jiang CR, Wang JL. Functional single index models for longitudinal data. Annals of Statistics. 2011; 39:362–388.

Lazar, N. The Statistical Analysis of Functional MRI Data. New York: Springer; 2008.

Li K-C. Sliced inverse regression for dimension reduction. Journal of the American Statistical Association. 1991; 86:316–327.

Li Y, Zhu H, Shen D, Lin W, Gilmore JH, Ibrahim JG. Multiscale adaptive regression models for neuroimaging data. Journal of the Royal Statistical Society: Series B. 2011; 73:559–578.

Ma Y, Zhu L. A semiparametric approach to dimension reduction. Journal of the American Statistical Association. 2012; 107:168–179. [PubMed: 23828688]

Ma Y, Zhu L. Efficient estimation in sufficient dimension reduction. Annals of statistics. 2013; 41:250–268. [PubMed: 24058219]

Ma Y, Zhu L. On estimation efficiency of the central mean subspace. J. R. Stat. Soc. Ser. B Stat. Methodol. 2014; 76:885–901.

Miranda MF, Zhu H, Ibrahim JG. Bayesian spatial transformation models with applications in neuroimaging data. Biometrics. 2013; 69:1074–1083. [PubMed: 24128143]

Morris JS. Functional regression. Annual Review of Statistics and Its Application. 2015; 2:321–359.

Penny WD, Trujillo-Barreto NJ, Friston KJ. Bayesian fmri time series analysis with spatial priors. NeuroImage. 2005; 24:350–362. [PubMed: 15627578]

Polzehl J, Spokoiny VG. Propagation-separation approach for local likelihood estimation. Probab. Theory Relat. Fields. 2006; 135:335–362.

Ramsay, JO.; Silverman, BW. Functional Data Analysis. second. New York: Springer; 2005.

Reiss PT, Huang L, Mennes M. Fast function-on-scalar regression with penalized basis expansions. The International Journal of Biostatistics. 2010; 6

Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. Journal of the Royal Statistical Society. Series B. 1991; 53:233–243.

Smith M, Fahrmeir L. Spatial bayesian variable selection with application to functional magnetic resonance imaging. Journal of the American Statistical Association. 2007; 102:417–431.

Smith S, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols T, Mackay C, Watkins K, Ciccarelli O, Cader M, Matthews P, Behrens T. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. NeuroImage. 2006; 31:1487–1505. [PubMed: 16624579]

Staicu A, Crainiceanu C, Carroll R. Fast analysis of spatially correlated multilevel functional data. Biostatistics. 2010; 11:177–194. [PubMed: 20089508]

Wang J-L, Chiou J-M, Mueller H-G. Review of functional data analysis. 2015

Wang X, van Eeden C, Zidek JV. Asymptotic properties of maximum weighted likelihood estimators. Journal of Statistical Planning and Inference. 2004; 119:37–54.

Worsley KJ, Taylor JE, Tomaiuolo F, Lerch J. Unified univariate and multivariate random field theory. NeuroImage. 2004; 23:189–195.

Xia Y. A constructive approach to the estimation of dimension reduction directions. The Annals of Statistics. 2007; 35:2654–2690.

Xia Y, Tong H, Li W, Zhu L-X. An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society: Series B. 2002; 64:363–410.

Yue Y, Loh JM, Lindquist MA. Adaptive spatial smoothing of fMRI images. Statistics and its Interface. 2010; 3:3–14.

Zhang J, Chen J. Statistical inferences for functional data. The Annals of Statistics. 2007; 35:1052–1079.

Zhu H, Fan J, Kong L. Spatially varying coefficient model for neuroimaging data with jump discontinuities. Journal of the American Statistical Association. 2014; 109:1084–1098. [PubMed: 25435598]

Zhu H, Li R, Kong L. Multivariate varying coefficient model for functional responses. Annals of Statistics. 2012; 40:2634–2666. [PubMed: 23645942]

