HHS Public Access

Author manuscript

J Am Stat Assoc. Author manuscript; available in PMC 2017 December 01.

Published in final edited form as:

J Am Stat Assoc. 2017 ; 112(519): 1169–1181. doi:10.1080/01621459.2016.1195742.

A Functional Varying-Coefficient Single-Index Model for Functional Response Data

Jialiang Li,

Associate Professor in Department of Statistics and Applied Probability in National University of Singapore, an Associate Professor in Duke-NUS Graduate Medical School and a Scientist in Singapore Eye Research Institute

Chao Huang, and

A doctoral student under the supervision of Dr. Hongtu Zhu

Hongtu Zhu

A Professor of Biostatistics, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, 77230, and University of North Carolina, Chapel Hill, NC, 27599

for the Alzheimer's Disease Neuroimaging Initiative‡

Abstract

Motivated by the analysis of imaging data, we propose a novel functional varying-coefficient single index model (FVCSIM) to carry out the regression analysis of functional response data on a set of covariates of interest. FVCSIM represents a new extension of varying-coefficient single index models for scalar responses collected from cross-sectional and longitudinal studies. An efficient estimation procedure is developed to iteratively estimate varying coefficient functions, link functions, index parameter vectors, and the covariance function of individual functions. We systematically examine the asymptotic properties of all estimators including the weak convergence of the estimated varying coefficient functions, the asymptotic distribution of the estimated index parameter vectors, and the uniform convergence rate of the estimated covariance function and their spectrum. Simulation studies are carried out to assess the finite-sample performance of the proposed procedure. We apply FVCSIM to investigating the development of white matter diffusivities along the corpus callosum skeleton obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

Keywords

Functional data analysis; Image data analysis; Single index model; Varying-coefficient model

Address for correspondence and reprints: Hongtu Zhu, Ph.D., hzhu@bios.unc.edu; Phone No: 919-966-7272.

‡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

1 Introduction

The aim of this paper is to develop a new functional regression model for the regression analysis of imaging (or functional) data collected over space and/or time, such as diffusion tensor imaging (DTI) and positron emission tomography (PET) (Towle et al. (1993); Niedereyer & da Silva (2005); Buzsaki (2011); Heywood et al. (2006); Zhu et al. (2007); Friston (2009)). A common feature of many imaging techniques is that massive functional data are observed/calculated at the same grid points, such as voxels in a three-dimensional space. Let $\mathbf{y}_i(s) = (y_{i1}(s), \ldots, y_{iJ}(s))^T$ be a J-dimensional functional response vector measured at a set of the same M location points, denoted as $\mathcal{S} = \{s_1, \ldots, s_M\}$, for subject i, $i = 1, \ldots, n$. We propose a novel functional varying-coefficient single index model given by

$$y_{ij}(s)=X_i^T\boldsymbol{\alpha}_j(s)+g_j(Z_i^T\boldsymbol{\beta}_j)+\eta_{ij}(s)+\varepsilon_{ij}(s)\text{ for }j=1,\ldots,J, \quad (1)$$

where X_i is a $p\times1$ covariate vector with varying coefficient functions $\boldsymbol{\alpha}_j(s) = (\alpha_{j1}(s), \cdots, \alpha_{jp}(s))^T$, $g_j(\cdot)$ is an unknown link function, Z_i is a $q \times 1$ vector of covariates with its index parameter vector $\boldsymbol{\beta}_j \in \mathbb{R}^q$, $\varepsilon_{ij}(s)$ is the measurement error, and $\eta_{ij}(s)$ characterizes individual curve variations. The random function $\{\eta_{ij}(s) : s \in [0, 1]\}$ is assumed to be a stochastic process with mean zero and covariance function $R(s, t) = cov\{\boldsymbol{\eta}_i(s), \boldsymbol{\eta}_i(t)\}$, where $\boldsymbol{\eta}_i(s) = (\eta_{i1}(s), \cdots, \eta_{iJ}(s))^T$. The error terms are mean zero process with covariance $\mathcal{E}(s, t) = cov\{\boldsymbol{\varepsilon}_i(s), \boldsymbol{\varepsilon}_i(t)\}$, where $\boldsymbol{\varepsilon}_i(s) = (\varepsilon_{i1}(s), \cdots, \varepsilon_{iJ}(s))^T$. Moreover, $\boldsymbol{\varepsilon}_i(s)$ and $\boldsymbol{\varepsilon}_i(t)$ are assumed to be independent for $s \neq t$ and $\mathcal{E}(s, t)$ takes the form of $\mathcal{E}(s, s)I(s = t)$, where $I(\cdot)$ is an indicator function. Model (1) allows a joint model of J different measures obtained from a single imaging modality or multiple modalties. For instance, for DTI, standard imaging measures include fractional anisotropy (FA) and mean diffusivity (MD).

Model (1) can be regarded as a novel integation of varying-coefficient models and single index models. When $g_j(\cdot) \equiv 0$, model (1) reduces to varying coefficient models widely adopted in the existing literature. This class of models has been thoroughly studied and developed for longitudinal, time series, and functional data. For varying coefficient models, it is of particular interest in data analysis to construct simultaneous confidence bands for $\boldsymbol{\alpha}_j(\cdot)$ and to develop global test statistics for the hypotheses regarding $\boldsymbol{\alpha}_j(\cdot)$. See Hoover et al. (1998), Fan & Zhang (1999), Wu & Chiang (2000), Ramsay & Silverman (2005), Fan et al. (2003), Morris & Carroll (2006), Fan & Zhang (2008), Wang et al. (2008), Cheng et al. (2009), Zhang (2011) and Zhu et al. (2012) for various statistical procedures proposed for different varying coefficient models. In particular, Morris & Carroll (2006) developed a Bayesian wavelet-based approach for a functional mixed effects modeling framework. Zhu et al. (2012) developed several statistical inference procedures for multivariate varying coefficient models for functional responses and systematically studied their theoretical properties.

In the absence of $\boldsymbol{\alpha}_j(s)$, that is, $\boldsymbol{\alpha}_j(s) \equiv 0$, model (1) reduces to single index models for functional responses with unknown link functions $g_j(\cdot)$. For identifiability, it is often assumed that $\beta_j^{\mathrm{T}}\beta_j=1$ and the first component of $\boldsymbol{\beta}_j$ is positive. Single index models have

been widely studied and developed for cross-sectional, longitudinal, and functional data. For single index models, one of the most important objectives in data analysis is to estimate and test the model index coefficients $\boldsymbol{\beta}_j$. See Horowitz (2009) for a comprehensive review of single index models. There is also a great interest in developing partial-linear single-index models via the integration of single-index models with linear regression models (Carroll et al., 1997; Wang et al., 2010). Most earlier references focus on univariate response observed from cross-sectional studies (Xia et al., 2002; Härdle et al., 1993; Xia, 2006; Wang et al., 2010; Cui et al., 2011; Ma & Zhu, 2013). Recently, Jiang & Wang (2010) developed functional single index models for functional/longitudinal response data and derived their associated estimation method and asymptotic theory.

Our FVCSIM can also be regarded as a novel extension of varying-coefficient single-index model (VCSIM) for scalar response data (Wong et al., 2008; Wang & Xue, 2011). There are at least three key differences between FVCSIM and VCSIM. (i) Our FVCSIM (1) is developed for multi-dimensional functional responses and explicitly incorporates long-range spatial correlation among functional data, whereas VCSIMs were developed for scalar responses without spatial correlation. (ii) Our FVCSIM requires a more sophisticated estimation method in order to account for both spatial correlation and spatial smoothness of imaging data. In contrast, in Wong et al. (2008), a bivariate kernel smoother was used to estimate the two types of functions, whereas in Wang & Xue (2011), a mean difference approach was proposed to reduce VCSIM to a varying-coefficient model. Bivariate smoothing usually produces less stable solutions and the mean difference approach relies on smoothing y and X first and may introduce additional biases. These two estimation methods are not directly applicable to correlated imaging data. (iii) The theoretical properties of various estimates in FVCSIM differ significantly from those in VCSIM. For FVCSIM, one must deal with the long-range spatial correlation of imaging data. First, such spatial correlation introduces a non-negligible effect into the asymptotic covariance of estimates, which leads to a different convergence rate for semiparametric modeling compared with the corresponding results for VCSIM. Second, the asymptotic behavior of estimated functional random effects and their covariance decomposition needs to be carefully evaluated and uniform consistency results must be established. Third, one can obtain much stronger theoretical properties associated with the estimated coefficients and functions, especially after we incorporate the estimated covariance of correlated image data. This property is very relevant to the facilitation of efficient inference.

Compared with the existing literature on functional data, our FVCSIM is a novel integration of single index models and varying-coefficient models for functional response data. At each grid point s, model (1) reduces to a partial-linear single-index model, whereas the functional single index model considered in Jiang & Wang (2010) reduces to a standard single-index model. Therefore, our model (1) differs from the functional single index model in Jiang & Wang (2010). Compared with Zhu et al. (2012) and Morris & Carroll (2006), model (1) is much more flexible in accommodating both the dynamic effects of X_i and the stationary effects of Z_i on functional data. Our estimation and inference procedures differ substantially from those in Jiang & Wang (2010), Zhu et al. (2012), and Morris & Carroll (2006). For instance, we can establish the weak convergence of the estimated varying co-efficient

functions and apply such a result to construct their confidence bands. Compared with the existing literature, we also establish the asymptotic distribution of single-index estimate in a totally different setting, while such extension requires non-trivial development in numerical implementation and theoretical derivation.

The rest of this article is organized as follows. In Section 2, an efficient estimation procedure is developed to iteratively estimate all unknown parameters and functions in model (1). In Section 3, we develop test statistics to test hypothesis of interest associated with varying-coefficient functions and/or index parameter vectors. Section 4 systematically investigates the asymptotic properties of various estimates and test statistics. In Section 5, simulation studies are used to examine the finite-sample performance of the proposed estimates and test statistics. In Section 6, we apply FVCSIM to investigate the development of white matter diffusivities along the corpus callosum skeleton obtained from ADNI. Section 7 concludes the paper with some discussions. Technical details are given in the supplementary document.

2 The Estimation Procedure

We develop an estimation procedure to estimate the varying coefficient functions $\boldsymbol{\alpha}_j(s)$, the single index functions $g_j(\cdot)$, the index parameter vectors $\boldsymbol{\beta}_j$, and the covariance function $R(s, s')$. For notational simplicity, we assume $s_m \in [0, 1]$ such that $s_1 \leq \cdots \leq s_M$, but our results can be easily extended to a compact subset of Euclidean space.

