

## Statistica Sinica Preprint No: SS-2023-0152

<b>Title</b>	Functional Joint Models for Imaging Genetic Data
<b>Manuscript ID</b>	SS-2023-0152
<b>URL</b>	<a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a>
<b>DOI</b>	10.5705/ss.202023.0152
<b>Complete List of Authors</b>	Qingzhi Zhong, Xinyuan Song and Hongtu Zhu
<b>Corresponding Authors</b>	Hongtu Zhu
<b>E-mails</b>	htzhu@email.unc.edu

## Functional Joint Models for Imaging Genetic Data

Qingzhi Zhong<sup>1</sup>, Xinyuan Song<sup>2\*</sup>, Hongtu Zhu<sup>3,4,5,6\*</sup>

<sup>1</sup>*School of Economics, Jinan University, Guangzhou, China.*

<sup>2</sup>*Department of Statistics, The Chinese University of Hong Kong,*

*Shatin, N.T., Hong Kong, P.R. China.*

*Departments of Biostatistics<sup>3</sup>, Genetics<sup>4</sup>, Computer Science<sup>5</sup>, and Statistics<sup>6</sup>*

*University of North Carolina at Chapel Hill, North Carolina, USA*

*Abstract:* We propose a functional joint modeling (FJM) framework for correlating imaging responses with genetic markers and clinical variables. Our FJM consists of a nonlinear multivariate functional principal component analysis (NMFPCA) and a functional multiple-index varying coefficient model (FMVCM). The NMFPCA, with unknown link functions, is used to extract meaningful functional principal component (FPC) scores of genetic markers, while the FMVCM identifies the varying association of the extracted FPC scores and clinical variables with imaging data. We propose an efficient estimation procedure to estimate unknown functions in our FJM and a regularization approach to simultaneously select relevant features from infinite-dimensional functional data and learn the model structure. The asymptotic convergence rate of estimators and model selection consistency are investigated. The proposed method is evaluated through simulation studies and applied to an imaging genetic data set extracted from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

---

\*These authors are jointly supervised this work: Xinyuan Song and Hongtu Zhu, Emails for Correspondence: xysong@cuhk.edu.hk and htzhu@email.unc.edu.

---

*Key words and phrases:* Functional joint model, Functional principal component, Imaging genetics, Neuroimaging data analysis, Varying coefficient.

## 1. Introduction

This study aims to integrate multimodal data to clarify the biological pathways that connect genetic factors to brain and, ultimately, to clinical outcomes (e.g., cognition, disease stage, and progression status) (Elliott et al., 2018; Zhu et al., 2023). By exploring these connections, we hope to gain valuable insights into brain development, healthy aging, and disease progression. One such pathway is reflected in Jack's hypothetical models for Alzheimer's disease (AD) (Jack Jr et al., 2013). This information could offer a comprehensive view of how behavioral outcomes relate to the genetic pathways of AD, as mediated by brain structure and function. Additionally, it could deepen our understanding of the interplay between genetic and environmental risk factors, and how modifying environmental influences might benefit individuals with genetic predispositions. Ultimately, this approach aims to inform more effective interventions and treatments (Veitch et al., 2022).

The hippocampus is particularly vulnerable to AD pathology and is among the first brain regions affected in early AD stages. The degree of hippocampal atrophy can vary between healthy individuals and those with neurological conditions, showing differences across age, gender, genotype, and APOE- $\epsilon$ 4 status. APOE- $\epsilon$ 4, a variant of the apolipoprotein E gene, is a major genetic risk factor for AD. Thus, understanding the relationship between hippocampal atrophy and genetic and behavioral factors is crucial. Due to the

vast number of single-nucleotide polymorphisms (SNPs) compared to sample sizes, and the categorical nature of SNP allele codes (0, 1, 2), associations between hippocampal morphometry and clinical or genetic data are likely nonlinear. As shown in Web Figure 3 and Tables 3 and 4 in Section 6, SNPs exhibit highly nonlinear patterns. Modeling these nonlinear relationships could improve the accuracy of predicting hippocampal atrophy. However, traditional linear regression cannot capture these complex associations between hippocampal measures and genetic/clinical predictors. To address this, we propose a novel FJM framework incorporating NMFPCA and FMVCM, offering a concise, informative, and interpretable approach for predicting hippocampal atrophy.

Several primary challenges arise in image response regression with genetic and clinical covariates. First, genetic markers are ultra-high dimensional and present a strong blockwise correlation (Wall and Pritchard, 2003). If all the SNPs within a specific linkage disequilibrium (LD) block are important but exhibit relatively weak signals, individually analyzing them may miss these SNPs. The second challenge arises from the infinite-dimensional nature of the image response, which complicates the characterization of the effects of genetic and clinical covariates with spatial structure. Conventional pixel- or voxel-wise approaches independently fit a model at each pixel or voxel and ignore their spatial structure, resulting in low power for identifying significant spatial patterns (Stein et al., 2010; Hibar et al., 2015). The third one is to smooth data over an irregular domain with complex boundaries in imaging data analysis.

Most existing regression methods with genetically measured covariates ignore the com-

bined significance of SNPs (Kang et al., 2019; Li et al., 2021, 2024). To appropriately utilize the LD block's structure information, we treat all the SNP values of each LD block as a functional variable and perform dimension reduction to deal with the ultra-high dimensionality of genetic markers. Popular linear and nonlinear unsupervised dimension reduction models (DRMs) such as multidimensional scaling, Laplacian eigenmaps, principal component analysis (PCA), kernel PCA, and functional PCA (FPCA) (Anowar et al., 2021; Zhu et al., 2023). Standard FPCA, which uses linear projection to find major sources of variation in functional data, has received much attention in the literature due to its computational simplicity and appealing theoretical properties (Ramsay and Silverman, 2005; Zhu et al., 2014, 2023). However, it requires FPCs to be linear combinations of original variables. When functional variables are nonlinearly correlated, FPCA fails to extract their nonlinear features and cannot satisfactorily represent the original functional data through leading FPCs, resulting in substantial bias or erroneous conclusions in estimation or classification.

Functional data analysis (FDA) methods treat pixels or voxels as discrete grids of a function, integrating information across and within functions to capture the spatial structure of imaging responses. However, existing FDA approaches for imaging responses primarily model linear relationships and low-dimensional covariates (Yu et al., 2021; Li et al., 2021; Zhu et al., 2023). While recent nonlinear variants of the functional varying coefficient model (FVCM) have emerged, they are limited to fixed-dimension covariates (Luo et al., 2016; Yu et al., 2019; Kim and Wang, 2021). This study extends these methods

to accommodate functional or high-dimensional latent covariates.

Furthermore, for many biomedical images, the regions of interest on images are usually irregularly shaped with complex domains and interior holes; the image contains only signals inside the region boundary. However, several popular smoothing methods, such as wavelet smoothing (Morris and Carroll, 2006), tensor product smoothing (Reiss and Ogden, 2010), and kernel smoothing (Zhu et al., 2023), suffer from the so-called “leakage” problem and perform poorly over complex domains (Sangalli et al., 2013). Therefore, developing reliable techniques to smooth data over an irregular domain with complex boundaries is particularly relevant in the current image response regression.

This paper aims to address the challenges of understanding the association between clinical and genetic variables and hippocampal atrophy in patients with Alzheimer’s disease. Suppose we observe a data set from  $n$  independent subjects consisting of image responses  $Y_i$ , a vector of functional covariates  $\mathbf{X}_i$ , and a vector of real-valued covariates  $\mathbf{Z}_i$  for  $i = 1, \dots, n$ . All images  $Y_i = \{Y_i(\mathbf{s}) : \mathbf{s} = (s_1, s_2)^T \in \mathcal{D}\}$  are registered and measured at the same set of points  $\mathbf{s}_1, \dots, \mathbf{s}_{N_s}$  in a compact domain  $\mathcal{D}$ ,  $\mathbf{X}_i(\mathbf{t}_i) = \{X_{i1}(t_{i1}), \dots, X_{ip}(t_{ip})\}^T \in \mathbb{R}^p$  with  $t_{ij} \in \mathcal{T}$ ,  $X_{ij}(t_{ij})$  is a functional variable observed at discrete points  $t_{ij1}, \dots, t_{ij, n_{ij}}$ , and  $\mathbf{Z}_i \in \mathbb{R}^{q_z}$  is a  $q_z \times 1$  covariate vector for the  $i$ th subject. Without loss of generality,  $\mathcal{D}$  and  $\mathcal{T}$  are assumed to be compact sets in  $\mathbb{R}^2$  and  $\mathbb{R}$ , respectively. In our ADNI data set,  $Y_i$  is the left (or right) hippocampus surface,  $X_{ij}(t_j)$  is the function of SNPs in a specific LD block, and  $\mathbf{Z}_i$  is the vector of covariates (e.g., age, gender, and education). To model  $Y_i(\mathbf{s})$  given  $(\mathbf{X}_i, \mathbf{Z}_i)$ , we consider an FJM

framework consisting of a DRM and an FVCM given by

$$\mathbf{X}_i = \mathbf{F}(\boldsymbol{\zeta}_i) + \mathbf{E}_{ix} \quad \text{and} \quad Y_i(\mathbf{s}) = m(\mathbf{s}, \boldsymbol{\zeta}_i, \mathbf{Z}_i) + \epsilon_i(\mathbf{s}), \quad (1.1)$$

respectively, where  $\boldsymbol{\zeta}_i$  is a relatively low-dimensional vector of latent variables,  $\mathbf{F}(\cdot)$  is the conditional mean function of  $\mathbf{X}_i$  given  $\boldsymbol{\zeta}_i$ ,  $m(\cdot, \cdot, \cdot)$  is the conditional mean function of  $Y_i(\mathbf{s})$  given  $(\mathbf{X}_i, \mathbf{Z}_i)$ ,  $\mathbf{E}_{ix}$  is a zero mean measurement error process independent of  $\boldsymbol{\zeta}_i$ , and  $\epsilon_i(\mathbf{s})$  is a zero mean measurement error independent of  $(\mathbf{X}_i, \mathbf{Z}_i)$ . Such FJM is a powerful tool for handling functional covariates  $\mathbf{X}_i$  and functional response data  $Y_i(\mathbf{s})$ .

Our proposal contributes to the existing literature in several aspects. First, for DRM, we propose to apply NMFPCA on functional covariates  $\mathbf{X}_i$  for nonlinear dimension reduction while allowing the number of latent variables to diverge to infinity. In contrast, many existing FJMs consider linear DRMs and truncate the first few FPCs with large eigenvalues (Müller and Yao, 2008; Zhu et al., 2023). Despite its simplicity, such a naive truncation procedure may encounter difficulties because the effect of latent scores on imaging measures may not coincide with the proportion of variation explained by functional variables (Bair et al., 2006). Second, for FVCM, we propose an FMVCM to delineate complex associations between imaging responses and covariates. When the FPC scores are observable, FMVCM reduces to the existing FVCMs. We adopt bivariate splines over triangulations (Li et al., 2021) to approximate coefficient images defined over a complex domain. We develop a two-step estimation procedure to integrate the advantages of B-spline, kernel,

and bivariate spline smoothing techniques. Third, we establish the theoretical properties of our proposed estimators, including model identifiability, the truncated nonlinear Karhunen–Loève approximation for  $\mathbf{X}_i(\mathbf{t})$ , estimation and selection consistency, and the convergence rates of estimation and selection.

The rest of this paper is organized as follows. Section 2 describes the key components of our FJM. Section 3 presents the estimation procedure, and Section 4 establishes the asymptotic properties of the proposed estimators. Sections 5 and 6 illustrate the empirical performance of the proposed methods through simulation studies and an application to the ADNI data set, respectively. Section 7 concludes the paper. The technical proofs are deferred to Web Appendices.