Zhu L, Wang T, Zhu L, Ferré L. Sufficient dimension reduction through discretization-expectation estimation. Biometrika. 2010; 97:295–304.
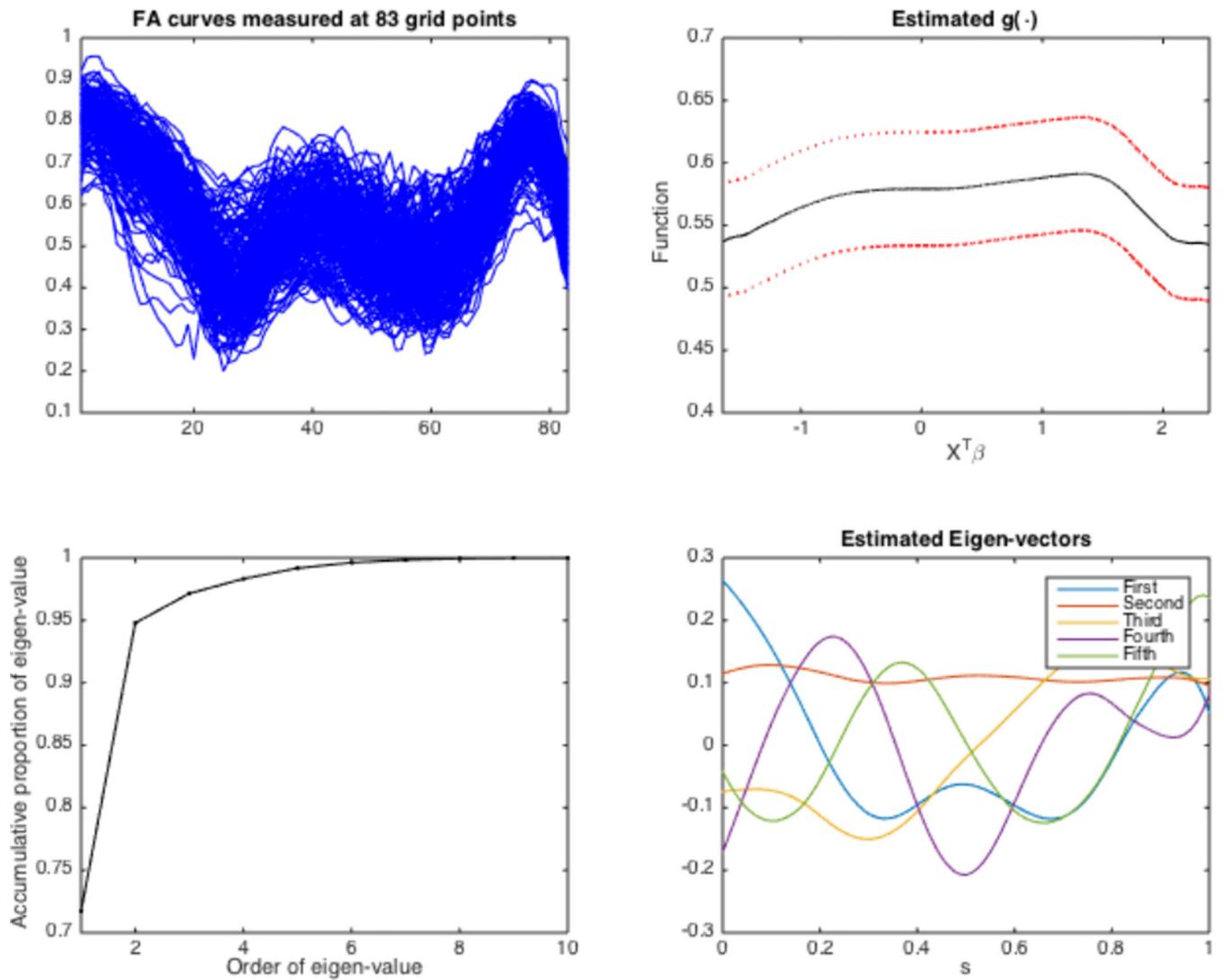
Figure 1.
ADNI data analysis: raw FA curves measured at 83 grid points (upper-left panel), the estimated index function with the broken red lines representing 95% simultaneous Confidence bands (upper-right panel), the estimated accumulative proportion of estimated eigenvalues (lower-left panel) and estimated eigenfunctions (lower-right panel) corresponding to the five largest eigenvalues.
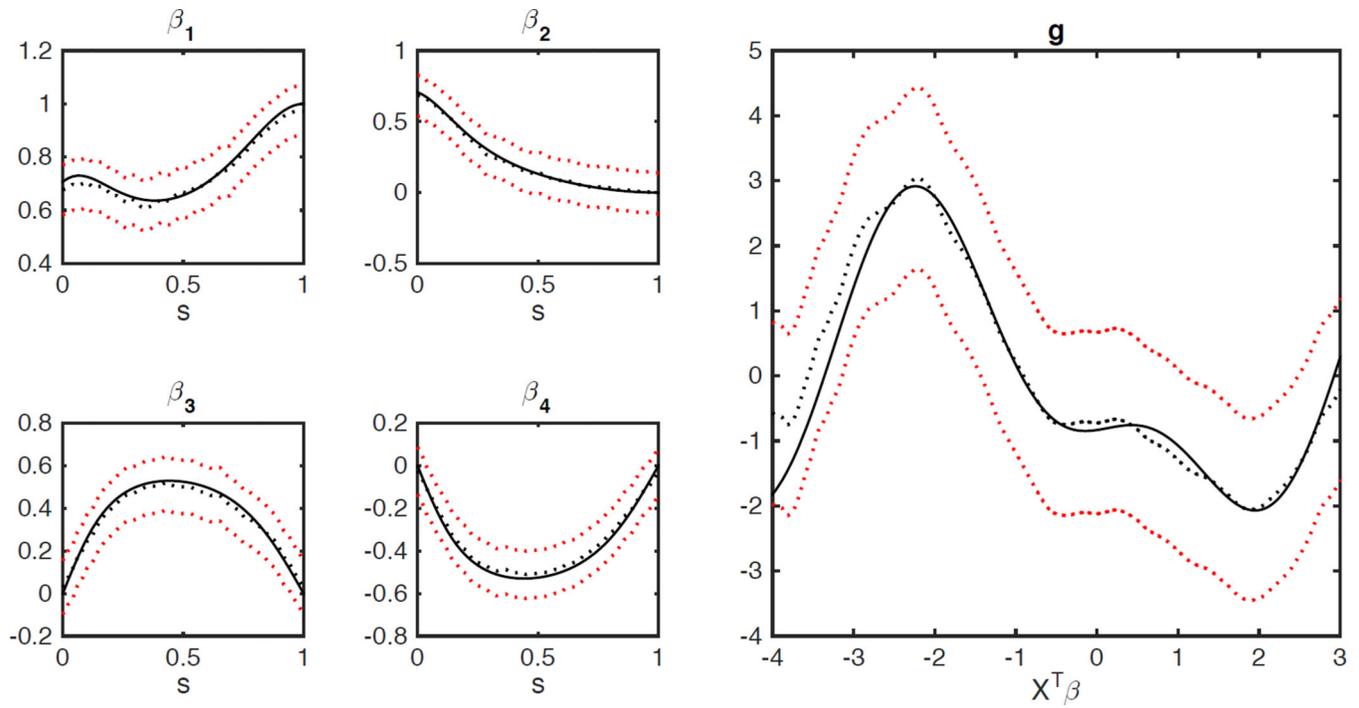
Figure 2.
Simulation results for model (3) with the first index function and $n = 200$: the true and estimated varying coefficient functions and the true and estimated index