2.1 Varying-Coefficient Functions and Index Parameter Vectors

We adopt the local linear approximation to estimate the varying coefficient functions and the single index functions. Specifically, we use the Taylor's expansion to obtain

$$\begin{aligned} \boldsymbol{\alpha}_j(s_m) &\approx \boldsymbol{\alpha}_j(s)+\dot{\boldsymbol{\alpha}}_j(s)(s_m-s)=\mathbf{a}_j+\mathbf{b}_j(s_m-s)/h_{j1}, \\ g_j(Z_i^T\boldsymbol{\beta}_j) &\approx g_j(z^T\boldsymbol{\beta}_j)+\dot{g}_j(Z_i^T\boldsymbol{\beta}_j)(Z_i^T\boldsymbol{\beta}_j-z^T\boldsymbol{\beta}_j)=c_j+d_j(Z_i^T\boldsymbol{\beta}_j-z^T\boldsymbol{\beta}_j)/h_{j2}, \end{aligned} \quad (2)$$

where h_{j1} and h_{j2} are bandwidths, $\dot{\boldsymbol{\alpha}}_j(s) = d\boldsymbol{\alpha}_j(s)/ds$ and $\dot{g}_j(t) = dg_j(t)/dt$, s_m is in a small neighborhood of s, and $z^T\boldsymbol{\beta}_j$ is in a small neighborhood of $Z_i^T\beta_j$. Moreover, we set $\mathbf{a}_j = \boldsymbol{\alpha}_j(s)$, $\mathbf{b}_j = \dot{\boldsymbol{\alpha}}_j(s)h_{j1}$, $c_j = g_j(z^T\boldsymbol{\beta}_j)$, and $d_j=\dot{g}_j(Z_i^T\beta_j)h_{j2}$. One may directly minimize an objective function given by

$$\sum_{m=1}^{M}\sum_{i=1}^{n}[y_{ij}(s_m)-\{X_i^T(\mathbf{a}_j+\mathbf{b}_j\frac{s_m-s}{h_{j1}})+c_j+d_j\frac{(Z_i-z)^T\boldsymbol{\beta}_j}{h_{j2}}\}]^2 K_{jim}(s,z), \quad (3)$$

where $K_{jim}(s, z)$s' are weights. A standard choice of the weight $K_{jim}(s, z)$ is a two-dimensional kernel given by

$$K_{jim}(s,z)=K_{h_{j1},h_{j2}}(s_m-s,(Z_i-z)^T\boldsymbol{\beta}_j),$$

where $K_{h_1,h_2}(x, y) = K(x/h_1, y/h_2)/(h_1h_2)$ is a scaled kernel with two bandwidths h_1 and h_2. The kernel that we choose in this paper is the product of two one-dimensional kernels such that $K(x, y) = K(x)K(y)$. However, since it is computationally difficult to directly minimize (3), we consider an iterative estimation approach as follows.

Step 0. *Initialization step:* At each s_m ($m = 1, \cdots, M$), we fit a partly linear single index model

$$y_{ij}(s_m)=X_i^T\boldsymbol{\alpha}_{jm}^*+g_{jm}(Z_i^T\boldsymbol{\beta}_{jm}^*)+\varepsilon_{ij}^*(s_m) \quad (4)$$

using data available at s_m. We then take $\hat{g}_j(\cdot)=M^{-1}\sum_{m=1}^M\hat{g}_{jm}(\cdot)$ and $\hat{\boldsymbol{\beta}}_j=\hat{\boldsymbol{\beta}}_j^*/\|\hat{\boldsymbol{\beta}}_j^*\|$ as the initial single index function and coefficient estimates, where $\hat{\boldsymbol{\beta}}_j^*=M^{-1}\sum_{m=1}^M\boldsymbol{\beta}_{jm}^*$. Write $y_{ij}^*(s_m)=y_{ij}(s_m)-\hat{g}_j(Z_i^T\hat{\boldsymbol{\beta}}_j)$. To obtain initial varying-coefficient estimates $\hat{\boldsymbol{\alpha}}_j(s)$, we then fit a varying-coefficient model (Zhu et al. (2012)) given by

$$y_{ij}^*(s)=X_i^T\boldsymbol{\alpha}_j(s)+\eta_{ij}(s)+\varepsilon_{ij}(s). \quad (5)$$

Step 1. Let $\mathscr{X}_{ji}(z) = (1, (Z_i-z)^T\hat{\boldsymbol{\beta}}_j/h_{j2})^T$. For a given $\hat{\boldsymbol{\beta}}_j$ and $\hat{\boldsymbol{\alpha}}_j(s)$, we can solve

$$\min_{c_j,d_j}\sum_{m=1}^M\sum_{i=1}^n\left[y_{ij}(s_m)-\{X_i^T\hat{\boldsymbol{\alpha}}_j(s_m)+(c_j,d_j)\mathscr{X}_{ji}(z)\}\right]^2K_{h_{j2}}((Z_i-z)^T\hat{\boldsymbol{\beta}}_j), \quad (6)$$

to obtain

$$\begin{aligned}&\left(\hat{g}_j(z^T\boldsymbol{\beta}_j),\widehat{\dot{g}_j}(z^T\boldsymbol{\beta}_j)h_{j2}\right)^T=\left(\hat{c}_j,\hat{d}_j\right)^T\\&=\textstyle\sum_{h_{j2}}(z)^{-1}\sum_{i=1}^n\sum_{m=1}^M K_{h_{j2}}((Z_i-z)^T\hat{\boldsymbol{\beta}}_j)\mathscr{X}_{ji}(z)\{y_{ij}(s_m)-X_i^T\hat{\boldsymbol{\alpha}}_j(s_m)\},\end{aligned} \quad (7)$$

where $\sum_{h_{j2}}(z)=M\sum_{i=1}^n K_{h_{j2}}((Z_i-z)^T\hat{\boldsymbol{\beta}}_j)\mathscr{X}_{ji}(z)\mathscr{X}_{ji}(z)^T$.

Step 2. Let $\mathscr{Z}_{im}(s)=(X_i^T, X_i^T(s_m-s)/h_{j1})^T$. Given $\hat{\boldsymbol{g}}_j(z^T\boldsymbol{\beta}_j)$, we update the estimate of $\boldsymbol{\alpha}_j(s)$ by minimizing

$$\min_{\mathbf{a}_j,\mathbf{b}_j}\sum_{m=1}^{M}\sum_{i=1}^{n}\Big(y_{ij}(s_m)-[X_i^T\{\mathbf{a}_j+\mathbf{b}_j(s_m-s)/h_{j1}\}+\hat{g}_j(Z_i^T\boldsymbol{\beta}_j)]\Big)^2K_{h_{j1}}(s_m-s), \tag{8}$$

which leads to

$$\begin{aligned}&\Big(\hat{\boldsymbol{\alpha}}_j(s),\widehat{\dot{\boldsymbol{\alpha}}_j}(s)h_{j1}\Big)^T=\Big(\hat{\mathbf{a}}_j,\hat{\mathbf{b}}_j\Big)^T\\&=W_{h_{j1}}(s)^{-1}\sum_{i=1}^{n}\sum_{m=1}^{M}K_{h_{j1}}(s_m-s)\mathscr{D}_{im}(s)\{y_{ij}(s_m)-\hat{g}_j(Z_i^T\boldsymbol{\beta}_j)\},\end{aligned} \tag{9}$$

where $W_{h_{j1}}(s)=\sum_{i=1}^{n}\sum_{m=1}^{M}K_{h_{j1}}(s_m-s)\mathscr{D}_{im}(s)\mathscr{D}_{im}(s)^T$.

Step 3. Let $\delta_{Z,i,i'}=Z_i-Z_{i'}$. Given $\hat{\boldsymbol{\alpha}}_j(s)=\hat{\mathbf{a}}_j$, $\hat{g}_j(Z_i^T\boldsymbol{\beta}_j)=\hat{c}_{ij}$, $\widehat{\dot{g}_j}(Z_i^T\boldsymbol{\beta}_j)h_{j2}=\hat{d}_{ij}$ and the estimate for $\boldsymbol{\beta}_j$ at the preceding step, denoted by $\hat{\beta}_j^o$, we update the estimate of $\boldsymbol{\beta}_j$ by minimizing

$$\min_{\boldsymbol{\beta}_j}\sum_{m=1}^{M}\sum_{i=1}^{n}\sum_{i\neq i,i'=1}^{n}[y_{ij}(s_m)-\{X_i^T\hat{\boldsymbol{\alpha}}_j(s_m)+\hat{c}_{i'j}+\hat{d}_{i'j}\frac{\delta_{Z,ii'}^T\boldsymbol{\beta}_j}{h_{j2}}\}]^2K_{h_{j2}}^*(\delta_{Z,ii'}^T\hat{\boldsymbol{\beta}}_j^o), \tag{10}$$

where $K_{h_{j2}}^*(\delta_{Z,ii'}^T\boldsymbol{\beta}_j)=K_{h_{j2}}(\delta_{Z,ii'}^T\boldsymbol{\beta}_j)/\{n^{-1}\sum_{i=1}^{n}K_{h_{j2}}(\delta_{Z,ii'}^T\boldsymbol{\beta}_j)\}$, which leads to

$$\hat{\boldsymbol{\beta}}_j=\Omega_{h_{j2}}^{-1}\sum_{i=1}^{n}\sum_{i'\neq i,i'=1}^{n}\sum_{m=1}^{M}\hat{d}_{i'j}\frac{\delta_{Z,ii'}}{h_{j2}}K_{h_{j2}}^*(\delta_{Z,ii'}^T\hat{\boldsymbol{\beta}}_j^o)\{y_{ij}(s_m)-X_i^T\hat{\boldsymbol{\alpha}}_j(s_m)-\hat{c}_{i'j}\}, \tag{11}$$

where $\Omega_{h_{j2}}=M\sum_{i=1}^{n}\sum_{i'\neq i,i'=1}^{n}\hat{d}_{i'j}^2K_{h_{j2}}^*(\delta_{Z,ii'}^T\hat{\boldsymbol{\beta}}_j^o)\delta_{Z,ii'}\delta_{Z,ii'}^T/h_{j2}^2$.

Step 4. Normalize the updated $\hat{\boldsymbol{\beta}}_j$ such that $\|\hat{\boldsymbol{\beta}}\|=1$ and then repeat Steps 1 to 3 until convergence.

The two bandwidths h_{j1} and h_{j2} involved in the above procedure are selected by using the cross-validation in our implementation. Since only univariate local linear regression is used in the estimation procedure, our procedure is numerically much more stable than those used in Wong et al. (2008) and Wang & Xue (2011). In our simulations, it usually takes less than 10 iterations for a stringent convergence criterion, in which the average of the squared differences between two consecutive estimates is less than 10^{-5}.

2.2 Individual Functions and Covariance Function

For clustered data with fixed M locations, we may use empirical estimates for the random effects and the covariance matrix. In this paper we consider a more general approach that

can meet this estimation goal and further adapt to functional data. Our methods may be appealing when M is large (as in our brain image example) and it is reasonable to consider the covariance to be a smooth function of the locations.

We employ the local linear regression to estimate all individual functions. Consider the Taylor's expansion for $\eta_{ij}(s_m)$ around s given by

$$\eta_{ij}(s_m)=\eta_{ij}(s)+\dot{\eta}_{ij}(s)(s_m-s)=\mathrm{a}_j+\mathrm{b}_j(s_m-s)/h_j^*, \quad (12)$$

where $\mathrm{a}_j = \eta_{ij}(s)$ and $\mathrm{b}_j=h_j^*\dot{\eta}_{ij}(s)$, in which $\dot{\eta}_{ij}(s) = d\eta_{ij}(s)/ds$. Let $y_{ij}^*(s)=y_{ij}(s)-X_i^T\hat{\boldsymbol{\alpha}}_j(s)-\hat{g}_j(Z_i^T\hat{\boldsymbol{\beta}}_j)$, we minimize

$$\sum_{m=1}^{M}[y_{ij}^*(s_m)-\{\mathrm{a}_j+\mathrm{b}_j(s_m-s)/h_j^*\}]^2K_{h_j^*}(s_m-s). \quad (13)$$

The estimates of a_j and b_j are given by

$$\begin{aligned}\begin{pmatrix}\hat{\eta}_{ij}(s)\\ \widehat{\dot{\eta}_{ij}(s)h_j^*}\end{pmatrix}&=\begin{pmatrix}\hat{\mathrm{a}}_j\\ \hat{\mathrm{b}}_j\end{pmatrix}=\{\sum_{m=1}^{M}K_{h_j^*}(s_m-s)(1,(s_m-s)/h_j^*)^T(1,(s_m-s)/h_j^*)\}^{-1}\\ &\times\sum_{m=1}^{M}K_{h_j^*}(s_m\\ &-s)(1,(s_m-s)/h_j^*)^Ty_{ij}^*(s_m).\end{aligned} \quad (14)$$

The minimizer $\hat{\mathrm{a}}_j$ gives the estimator $\hat{\eta}_{ij}(s_m)$. We obtain $\hat{\boldsymbol{\eta}}_i(s) = (\hat{\eta}_{i1}(s), \cdots, \hat{\eta}_{iJ}(s))^T$. The bandwidth h_j^* can be selected by the usual cross-validation or the generalized cross-validation.