## 2. Model Setup

We introduce the two components of FJM, including NMFPCA and FMVCM, as follows.

### 2.1 NMFPCA

We jointly model multiple functional curves from the realization of multiple correlated random functions with mean function  $\boldsymbol{\mu}(t) = \{\mu_1(t), \dots, \mu_p(t)\}^T = E\{\mathbf{X}_i(t)\}$  and matrix-valued covariance function  $\mathbf{C}(s, t) = \text{cov}\{\mathbf{X}_i(s), \mathbf{X}_i(t)\} \in \mathbb{R}^p \times \mathbb{R}^p$ . Conventional MFPCA methods assume an orthogonal expansion  $\mathbf{C}(s, t) = \sum_{k=1}^{\infty} \lambda_k \boldsymbol{\varphi}_k(s) \boldsymbol{\varphi}_k(t)^T$  and develop DRM for  $\mathbf{X}_i(\cdot)$  using the truncated Karhunen–Loève representation (Happ and Greven,

2018) given by

$$\mathbf{X}_i(t) = \boldsymbol{\mu}(t) + \sum_{k=1}^{K_n} \xi_{ik} \boldsymbol{\varphi}_k(t) + \boldsymbol{\varepsilon}_i(t), \quad t \in \mathcal{T}, \quad (2.1)$$

where  $K_n$  is a positive integer,  $\boldsymbol{\varphi}_k(t)$  is the  $k$ th orthonormal eigenfunction of  $\mathbf{C}(s, t)$ ,  $\lambda_k$  is the eigenvalue corresponding to  $\boldsymbol{\varphi}_k$ ,  $\xi_{ik} = \int_{\mathcal{T}} \{\mathbf{X}_i(t) - \boldsymbol{\mu}(t)\}^T \boldsymbol{\varphi}_k(t) dt$  are uncorrelated over  $i$  and  $k$ , and  $\boldsymbol{\varepsilon}_i(t)$  is measurement error independent of  $\xi_{ik}$ . Model (2.1) has received much attention in the literature (Müller and Yao, 2008; Happ and Greven, 2018). Moreover, multivariate functional data are projected into the linear space spanned by multivariate FPCs  $(\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots)$  and represented as a linear function of the multivariate FPC scores  $(\xi_{i1}, \xi_{i2}, \dots)$ . The FPC scores are typically used to characterize the functional curve  $\mathbf{X}_i(\cdot)$  for downstream regression or clustering analysis. Although using  $\boldsymbol{\mu}(t) + \sum_{k=1}^{K_n} \xi_{ik} \boldsymbol{\varphi}_k(t)$  to approximate functional curves in model (2.1) implies that the functional variables are linearly dependent, such a linearity assumption may be too restrictive to reflect various realistic scenarios. Therefore, it is of scientific interest and practical value to identify the complex dependence structure of functional variables and construct a highly informative vectorized representation for functional data with minimal loss.

We propose an NMFPCA procedure with unknown link functions to flexibly model the data dependence and extract nonlinear information from functional data. Suppose functional variables are correlated by sharing latent scores. The relationships between

functional variables and projection spaces are modeled through unknown link functions:

$$X_{ij}(t) = \mu_j(t) + f_j \left\{ \sum_{k=1}^{\infty} \xi_{ik} \phi_{jk}(t) \right\} + \varepsilon_{ij}(t), \quad t \in \mathcal{T}, \quad (2.2)$$

where  $\mu_j(t) = E \{X_{ij}(t)\}$  is the mean function,  $\phi_{jk}(t)$  is the  $k$ th nonlinear FPC for the  $j$ th functional variable,  $\xi_{ik}$  is the  $k$ th nonlinear FPC score with mean zero and  $\text{cov}(\xi_{ik'}, \xi_{ik}) = 0$  if  $k' \neq k$ , and  $\lambda_k$  otherwise, where  $\lambda_1 \geq \lambda_2 \geq \dots > 0$ ,  $\sum_k \lambda_k < \infty$ , and  $\varepsilon_{ij}(t)$  is measurement error independent of  $\xi_{ik}$ . Moreover,  $f_j(\cdot)$  is a curve-specific link function that characterizes a possible nonlinear relationship between  $X_{ij}(t)$  and  $\sum_{k=1}^{\infty} \xi_{ik} \phi_{jk}(t)$ . If  $f_j(\cdot)$  is a nonlinear function, then  $X_{ij}(\cdot)$  is projected into a nonlinear space of  $\{\phi_{j1}(t), \phi_{j2}(t), \dots\}$  and characterized by a nonlinear score vector  $(\xi_{i1}, \xi_{i2}, \dots)$ . Such nonlinear projection allows functional variables to be nonlinearly dependent. If  $f_j(\cdot)$  is linear, model (2.2) reduces to the classical MFPCA. The link function  $f_j(\cdot)$  is selected based on the observed data; thus, its choice is flexible and data-driven.

The interpretation of the NMFPCA is different from that of the conventional FPCA. First, the variance  $\lambda_k$  represents the degree of variability of  $X_{ij}(t)$  in the  $k$ th direction, while its magnitude is not specified by the covariance operator of  $\mathbf{X}_i(t)$ . Second, when  $f_j(\cdot)$  is a nonlinear function, the nonlinear FPCs are the basis functions of the nonlinear dimension reduction space of  $\mathbf{X}_i(t)$  instead of the eigenfunctions of the covariance operator of  $\mathbf{X}_i(t)$ . Third, FPC score  $\xi_{ik}$  is the weight of  $\phi_{jk}(t)$  in the nonlinear dimension reduction representation of  $\mathbf{X}_i(t)$ . It does not necessarily satisfy

$\xi_{ik} = \sum_{j=1}^p \int_{\mathcal{T}} \{X_{ij}(t) - \mu_j(t) - \varepsilon_{ij}(t)\} \phi_{jk}(t) dt$ . Model (2.2) is unidentifiable without imposing appropriate identifiability constraints. The following proposition presents conditions for identifying model (2.2) while not restricting model flexibility.

**Proposition 1.** Under Conditions (I1) and (I2) stated in Web Appendix B, for all  $i, j$  and  $k$ ,  $\mu_j(\cdot)$ ,  $f_j(\cdot)$ ,  $\phi_{jk}(\cdot)$ , and  $\xi_{ik}$  are unique.

Even though there is an infinite number of FPCs, the number of FPCs estimated consistently from the sample is much fewer, as shown in our theory in Section 4. Therefore, as the variance  $\lambda_k$  decreases toward 0, a truncated approximation to  $X_{ij}(t)$  is often used in functional data analysis.

**Proposition 2.** Suppose  $\sup_t \phi_{jk}(t)^2 < \infty$  and  $f_j(\cdot)$  has bounded first derivative for all  $j, k$ . Then, as  $K_n \rightarrow \infty$ , we have

$$\sum_{j=1}^p \sup_{t \in \mathcal{T}} E \left[ X_{ij}(t) - \mu_j(t) - f_j \left\{ \sum_{k=1}^{K_n} \xi_{ik} \phi_{jk}(t) \right\} - \varepsilon_{ij}(t) \right]^2 \rightarrow 0. \quad (2.3)$$

## 2.2 FMVCM for Imaging Response

We propose FMVCM to delineate the relationship between the imaging response and scalar and functional predictors. We consider the latent variable  $\zeta_i$  related to the imaging response in a multiple-index varying coefficient form and the scalar vector  $\mathbf{Z}_i$  in a spatially

varying coefficient form as follows:

$$Y_i(\mathbf{s}) = m(\mathbf{s}, \boldsymbol{\zeta}_i, \mathbf{Z}_i) + \epsilon_i(\mathbf{s}) = \psi \left\{ \sum_{k=1}^{K_n} \mathbf{g}_k(\mathbf{s}) \zeta_{ik} \right\} + \boldsymbol{\theta}(\mathbf{s})^T \mathbf{Z}_i + \epsilon_i(\mathbf{s}), \quad (2.4)$$

where  $\sum_{k=1}^{K_n} \mathbf{g}_k(\mathbf{s}) \zeta_{ik} = \left\{ \sum_{k=1}^{K_n} g_{k1}(\mathbf{s}) \zeta_{ik}, \dots, \sum_{k=1}^{K_n} g_{kq}(\mathbf{s}) \zeta_{ik} \right\}^T$ ,  $\psi: \mathbb{R}^q \rightarrow \mathbb{R}$  is an unknown link function and  $q < K_n$ ,  $\boldsymbol{\zeta}_i = (\zeta_{i1}, \dots, \zeta_{iK_n})^T$  with  $\zeta_{ik}$  being a transformation of  $\xi_{ik}$  defined below,  $\boldsymbol{\theta}(\cdot) = \{\theta_1(\cdot), \dots, \theta_{q_z}(\cdot)\}^T$  is the spatial varying coefficient vector, and  $\epsilon_i(\mathbf{s})$  is a measurement error with mean zero and independent of  $\boldsymbol{\zeta}_i, \mathbf{Z}_i$ , and  $\varepsilon_{ij}(t)$ . Model (2.4) describes the effect of functional variables via the nonlinear multivariate FPC scores extracted in Section 2.1. We use the standard normal CDF and the transformed variable  $\zeta_{ik} = \Phi\left(\lambda_k^{-1/2} \xi_{ik}\right) - 0.5$  in what follows, leading to a uniformly distributed  $\zeta_{ik}$  on  $[-0.5, 0.5]$  when  $\xi_{ik} \sim N(0, 1)$ . The FPC score  $\xi_{ik}$  is then transformed to take values in a compact and bounded subset of the real line to avoid scale issues.

The multiple-index model is powerful for modeling high-dimensional predictors. Under model (2.4), the imaging measures relate to latent variables in  $\boldsymbol{\zeta}_i$  only through  $q$  spatial indices, thereby reducing the model dimension from diverging  $K_n$  to fixed  $q$  and significantly alleviating the curse of dimensionality. Furthermore, the imaging response depends on scalar covariates through functional linear regression, which provides a straightforward interpretation of the covariate effects. Therefore, model (2.4) not only retains the advantages of the standard multiple-index model (Xia, 2008), such as sufficient dimension reduction, flexible modeling, and easy interpretation, but also extends it in the following

directions. First, the proposed FMVCM replaces constant coefficients with varying coefficients to allow for the dynamic association among variables. Second, by introducing latent variables, the model accommodates unobserved covariates. Third, model (2.4) allows the number of latent covariates  $\zeta_i$  to diverge to infinity with the sample size. Notably, SVMCM and the single-index varying coefficient model are exceptional cases of model (2.4). However, all the covariates in these models are observable. In contrast, our model includes the estimated NMFPCA scores extracted from  $\mathbf{X}_i(t)$  as predictors, which are not directly observable. Moreover,  $\mathbf{X}_i(t)$  is collected discretely at irregular sample points with measurement errors. The impact of unobservable FPC scores and measurement errors on the resulting estimator elicits additional challenges for estimation and theoretical exploration.

To sufficiently capture the relationship between imaging responses and functional covariates by the first  $K_n$  latent scores, the truncation point  $K_n$  should be chosen as a large number. However, simply truncating FPCs to explain most of the variation in  $\mathbf{X}_i(t)$  is inappropriate while retaining more than needed predictors causes overfitting. Therefore, we use a sparse penalized method to identify important FPCs. The relationship between imaging responses and functional covariates guides the penalization procedure. The resulting parsimonious model enables improvement of estimation efficiency and accuracy.