functions. In each panel, the solid line represents the true function, the broken line represents the estimated function, and the red broken lines are the corresponding 95% simultaneous Confidence bands.
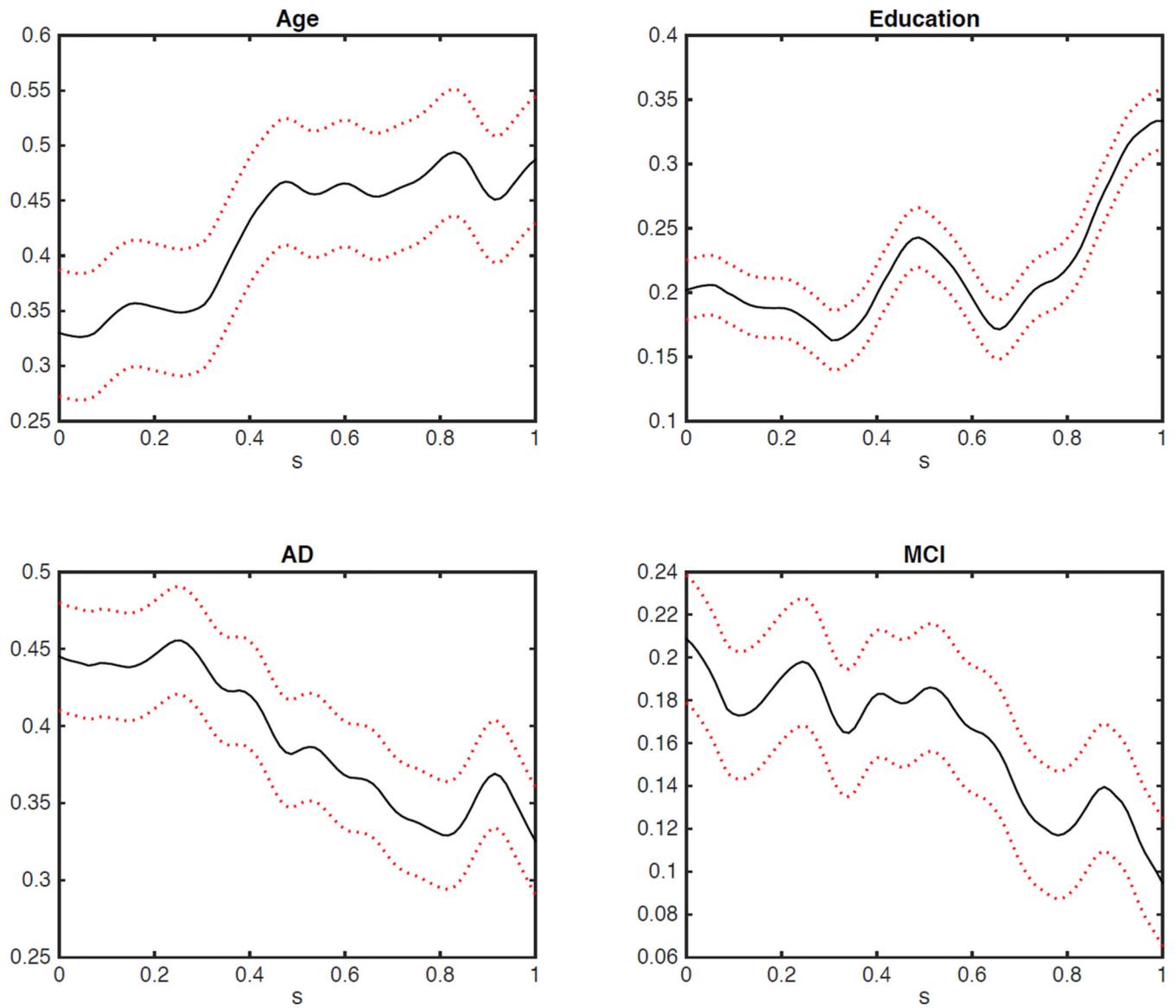
Figure 3.
ADNI data analysis: the four estimated varying coefficients for Age, Education, AD, and MCI. The black solid lines are estimated coefficients and the red broken lines are their corresponding 95% simultaneous Confidence bands.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Estimation results from 200 simulated data sets corresponding to the index function $g_1(X^T \beta(s))$ with n = 200 and 40. MAE is mean absolute error, RMSE is root mean square error, MIAE is mean integrated absolute error, and MISE is mean integrated square error