The covariance matrix $R(s, t)$ can thus be estimated by an empirical covariance estimator

$$\hat{R}(s,t)=n^{-1}\sum_{i=1}^{n}\hat{\boldsymbol{\eta}}_i(s)\hat{\boldsymbol{\eta}}_i(t)^T. \quad (15)$$

Subsequently, we can calculate the spectral decomposition of $\hat{R}_{jj}(s, t)$ for each j as follows:

$$\hat{R}_{jj}(s,t)=\sum_{l}\hat{\lambda}_{jl}\hat{\psi}_{jl}(s)\hat{\psi}_{jl}(t), \quad (16)$$

where $\hat{\lambda}_{j1} \geq \hat{\lambda}_{j2} \geq \ldots \geq 0$ are estimated eigenvalues and the $\hat{\psi}_{jl}(t)$'s are the corresponding estimated principal components. Furthermore, the (j, l)-th functional principal component scores can be computed using $\hat{\xi}_{ijl}=\sum_{m=1}^{M}\hat{\eta}_{ij}(s_m)\hat{\psi}_{jl}(s_m)(s_m-s_{m-1})$ for $i = 1, \ldots, n$.

The noise variance function $\mathcal{E}(s, s)$ measures the variation of $\boldsymbol{\varepsilon}(s)$ and its estimation can be based on the residuals $\hat{\varepsilon}_{ij}(s)=y_{ij}^{*}(s)-\hat{\eta}_{ij}(s)$ by following Hall & Marron (1990) and Fan & Yao (1998). The kernel estimator for $\mathcal{E}_{jj}(s, s)$ with a bandwidth h^{**} is given by

$$\hat{\mathcal{E}}_{jj}(s,s)=\frac{\sum_{i=1}^{n}\sum_{m=1}^{M}K_{h^{**}}(s_m-s)\hat{\varepsilon}_{ij}^{2}(s_m)}{n\sum_{m=1}^{M}K_{h^{**}}(s_m-s)}. \quad (17)$$

2.3 Refined Function amd Parameter Estimation

The estimated covariance structure may be incorporated to account for the spatial dependence in the model and further improve the estimation results. In this section, we adopt the estimated covariance matrix to obtain more efficient estimates for regression coefficients. Let $\Phi_j = R_{jj}(s_k, s_l) + \mathcal{E}_{jj}(s_k, s_l)_{k,l=1, \ldots, M}$ be the covariance matrix for the j–th response observed at M grid points in the sample. When Φ_j is given, we may obtain the following refined estimates $\tilde{g}_j(\cdot)$, $\tilde{\boldsymbol{\alpha}}_j(\cdot)$ and $\tilde{\boldsymbol{\beta}}_j$ for $g_j(\cdot)$, $\boldsymbol{\alpha}_j(\cdot)$ and $\boldsymbol{\beta}_j$, respectively. We modify the three iterative steps in Section 2.1 as follows.

Step 1. Let $\mathbf{1}_M$ be an $M \times 1$ vector of ones, $\mathbf{s} = (s_1, \cdots, s_M)$, $y_{ij}(\mathbf{s}) = (y_{ij}(s_1), \cdots, y_{ij}(s_M))^T$, and $\boldsymbol{\alpha}_j(\mathbf{s}) = (\boldsymbol{\alpha}_j(s_1), \cdots, \boldsymbol{\alpha}_j(s_M))$. Given $\hat{\boldsymbol{\beta}}_j$ and $\hat{\boldsymbol{\alpha}}_j(\mathbf{s})$, we can solve

$$\min_{c_j,d_j}\sum_{i=1}^{n}[y_{ij}(\mathbf{s})-\{\hat{\boldsymbol{\alpha}}_j(\mathbf{s})^T X_i+(c_j,d_j)\mathcal{X}_{ji}(z)\mathbf{1}_M\}]^T\Phi_j^{-1}\times[y_{ij}(\mathbf{s})-\{\hat{\boldsymbol{\alpha}}_j(\mathbf{s})^T X_i+(c_j,d_j)\mathcal{X}_{ji}(z)\mathbf{1}_M\}]\,K_{h_{j2}}((Z_i-z)^T\hat{\boldsymbol{\beta}}_j), \quad (18)$$

which yields that $(\tilde{g}_j(z^T\boldsymbol{\beta}_j), \widetilde{\dot{g}_j(z^T\boldsymbol{\beta}_j)h_{j2}})^T=(\tilde{c}_j,\tilde{d}_j)^T$ is equal to

$$\{\sum_{i=1}^{n}K_{h_{j2}}((Z_i-z)^T\hat{\boldsymbol{\beta}}_j)\mathcal{X}_{ji}(z)\mathcal{X}_{ji}(z)^T\mathbf{1}_M^T\Phi_j^{-1}\mathbf{1}_M\}^{-1}\times\sum_{i=1}^{n}K_{h_{j2}}((Z_i-z)^T\hat{\boldsymbol{\beta}}_j)\mathcal{X}_{ji}(z)\mathbf{1}_M^T\Phi_j^{-1}\{y_{ij}(\mathbf{s})-X_i^T\hat{\boldsymbol{\alpha}}_j(\mathbf{s})\mathbf{1}_M\}. \quad (19)$$

Step 2. Let $\mathscr{D}_i(s)=(\mathbf{1}_M X_i^T, (\mathbf{s}-s)X_i^T/h_{j1})$. Given $\tilde{g}_j(z^T\boldsymbol{\beta}_j)$, we update the estimate of $\boldsymbol{\alpha}_j(s)$ by minimizing

$$\min_{\mathbf{a}_j,\mathbf{b}_j}\sum_{i=1}^{n}[y_{ij}(\mathbf{s})-\{\mathscr{D}_i(s)(\mathbf{a}_j^T,\mathbf{b}_j^T)^T+\tilde{g}_j(Z_i^T\boldsymbol{\beta}_j)\mathbf{1}_M\}]^T\Phi_j^{-1/2}\tilde{K}_j\Phi_j^{-1/2}\times[y_{ij}(\mathbf{s})-\{\mathscr{D}_i(s)(\mathbf{a}_j^T,\mathbf{b}_j^T)^T+\tilde{g}_j(Z_j^T\boldsymbol{\beta}_j)\mathbf{1}_M\}], \tag{20}$$

where $\tilde{K}_j = \mathrm{diag}(K_{h_{j1}}(s_1-s), \cdots, K_{h_{j1}}(s_1-s))$. This leads to

$$\begin{aligned}&(\tilde{\boldsymbol{\alpha}}_j(s), \widetilde{\dot{\boldsymbol{\alpha}}_j}(s)h_{j1})^T=(\tilde{\mathbf{a}}_j,\tilde{\mathbf{b}}_j)^T\\&=\{\sum_{i=1}^{n}\mathscr{D}_i(s)^T\Phi_j^{-1/2}\tilde{K}_j\Phi_j^{-1/2}\mathscr{D}_i(s)\}^{-1}\times\sum_{i=1}^{n}\mathscr{D}_i(s)^T\Phi_j^{-1/2}\tilde{K}_j\Phi_j^{-1/2}\{y_{ij}(\mathbf{s})-\tilde{g}_j(Z_i^T\boldsymbol{\beta}_j)\mathbf{1}_M\}.\end{aligned} \tag{21}$$

Step 3. Given $\tilde{\boldsymbol{\alpha}}_j(s) = \tilde{\mathbf{a}}_j$, $\tilde{g}_j(Z_i^T\beta_j)=\tilde{c}_{ij}, \widetilde{\dot{g}_j}(Z_i^T\beta_j)h_{j2}=\tilde{d}_{ij}$ and the initial estimate $\hat{\boldsymbol{\beta}}_j$, we update the estimate of $\boldsymbol{\beta}_j$ by minimizing

$$\begin{aligned}\min_{\boldsymbol{\beta}_j}\sum_{i=1}^{n}\sum_{i'\neq i,i'=1}^{n}&[y_{ij}(\mathbf{s})-\tilde{\boldsymbol{\alpha}}_j(\mathbf{s})^T X_i-\{\tilde{c}_{i'j}+\tilde{d}_{i'j}\frac{\delta_{Z,ii'}^T\boldsymbol{\beta}_j}{h_{j2}}\}\mathbf{1}_M]^T\Phi_j^{-1}\\&\times[y_{ij}(\mathbf{s})-\tilde{\boldsymbol{\alpha}}_j(\mathbf{s})^T X_i-\{\tilde{c}_{i'j}+\tilde{d}_{i'j}\frac{\delta_{Z,ii'}^T\boldsymbol{\beta}_j}{h_{j2}}\}\mathbf{1}_M]K_{h_{j2}}^*(\delta_{Z,ii'}^T\hat{\boldsymbol{\beta}}_j),\end{aligned} \tag{22}$$

which leads to

$$\begin{aligned}\tilde{\boldsymbol{\beta}}_j=&\{\sum_{i=1}^{n}\sum_{i'\neq i,i'=1}^{n}\tilde{d}_{i'j}^2K_{h_{j2}}^*(\delta_{Z,ii'}^T\hat{\boldsymbol{\beta}}_j)\delta_{Z,ii'}\delta_{Z,ii'}^T\mathbf{1}_M^T\Phi_j^{-1}\mathbf{1}_M\}^{-1}/h_{j2}^2\\&\times\sum_{i=1}^{n}\sum_{i'\neq i,i'=1}^{n}\tilde{d}_{i'j}\frac{\delta_{Z,ii'}}{h_{j2}}K_{h_{j2}}^*(\delta_{Z,ii'}^T\hat{\boldsymbol{\beta}}_j)\mathbf{1}_M^T\Phi_j^{-1}\{y_{ij}(\mathbf{s})-\tilde{\boldsymbol{\alpha}}_j(\mathbf{s})^T X_i-\tilde{c}_{i'j}\mathbf{1}_M\}.\end{aligned} \tag{23}$$

The bandwidths h_{j1} and h_{j2} are also selected by the cross-validation. To differentiate from the initial bandwidths, we denote the selected bandwidths as $h_{j1}^{\dagger}$ and $h_{j2}^{\dagger}$.

We provide the asymptotic properties of the refined estimates in Theorem 4 of this paper. Lemma 5 in the Appendix justifies their optimality for the estimation of Euclidean model

parameters. In practice, the unknown Φ_j may be estimated from $(\hat{R}_{jj}(s_k, s_l) + \hat{\hat{\varepsilon}}_{jj}(s_k, s_l))_{k,l=1,\ldots,M}$ using (15) and (17).