Model (2.4) is unidentifiable without imposing some conditions. One possible approach is to adopt the identification methods in Xia (2008) since for each location  $\mathbf{s}$ , model (2.4) reduces to a standard multiple-index model. However, these methods ignore the inherent spatial correlation in neuroimaging data (Zhu et al., 2023). Moreover, the

restrictions on  $\psi(\cdot)$  and  $\mathbf{G}(\mathbf{s}) = \{\mathbf{g}_1(\mathbf{s}), \dots, \mathbf{g}_{K_n}(\mathbf{s})\}$  are stringent, leading to difficulties in finding such a link function and varying coefficients for all  $\mathbf{s}$ . Instead of extending these methods to imaging data, we propose an alternative approach to address the model identification problem. We state this approach in the following proposition.

**Proposition 3.** Under Condition (I3) stated in Web Appendix B,  $\psi(\cdot)$ ,  $\mathbf{G}(\cdot)$ , and  $\boldsymbol{\theta}(\cdot)$  are unique.

### 3. Estimation Procedure

The estimation procedure for our FJM is very challenging since it involves estimating nonlinear link functions,  $f_j(\cdot)$  and  $\psi(\cdot)$ , latent scores  $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{iK_n})^T$ , high dimensional nonlinear FPCs,  $\boldsymbol{\phi}_j(\cdot) = \{\phi_{j1}(\cdot), \dots, \phi_{jK_n}(\cdot)\}^T$ , and the spatial varying coefficient functions over irregular domains. Its associated computational burden is much heavier than the standard FPCA, SVCM, and multiple-index model. We use univariate and bivariate spline and local polynomial smoothing to implement the estimation procedure and ease the computational burden. We fit the proposed NMFPCA by minimizing the residual sum of squares,

$$\sum_{i=1}^n \sum_{j=1}^p \sum_{d=1}^{n_{ij}} [X_{ij}(t_{ijd}) - \mu_j(t_{ijd}) - f_j \{ \boldsymbol{\xi}_i^T \boldsymbol{\phi}_j(t_{ijd}) \}]^2. \quad (3.1)$$

For ease of presentation, suppose all the mean functions and nonlinear FPCs have common compact support. We approximate  $\mu_j(t)$  and  $\boldsymbol{\phi}_j(t)$  by  $\mu_{nj}(t) = \mathbf{u}_j^T \mathbf{B}_n(t)$  and  $\boldsymbol{\phi}_{nj}(t) = \boldsymbol{\Gamma}_j \mathbf{B}_n(t)$ , respectively, and use local linear smoother to approximate  $f_j(\cdot)$ , where  $\mathbf{B}_n(\cdot) =$

$\{b_1(\cdot), \dots, b_{k_n}(\cdot)\}^T$  is a set of B-spline basis functions of order  $l + 1$  with knots  $0 = d_0 < d_1 < \dots < d_{M_n} < d_{M_n+1} = 1$ , satisfying  $\max_{1 \leq j \leq M_n} |d_j - d_{j+1}| = O(n^{-\nu_1})$  with  $k_n = M_n + l + 1$  and  $\nu_1 \in (0, 0.5)$ . The formula for estimating NMFPCA is presented in Web Appendix A.

To implement NMFPCA, we must select the number of FPCs  $K_n$  and B-spline basis  $k_n$ , and bandwidths  $h_1$ . Under a high-dimensional setting, Li et al. (2011) proposed the cross-validated BIC based on penalized eigenvalues to consistently estimate the dimension of the central subspace for sufficient dimension reduction. We modify the cross-validated BIC rule to estimate the divergent number of FPCs,  $K_n$ , as follows:

$$\hat{K}_n = \arg \max_{K \leq K_{\max}} \left( \sum_{k=1}^K \hat{\lambda}_k - \frac{K \hat{\lambda}_1}{4n^{1/4}} \right), \quad (3.2)$$

where  $\hat{\lambda}_k = \text{var}(\hat{\xi}_{ik})$ ,  $\hat{\xi}_{ik}$  is the estimator of  $\xi_{ik}$  defined in Section 3, and  $K_{\max}$  is the upper bound for the number of FPCs. Proposition 4 shows that (3.2) can consistently select the number of FPCs. In the simulation studies and real data analysis, we set  $K_{\max} = 50$  and estimate  $\lambda_k$  by  $\hat{\lambda}_k = \sum_{i=1}^n \hat{\xi}_{ik}^2 / n$ , which is confirmed to work well in all the cases considered. Compared to the traditional FPCA, the proposed estimation is less sensitive to the choice of  $K_n$  since we further choose the FPC scores by sparse penalty in subsequent FMVCM.

We use the cubic B-spline basis functions, then,  $k_n = M_n + 4$ , where  $M_n$  is the number of interior knots. A small  $M_n$  (e.g., 2 to 6) is usually good enough for smooth, monotonic, or unimodal functions.

Given  $k_n$ , we adopt the Epanechnikov kernel function and minimize BIC to select  $h_1$ :

$$\begin{aligned} \text{BIC}(h_1) = & \log \left( \frac{1}{\sum_{i=1}^n \sum_{j=1}^p n_{ij}} \sum_{i=1}^n \sum_{j=1}^p \sum_{d=1}^{n_{ij}} \left[ X_{ij}(t_{ijd}) - \hat{\mu}_j(t_{ijd}) - \hat{f}_j \left\{ \sum_{k=1}^{\hat{K}_n} \hat{\xi}_{ik} \hat{\phi}_{jk}(t_{ijd}) \right\} \right]^2 \right) \\ & + df(h_1) \frac{\log \left( \sum_{i=1}^n \sum_{j=1}^p n_{ij} \right)}{\sum_{i=1}^n \sum_{j=1}^p n_{ij}}, \end{aligned} \quad (3.3)$$

where  $\hat{\mu}_j(\cdot)$ ,  $\hat{f}_j(\cdot)$ ,  $\hat{\xi}_{ik}$ , and  $\hat{\phi}_{jk}(\cdot)$  are estimators, and  $df(h_1)$  is the number of estimated parameters for each observation point. The minimization of (3.3) can be solved through grid search.

Next, we develop estimation for FMVCM. Let  $\hat{\boldsymbol{\zeta}}_i = (\hat{\zeta}_{i1}, \dots, \hat{\zeta}_{iK_n})^T$  be the estimator of  $\boldsymbol{\zeta}_i$  obtained from the NMFPCA. To fit the proposed FMVCM and select relevant latent variables, we propose to minimize

$$\frac{1}{nN_s} \sum_{i=1}^n \sum_{j=1}^{N_s} \left[ Y_i(\mathbf{s}_j) - \psi \left\{ \sum_{k=1}^{K_n} \mathbf{g}_k(\mathbf{s}_j) \hat{\zeta}_{ik} \right\} - \boldsymbol{\theta}(\mathbf{s}_j)^T \mathbf{Z}_i \right]^2 + \sum_{k=1}^{K_n} \lambda \frac{\|\mathbf{g}_k\|_2}{\|\tilde{\mathbf{g}}_k\|_2}, \quad (3.4)$$

where  $\|\cdot\|_2$  is the  $L^2$  norm,  $\tilde{\mathbf{g}}_k$  is a known adaptive weight such as the last iterative estimator, and  $\lambda$  is a regularization parameter. We use local linear and bivariate splines over triangulation (Li et al., 2021) to approximate  $\psi(\cdot)$  with bandwidth  $h_2$ ,  $\boldsymbol{\theta}(\cdot)$ , and  $\mathbf{g}_k(\cdot)$  for all  $k$ .

The bivariate splines over triangulations are piecewise polynomial bivariate functions over a 2D triangulated domain. They can efficiently handle the data distributed in irregular regions with complicated boundaries. Let  $\Delta = \{\tau_1, \dots, \tau_{N_\Delta}\}$  be a triangle of  $\mathcal{D}$ ,

which can be constructed by Delaunay triangulation (Li et al., 2021). Given  $\tau \in \Delta$ , let  $|\tau|$  be its longest edge length, and  $R_\tau$  be the radius of the largest disk that can be inscribed in  $\tau$ . Define  $\beta_\tau = |\tau|/R_\tau$  as the shape parameter of  $\tau$ , and  $|\Delta| := \max\{|\tau|, \tau \in \Delta\}$  is the size of  $\Delta$ . Furthermore, define  $\mathcal{G}_\varpi^v(\Delta) = \{g \in \mathbb{C}^v(\mathcal{D}) : g|_\tau \in \mathcal{P}_\varpi(\tau), \tau \in \Delta\}$  as the spline space of degree  $\varpi$  and smoothness  $v$  over triangulation  $\Delta$ , where  $\mathbb{C}^v(\mathcal{D})$  is the collection of all  $v$ th continuously differentiable functions over  $\mathcal{D}$ ,  $g|_\tau$  is the polynomial in  $\mathcal{P}_\varpi(\tau)$ , and  $\mathcal{P}_\varpi(\tau)$  is the space of all polynomials of degree less than or equal to  $\varpi$  on  $\tau$ . Let  $\mathbf{S}_n^*(\cdot) = \{S_1^*(\cdot), \dots, S_{J_n}^*(\cdot)\}^T$  be the set of bivariate Bernstein basis polynomials for  $\mathcal{G}_\varpi^v(\Delta)$ , where  $J_n = N_\Delta(\varpi + 1)(\varpi + 2)/2$ . Then,  $\mathbf{g}_k(\mathbf{s})$  and  $\boldsymbol{\theta}(\mathbf{s})$  can be approximated by  $\mathbf{g}_{nk}(\mathbf{s}) = \boldsymbol{\alpha}_k^* \mathbf{S}_n^*(\mathbf{s})$  and  $\boldsymbol{\theta}_n(\mathbf{s}) = \boldsymbol{\beta}^* \mathbf{S}_n^*(\mathbf{s})$ , respectively. We develop an estimation procedure for FMVCM with the sparse penalty stated in Web Appendix A, which is equivalent to solving a series of linear equations that each update has a close form or is easy to implement through the existing package. For the index dimension, we do not use the cross-validation (CV) criterion proposed by Xia (2008) since this criterion is computationally burdensome, especially in image regression. We instead select  $q$  and  $h_2$  by minimizing BIC, which has been widely used in high-dimensional settings (Wang et al., 2007) and defined as follows:

$$\begin{aligned} \text{BIC}(q, h_2) = & \log \left( \frac{1}{nN_s} \sum_{i=1}^n \sum_{j=1}^{N_s} \left[ Y_i(\mathbf{s}_j) - \hat{\psi} \left\{ \sum_{k=1}^{\hat{K}_n} \hat{\mathbf{g}}_k(\mathbf{s}_j) \hat{\zeta}_{ik} \right\} - \hat{\boldsymbol{\theta}}(\mathbf{s}_j)^T \mathbf{Z}_i \right]^2 \right) \\ & + df(q, h_2) \frac{\log(nN_s)}{nN_s}, \end{aligned} \quad (3.5)$$

where  $df(q, h_2)$  is the number of estimated non-zero parameters at each observation point.

The minimization of (3.5) can be solved through a two-dimensional grid search. The simulation studies in Section 5.2 show that this selection procedures performs satisfactorily.