	n = 200			n = 40		
Parameters	λ_1	λ_2	σ^2	λ_1	λ_2	σ^2
MAE	0.1551	0.0621	0.0202	0.1773	0.1026	0.0594
RMSE	0.1724	0.0726	0.0267	0.2108	0.1212	0.0667

n = 200							
Functions	$\beta_1(s)$	$\beta_2(s)$	$\beta_3(s)$	$\beta_4(s)$	$\psi_1(s)$	$\psi_2(s)$	$g(\cdot)$
MIAE	0.0288	0.0440	0.0403	0.0119	0.0137	0.0138	0.1063
MISE	0.0042	0.0077	0.0093	0.0039	0.0033	0.0034	0.0412

n = 40							
Functions	$\beta_1(s)$	$\beta_2(s)$	$\beta_3(s)$	$\beta_4(s)$	$\psi_1(s)$	$\psi_2(s)$	$g(\cdot)$
MIAE	0.0488	0.0751	0.0686	0.0597	0.0291	0.0299	0.2536
MISE	0.0121	0.0249	0.0251	0.0241	0.0061	0.0060	0.1530

Table 2

Coverage probabilities of simultaneous confidence bands for n = 40 and 200 based on 500 simulated data sets. The confidence level is 95%.

n	$\beta_1(s)$	$\beta_2(s)$	$\beta_3(s)$	$\beta_4(s)$	$g(\cdot)$
40	0.936	0.932	0.944	0.934	0.924
200	0.946	0.954	0.948	0.946	0.942

Table 4

Prediction results for the ADNI data analysis. EST is estimate, SE is standard error, MAE is mean absolute error, and RMSE is root mean square error.

| | | $\pi = 0.3$ | | $\pi = 0.5$ | | $\pi = 0.7$ | |
		MAE	RMSE	MAE	RMSE	MAE	RMSE
SIVC	EST	0.114	0.1374	0.113	0.136	0.112	0.135
	SE	0.001	0.001	0.002	0.002	0.002	0.002
MVCM	EST	0.599	0.650	0.599	0.627	0.600	0.628
	SE	0.012	0.028	0.014	0.022	0.013	0.016