3 Hypothesis Test and Inference

It is of interest to carry out hypothesis test and other inference procedures after estimation. We provide a general guideline on how to test hypothesis of interest and construct confidence intervals for parametric and nonparametric components. Let $\boldsymbol{\alpha}(s) = (\alpha_1(s)^T, \ldots, \alpha_J(s)^T)^T$ and $\beta = (\beta_1^T, \ldots, \boldsymbol{\beta}_j^T)^T$. Let $\tilde{\alpha}_{j0}(\cdot)$, $\tilde{g}_{j0}(\cdot)$ and $\tilde{\boldsymbol{\beta}}_{j0}$ be the estimators of unknown functions and parameters under the null hypothesis and $\tilde{\alpha}_{j1}(\cdot)$, $\tilde{g}_{j1}(\cdot)$ and $\tilde{\boldsymbol{\beta}}_{j1}$ be the corresponding estimators under the alternative hypothesis.

First, we consider the test of the hypotheses for the varying coefficients functions as follows:

$$H_0: \mathrm{A} \times \mathrm{vec}(\boldsymbol{\alpha}(s)) = \mathrm{a}_0(s) \quad \text{v. s.} \quad H_1: \mathrm{A} \times \mathrm{vec}(\boldsymbol{\alpha}(s)) \neq \mathrm{a}_0(s), \tag{24}$$

where A is an $l \times Jp$ full rank matrix and $\mathrm{a}_0(s)$ is an $l \times 1$ vector of known functions.

Write $\mathrm{d}(s) = \mathrm{A} \times \mathrm{vec}(\tilde{\boldsymbol{\alpha}}_{(1)}(s)) - \mathrm{a}_0(s)$, where $\mathrm{vec}(\tilde{\boldsymbol{\alpha}}_{(1)}(s)) = (\tilde{\alpha}_{11}(s)^T, \ldots, \tilde{\alpha}_{J1}(s)^T)^T$. We may construct a global test as follows:

$$\mathrm{T}_\alpha = \int_0^1 \mathrm{d}(s)^T \{\mathrm{A}[\hat{R}(s,s) \otimes \hat{\Gamma}_1^{-1}]\mathrm{A}^T\}^{-1} \mathrm{d}(s) ds, \tag{25}$$

where $\hat{\Gamma}_1 = \sum_{i=1}^n X_i X_i^T$. Although we will prove the asymptotic null distribution of T_α in Theorem 4, such theoretical result does not produce a good approximation when the sample size n is relatively small. We thus propose the following bootstrap test procedure.

Step 1. Fit model (1) under H_0 to obtain $\tilde{\alpha}_{j0}(\cdot)$, $\tilde{g}_{j0}(\cdot)$, $\tilde{\boldsymbol{\beta}}_{j0}$, $\tilde{\eta}_{ij0}(s_m)$, and $\tilde{\varepsilon}_{ij0}(s_m)$.

Step 2. Draw $\mathrm{z}_i^{(b)}$ and $\mathrm{z}_{im}^{(b)}$ for $i = 1, \ldots, n$ and $m = 1, \ldots, M$ independently from the standard normal distribution and construct

$$\hat{y}_{ij}(s_m)^{(b)} = X_i^T \hat{\alpha}_{j0}(s_m) + \tilde{g}_{j0}(Z_i^T \hat{\boldsymbol{\beta}}_{j0}) + \tilde{\eta}_{ij0}(s_m)\mathrm{z}_i^{(b)} + \tilde{\varepsilon}_{ij0}(s_m)\mathrm{z}_{im}^{(b)}. \tag{26}$$

Step 3. Refit model (1) using $\hat{y}_{ij}(s_m)^{(b)}$ as the response values and calculate $\mathrm{T}_{\alpha,b}$ using formula (25) from this bootstrap sample.

Step 4. Repeat Steps 2 and 3 B times to obtain $\{\mathrm{T}_{\alpha,b} : b = 1, \cdots, B\}$ and then approximate the test p-value as $B^{-1}\sum_{b=1}^B I(\mathrm{T}_{\alpha,b} \geq \mathrm{T}_\alpha)$. Reject H_0 if the p-value is lower than a pre-specified significance level α, say 0.05.

The bootstrap test can be also applied to test the linear hypotheses for $\boldsymbol{\beta}$ as follows:

$$H_0: \mathrm{C}\boldsymbol{\beta} = \mathrm{c}_0 \quad \text{v. s.} \quad H_1: \mathrm{C}\boldsymbol{\beta} \neq \mathrm{c}_0, \quad (27)$$

where C is a given $l \times Jq$ full rank matrix and c_0 is a given $l \times 1$ vector. An F-type test can be constructed by

$$\mathrm{T}_\beta = 0.5n \frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}, \quad (28)$$

where $\mathrm{RSS}_0 = \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{m=1}^{M} \{y_{ij}(s_m) - X_i^T \tilde{\alpha}_{j0}(s_m) - \tilde{g}_{j0}(Z_i^T \tilde{\boldsymbol{\beta}}_{j0})\}^2$ and

$$\mathrm{RSS}_1 = \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{m=1}^{M} \{y_{ij}(s_m) - X_i^T \tilde{\alpha}_{j1}(s_m) - \tilde{g}_{j1}(Z_i^T \tilde{\boldsymbol{\beta}}_{j1})\}^2$$

are the residual sum of squares under H_0 and H_1, respectively. The asymptotic distribution of (28) can be established by following Fan & Huang (2005) under some mild conditions. We omit it for simplicity.

Besides hypothesis tests, one may also be interested in computing the confidence intervals of the estimated parameters and the confidence bands of the estimated functions. Although we will prove their theoretical distributions as $n \to \infty$, such asymptotic results may not be accurate enough in practice. We suggest the use of bootstrap methods. For instance, we may modify the semiparametric bootstrap procedure described in this section to generate bootstrapped samples and then calculate their corresponding estimates $\{(\hat{\alpha}_j^{(b)}, \hat{\beta}_j^{(b)}, \hat{g}_j(\cdot)^{(b)}) : b = 1, \cdots, B\}$. Subsequently, we can compute the confidence intervals and bands.

In real applications, the specific assignment of variables of interest into model (1) is usually guided by subjective matter experts in order to achieve meaningful practical interpretations. One may objectively allocate variables of interest into X and Z. Specifically, Cheng et al. (2009) considered a Bayesian information criterion on structural selection, whereas Zhang (2011) used the cross-validation criterion. Moreover, group penalization for functional estimation can also be used for structural selection (Cheng et al. (2014)). These ideas may be readily adopted for model (1).

4 Asymptotic Properties

We first define some notations. Let $u_r(K) = \int t^r K(t)dt$ and $v_r(K) = \int t^r K^2(t)dt$ for any integer r. For any smooth functions $f(s)$ and $g(s, t)$, define $\dot{f}(s) = df(s)/ds$, $\ddot{f}(s) = d^2 f(s)/ds^2$, $\dddot{f}(s) = d^3 f(s)/ds^3$, and $g^{(a,b)}(s, t) = \partial^{a+b} g(s, t)/\partial^a s \partial^b t$, where a and b are any nonnegative integers. The true values of $\boldsymbol{\alpha}(s)$, $\boldsymbol{\beta}$, and $g_j(\cdot)$ are denoted by $\boldsymbol{\alpha}_0(\cdot)$, $\boldsymbol{\beta}_0$ and $g_{j0}(\cdot)$, respectively. Let $\mu_{\beta j0}(z) = E(Z|Z^T\boldsymbol{\beta}_{j0} = z^T\boldsymbol{\beta}_{j0})$, $\nu_{\beta j0}(z) = \mu_{\beta j0}(z) - z$, $\omega_{\beta j0}(z) = E(ZZ^T|Z^T\boldsymbol{\beta}_{j0} = z^T\boldsymbol{\beta}_{j0})$, $\omega_{1jj'}(z_1,$

$z_2) = \nu_{\beta j0}(z_1)\nu_{\beta j' 0}(z_2)^T$, and $\omega_{2j}(z) = \omega_{\beta j0}(z) - \mu_{\beta j0}(z)\mu_{\beta j0}(z)^T$ for $j = 1, \ldots, J$. Let A^+ be the Moore-Penrose inverse of a symmetric matrix A and $f_j(\cdot)$ be the density function of $Z^T\boldsymbol{\beta}_j$. It is assumed that the initial value for estimating $\boldsymbol{\beta}_j$ is in an $O(1/\sqrt{n})$ neighbor of $\boldsymbol{\beta}_{j0}$, denoted as $B_{j,n}$ such that $B_{j,n} = \{\boldsymbol{\beta}_j : |\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j0}| \le \delta_{\beta_j} = C_0 n^{-1/2}\}$ (Härdle et al. (1993); Xia (2006)), where δ_{β_j} is the radius of $B_{j,n}$. We also assume that the initial estimate for $\alpha_j(s)$ is in an $O(1/\sqrt{Mh_{j1}})$ neighbor of $\alpha_{j0}(s)$, which can be achieved by using the estimates in Fan & Huang (2005) and Zhu et al. (2012).

We state the following theorems, whose detailed conditions and proofs can be found in the Appendix. The first theorem establishes the weak convergence of $(\hat{\boldsymbol{\beta}}, \hat{g}_j(z^T\hat{\boldsymbol{\beta}}_j), \hat{\boldsymbol{\alpha}}(s))$, which characterizes the asymptotic behavior of our estimators and is essential for constructing valid inferences based on these estimators.

Theorem 1—Suppose that Assumptions (1)–(9) hold. As $n \to \infty$, we have the following results:

i. $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to_d N(0, \Psi_\beta)$, *where* $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1^T, \ldots, \hat{\boldsymbol{\beta}}_J^T)^T$ *and the* (j, j')*th block in* Ψ_β *is*

$$[E\{\dot{g}_{j0}(Z^T\boldsymbol{\beta}_{j0})^2\omega_{2j}(Z)\}]^+ E[\dot{g}_{j0}(Z_1^T\boldsymbol{\beta}_{j0})\dot{g}_{j'0}(Z_2^T\boldsymbol{\beta}_{j'0})\omega_{1jj'}(Z_1, Z_2)\sigma^2_{\eta_{jj'}}] \times [E\{\dot{g}_{j'0}(Z^T\boldsymbol{\beta}_{j'0})^2\omega_{2j'}(Z)\}]^+,$$

in which $\sigma^2_{\eta_{jj'}} = \int_0^1\int_0^1 R_{jj'}(s,t)\,ds\,dt$.

ii. $\sqrt{nh_{j2}}\{\hat{g}_j(z^T\hat{\boldsymbol{\beta}}_j) - g_{j0}(z^T\boldsymbol{\beta}_{j0}) - 0.5u_2\ddot{g}_{j0}(z^T\boldsymbol{\beta}_{j0})h_{j2}^2\} \to_d N(0, \Psi_{g_j})$, *where* $\Psi_{g_j} = v_0\sigma^2_{\eta_{jj}}/f_j(z^T\boldsymbol{\beta}_{j0})$.

iii. $\sqrt{n}\{vec(\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_0(s) - 0.5u_2[\ddot{\boldsymbol{\alpha}}_0(s)\mathbf{H}_1^2 + \Gamma_2\mathbf{H}_2^2]) : s \in [0,1]\}$ *converges weakly to a mean-zero Gaussian process with covariance matrix* $R(s,s') \otimes \Gamma_1^{-1}$, *where* $\hat{\boldsymbol{\alpha}}(s) = [\hat{\alpha}_1(s), \ldots, \hat{\alpha}_J(s)]$, $\ddot{\boldsymbol{\alpha}}(s) = [\ddot{\alpha}_1(s), \ldots, \ddot{\alpha}_J(s)]$, $\Gamma_1 = E[X_1^{\otimes 2}]$, $\mathbf{H}_1 = diag(h_{11}, \ldots, h_{J1})$, $\Gamma_2 = \Gamma_1^{-1}\,E[X_1\ddot{g}_{10}(Z_1^T\beta_{10}), \ldots, X_1\ddot{g}_{J_0}(Z_1^T\beta_{J_0})]$, *and* $\mathbf{H}_2 = diag(h_{12}, \ldots, h_{J2})$.