#### 4. Asymptotic Properties

In this section, we establish the asymptotic properties for the proposed estimators, but we defer their proofs to Web Appendix B. We put a subscript 0 on a parameter/function to denote its true value, e.g.,  $\mathbf{g}_{k0}$  is the true value of  $\mathbf{g}_k$ . We first introduce some notations and definitions. Denote the supremum norm by  $\|\cdot\|_\infty$ . Let  $\mathcal{A} = \{k, \|\mathbf{g}_{k0}\|_2 \neq 0, k = 1, \dots, K_n\}$ , and  $|\mathcal{A}|$  denotes the number of elements in  $\mathcal{A}$ . The adaptive weight satisfies  $\max_{k \notin \mathcal{A}} \|\tilde{\mathbf{g}}_k\|_2 = O_p(n^{-\delta})$  for  $\delta > 0$ , and there exists a constant  $M > 0$ , such that for any  $\epsilon > 0$ ,  $P(\min_{k \in \mathcal{A}} \|\tilde{\mathbf{g}}_k\|_2 > M\varrho) > 1 - \epsilon$ , where  $\varrho = \min_{k \in \mathcal{A}} \|\mathbf{g}_{k0}\|_2$ .

##### Theorem 1. (Convergent rate on NMFPCA)

Let  $\delta_k = \min_{k' \leq k} (\lambda_{k'-1,0} - \lambda_{k',0}, \lambda_{k',0} - \lambda_{k'+1,0})$ , and  $\boldsymbol{\eta}_i(t) = \{\eta_{i1}(t), \dots, \eta_{ip}(t)\}^T$  with  $\eta_{ij}(t) = \boldsymbol{\xi}_i^T \boldsymbol{\phi}_j(t)$ . Denote  $r$  the order of the Hölder space, which  $\mu_{j0}(\cdot)$  and  $\phi_{jk,0}(\cdot)$  belong to. Under Conditions (C1)–(C7) stated in Web Appendix B, if  $k_n \rightarrow \infty, k_n/n_{ij} \rightarrow 0$ , and  $h_1 \rightarrow 0, \sum_{i=1}^n h_1 n_{ij} \rightarrow \infty$  as  $n_{ij} \rightarrow \infty, n \rightarrow \infty$ , we have

- (a)  $\|\hat{\boldsymbol{\eta}}_i(t) - \boldsymbol{\eta}_{i0}(t)\|_2 = O_p(r_{\eta,i})$ ;
- (b)  $\|n^{-1} \sum_{i=1}^n \hat{\boldsymbol{\eta}}_i(s) \hat{\boldsymbol{\eta}}_i(t)^T - E\{\boldsymbol{\eta}_{i0}(s) \boldsymbol{\eta}_{i0}(t)^T\}\|_2 = O_p(r_{\mathbf{G}})$ ;
- (c)  $\hat{\xi}_{ik} - \xi_{ik,0} = O_p\{\max(r_{\eta,i}, \delta_k^{-1} r_{\mathbf{G}})\}$ ,

where  $n_{ij}$  is the number of observation points for  $X_{ij}(\cdot)$ ,  $\|\boldsymbol{\eta}_i(t)\|_2 = \sqrt{\sum_{j=1}^p \int_{\mathcal{T}} \eta_{ij}^2(t) dt}$ ,

$$\|\boldsymbol{\eta}_i(t)\boldsymbol{\eta}_i(t)^T\|_2 = \sqrt{\sum_{j=1}^p \sum_{j'=1}^p \int_{\mathcal{T}} \eta_{ij}^2(t)\eta_{ij'}^2(t)dt},$$

$$r_{\eta,i} \equiv \left(\sum_{j=1}^p k_n n_{ij}^{-1}\right)^{1/2} + k_n^{-r} + h_1^2 + \left\{\sum_{j=1}^p \left(\sum_{i=1}^n h_1 n_{ij}\right)^{-1}\right\}^{1/2} + n^{-1/2}, \quad \text{and}$$

$$r_{\mathbf{G}} \equiv \left\{\sum_{i=1}^n \sum_{j=1}^p k_n (n n_{ij})^{-1}\right\}^{1/2} + k_n^{-r} + h_1^2 + \left\{\sum_{j=1}^p \left(\sum_{i=1}^n h_1 n_{ij}\right)^{-1}\right\}^{1/2} + n^{-1/2}.$$

Theorem 1 incorporates the results developed by Bosq (2000), Zhu et al. (2014), and many other works on nonparametric estimators. Theorem 1 (a) and (b) present the  $L_2$  convergence rate of  $\widehat{\eta}_{ij}(t) = \widehat{\boldsymbol{\xi}}_i^T \widehat{\boldsymbol{\phi}}_j(t)$  and its sample autocovariance function, respectively. The uniform convergence rate of Luo et al. (2016) and Li et al. (2017) is slightly stronger since we consider a different estimation approach, and their method cannot be extended to nonlinearly related functional data. Our results are more appealing to functional data analysis, where acquiring a proper relationship estimation is critical to the subsequent regression analysis.

In Theorem 1, the terms  $k_n^{-r}$  and  $h_1^2$  represent the approximation errors incurred by employing spline and local linear approximations for  $(\mu_j, \phi_{jk})$  and  $f_j$ , respectively. The expressions  $\left(\sum_{j=1}^p k_n n_{ij}^{-1}\right)^{1/2}$  and  $\left\{\sum_{i=1}^n \sum_{j=1}^p k_n (n n_{ij})^{-1}\right\}^{1/2}$  denote the estimation errors arising from the spline expansion for nonparametric functions, where  $n_{ij}$  acts as the effective sample size in the estimation process. Similarly,  $\left\{\sum_{j=1}^p \left(\sum_{i=1}^n h_1 n_{ij}\right)^{-1}\right\}^{1/2}$  captures the estimation error stemming from local linear estimation of  $(f_j, = 1, \dots, p)$ , while  $n^{-1/2}$  accounts for the measurement error. The convergence rates of the estimated scores

are contingent upon the decay rate of  $\lambda_{k_0}$ , implying that a rapid decay of  $\lambda_{k_0}$  hastens the convergence of  $\widehat{\xi}_{ik}$ .

We consider two notable scenarios to delve deep into the asymptotic results. For ease of discussion, we assume that  $n_{ij} \equiv \nu$ . In this case, the convergence rate of  $\widehat{\boldsymbol{\eta}}_i(t)$  and  $n^{-1} \sum_{i=1}^n \widehat{\boldsymbol{\eta}}_i(s) \widehat{\boldsymbol{\eta}}_i(t)^T$  reduces to  $\sqrt{k_n/\nu} + k_n^{-r} + h_1^2 + 1/\sqrt{nh_1\nu} + n^{-1/2}$ , while the convergence rate of  $\widehat{\xi}_{ik}$  reduces to  $\delta_k^{-1} \left( \sqrt{k_n/\nu} + k_n^{-r} + h_1^2 + 1/\sqrt{nh_1\nu} + n^{-1/2} \right)$ .

1. When  $h_1^2 + 1/\sqrt{nh_1\nu} + n^{-1/2} = o\left(\sqrt{k_n/\nu} + k_n^{-r}\right)$  and  $\nu = n$ , the convergence rate of  $\widehat{\boldsymbol{\eta}}_i(t)$  and  $n^{-1} \sum_{i=1}^n \widehat{\boldsymbol{\eta}}_i(s) \widehat{\boldsymbol{\eta}}_i(t)^T$  reduces to the well-known rate  $O_p\left(\sqrt{\frac{k_n}{n}} + k_n^{-r}\right)$  for estimating a one-dimensional function. Additionally, it achieves the optimal convergence rate  $O_p\left\{n^{-r/(2r+1)}\right\}$  (Stone, 1980) when further setting  $k_n = n^{1/(2r+1)}$ . In this scenario, the convergence rate of  $\widehat{\xi}_{ik}$  reduces to  $O_p\left\{\delta_k^{-1} n^{-r/(2r+1)}\right\}$ , where the influence of estimating  $f_j$  can be deemed negligible.
2. When the number of observation points is sufficiently large, and the bandwidth is adequately small, linear FPCA methods establish  $\sqrt{n}$ -consistency for the estimators of  $\boldsymbol{\eta}_i(t)$  and its covariance function (Hall et al., 2006; Zhu et al., 2014). Furthermore, Zhu et al. (2014) derived that  $\widehat{\xi}_{ik} - \xi_{ik,0} = O_p\left(\delta_k^{-1} n^{-1/2}\right)$  for linear FPC scores. These results can be attained in our nonlinear framework if  $\nu$  and  $k_n$  are sufficiently large, and  $h_1$  is adequately small, ensuring  $h_1^2 + 1/\sqrt{nh_1\nu} + \sqrt{k_n/\nu} + k_n^{-r} = O\left(n^{-1/2}\right)$ .

**Proposition 4.** Under Conditions (C1)–(C7) stated in Web Appendix B, if  $K_n \leq K_{\max} = o(n^{1/4})$  and  $r_{\mathbf{G}} = o(n^{-1/4})$ , we have  $\widehat{K}_n \rightarrow K_n$  with probability tending to 1.

**Theorem 2. (Convergent rate on transformed FPC scores)** *Suppose that the transformation function  $\Phi(\cdot)$  has a bounded derivative, and the conditions in Theorem 1 hold, we have  $\widehat{\zeta}_{ik} - \zeta_{ik,0} = \lambda_{k0}^{-1/2} O_p \left\{ \max \left( r_{\eta,i}, \delta_k^{-1} r_{\mathbf{G}} \right) \right\}$ .*

Theorem 2 shows the estimation accuracy of FPC scores decreases as the order of principal components gets high. Combining the results of Theorem 2 and Proposition 2, we conclude that a high value of the truncation point  $K_n$  can reduce the approximation error but increase the estimation error. These results may provide theoretical guidance on choosing an appropriate  $K_n$  for downstream data analysis. Nevertheless, the problem is relatively insensitive to the selection of  $K_n$  because our subsequent FMVCM will automatically select important principal components and rule out unimportant ones.

**Theorem 3. (Convergent rate on FMVCM)** *Under the conditions in Theorem 2 and Condition (C8) stated in Web Appendix B, if  $h_2 \rightarrow 0, nN_s h_2^q \rightarrow \infty$ , and  $h_2^4 + \frac{1}{nN_s h_2^q} + \lambda \frac{|A|^{1/2}}{\varrho} + K_n^{c_0+1} \max \left( n^{-1} \sum_{i=1}^n r_{\eta,i}^2, K_n^{2c_0+2} r_{\mathbf{G}}^2 \right) = o \left\{ \frac{K_n}{nN_s |\Delta|^2} + K_n |\Delta|^{2(\varpi+1)} \right\}$ , with  $c_0$  being the polynomial decay rate of  $\lambda_{k0}$ , we have  $\|\widehat{\mathbf{G}}(\mathbf{s}) - \mathbf{G}_0(\mathbf{s})\|_2^2 = O_p \left\{ K_n (nN_s |\Delta|^2)^{-1} + K_n |\Delta|^{2(\varpi+1)} \right\}$  and  $\|\widehat{\boldsymbol{\theta}}(\mathbf{s}) - \boldsymbol{\theta}_0(\mathbf{s})\|_2^2 = O_p \left\{ K_n (nN_s |\Delta|^2)^{-1} + K_n |\Delta|^{2(\varpi+1)} \right\}$ .*

In the penalized FMVCM, the convergence rate of  $(\widehat{\mathbf{G}}(\mathbf{s}), \widehat{\boldsymbol{\theta}}(\mathbf{s}))$  depends not only on the spline estimation error  $K_n (nN_s |\Delta|^2)^{-1}$  and the approximation error  $K_n |\Delta|^{2(\varpi+1)}$  but is also affected by the selection and estimation of transformed scores  $\zeta_i$  and the estimation of the link function  $\psi(\cdot)$ . These uncertainties are reflected in terms  $\lambda \frac{|A|^{1/2}}{\varrho}$ ,  $K_n^{c_0+1} \max \left( n^{-1} \sum_{i=1}^n r_{\eta,i}^2, K_n^{2c_0+2} r_{\mathbf{G}}^2 \right)$  and  $h_2^4 + \frac{1}{nN_s h_2^q}$ . When the bandwidths  $h_1$  and  $h_2$  and the penalized parameter  $\lambda$  are sufficiently small, and  $n_{ij}$ ,  $k_n$ , and  $nN_s$  are sufficiently large,

the expression  $h_2^4 + \frac{1}{nN_s h_2^q} + \lambda \frac{|\mathcal{A}|^{1/2}}{\rho} + K_n^{c_0+1} \max(n^{-1} \sum_{i=1}^n r_{\eta,i}^2, K_n^{2c_0+2} r_{\mathbf{G}}^2)$  becomes negligible compared to  $\frac{K_n}{nN_s |\Delta|^2} + K_n |\Delta|^{2(\varpi+1)}$ . This implies that the uncertainties introduced by  $\widehat{\zeta}_i$  and  $\widehat{\psi}(\cdot)$  can be disregarded. Furthermore, if  $K_n$  is finite, Theorem 3 simplifies to the well-known convergence rate of bivariate spline over triangulation (Li et al., 2021):  $\|\widehat{\mathbf{G}}(\mathbf{s}) - \mathbf{G}_0(\mathbf{s})\|_2^2 = O_p \left\{ \frac{1}{nN_s |\Delta|^2} + |\Delta|^{2(\varpi+1)} \right\}$  and  $\|\widehat{\boldsymbol{\theta}}(\mathbf{s}) - \boldsymbol{\theta}_0(\mathbf{s})\|_2^2 = O_p \left\{ \frac{1}{nN_s |\Delta|^2} + |\Delta|^{2(\varpi+1)} \right\}$ .