The key challenge in proving Theorem 1 is to deal with within-subject dependence, the unknown link function, and the index parameter vector. In fact the covariance between $\boldsymbol{\eta}(s)$ and $\boldsymbol{\eta}(s')$ in the newly proposed multivariate varying coefficient model does not converge to zero under the within-curve dependence structure and thus requires special care in the derivation of the asymptotic distribution. Also, because of the iterative nature of the proposed estimation procedure, the proof of the single index component requires an explicit induction argument. In practice, the estimation of the unknown moment quantities involved in this theorem is highly non-trivial. We therefore recommend the bootstrap resampling method given in the previous section.

The consistency and asymptotic normality of the minimum average variance estimator (MAVE) for single index model were established in Xia et al. (2002) and Xia (2006). The

asymptotic results for the estimation of the single index component in our model are more general since our estimation reduces to MAVE under simpler settings. Similar to Xia (2006), we can show that Ψ_β achieves the semiparametric information lower bound under the exponential family when there is no within-subject dependence for the data. Although research interest is usually restricted on the coefficients, we also provide the asymptotic result for single index estimate $\hat{g}_j(\cdot)$ for the sake of completeness. Such a result may be used to quantify the estimation variability from a theoretical point of view. Finally, the estimation for the varying-coefficients functions in model (1) has a similar distributional result to Zhu et al. (2012) for multivariate functional data. We note because of the within-curve dependence the convergence speed for the estimated varying-coefficients is at the order $O(n^{-1/2})$ instead of $O((nMh_{j1})^{-1/2})$, which is usually the expected rate for cross-sectional data.

We next study the asymptotic bias and covariance of $\hat{\eta}_{ij}(s)$ as follows. We distinguish between two cases. The first case is to condition on the design points in $\mathcal{S}$, $\mathbf{X}$, $\mathbf{Z}$ and $\boldsymbol{\eta}$. The second case is to condition on the design points in $\mathcal{S}$, $\mathbf{X}$ and $\mathbf{Z}$. For the first case the conditional events involve the subject specific random effects. The bias and variance of the estimated random effects thus reflect individual heterogeneity. On the other hand, the second case does not condition on the sets of random effects and thus the bias and variance of the estimated random effects are considered for subjects randomly selected from a homogeneous population. The two cases may both be of interest for practitioners. We define $K^{\#}(t) = \int K(u)K(u+t)du$.

Theorem 2—Under Assumptions (1) and (3)–(9), the following results hold for all $s \in (0, 1)$.

i. *Conditioning on* $(\mathcal{S}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\eta})$*, we have*

$$Bias[\hat{\eta}_{ij}(s)|\mathcal{S},\boldsymbol{\eta},\mathbf{X},\boldsymbol{Z}]=0.5u_2\{\ddot{\eta}_{ij}(s)h_j^{*2}+X_i^T\ddot{\alpha}_{j0}(s)h_{j1}^2+\ddot{g}_{j0}(Z_i^T\boldsymbol{\beta}_{j0})h_{j2}^2\}\{1+o_p(1)\}+O_p(n^{-1/2}),$$

$$Cov[\hat{\eta}_{ij}(s),\hat{\eta}_{ij}(t)|\mathcal{S},\boldsymbol{\eta},\mathbf{X},\boldsymbol{Z}]$$
$$=K^{\#}((s-t)/h_j^*)\mathscr{E}_{jj}(s,s)\pi(t)^{-1}(Mh_j^*)^{-1}O_p(1)-X_i^T\Gamma_1^{-1}X_i(nMh_{j1})^{-1}O_p(1)-v_0(f(Z_i^T\boldsymbol{\beta}_{j0}))^{-1}\pi(s)^{-2}(nMh_{j2})^{-1}O_p(1).$$

ii. *The asymptotic bias and covariance of* $\hat{\eta}_{ij}(s)$ *conditional on* $\mathcal{S}$*,* $\mathbf{X}$ *and* $\mathbf{Z}$ *are given by*

$$Bias[\hat{\eta}_{ij}(s)|\mathcal{S},\mathbf{X},\boldsymbol{Z}]=0.5u_2\{X_i^T\ddot{\alpha}_{j0}(s)h_{j1}^2+\ddot{g}_{j0}(Z_i^T\boldsymbol{\beta}_{j0})h_{j2}^2\}\{1+o_p(1)\},$$
$$Cov(\hat{\eta}_{ij}(s)-\eta_{ij}(s),\hat{\eta}_{ij}(t)-\eta_{ij}(t)|\mathcal{S},\mathbf{X},\boldsymbol{Z})=\{1+o_p(1)\}\{0.25u_2^2h_j^{*4}R_{jj}^{(2,2)}(s,t)+$$
$$K^{\#}((s-t)/h_j^*)\times\pi(t)^{-1}(Mh_j^*)^{-1}O_p(1)+n^{-1}(X_i^T\Gamma_1^{-1}X_i)R_{jj}(s,t)\}.$$

iii. *The mean integrated squared error (MISE) of* $\hat{\eta}_{ij}$ *is given by*

$$n^{-1}\sum_{i=1}^{n}\int_0^1 E\{[\hat{\eta}_{ij}(s)-\eta_{ij}(s)]^2|\mathscr{S}\}\pi(s)ds$$
$$=\{1+o_p(1)\}\times[O((Mh_j^*)^{-1})+O_p(h_{j2}^4)+n^{-1}\int_0^1 R_{jj}(s,s)\pi(s)ds$$
$$+0.25u_2^2\int_0^1\{\ddot{\boldsymbol{\alpha}}_j(s)^T\Gamma_1\ddot{\boldsymbol{\alpha}}_j(s)h_{j1}^4+R_{jj}^{(2,2)}(s,s)h_j^{*4}\}\pi(s)ds]. \quad (29)$$

iv. *The optimal bandwidth to minimize MISE (29) is given by*

$$\hat{h}_j^*=O(M^{-1/5}). \quad (30)$$

v. *The first order local polynomial kernel reconstructions* $\hat{\eta}_{ij}(s)$ *using* $\hat{h}_j^*$ *in (30) satisfy*

$$\sup_{s\in[0,1]}|\hat{\eta}_{ij}(s)-\eta_{ij}(s)|=O_p(|\log(M)|^{1/2}M^{-2/5}+h_{j1}^2+h_{j2}^2+n^{-1/2}) \quad (31)$$

for $i = 1, \cdots, n$.

This theorem may be used to study the statistical property of the entity-specific effects in the functional regression analysis. Because the estimation of $\eta_{ij}(s)$ succeeds the estimation of the regression coefficients, the MISE of $\hat{\eta}_{ij}(s)$ appears to be a function of the three bandwidths h_{j1}, h_{j2} and h_j^*. If the optimal bandwidth in Theorem 2 (iv) is used, the resulting MISE can achieve the order of $M^{-4/5}+h_{j1}^4+h_{j2}^4+n^{-1}$. Practitioners may interpret the estimated random effects as how differently individual subjects behave from the population average.

We then present the asymptotic results for the refined estimates which fully acknowledge the dependence in the error process.

Theorem 3—Suppose that Assumptions (1)–(10) and (12) hold and $h_{j1}^\dagger$ and $h_{j2}^\dagger$ are of the same order as h_{j1} and h_{j2}, respectively. As $n \to \infty$, we have the following results:

i. $\sqrt{n}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\to_d N(0,\Psi_\beta^*)$, *where* $\tilde{\boldsymbol{\beta}}=(\tilde{\boldsymbol{\beta}}_1^T,\ldots,\tilde{\boldsymbol{\beta}}_J^T)^T$, *and the* (j, j')*th block in* Ψ_β^* *is*

$$[E\{\dot{g}_{j0}(Z^T\boldsymbol{\beta}_{j0})^2\omega_{2j}(Z)\int\int\mathscr{W}_j(s,t)dsdt\}]^+E\{\dot{g}_{j0}(Z_1^T\boldsymbol{\beta}_{j0})\dot{g}_{j'0}(Z_2^T\boldsymbol{\beta}_{j'0})$$
$$\times\omega_{1jj'}(Z_1,Z_2)\int\int\mathscr{W}_{jj'}(s,t)dsdt\}[E\{\dot{g}_{j'0}(Z^T\boldsymbol{\beta}_{j'0})^2\omega_{2j'}(Z)\int\int\mathscr{W}_{j'}(s,t)dsdt\}]^+,$$

where $\mathscr{W}_j(s,t)$ *and* $\mathscr{W}_{jj'}(s,t)$ *are defined in Assumption (12).*

ii. $\sqrt{nh_{j2}^\dagger}\{\tilde{g}_j(z^T\tilde{\boldsymbol{\beta}}_j)-g_{j0}(z^T\boldsymbol{\beta}_{j0})-0.5u_2\ddot{g}_{j0}(z^T\boldsymbol{\beta}_{j0})h_{j2}^{\dagger 2}\}\to_d N(0,v_0/\{f_j(z^T\boldsymbol{\beta}_{j0})\int\int\mathscr{W}_j(s,t)dsdt\})$.

iii. $\sqrt{n}\{vec(\tilde{\boldsymbol{\alpha}}(s)-\boldsymbol{\alpha}_0(s)-0.5u_2[\ddot{\boldsymbol{\alpha}}_0(s)\mathbf{H}_1^{\dagger 2}+\Gamma_2\mathbf{H}_2^{\dagger 2}]):s\in[0,1]\}$ *converges weakly to a mean-zero Gaussian process with covariance matrix* $R(s,s')\otimes\Gamma_1^{-1}$*, where* $\tilde{\boldsymbol{\alpha}}(s)=[\tilde{\alpha}_1(s),\ldots,\tilde{\alpha}_J(s)]$, $\mathbf{H}_1^{\dagger}=diag(h_{11}^{\dagger},\ldots,h_{J_1}^{\dagger})$, *and* $\mathbf{H}_2^{\dagger}=diag(h_{12}^{\dagger},\ldots,h_{J_2}^{\dagger})$

An important implication of Theorem 3 (i) is the efficiency gain in the use of the refined estimator $\tilde{\boldsymbol{\beta}}$ compared with $\hat{\boldsymbol{\beta}}$. That is, the individual elements of $\tilde{\boldsymbol{\beta}}$ have smaller asymptotic variance than the corresponding elements of $\hat{\boldsymbol{\beta}}$. See the proof of Lemma 5 in the Appendix. The refined procedure is thus more efficient in practice, especially when the goal of interest is to achieve an accurate estimation of the covariate effect.

Finally, we have the following asymptotic results for the test statistic T_a. By Theorem 1 (iii), $\sqrt{n}[\mathrm{A}\{R(s,s)\otimes\Gamma_1^{-1}\}\mathrm{A}^T]^{-1/2}\mathrm{d}(s)$ converges weakly to a Gaussian process. The results in Theorem 4 follow directly from Theorem 1 and Fubini theorem.

Theorem 4—Under Assumptions (1)–(10), we have that T_a converges weakly to $\int_0^1(s)^T(s)ds$ as $n\to\infty$, where $\mathbb{G}(s)$ is an l–dimensional Gaussian process. In particular, under H_0, $\mathbb{G}(s)$ is a zero-mean Gaussian process.

5 Simulation Study

In the following simulation study, the data were generated from model (1) with $J=2$, in which $s_m\sim U[0,1]$, $(\varepsilon_{i1}(s_m),\varepsilon_{i2}(s_m))^T\sim N((0,0)^T,\mathscr{E}(s_m,s_m)=\mathrm{diag}(\sigma_1^2,\sigma_2^2)$, and $X_i=(1, X_{i1}, X_{i2})$ for all $i=1,\ldots,n$ and $m=1,\ldots,M$. Moreover, X_{i1} and X_{i2} were independently generated from $N(0,1)$ and $\eta_{ij}(s)=\xi_{ij1}\psi_{j1}(s)+\xi_{ij2}\psi_{j2}(s)$, where $\xi_{ijl}\sim N(0,\lambda_{jl})$ for $j=1,2$ and $l=1,2$. We generated Z_i from a 4-dimensional standard normal distribution with the index parameter vectors $\beta_1=(2,1,3,-1)^T/\sqrt{15}$ and $\beta_2=(1,2,-3,1)^T/\sqrt{15}$. Furthermore, s_m, X_{i1}, X_{i2}, ξ_{i11}, ξ_{i12}, ξ_{i21}, ξ_{i22}, $\varepsilon_{i1}(s_m)$, and $\varepsilon_{i2}(s_m)$ are independent random variables. We set $(\lambda_{11},\lambda_{12},\sigma_1^2,\lambda_{21},\lambda_{22},\sigma_2^2)=(1.2,0.6,0.2,1,0.5,0.1)$ and the single index functions, eigenfunctions, and varying coefficient functions as follows:

$$g_1(z)=\cos(z),\ \ g_2(z)=\sin(z);\ \psi_{11}(s)=\sqrt{2}\sin(2\pi s),\ \ \psi_{12}(s)=\sqrt{2}\cos(2\pi s);$$
$$\psi_{21}(s)=\sqrt{2}\cos(2\pi s),\ \ \psi_{22}(s)=\sqrt{2}\sin(2\pi s);\alpha_1(s)=(s^2,(1-s)^2,-4s(1-s))^T,$$
$$\alpha_2(s)=(5(s-0.5)^2,s^{0.5},1.75\{\exp(-(3s-1)^2)+\exp(-(4s-3)^2)\}-0.75)^T.$$

We conducted extensive simulation studies under different settings and reported the estimation results for $n=200$ and 400 and $M=50$ after 500 simulations. We applied the estimation procedure in Section 2 to each simulated data set and calculated all unknown quantities. Table 1 summarizes the numerical performance of our estimators with sample size $n=200$, where we report mean absolute error (MAE) and root mean square error (RMSE) for estimated parameters and mean integrated absolute error (MIAE) and mean integrated squared error (MISE) for estimated functions. The results indicate satisfactory performance of our estimators since all MAE, RMSE, MIAE and MISE values are quite small. We notice that the refined estimators achieve smaller estimation error compared with the initial estimators. Typical estimated functions with median performance are displayed in

Figures 1 and 2. The estimated curves (broken lines) closely resemble the corresponding true functions (solid lines) in these figures. We also carried out some additional experiments under different sample sizes and obtained similar findings. Table 2 presents the estimation results for $n = 400$. As expected, increasing sample size decreases estimation error.

We then assess the finite-sample performance of T_a proposed in Section 3. We are interested in testing whether any varying-coefficient function can be constant. We consider the same data generating mechanism as in the previous simulations except that the varying coefficients are generated as follows:

$$\mathrm{vec}(\boldsymbol{\alpha}(s)) = \bar{\boldsymbol{\alpha}}_0 + \mathrm{k}(\boldsymbol{\alpha}_0(s) - \bar{\boldsymbol{\alpha}}_0), \quad (32)$$

where $\boldsymbol{\alpha}_0(s) = (a_{10}(s)^T, a_{20}(s)^T)^T$, in which $a_{10}(s) = (s^2, (1-s)^2, -4s(1-s))^T$ and $a_{20}(s) = (5(s-0.5)^2, s^{0.5}, 1.75\{\exp(-(3s-1)^2) + \exp(-(4s-3)^2)\} - 0.75)^T$, $\bar{\boldsymbol{\alpha}}_0 = E\{\boldsymbol{\alpha}_0(S)\}$, and k is a tuning parameter characterizing the distance between the null hypothesis and the alternative one. When k = 0, the model reduces to the null hypothesis, when all functions are constants. We set $B = 500$ in the proposed bootstrap test and report the power function of T_a based on 500 simulations in the left panel of Figure 3. Moreover, T_a attains the significance level under the null hypothesis and the power increases as k increases.

Finally, we consider the other example on the test of varying-coefficient functions. Specifically, we generated the coefficient functions via the following alternative model

$$\mathrm{vec}(\boldsymbol{\alpha}(s)) = \mathrm{k}\boldsymbol{\alpha}_0(s), \quad (33)$$

where $\boldsymbol{\alpha}_0(s)$ is the same as that in (32). Under the null hypothesis, all varying-coefficient functions are equal to zero as k = 0. We examine the test performance over a range of k values and also plot the power curve of the bootstrap test in the right panel of Figure 3. As expected, the proposed test preserves the specified significance level under the null hypothesis and its power increases with k.

6 Real Data Analysis

We used model (1) to analyze a real DTI data set with $n = 214$ subjects collected from NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) study[1]. The NIH ADNI is an ongoing public-private partnership to test whether genetic, structural and functional neuroimaging,

[1]Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year publicprivate partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over

and clinical data can be integrated to assess the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The structural brain MRI data and corresponding clinical and genetic data from baseline and follow-up were downloaded from the ADNI publicly available database (https://ida.loni.usc.edu/). The demographic information about the data set in this paper is presented in Table 3.

The DTI data were processed by two key steps including a weighted least squares estimation method (Basser et al., 1994; Zhu et al., 2007) to construct the diffusion tensors and a FSL TBSS pipeline (Smith et al., 2006) to register DTIs from multiple subjects to create a mean image and a mean skeleton. Specifically, maps of fractional anisotropy (FA) were computed for all subjects from the DTI after eddy current correction and automatic brain extraction using FMRIB software library. FA maps were then fed into the TBSS tool, which is also part of the FSL. In the TBSS analysis, the FA data of all the subjects were aligned into a common space by non-linear registration and the mean FA image were created and thinned to obtain a mean FA skeleton, which represents the centers of all WM tracts common to the group. Subsequently, each subjects aligned FA data were projected onto this skeleton.

We focus on the midsagittal corpus callosum skeleton and associated FA curves from all subjects at $M = 83$ location points as shown in Fig. 4. The corpus callosum (CC) is the largest fiber tract in the human brain and is a topographically organized structure. It is responsible for much of the communication between the two hemispheres and connects homologous areas in the two cerebral hemispheres. It is important in the transfer of visual, motoric, somatosensory, and auditory information.

We are interested in testing the association between FA and diagnostic groups (MCI and AD) using model (1). Specifically, the X_i vector includes an intercept term and the gender variable (coded by a dummy variable indicating for male) and the Z_i vector includes the age of the subject (years), an indicator for handiness (coded by a dummy variable indicating for left-hand), the education level (years), an indicator for Alzheimer's disease (AD) status (19.6%) and an indicator for mild cognitive impairment (MCI) status (55.1%). We standardized all continuous variables to be mean zero and variance one.

The estimated varying-coefficients models are shown in Figure 5, along with 95% bootstrap confidence bands. The intercept function characterizes the nonlinear trend of FA values. The estimated coefficient function for sex suggests that the difference of FA values between men and women is not a constant. Since the coefficient curve is positive at most of the grid points, it may indicate that men tend to have higher FA values than women. This finding is consistent with the previous analysis with functional varying-coefficient models (Zhu et al. (2012)). On the other hand, when we include sex in Z for the single index, its estimated coefficient is of size 0.2236 and in the same direction.

50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

The estimated single index function is presented in Figure 5. Table 4 presents the estimated coefficients for the single-index variables and their 95% confidence intervals. Since the estimated single index function is decreasing, the estimated positive coefficients in Table 4 indicate negative association between FA and the five covariates of interest in Z_i. Being older, left-hand or more educated may lead to smaller FA values. AD and MCI patients tend to have lower FA values than normal subjects, while AD patients decreases more than MCI patients. For comparison, we also present the estimated coefficients and 95% bootstrap confidence interval for the refined estimates. The regression coefficients are of similar directions and magnitudes and their confidence intervals are all much narrower than those without the covariance adjustment.

Six non-zero eigenvalues of the covariance were estimated in a descending order to be 0.2294, 0.0290, 0.0255, 0.0107, 0.0086, and 0.0041, respectively. The first three eigenvalues account for 92.3% of the total variability and the remaining eigenvalues rapidly drop to zero. The estimated eigenfunctions corresponding to these eigen-values are shown in Figure 6. The first eigen-function, with a dominant eigenvalue accounting for 74.6% of the total variation, is simple in structure and resembles a single cycle of a sine wave. While the remaining eigen-functions are also quite simple and roughly sinusoidal, they contain more and more cycles.

7 Discussion

We have proposed a novel functional varying-coefficient single index model (FVCSIM) to carry out the regression analysis of functional responses on a set of covariates of interest (e.g., time). The varying-coefficient and single index components in FVCSIM allow us to accommodate the dynamic effect of X and the stationary effect of Z on functional data. We have developed an efficient estimation procedure to iteratively estimate varying coefficient functions, link functions, index parameter vectors, and the covariance function of individual functions. We have systematically examined the asymptotic properties of all estimators in FVCSIM. Through simulation studies and a real data example, we have shown that FVCSIM is a valuable statistical model for quantifying the complex relationships between imaging data and clinical variables of interest.

Several important issues need to be addressed in future research. First, we will extend model (1) to FVCSIM with high-dimensional covariates X and/or Z. Such extension is extremely critical for quantifying the effects of a huge number of genetic markers on imaging phetotype data. Second, we will extend model (1) to functional single index models with varying index parameters $\boldsymbol{\beta}_j(s)$. Moreover, we will consider high-dimensional covariate vector Z under such models and include regularization terms to incorporate the spatial smoothness and sparsity. This allows us to achieve dimension reduction and variable selection for complicated functional responses. Third, it is meaningful to extend model (1) from cross-sectional functional data to longitudinal functional data. Many complexities and new statistical tools will definitely emerge from these new developments.

Acknowledgments

Dr. Zhu's work was partially supported by NIH grants MH086633 and 1UL1TR001111 and NSF grants SES-1357666 and DMS-1407655. Dr. Li's research work was partially supported by National Medical Research Council NMRC/CBRG/0014/2012 and Academic Research Funding R-155-000-152-112. This material was based upon work partially supported by the NSF grant DMS-1127914 to the Statistical and Applied Mathematical Science Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We are grateful for the many valuable suggestions from referees, associated editor, and editor.