**Theorem 4. (Selection consistency)** *Under the conditions in Theorem 3, if  $\frac{K_n}{nN_s |\Delta|^2} + K_n |\Delta|^{2(\varpi+1)} \rightarrow 0$  and  $\left\{ \frac{K_n}{nN_s |\Delta|^2} + K_n |\Delta|^{2(\varpi+1)} \right\} (\lambda n^\delta)^{-1} \rightarrow 0$ , then we have*

$$\lim_{n \rightarrow \infty} P(\|\widehat{\mathbf{g}}_k\|_2 \neq 0 \text{ for } k \in \mathcal{A} \text{ and } \|\widehat{\mathbf{g}}_k\|_2 = 0 \text{ for } k \notin \mathcal{A}) = 1.$$

The condition  $\left\{ \frac{K_n}{nN_s |\Delta|^2} + K_n |\Delta|^{2(\varpi+1)} \right\} (\lambda n^\delta)^{-1} \rightarrow 0$  requires that  $\lambda$  converges to 0 not too fast to ensure the consistency of variable selection. This lower bound, together with the upper bound in Theorem 3 for  $\lambda$ , yields  $n^{-\delta} \left\{ \frac{K_n}{nN_s |\Delta|^2} + K_n |\Delta|^{2(\varpi+1)} \right\} \ll \lambda \ll \frac{\rho}{|\mathcal{A}|^{1/2}} \left\{ \frac{K_n}{nN_s |\Delta|^2} + K_n |\Delta|^{2(\varpi+1)} \right\}$ , where  $a_n \ll b_n$  stands for  $\frac{a_n}{b_n} \rightarrow 0$ . The existence of such  $\lambda$  can be guaranteed by  $n^{-\delta} \ll \frac{\rho}{|\mathcal{A}|^{1/2}}$ . As a result, we can simultaneously obtain the consistency of estimation and selection for the parameters.

## 5. Numerical Studies

In this section, we examine the finite sample performance of FJM. We compare the NMF-PCA with classical MFPCA proposed by Happ and Greven (2018) and compare the proposed FMVCM with the SVCM offered by Yu et al. (2021). The MFPCA and SVCM are obtained from *MFPCA* and *SVCMimage* packages in R, respectively. We evaluate

the accuracy of reconstructed curves and extracted scores in FPCA and the accuracy of estimated coefficient functions in the image regression. Specifically, we compute the mean squared errors (MSEs) of estimators, including  $\text{MSE}(\widehat{\mathbf{g}}_k) = N_s^{-1} \sum_{j=1}^{N_s} \{\widehat{\mathbf{g}}_k(\mathbf{s}_j) - \mathbf{g}_k(\mathbf{s}_j)\}^2$ ,  $\text{MSE}(\widehat{X}_j) = (\sum_{i=1}^n n_{ij})^{-1} \sum_{i=1}^n \sum_{d=1}^{n_{ij}} \{\widehat{X}_{ij}(t_{ijd}) - X_{ij}(t_{ijd})\}^2$ , and  $\text{MSE}(\widehat{\xi}_k) = n^{-1} \sum_{i=1}^n (\widehat{\xi}_{ik} - \xi_{ik})^2$  for  $j = 1, \dots, p$  and  $k = 1, \dots, K_n$ .

### 5.1 Performance of NMFPCA

**Xmodel.** We set  $p = 2$  and generate the multivariate functional variable from  $X_{ij}(t) = \mu_j(t) + f_j \{ \sum_{k=1}^4 \xi_{ik} \phi_{jk}(t) \} + \varepsilon_{ij,t}$ , where  $\mu_1(t) = t + \sin(\pi t)$ ,  $\mu_2(t) = \exp(t)$ ,  $\xi_{ik} \sim N(0, \lambda_k)$  with  $\lambda_k = \frac{\exp\{-(k+1)/2\}}{\sum_{k=1}^4 \exp\{-(k+1)/2\}}$ , and  $\varepsilon_{ij,t} \sim N(0, \sigma^2)$ . The construction of eigenfunctions  $\phi_{jk}(t)$  refers to the online appendix of Happ and Greven (2018). We consider link functions: Case I with  $f_1(u) = u^3/2 + 2.5u$  and  $f_2(u) = \sin(\pi u/5)$ ; Case II with  $f_j(u) = u$  for  $j = 1, 2$ . We take sample size  $n = 50, 100$ , and  $150$ , and the number of observations  $n_{ij} = 10, 40$ , and  $80$ , with the observation points randomly sampled from  $U(0, 1)$ .

To implement NMFPCA, we select bandwidth  $h_1 = 1$  and the number of splines  $k_n = 8$  by the approach in Section 3. Web Figure 1 shows the values of  $\text{BIC}(h_1)$  for Case I with  $n = 50, 100$ , and  $150$ ,  $n_{ij} = 80$ , and  $\sigma^2 = 0.1$ . Table 1 presents the MSEs of the  $k$ th score  $\xi_k$  and the  $j$ th functional variable  $X_j(\cdot)$  based on 100 Monte Carlo simulations under Case I of Xmodel. The MSEs of NMFPCA are consistently smaller than those of MFPCA. The improvement becomes increasingly remarkable as the noise level increases, indicating that modeling the nonlinear relationship between  $X_{ij}(t)$  and  $\xi_i^T \phi_j(t)$  leads to better-extracted information and a more accurate recovery curve. We

Table 1: MSEs of  $k$ th score  $\xi_k$  and  $j$ th functional variable  $X_j(\cdot)$  in Case I (nonlinear  $f_j$ ) of Xmodel with various sample sizes, observation levels, and noise levels.

$\sigma^2$	$(n, n_{ij})$	MFPCA						NMFPCA					
		$MSE(\hat{\xi}_1)$	$MSE(\hat{\xi}_2)$	$MSE(\hat{\xi}_3)$	$MSE(\hat{\xi}_4)$	$MSPE(\hat{X}_1)$	$MSPE(\hat{X}_2)$	$MSE(\hat{\xi}_1)$	$MSE(\hat{\xi}_2)$	$MSE(\hat{\xi}_3)$	$MSE(\hat{\xi}_4)$	$MSPE(\hat{X}_1)$	$MSPE(\hat{X}_2)$
0.1	(50,10)	4.94	2.52	1.58	1.15	1.87	0.67	0.19	0.19	0.14	0.11	1.41	0.08
	(50,40)	1.09	0.56	0.43	0.28	0.27	0.07	0.02	0.05	0.03	0.02	0.21	0.02
	(50,80)	0.92	0.49	0.33	0.19	0.05	0.01	0.01	0.03	0.01	0.01	0.04	0.01
	(100,10)	5.14	2.70	1.80	1.34	2.23	0.72	0.18	0.19	0.13	0.10	1.38	0.06
	(100,40)	1.00	0.53	0.42	0.24	0.24	0.05	0.01	0.03	0.02	0.01	0.14	0.02
	(100,80)	0.92	0.49	0.34	0.20	0.05	0.01	0.01	0.02	0.01	0.00	0.03	0.01
0.5	(50,10)	23.13	10.69	6.09	4.13	5.45	3.02	0.32	0.24	0.16	0.14	2.00	0.30
	(50,40)	2.34	1.25	0.90	0.73	1.23	0.32	0.08	0.13	0.09	0.07	0.48	0.05
	(50,80)	0.97	0.50	0.36	0.21	0.14	0.04	0.01	0.04	0.01	0.01	0.08	0.02
	(100,10)	25.50	11.40	7.35	4.93	5.90	3.41	0.30	0.22	0.14	0.12	1.85	0.23
	(100,40)	1.70	1.07	0.90	0.74	1.29	0.28	0.05	0.11	0.08	0.07	0.38	0.04
	(100,80)	0.95	0.50	0.37	0.21	0.13	0.02	0.01	0.03	0.01	0.01	0.08	0.01
1.5	(50,10)	21.20	11.83	7.28	4.96	6.07	3.27	0.27	0.23	0.15	0.13	1.80	0.14
	(50,40)	1.580	1.04	0.91	0.75	1.31	0.25	0.04	0.10	0.08	0.06	0.36	0.03
	(50,80)	0.940	0.50	0.36	0.22	0.14	0.02	0.01	0.02	0.01	0.01	0.07	0.01
	(100,10)	72.10	30.85	17.51	11.53	11.18	8.69	0.34	0.25	0.18	0.14	2.63	0.50
	(100,40)	6.22	2.89	1.82	1.38	2.93	0.96	0.19	0.18	0.13	0.10	0.95	0.12
	(100,80)	1.131	0.63	0.53	0.36	0.48	0.13	0.02	0.06	0.03	0.03	0.34	0.04
1.5	(100,10)	71.69	32.91	20.80	13.95	12.43	9.44	0.34	0.24	0.16	0.13	2.60	0.40
	(100,40)	4.76	2.71	1.92	1.51	3.16	0.84	0.15	0.18	0.12	0.11	0.81	0.08
	(100,80)	1.05	0.59	0.49	0.36	0.47	0.10	0.02	0.05	0.03	0.02	0.32	0.03
	(150,10)	61.25	34.04	21.34	14.58	12.56	9.64	0.34	0.22	0.16	0.13	2.48	0.34
	(150,40)	4.35	2.70	1.96	1.56	3.29	0.78	0.15	0.17	0.12	0.11	0.81	0.07
	(150,80)	1.05	0.58	0.50	0.37	0.51	0.09	0.02	0.05	0.02	0.02	0.32	0.03

also compare NMFPCA and MFPCA when the link function is linear for fairness. Web Table 1 summarizes the results. The NMFPCA outperforms MFPCA in most cases except for the case without measurement errors. This is due to the fact that directly performing MFPCA on the contaminated observations is sensitive to measurement errors, thereby leading to the poor performance of MFPCA in the presence of measurement errors. In contrast, NMFPCA avoids this issue by directly estimating FPC scores. Notably, the nonparametric estimation of the link function leads to a slightly worse performance of the proposed procedure than MFPCA in the absence of measurement errors. However, increasing the number of observations can compensate for this slight deficiency. Therefore, we conclude that the gains obtained from NMFPCA are substantial regardless of linearly or nonlinearly correlated functional data. In addition, as shown in Table 1 and Web Table 1, the estimation accuracy of NMFPCA improves with the sample size or the number of observation points and reduces as the noise level increases.