References

Basser PJ, Mattiello J, LeBihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. Journal of Magnetic Resonance, Ser B. 1994; 103:247–254.

Buzsaki, G. Rhythms of The Brain. Oxford University Press; 2011.

Carroll RJ, Fan J, Gijbels I, Wand MP. Generalized partially linear single-index models. J Amer Statist Assoc. 1997; 92:477–489.

Cheng MY, Toshio H, Li J, Peng H. Nonparametric independence screening and structural identification for ultra-high dimensional longitudinal data. Annals of Statistics. 2014; 42:1819–1849.

Cheng MY, Zhang W, Chen L. Statistical estimation in generalized multi-parameter likelihood models. Journal of the American Statistical Association. 2009; 104:1179–1191.

Cui X, Härdle WK, Zhu LX. The EFM approach for single-index models. Annals of Statistics. 2011; 39:1658–1688.

Fan, J., Gijbels, I. Local Polynomial Modelling and Its Applications. Chapman and Hall; London: 1996.

Fan J, Huang T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli. 2005; 11:1031–1057.

Fan J, Yao Q. Efficient estimation of conditional variance functions in stochastic regression. Biometrika. 1998; 85:645–660.

Fan J, Yao Q, Cai Z. Adaptive varying-coefficient linear models. J R Stat Soc Ser B Stat Methodol. 2003; 65:57–80.

Fan J, Zhang W. Statistical estimation in varying coefficient models. The Annals of Statistics. 1999; 27:1491–1518.

Fan J, Zhang W. Statistical methods with varying coefficient models. Stat Interface. 2008; 1:179–195. [PubMed: 18978950]

Friston KJ. Modalities, modes, and models in functional neuroimaging. Science. 2009; 326:399–403. [PubMed: 19833961]

Hall P, Marron JS. On variance estimation in nonparametric regression. Biometrika. 1990; 77:415–419.

Hall P, Müller HG, Wang JL. Properties of principal component methods for functional and longitudinal data analysis. The Annals of Statistics. 2006; 34:1493–1517.

Härdle W, Hall P, Ichimura H. Optimal smoothing in single-index models. Ann Statist. 1993; 21:157–178.

Heywood, I., Cornelius, S., Carver, S. An Introduction to Geographical Information Systems. Pearson Prentice Hall; 2006.

Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika. 1998; 85:809–822.

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

Jiang CR, Wang JL. Covariate-adjusted functional principal components analysis for longitudinal data. The Annals of Statistics. 2010; 38:1194–1226.

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

Li Y, Hsing T. Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. The Annals of Statistics. 2010b; 38:3028–3062.

Ma Y, Zhu L. A review on dimension reduction. International Statistics Review. 2013; 81:134–150.

Mack YP, Silverman BW. Weak and strong uniform consistency of kernel regression estimates. Z Wahrsch Verw Gebiete. 1982; 61:405–415.

Morris JS, Carroll RJ. Wavelet-based functional mixed models. J R Stat Soc Ser B. 2006; 68:179–199.

Niedermeyer, E., da Silva, FL. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Lippincot Williams & Wilkins; 2005.

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

Rosasco L, Belkin M, Vito ED. On learning with integral operators. Journal of Machine Learning. 2010; 11:905–934.

Shao, J. Mathematical Statistics. Springer-Verlag; New York: 1999.

Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TEJ. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. NeuroImage. 2006; 31:1487–1505. [PubMed: 16624579]

Towle VL, Bolaños J, Suarez D, Tan K, Grzeszczuk R, Levin DN, Cakmur R, Frank SA, Spire JP. The spatial location of EEG electrodes: locating the best-fitting sphere relative to cortical anatomy. Electroencephalogr Clin Neurophysiol. 1993; 86:1–6. [PubMed: 7678386]

van der Vaart, AW. Asymptotic Statistics. Cambridge University Press; 1998.

Wang JL, Xue L, Zhu L, Chong YS. Estimation for a partial-linear single-index model. The Annals of Statistics. 2010; 1:246–274.

Wang L, Li H, Huang JZ. Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. Journal of the American Statistical Association. 2008; 103:1556–1569. [PubMed: 20054431]

Wang Q, Xue L. Statistical inference in partially-varying-coefficient single-index model. Journal of Multivariate Analysis. 2011; 102:1–19.

Wong H, Ip W, Zhang R. Varying-coefficient single-index model. Computational Statistics and Data Analysis. 2008; 52:1458–1476.

Wu CO, Chiang CT. Kernel smoothing on varying coefficient models with longitudinal dependent variable. Statist Sinica. 2000; 10:433–456.

Xia Y. Asymptotic distributions for two estimators of the single-index model. Econometric Theory. 2006; 22:1112–1137.

Xia Y, Tong H, Li W, Zhu LX. An adaptive estimation of dimension reduction space. J R Stat Soc Ser B Stat Methodol. 2002; 64:363–410.

Zhang W. Identification of the constant components in generalised semivarying coefficient models by cross–validation. Statistica Sinica. 2011; 21:1913–1929.

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

Zhu H, Zhang H, Ibrahim JG, Peterson BG. Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data (with discussion). Journal of the American Statistical Association. 2007; 102:1085–1102.

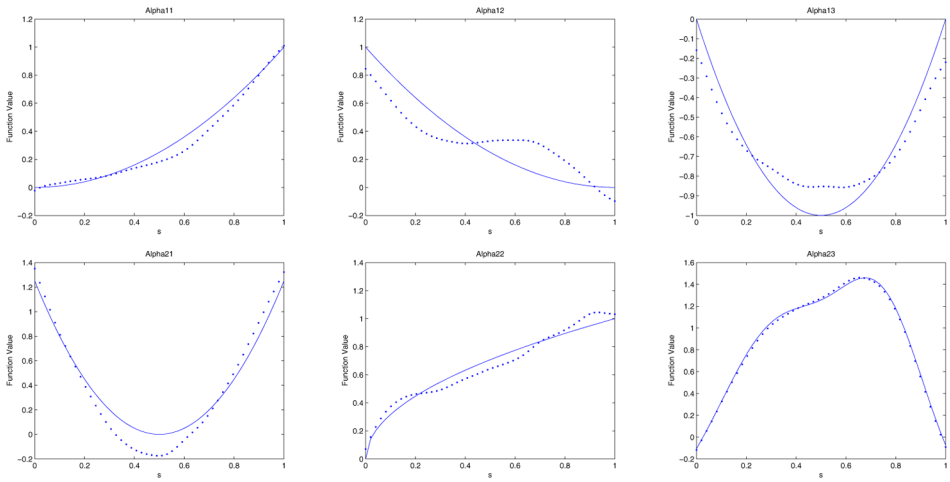

Figure 1.

Estimated varying coefficients with median MISE over 500 simulations. The top panel is for $\alpha_1(s) = (s^2, (1-s)^2, -4s(1-s))^T$, whereas the bottom panel is based on $\alpha_2(s) = (5(s-0.5)^2, s^{0.5}, 1.75(\exp(-(3s-1)^2) + \exp(-(4s-3)^2)) - 0.75)^T$. The solid lines are true functions and the broken lines are estimates.

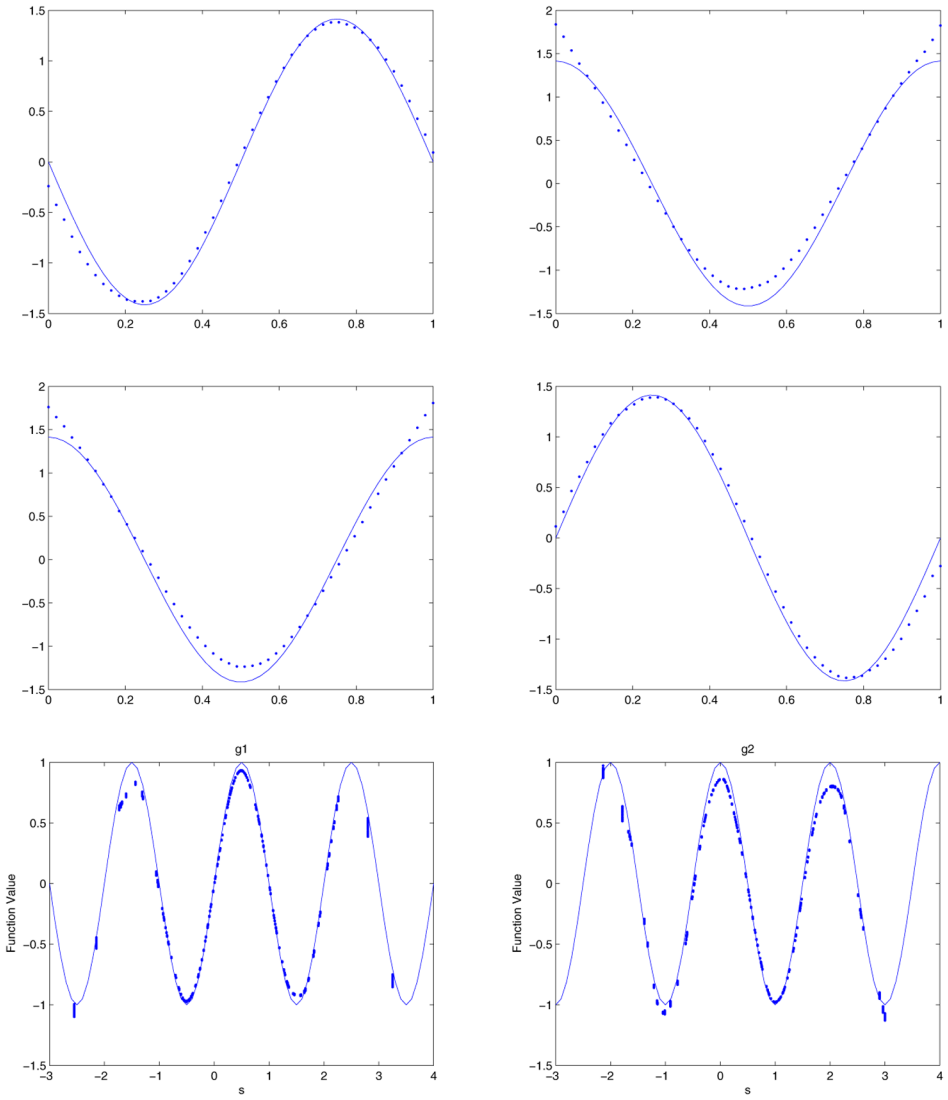

Figure 2.

The estimated eigen-functions (top two panels) and single-index functions (bottom panel) with median MISE over 500 simulations. The solid lines are the true functions and the broken lines are the estimated ones.

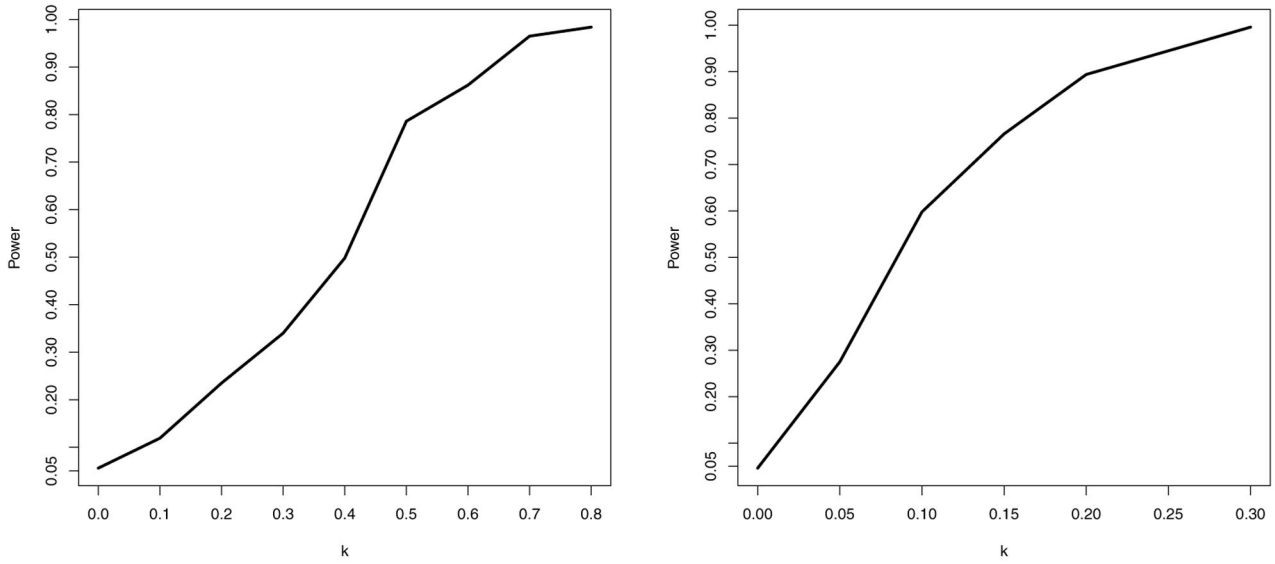

Figure 3.

The power curves for testing hypotheses (32) (left panel) and (33) (right panel), respectively. In both panels, k = 0 corresponds to the null hypothesis. The significance level is set at α = 0.05. Bootsrap sample size B = 500 and results are based on 500 simulations.

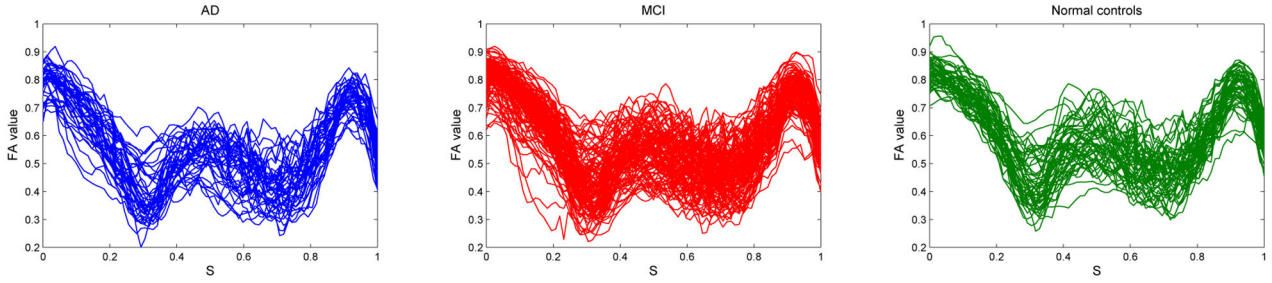

Figure 4.

Representative FA imaging data for ADNI study: FA template curves measured at 83 grid points along the midsagittal skeleton of the corpus callosum (black) of subjects in 3 groups (from left to right): AD (left), MCI (middle) and NC (right).

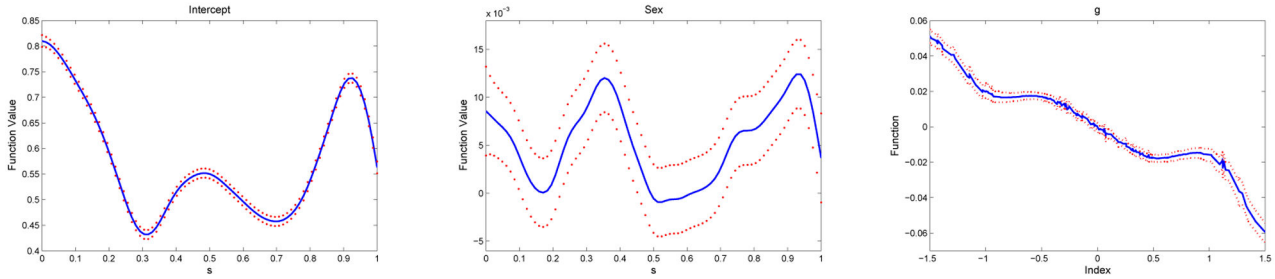

Figure 5.

ADNI data analysis: the estimated varying-coefficient functions for interecept and sex (left and middle) and the estimated single-index function (right) for the ADNI data. Solid lines are the estimates and dashed lines are the 95% bootstrap confidence intervals.

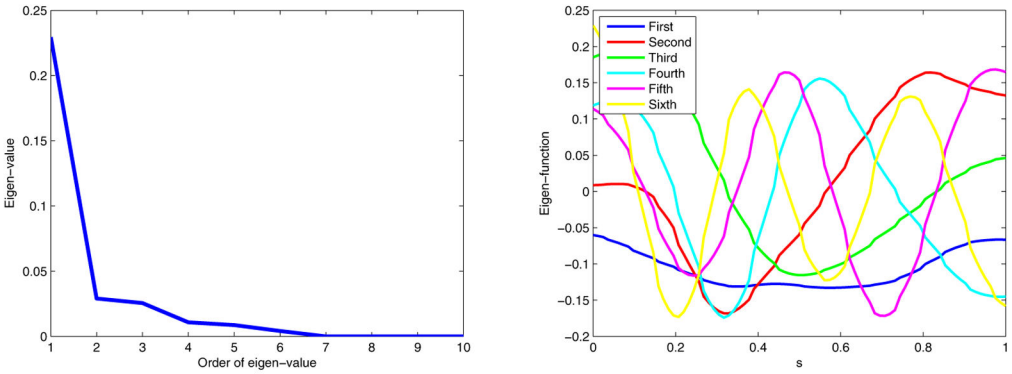

Figure 6.

ADNI data analysis: the estimated eigen-values in descending order (left) and estimated eigen-functions (right) corresponding to the six largest eigen-values.

Table 1

Estimation results for $n = 200$ obtained from 500 simulations: MAE is mean absolute error; RMSE is root mean square error; MIAE is mean integrated absolute error; and MISE is mean integrated square error. The values in bold-face are obtained from the refined estimates incorporating the estimated covariance.

Parameters	β_{11}	β_{12}	β_{13}	β_{14}		
MAE	0.0030	0.0033	0.0025	0.0031		
	0.0026	**0.0027**	**0.0019**	**0.0025**		
RMSE	0.0036	0.0043	0.0030	0.0044		
	0.0028	**0.0034**	**0.0025**	**0.0033**		
Parameters	β_{21}	β_{22}	β_{23}	β_{24}		
MAE	0.0032	0.0019	0.0017	0.0040		
	0.0028	**0.0015**	**0.0013**	**0.0032**		
RMSE	0.0039	0.0031	0.0024	0.0043		
	0.0031	**0.0029**	**0.0019**	**0.0032**		
Parameters	λ_{11}	λ_{12}	λ_{21}	λ_{22}	σ_1^2	σ_2^2
MAE	0.038	0.058	0.014	0.055	0.047	0.042
RMSE	0.052	0.048	0.041	0.032	0.038	0.034
Functions	$a_{11}(s)$	$a_{12}(s)$	$a_{13}(s)$	$a_{21}(s)$	$a_{22}(s)$	$a_{23}(s)$
MIAE	0.0466	0.0352	0.0480	0.0495	0.0279	0.0357
MISE	0.0102	0.0076	0.0265	0.0272	0.0057	0.0105
Functions	$\psi_{11}(s)$	$\psi_{12}(s)$	$\psi_{21}(s)$	$\psi_{22}(s)$	$g_1(z)$	$g_2(z)$
MIAE	0.0582	0.0355	0.0649	0.0562	0.0542	0.0558
MISE	0.0245	0.0167	0.0182	0.0119	0.0204	0.0213

Table 2

Estimation results for $n = 400$ obtained from 500 simulations. MAE is mean absolute error; RMSE is root mean square error; MIAE is mean integrated absolute error; MISE is mean integrated square error. The values in bold-face are obtained from the refined estimates incorporating the estimated covariance.

Parameters	β_{11}	β_{12}	β_{13}	β_{14}		
MAE	0.0022	0.0023	0.0017	0.0020		
	0.0019	**0.0021**	**0.0012**	**0.0014**		
RMSE	0.0024	0.0036	0.0021	0.0035		
	0.0022	**0.0027**	**0.0016**	**0.0024**		
Parameters	β_{21}	β_{22}	β_{23}	β_{24}		
MAE	0.0020	0.0014	0.0012	0.0031		
	0.0015	**0.0013**	**0.0011**	**0.0015**		
RMSE	0.0031	0.0026	0.0018	0.0035		
	0.0022	**0.0021**	**0.0013**	**0.0021**		
Parameters	λ_{11}	λ_{12}	λ_{21}	λ_{22}	σ_1^2	σ_2^2
MAE	0.026	0.044	0.011	0.048	0.035	0.031
RMSE	0.043	0.038	0.036	0.024	0.032	0.025
Functions	$a_{11}(s)$	$a_{12}(s)$	$a_{13}(s)$	$a_{21}(s)$	$a_{22}(s)$	$a_{23}(s)$
MIAE	0.0367	0.0246	0.0291	0.0368	0.0225	0.0269
MISE	0.0074	0.0052	0.0184	0.0193	0.0036	0.0085
Functions	$\psi_{11}(s)$	$\psi_{12}(s)$	$\psi_{21}(s)$	$\psi_{22}(s)$	$g_1(z)$	$g_2(z)$
MIAE	0.0462	0.0270	0.0558	0.0449	0.0436	0.0472
MISE	0.0158	0.0106	0.0113	0.0086	0.0152	0.0169

Table 3

Demographic information about ADNI data, including disease status, range of age (RA), gender, handiness and Range of education level (REL).

Disease status	Num.	RA(mean)	Gender(F/M*)	Handiness(R/L**)	REL in years (mean)
AD	42	55.7–90.4 (75.5)	15/27	38/4	9–20 (15.4)
MCI	118	48.4–88.6 (72.6)	46/72	106/12	11–20 (15.9)
Normal control	54	59.9–89.1 (72.5)	30/24	50/4	12–20 (16.4)
All data	214	48.4–90.4(73.2)	91/123	194/20	9–20 (15.9)

*F=Female; M=Male.

**R=Right; L=Left.

J Am Stat Assoc. Author manuscript; available in PMC 2017 December 01.

Table 4

Estimation results of regression coefficients for ADNI data. 95% confidence intervals are based on 500 bootstrap resamples. The values in bold-face are obtained from the refined estimation incorporating the estimated covariance.

	Age	**Handiness**	**Education**	**AD**	**MCI**
Coefficients	.6589	.2182	.3053	.5669	.3222
	.6756	**.2251**	**.2824**	**.5772**	**.2828**
95% CI	[.636, .682]	[.195, .236]	[.276, .334]	[.547, .583]	[.302, .346]
	[.655, .678]	**[.215, .241]**	**[.270, .297]**	**[.573, .592]**	**[.270, .295]**