We also investigate the performance of (3.2) in selecting the number of FPCs. Web Table 2 reports the bias and standard deviation (SD) of  $\hat{K}_n$ . Most of the bias and SD values are reasonably small, demonstrating the effectiveness of our approach.

## 5.2 Performance of FMVCM

We conduct two simulation studies for correlating image with scalar and functional covariates. The imaging response is generated with multiple-index and linear link functions, as described below.

**Ymodel I.** Let  $q = 2, q_z = 2$  and  $\mathcal{D} = [-1, 4] \times [-1, 1]$ .  $\mathbf{X}_i(t)$  are obtained according

to the setting similar to Case I of Xmodel, except that  $\sigma^2 = 0.25$  and  $n_{ij}$  is generated from a discrete uniform distribution on  $\{70, \dots, 80\}$ . The imaging responses are simulated from  $Y_i(\mathbf{s}_j) = \psi \left\{ \sum_{k=1}^4 g_{k1}(\mathbf{s}_j) \zeta_{ik}, \sum_{k=1}^4 g_{k2}(\mathbf{s}_j) \zeta_{ik} \right\} + \theta_1(\mathbf{s}_j) Z_{i1} + \theta_2(\mathbf{s}_j) Z_{i2} + \epsilon_i(\mathbf{s}_j)$ , where  $\theta_1(\mathbf{s})$  is generated by function *fs.test* in R package *mgcv*,  $\theta_2(\mathbf{s}) = \theta_1(\mathbf{s}) + \sin(5s_1)$ ,  $g_{31}(\mathbf{s}) = \theta_1(\mathbf{s}) + \sin(5s_1) + \cos(5s_2)$ ,  $g_{41}(\mathbf{s}) = \theta_1(\mathbf{s}) + 5s_2$ ,  $g_{32}(\mathbf{s}) = \theta_1(\mathbf{s}) + 5s_1$ ,  $g_{42}(\mathbf{s}) = \theta_1(\mathbf{s}) + \cos(5s_2)$ ,  $\psi \left\{ \sum_{k=1}^4 g_{k1}(\mathbf{s}) \zeta_{ik}, \sum_{k=1}^4 g_{k2}(\mathbf{s}) \zeta_{ik} \right\} = \sin \left\{ \sum_{k=1}^4 g_{k1}(\mathbf{s}) \zeta_{ik} \right\} + 2 \cos \left\{ \sum_{k=1}^4 g_{k2}(\mathbf{s}) \zeta_{ik} \right\}$ ,  $g_{kj}(\mathbf{s}) \equiv 0$  for  $k = 1, 2$ ,  $Z_{i1}$  and  $Z_{i2}$  are independently generated from  $U[-1, 1]$ , and  $\epsilon_i(\mathbf{s}_j)$  is generated from a Gaussian process with zero mean and covariance function  $\text{Cov}\{\epsilon_i(\mathbf{s}_j), \epsilon_i(\mathbf{s}_{j'})\} = \sigma_\epsilon^2 \times 0.3^{\|\mathbf{s}_j - \mathbf{s}_{j'}\|}$ . For each image, we consider the horseshoe domain (Sangalli et al., 2013) and generate data at a grid of  $N_s = 101 \times 41$  pixels with 2611 pixels fall within the horseshoe domain.

**Ymodel II.** The setting is similar to Ymodel I, except that  $q = 1$ , and  $Y_i(\mathbf{s}_j) = \psi \left\{ \sum_{k=1}^4 g_{k1}(\mathbf{s}_j) \zeta_{ik} \right\} + \theta_1(\mathbf{s}_j) Z_{i1} + \theta_2(\mathbf{s}_j) Z_{i2} + \epsilon_i(\mathbf{s}_j)$ , where  $\psi \left\{ \sum_{k=1}^4 g_{k1}(\mathbf{s}_j) \zeta_{ik} \right\} = \sum_{k=1}^4 g_{k1}(\mathbf{s}_j) \zeta_{ik}$ .

To apply the bivariate spline smoothing, we consider a triangulation with 109 triangles and 95 vertices and bivariate spline basis functions with degree  $\varpi = 2$  and smoothness  $v = 1$ . We select  $\{h_2, q\} = \{0.6, 2\}$  for Ymodel I and  $\{h_2, q\} = \{0.1, 1\}$  for Ymodel II by the approach in Section 3. Table 2 and Web Table 3 present the MSEs of nonzero coefficient functions for the FMVCM and SVCm methods based on 100 Monte Carlo simulations under different combinations of  $n = (50, 100, 150)$  and noise levels  $\sigma_\epsilon^2 = (0.1, 0.25)$ . When the link function is nonlinear, the proposed FMVCM outperforms SVCm in all the settings considered. This result implies that SVCm is restrictive and cannot ade-

5.2 Performance of FMVCM26

Table 2: MSEs of nonzero coefficient functions and prediction errors using FMVCM (the proposed) and SVCM for Ymodel I (nonlinear  $\psi(\cdot)$ ).

$\sigma_\epsilon^2$	$n$	FMVCM							SVCM		
		$MSPE(Y)$	$MSE(\hat{\theta}_1)$	$MSE(\hat{\theta}_2)$	$MSE(\hat{g}_{31})$	$MSE(\hat{g}_{41})$	$MSE(\hat{g}_{32})$	$MSE(\hat{g}_{42})$	$MSPE(Y)$	$MSE(\hat{\theta}_1)$	$MSE(\hat{\theta}_2)$
0.1	50	1.57	0.02	0.03	0.18	0.45	0.62	0.46	4.47	0.40	0.30
	100	1.44	0.01	0.03	0.13	0.40	0.60	0.39	3.89	0.22	0.18
	150	1.44	0.01	0.02	0.11	0.32	0.47	0.33	3.55	0.14	0.16
0.25	50	1.83	0.03	0.04	0.26	0.48	0.69	0.48	4.61	0.40	0.31
	100	1.69	0.01	0.03	0.15	0.46	0.67	0.46	4.02	0.22	0.18
	150	1.61	0.01	0.02	0.12	0.32	0.61	0.44	3.66	0.14	0.16

quately reveal the nonlinear association between imaging responses and covariates. In contrast, FMVCM adapts the image-on-scalar regression to a multiple-index framework, thereby capturing the nonlinear features and being highly flexible for general situations. When the link function is linear, FMVCM performs similarly to or slightly worse than SVCM due to the estimated link function and group penalty. Moreover, FMVCM can account for the influence of the estimation and selection of NMFPCA scores contaminated with measurement errors and thus significantly outperforms SVCM. Likewise, the estimation performance improves as  $n$  increases. We also compare the prediction errors,  $MSPE(Y) = (100 \times 2611)^{-1} \sum_{i=1}^{100} \sum_{j=1}^{2611} \{\hat{Y}_i(\mathbf{s}_j) - Y_i(\mathbf{s}_j)\}^2$ , on a test data set. We first estimate the transformed NMFPCA scores from the test set based on the trained NMFPCA, then plug these estimated scores and scalar predictors from the test data set into the trained FMVCM and compute the predicted values  $\hat{Y}_i(\mathbf{s}_j)$ . Table 2 and Web Table 3 show that putting a sparsity penalty on the imaging regression model reduces prediction errors.

Figure 1 and Web Figure 2 depict the true surfaces and their estimated bivariate functions for Ymodels I and II. For an easy comparison of FMVCM and SVCM, we only

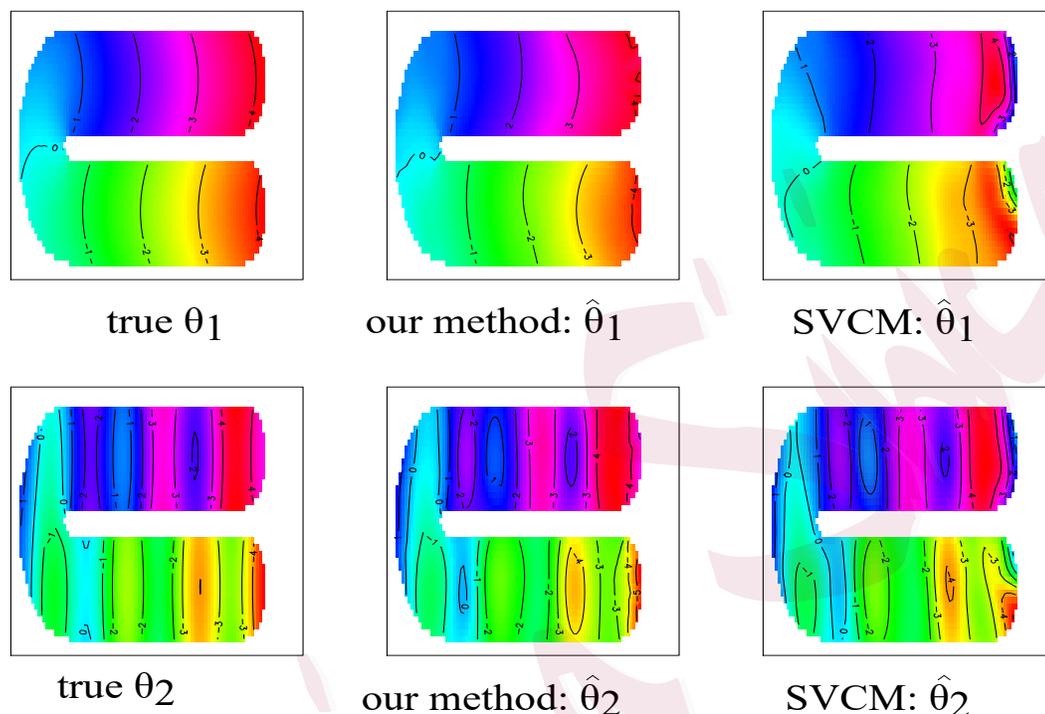


Figure 1: The true coefficient functions and average estimators of linear part for Ymodel I (nonlinear  $\psi(\cdot)$ ) with  $n = 100$  and  $\sigma_\epsilon^2 = 0.25$ .

present the bivariate functions of the linear parts. The estimated coefficient functions based on FMVCM are close to the truth and can capture the main features of the true bivariate functions in Ymodels I and II, whereas SVCM fails to capture the spatial pattern in Ymodel I. Web Table 4 presents the model selection results, including the selection percentages of each  $\mathbf{g}_k$ , the number of nonzero coefficient functions correctly identified as nonzero functions ( $\# \text{correct nonzero}$ ), and the number of zero coefficient functions correctly identified as zero functions ( $\# \text{correct zero}$ ). For each group of nonzero coefficient functions, the selection percentage is high and increases to 100% as the sample

size increases or the noise level decreases. Hence, our procedure can effectively identify important FPCs and rule out unimportant ones.

We examine the performance of BIC for selecting the number of index  $q$ . We calculate the frequency of  $q$  selected by BIC based on 100 repetitions in Ymodel I and Ymodel II. Web Table 5 reports the results under  $(n, \sigma_\epsilon^2) = (100, 0.25)$ . It appears that the approach in Section 3 works well in identifying the true  $q = 2$  in Ymodel I and  $q = 1$  in Ymodel II.

To investigate our method's robustness and efficiency, we compare FJMs with MFPCA + FMVCM, FMVCM with true  $\zeta_i$  (FMVCM+T), and FMVCM with true and important  $\zeta_i$  (FMVCM+TI). Results in Web Table 6 show that MFPCA + FMVCM produces large MSE and SD for  $\hat{g}_{31}$ ,  $\hat{g}_{41}$ ,  $\hat{g}_{32}$ , and  $\hat{g}_{42}$ . Thus, inadequate information extraction in the dimension reduction leads to severe estimation errors and unstable estimates. In contrast, the MSE and SD of FJM are close to those of the FMVCM+T and FMVCM+TI estimators, indicating that our method is robust with negligible loss of efficiency.

We also investigate the performance of the proposed estimators with misspecified  $K_n = 2, 7$ , and 12. Web Table 7 presents the MSE and SD of the estimators for Ymodel I with  $n = 100$  and  $\sigma_\epsilon^2 = 0.1$ , and the true value of  $K_n$  is 4. The results show that the estimation accuracy is worse when  $K_n$  is under-specified but similar to that under true  $K_n$  when the number of FPCs is over-specified. This finding implies that the proposed FJMs are insensitive to misspecified  $K_n$  when the extracted information in the dimension reduction procedure is sufficient.

Additionally, based on an anonymous referee's suggestion, we designed a setting to

mimic the ADNI dataset. We set  $n = 606$ ,  $p = 100$ , and  $K_n = 20$  with  $\mu_j(t) = t + \sin(\pi t)$  and  $f_j(u) = u^3/2 + 2.5u$  for  $j \leq 50$ , and  $\mu_j(t) = \exp(t)$  and  $f_j(u) = \sin(\pi u/5)$  for  $50 < j \leq 100$ , with  $\sigma^2$  values of 0.1 and 0.5. The  $n_{ij}$  values were drawn from a discrete uniform distribution over  $\{1632, \dots, 7787\}$ . We also set  $q = 2$ ,  $q_z = 6$ , and  $\sigma_\epsilon^2 = 0.1$ , using the estimated coefficient functions from the ADNI analysis for the left hippocampus without APOE- $\epsilon 4$  and disease status, denoted as  $g_{11}, \dots, g_{20,2}$  and  $\theta_1, \dots, \theta_6$ , respectively. Further,  $Z_{i1} = 1$ , while  $Z_{i2}$  and  $Z_{i3}$  were generated from  $U[1, 10]$ , and  $Z_{i4}$ ,  $Z_{i5}$ , and  $Z_{i6}$  from  $Bernoulli(0.5)$ . Other settings for generating  $\mathbf{X}_i(t)$  and  $Y_i(\mathbf{s})$  follow Ymodel I. Web Tables 8 and 9 show that our proposed method outperforms competitors and accurately identifies key covariates.

## 6. ADNI Data Analysis

### 6.1 Data Description and Preprocessing Pipeline

We applied the proposed FJM to the ADNI data set with imaging, genetic, and clinical variables collected by the ADNI study ([adni.loni.usc.edu](http://adni.loni.usc.edu)). The ADNI study started in 2004 with the primary objective of identifying biomarkers for accurate AD diagnosis in an early stage. In this data analysis, our ADNI data set includes hippocampal surface data, SNPs data, and clinical variables from  $n = 606$  subjects in ADNI1, including 113 AD, 316 mild cognitive impairment (MCI), and 177 cognitive-normal (CN). There are 361 males (mean age,  $76.03 \pm 6.65$  years old) and 245 females (mean age,  $75.04 \pm 6.57$  years old). We extracted hippocampal surface data from raw MRI data. These MRIs were collected

---

## 6.1 Data Description and Preprocessing Pipeline

from 1.5 Tesla MRI scanners with individualized protocols, and MRI protocol repetition time = 2400 ms, inversion time = 1000 ms, field of view = 24 cm, flip angle =  $8^\circ$ , and  $256 \times 256 \times 170$  acquisition matrix, yielding  $1.25 \times 1.26 \times 1.2$  mm<sup>3</sup> voxel size.

The processing pipeline for the MRI data includes standard steps and automatic regional labeling. Then, we adopted a hippocampal subregional analysis package (Shi et al., 2013) based on surface fluid registration. In the surface fluid registration, two cuts were introduced at the front and back of the hippocampal surface so that it could be converted into a genus zero surface with two open boundaries. The use of conformal parameterization can convert a 3D surface registration problem into a 2D image registration problem, and then the flow induced in the parameter domain establishes high-order correspondences among 3D surfaces. Hence, the radial distance map on the registered surface was generated for each subject, generating a  $N_s = 100 \times 150$  image response. Such hippocampus surface measures retain information on the deformation along the surface normal direction (Wang et al., 2011).

We aimed to examine the effects of clinical, genetic, and demographic variables on the left and right hippocampi. There are 6,087,025 genotyped and imputed SNPs on all of the 22 chromosomes and clinical covariates, including age, gender (1 = Female), education length, retirement (1 = Yes), and handedness (1 = Left). Since the number of SNPs is significantly larger than the sample size and the genetic variants present a strong blockwise correlation structure, we adopted a blockwise screening approach to reduce the number of candidate SNPs. Specifically, after using the default method (Gabriel et al.,

2002), Haploview (Barrett et al., 2005), and PLINK (Purcell et al., 2007) to form LD blocks, we screened the LD blocks while controlling for the clinical variables and the top five principal components of the whole genome data. We kept the top 100 Bonferroni significant LD blocks for the left and right hippocampi. As the left and right hippocampi are asymmetric (Pedraza et al., 2004), we apply FJM to each of them separately.

## 6.2 Analysis and Results

We performed NMFPCA on the top  $p = 100$  LD blocks from either the left or right hippocampus, controlling for covariates excluding the number of APOE- $\epsilon 4$  alleles and baseline diagnosis status. Individuals inheriting one or two copies of the APOE- $\epsilon 4$  allele face an elevated risk of developing Alzheimer's disease compared to non-carriers. Web Figure 3 presents the estimates of  $\hat{f}_j(\cdot)$ , revealing that the  $\hat{f}_j(\cdot)$  has a clear nonlinear pattern and justifying the necessity of delineating the nonlinear relationships of the top 100 LD blocks. To apply the proposed FMVCM and SVCVM to analyze the hippocampus surface data, we first fitted the left or right top 100 LD blocks using the conventional MFPCA and the proposed NMFPCA. The functional covariate  $X_{ij}(t_{ijd})$  was the SNP values at the  $d$ th genomic location on the  $j$ th LD block for the  $i$ th individual, with  $i = 1, \dots, 606, j = 1, \dots, 100$ , and  $d = 1, \dots, n_{ij}$ .  $n_{ij}$  is the number of genomic locations on the  $j$ th LD block for the  $i$ th individual, ranging from 1632 to 7787. We used the maxima and minima of locations in each LD block as 0 and 1 and scaled the locations into  $[0, 1]$  before analysis. Then, we selected  $(K_n, h_1, k_n) = (20, 0.07, 20)$  based on the proposed criterion in Section 3.

Next, we treated the extracted FPC scores as known covariates with the clinical covariates detailed in Section 6.1, and applied FMVCM and SVCM to the left and right hippocampal surface data separately.  $Y_i(\mathbf{s}_j)$  was the left or right hippocampus surface measure at point  $\mathbf{s}_j$  for the  $i$ th individual, with  $j = 1, \dots, 15000$ .  $\zeta_i$  was the transformed 20-dimensional latent score vector.  $\mathbf{Z}_i$  was 6-dimensional real-valued covariates including five clinical variables and an intercept term. In addition, we chose the number of indices  $q = 2$  by BIC (3.5) and used triangulation with 30 triangles and 24 vertices. Based on this triangulation, we generated the bivariate spline basis functions with  $(\varpi, v, h_2) = (2, 1, 0.16)$  and  $(\varpi, v, h_2) = (2, 1, 0.11)$  for the left and right hippocampi, respectively. We performed the computation on a single 14-core machine with 98GB of RAM. Web Table 10 reports the computation time, and Web Table 11 presents the selected FPCs for FMVCM with MFPCAs and NMFPCAs. The NMFPCA+FMVCM (FJM) selects fewer FPC scores than MFPCA+FMVCM.

We randomly split the data into two equal parts, the training and test sets, and evaluate the performance of MFPCA and NMFPCA in terms of the prediction error of 100 LD blocks (functional covariates). Table 3 presents the prediction errors and SDs for the functional covariates under  $K_n = 4, 20, 30$ , and 40. The results show that NMFPCA consistently yields smaller prediction errors than MFPCA. Similarly, we compared the performance of SVCM and FMVCM in predicting the radial distance of both hippocampi (image response). We considered four settings with different combinations of {MFPCA, NMFPCA} and {SVCM, FMVCM}. Table 4 presents the prediction errors and SDs for

the image response with large  $K_n = 20, 30,$  and  $40$ . We summarize our findings as follows. First, MFPCA+FMVCM and NMFPCA+FMVCM outperform MFPCA+SVCM and NMFPCA+SVCM, respectively. When the dimensional reduction procedure is fixed to MFPCA or NMFPCA, FMVCM exhibits better prediction performance than SVCM, implying that SVCM cannot adequately uncover the complex associations between the imaging responses and scalar covariates. Second, NMFPCA+FMVCM significantly outperforms MFPCA+FMVCM. When the scalar covariate model is fixed to FMVCM, NMFPCA is superior to MFPCA. These findings indicate that the nonlinear dependence of functional variables and the nonlinearly varying association between images and covariates in this application are essential. Hence, the proposed FJM (NMFPCA+FMVCM) model produces the smallest prediction errors among all the competing models. Furthermore, FJM demonstrates insensitivity to the increased number of FPCs, indicating that the BIC rule selects an appropriate number of FPCs that extract sufficient information about the functional covariates.

Table 3: Prediction errors (PE) and SD of the left and right top 100 LD blocks.

	left LD blocks				right LD blocks			
	$K_n = 4$	$K_n = 20$	$K_n = 30$	$K_n = 40$	$K_n = 4$	$K_n = 20$	$K_n = 30$	$K_n = 40$
MFPCA PE	0.8552	0.5359	0.5366	0.5354	0.8593	0.5353	0.5368	0.5418
SD	0.0028	0.0002	0.0013	0.0001	0.0019	0.0002	0.0038	0.0035
NMFPCA PE	0.5289	0.3331	0.3047	0.3487	0.5332	0.3476	0.3203	0.3435
SD	0.0021	0.0007	0.0001	0.0012	0.0015	0.0007	0.0009	0.0003

Next, we provide explanations for the estimated coefficient functions. Web Figures 4 and 5 show the estimated coefficient functions using FJM for the left and right hippocampi.

Table 4: Prediction errors (PE) and SD for the right and left hippocampi.

	without APOE- $\epsilon 4$ and disease status						given APOE- $\epsilon 4$ and disease status					
	left hippocampus			right hippocampus			left hippocampus			right hippocampus		
$K_n$	20	30	40	20	30	40	20	30	40	20	30	40
MFPCA + SVC	0.9849	0.9079	0.9100	0.9681	0.9761	0.9846	0.9626	0.8873	0.8891	0.9484	0.9465	0.9555
SD	0.0248	0.0516	0.0565	0.0099	0.0355	0.0515	0.0092	0.0509	0.0541	0.0243	0.0233	0.0362
MFPCA + FMVCM	0.9060	0.9465	0.9447	0.9041	0.9719	1.0224	0.8473	0.9534	0.9523	0.8751	0.9788	1.0042
SD	0.0366	0.0145	0.0083	0.0100	0.0396	0.0247	0.0049	0.0080	0.0007	0.0130	0.0055	0.0037
NMFPCA + SVC	0.9833	0.9702	0.9794	0.9814	0.9726	0.9790	0.9808	0.9692	0.9851	0.9620	0.9422	0.9675
SD	0.0165	0.0100	0.0200	0.0200	0.0159	0.0103	0.0177	0.0205	0.0163	0.0081	0.0026	0.0084
FJM (NMFPCA	0.7297	0.7220	0.7251	0.6795	0.6775	0.6763	0.6442	0.6414	0.6439	0.6454	0.6638	0.6502
+ FMVCM)	0.0252	0.0272	0.0267	0.0200	0.0212	0.0255	0.0210	0.0180	0.0244	0.0141	0.0196	0.0188

Let “Intercept” represent the image parameter of the intercept term, “FPC score  $k$ ” denote the regression coefficient function of the  $k$ th FPC score, and “Age” indicate the coefficient function corresponding to scalar covariate age. We observe an asymmetric clinical and genetic effect on the left and right hippocampi. Web Figure 4 presents four hippocampal subfields. The subfields Cornu Ammonis region 1 (CA1) and subiculum (Sub) show higher sensitivity than Cornu Ammonis region 2 (CA2) and Cornu Ammonis region 3 (CA3). The AD progression initially affects CA1 and subiculum and then spreads to CA2 and CA3 subfields (De Flores et al., 2015).

Web Table 12 presents the number of negative entries for the estimated coefficient functions of clinical variables. Web Figure 4 shows more than 13,800 pixels (over 15000 pixels) of the estimated coefficient function for age are negative, implying that hippocampus atrophy with age. On the other hand, more than 14,000 pixels of the estimated gender image are positive, indicating that hippocampus atrophy is more severe for men than women. Published medical reports (e.g., Mielke et al., 2014) found that men have a greater risk of MCI, and more than half of the individuals in this data are MCI, partially explaining why gender positively affects the radial distance of both hippocampi. The effect of retirement is negative in the CA1 and subiculum subregions, suggesting an excess risk of cognition deficit among retired individuals. Education produces a negative impact on the radial distance of both hippocampi. Existing literature (Olazarán et al., 2010) has found that as AD progresses, high educational attainment helps individuals maintain more efficient cognitive function over a short period, afterward leading to faster

dementia progression. A total of 13,429 and 5,079 entries of the estimated handedness image corresponding to the left and right hippocampi are negative, suggesting handedness is negatively (positively) associated with most pixels of the radial distances of the left (right) hippocampi. The cross-distributed control of the human brain partially explains this result.

We also analyzed the ADNI data set, including covariates such as the number of APOE- $\epsilon$ 4 alleles and baseline disease status. The details of the analysis and results are provided in Web Appendix C.

## 7. Concluding Remarks

In this paper, we considered the images on scalar and functional variables regression. We proposed an NMFPCA to accommodate the nonlinear relationships among functional variables. Then, we modeled the effects of functional covariates via the extracted nonlinear FPC scores. Moreover, we proposed FMVCM to investigate the spatially varying association between the imaging response and the FPC scores, as well as multiple covariates, allowing for highly flexible and complex association relationships. To our knowledge, little has been done in the joint analysis of complex associations among imaging measures, functional covariates, and multivariate scalar variables. In addition, we developed an estimation algorithm that combines the advantages of B-spline, bivariate splines, and kernel smoothing. Our method allows the number of principal components to diverge to infinity and performs the selection and estimation of nonlinear multivariate FPC scores through the penalized least square approach. These features enable our method to become a widely

applicable and flexible framework for image data analysis in high-dimensional settings.

This study can be extended in the following aspects. We may consider estimating the bivariate coefficient functions without estimating the link function. That is, treating the link function as a nuisance parameter, the interest is to find sufficient dimension reduction space. Furthermore, we will derive the optimal convergence rates and inference procedure for FJM. Whether the rates are optimal is worth further investigating.

### **Supplementary Material**

The Supplementary Material contains the theoretical proofs in Section 4, the estimation procedure in Section 3, ADNI analysis given APOE- $\epsilon 4$  and disease status in Section 6, and Web Tables 1–11 and Figures 1–7 in Sections 5 and 6.

### **Acknowledgements**

The authors would like to express their gratitude to the editor, and the anonymous referees for their careful reading and useful comments which led to an improved presentation of the paper. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://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](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf). This work was partially sup-

ported by GRF grant 14302519 from the Research Grant Council of the HKSAR for Dr. Song. Zhong's work was supported by National Natural Science Foundation of China (12401349) and China Postdoctoral Science Foundations (2024M751116).

## References

- Anowar, F., S. Sadaoui, and B. Selim (2021). Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne). *Computer Science Review* 40, 100378.
- Bair, E., T. Hastie, D. Paul, and R. Tibshirani (2006). Prediction by supervised principal components. *Journal of the American Statistical Association* 101(473), 119–137.
- Barrett, J. C., B. Fry, J. Maller, and M. J. Daly (2005). Haploview: analysis and visualization of ld and haplotype maps. *Bioinformatics* 21(2), 263–265.
- Bosq, D. (2000). *Linear processes in function spaces: theory and applications*. Springer, New York.
- De Flores, R., R. La Joie, and G. Chételat (2015). Structural imaging of hippocampal subfields in healthy aging and alzheimer's disease. *Neuroscience* 309, 29–50.
- Elliott, L. T., K. Sharp, F. Alfaro-Almagro, S. Shi, K. L. Miller, G. Douaud, J. Marchini, and S. M. Smith (2018). Genome-wide association studies of brain imaging phenotypes in uk biobank. *Nature* 562(7726), 210–216.
- Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, et al. (2002). The structure of haplotype blocks in

- the human genome. *science* 296(5576), 2225–2229.
- Hall, P., H.-G. Müller, and J.-L. Wang (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann Stat* 34(3), 1493–1517.
- Happ, C. and S. Greven (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association* 113(522), 649–659.
- Hibar, D. P., J. L. Stein, M. E. Renteria, A. Arias-Vasquez, S. Desrivieres, N. Jahanshad, R. Toro, K. Wittfeld, L. Abramovic, M. Andersson, et al. (2015). Common genetic variants influence human subcortical brain structures. *Nature* 520(7546), 224–229.
- Jack Jr, C. R., D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, et al. (2013). Tracking pathophysiological processes in alzheimer’s disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology* 12(2), 207–216.
- Kang, K., J. Cai, X. Song, and H. Zhu (2019). Bayesian hidden markov models for delineating the pathology of alzheimer’s disease. *Statistical methods in medical research* 28(7), 2112–2124.
- Kim, M. and L. Wang (2021). Generalized spatially varying coefficient models. *Journal of Computational and Graphical Statistics* 30(1), 1–10.
- Li, B., A. Artemiou, and L. Li (2011). Principal support vector machines for linear and nonlinear sufficient dimension reduction. *The Annals of Statistics* 39(6), 3182–3210.
- Li, J., C. Huang, Z. Hongtu, and A. D. N. Initiative (2017). A functional varying-coefficient single-index model for functional response data. *Journal of the American Statistical Association*

- ciation* 112(519), 1169–1181.
- Li, T., Y. Yu, J. Marron, and H. Zhu (2024). A partially functional linear regression framework for integrating genetic, imaging, and clinical data. *The Annals of Applied Statistics* 18(1), 704–728.
- Li, X., L. Wang, H. J. Wang, and A. D. N. Initiative (2021). Sparse learning and structure identification for ultrahigh-dimensional image-on-scalar regression. *Journal of the American Statistical Association* 116(536), 1994–2008.
- Luo, X., L. Zhu, and H. Zhu (2016). Single-index varying coefficient model for functional responses. *Biometrics* 72(4), 1275–1284.
- Mielke, M. M., P. Vemuri, and W. A. Rocca (2014). Clinical epidemiology of alzheimer’s disease: assessing sex and gender differences. *Clinical epidemiology* 6, 37–48.
- Morris, J. S. and R. J. Carroll (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B* 68(2), 179–199.
- Müller, H.-G. and F. Yao (2008). Functional additive models. *Journal of the American Statistical Association* 103(484), 1534–1544.
- Olazarán, J., B. Reisberg, L. Clare, I. Cruz, J. Peña-Casanova, T. Del Ser, B. Woods, C. Beck, S. Auer, C. Lai, et al. (2010). Nonpharmacological therapies in alzheimer’s disease: a systematic review of efficacy. *Dementia and geriatric cognitive disorders* 30(2), 161–178.
- Pedraza, O., D. Bowers, and R. Gilmore (2004). Asymmetry of the hippocampus and amygdala in mri volumetric measurements of normal adults. *Journal of the International Neuropsychological Society* 10(5), 664–678.

- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, M. J. Daly, et al. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* 81(3), 559–575.
- Ramsay, J. O. and B. W. Silverman (2005). *Functional data analysis*. Springer.
- Reiss, P. T. and R. T. Ogden (2010). Functional generalized linear models with images as predictors. *Biometrics* 66(1), 61–69.
- Sangalli, L. M., J. O. Ramsay, and T. O. Ramsay (2013). Spatial spline regression models. *Journal of the Royal Statistical Society: Series B* 75(4), 681–703.
- Shi, J., P. M. Thompson, B. Gutman, Y. Wang, A. D. N. Initiative, et al. (2013). Surface fluid registration of conformal representation: Application to detect disease burden and genetic influence on hippocampus. *NeuroImage* 78, 111–134.
- Stein, J. L., X. Hua, S. Lee, A. J. Ho, A. D. Leow, A. W. Toga, A. J. Saykin, L. Shen, T. Foroud, N. Pankratz, et al. (2010). Voxelwise genome-wide association study (vgwas). *neuroimage* 53(3), 1160–1174.
- Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The annals of Statistics*, 1348–1360.
- Veitch, D. P., M. W. Weiner, P. S. Aisen, L. A. Beckett, C. DeCarli, R. C. Green, D. Harvey, C. R. Jack Jr, W. Jagust, S. M. Landau, et al. (2022). Using the alzheimer’s disease neuroimaging initiative to improve early detection, diagnosis, and treatment of alzheimer’s disease. *Alzheimer’s & Dementia* 18(4), 824–857.

- Wall, J. D. and J. K. Pritchard (2003). Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics* 4(8), 587–597.
- Wang, H., R. Li, and C.-L. Tsai (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika* 94(3), 553–568.
- Wang, Y., Y. Song, P. Rajagopalan, T. An, K. Liu, Y.-Y. Chou, B. Gutman, A. W. Toga, P. M. Thompson, A. D. N. Initiative, et al. (2011). Surface-based tbm boosts power to detect disease effects on the brain: an n= 804 adni study. *Neuroimage* 56(4), 1993–2010.
- Xia, Y. (2008). A multiple-index model and dimension reduction. *Journal of the American Statistical Association* 103(484), 1631–1640.
- Yu, S., G. Wang, L. Wang, C. Liu, and L. Yang (2019). Estimation and inference for generalized geoadditive models. *Journal of the American Statistical Association* 115(530), 761–774.
- Yu, S., G. Wang, L. Wang, and L. Yang (2021). Multivariate spline estimation and inference for image-on-scalar regression. *Statistica Sinica* 31(3), 1463–1487.
- Zhu, H., T. Li, and B. Zhao (2023). Statistical learning methods for neuroimaging data analysis with applications. *Annual Review of Biomedical Data Science* 6(1), 73–104.
- Zhu, H., F. Yao, and H. H. Zhang (2014). Structured functional additive regression in reproducing kernel hilbert spaces. *Journal of the Royal Statistical Society: Series B* 76(3), 581–603.