ABSTRACT

MA, WANYING. New Functional Data Methods for Neuroimaging Data. (Under the direction of Luo Xiao.)

Because of technology advancement in both the processing and storage of large data sets, functional data has become common in many scientific areas nowadays. Neuroimaging techniques which allows for the further understanding of the brain inner working with disease diagnosis, cognitive process and psychological process have seen a explosive growth of cross-disciplinary research work in this field. The complex structures of neuroimaging data have posed special challenges and opportunities for functional data analysis. In this thesis, we are motivated by neuroimaging studies and propose new methods for functional data analysis.

In Chapter 2, motivated by a functional magnetic resonance imaging (MRI) study on pain caused by thermal stimuli applied to human subjects, we propose a new functional mixed model for scalar on function regression. The model extends the standard scalar on function regression for repeated outcomes by incorporating subject-specific random functional effects. Using functional principal component analysis, the new model can be reformulated as a mixed effects model and thus easily fit. A test is also proposed to assess the existence of the subject-specific random functional effects. We evaluate the performance of the model and test via a simulation study, as well as on data from the motivating fMRI study of thermal pain. The data application indicates significant subject-specific effects of the human brain hemodynamics related to pain and provides insights on how the effects might differ across subjects.

In Chapter 3, motivated by a diffusion tensor imaging (DTI) study of sports-related concussion (SRC) on college football athletes, we propose a combination of functional data methods and feature selection methods to differentiate two classes of athletes, players with concussion and players without concussion. To deal with the inherent difficulty of aligning the brain tracts across human subjects, we propose to consider the densities of scalar observations along the tracts. Then we transform the densities to functional data, for which we apply functional principal component analysis to extract features. As there is a large number of features, we propose to use a K-means based probabilistic subset search (PSS) method to conduct feature selection. In particular, the proposed PSS method is adapted to allow for homogeneous subgroup

clustering, which is essential due to the heterogeneous nature of SRC. The proposed methods are applied to the motivating DTI study and identify a small subset of features that provides good out-of-sample classification accuracy. Moreover, 2 subgroup clusters are identified for the concussed group, which confirms the heterogeneity of SRC.

New Functional Data Methods for Neuroimaging Data

by
Wanying Ma

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2019

APPROVED BY:

Anarb Maity	Rui Song
Xinge Jeng	Luo Xiao Chair of Advisory Committee

DEDICATION

To my dear parents and brother.

BIOGRAPHY

The author was born in November 1993 and grew up in Henan, China. Before she came to United States, she earned her Bachelor of Science degree in Statistics from School of Mathematical Sciences at Beijing Normal University in May of 2014. Later in August, she went to North Carolina State University to pursue the Ph.D. degree in Statistics at Department of Statistics. In May of 2017, she earned the Masters degree in Statistics at NC State, and she will defend her Ph.D. degree in 2019.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest and sincerest gratitude to my advisor, Dr. Luo Xiao for his infinite support and professional guidance throughout my graduate research. For the past three and a half years, Dr. Xiao devoted the endless patience and time to prepare me well for the projects, help me get through the challenges, provide insightful suggestions, sharpen my presentation skills and academic writing skills, and show me great examples for both scientific research and co-working. His great passion to research, earnest attitude, broad knowledge in statistics, professionalism have deeply inspired me and will always be the motivation to further improve myself in my career development. I feel truly grateful to ever have him as my advisor.

I would also like to thank Dr. Arnarb Maity, Dr. Rui Song, Dr. Jessie Jeng and Dr. Lokendra Pal for being my committee members and their valuable comments with regard to my dissertation. I also feel really fortunate to be a member of the big Statistics family at North Carolina State University, and I would like to extend my sincere appreciation to the whole Statistics graduate program. This will be the most important experience in my life, and it is such a wonderful thing to have you in this memory. In particular, I am very thankful to Dr. Wenbin Lu for his help on my graduate study as the director of graduate program, and I am very grateful to both Lanakila Alexander and Alyson McCoy for answering my countless questions. Many thanks to my dear friends, Benjamin Hu, Dr. Cai Li, Dr. Stephanie Chen, Dr. Munir Winkle, Dr. Yuan Feng, Ruonan Li for their supports and companion.

I would like to thank my mentors Dr. Mona Liu and Dr. Jian Zhu from Takeda, Dr. Radha Mohanty from Bayer Crop Science, and Dr. Youlan Rao from United Therapeutics. Thank you for providing these great internship opportunities and sharing your life experience with me. Your training, help and advice are one of the most important assets for me to prepare for my future career.

Last but not least, I want to thank my dear parents and brother. You give me the unconditional love and supports, cheer me up whenever I feel down in my life, and give me the strength to face every challenge. Without your trust, understanding and supports, I can never make it this far. Thank you for giving me such a lovely family that is full of love, faith and happiness.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

1.1 General Description

Functional data analysis (FDA) has kept raising research interests in recent decades. As the technology developments in the measurement accuracy and storage of large data sets, a variety of machines and devices across many scientific domains have been well designed to enable the continuous monitoring of data observations, which results in the functional or imaging data types. For example, one can consider the functional magnetic resonance imaging (fMRI) brain scan data set, where a continuous brain scan was made for each of 20 participants over several visits in a fMRI pain study, producing fMRI time series that consist of 23 equidistant temporal measurements of brain activation. These fMRI time series were recorded for each of 21 pain-responsive regions of interest over the brain for each participant at each visit. In FDA, the fMRI time series of the brain activation measurements are viewed as a function of time. Different from the traditional data format, functional data treats the function or curve as a basic unit instead of the single point, and FDA deals with the theory and methodologies for

functional data. The intrinsically infinite dimensional property of functional data has naturally posed challenges for the purposes of testing, regression and clustering in many scientific areas, and researchers have made great contribution in solving these problems and further developing the methods and tools in FDA; see [Ram06; Wan16] for a comprehensive overview.

Functional data is often observed in a continuous manner on a fixed or random (time or location) grid, and typically consists of a series of curves. The core assumption in FDA is smoothness, that is, it is assumed that the actual observed functional data is the noisy observation of a true underlying unknown smooth trajectory $X(\cdot)$, contaminated with some measurement errors. Since data can only be recorded at discrete time points, functional data can also be viewed as sampled realizations from the full infinite-dimensional trajectory $X(\cdot)$ at finite grids. In FDA, there are mainly two types of functional data depending on the different types of sampling schemes: (1) dense functional data, where functional data for all subjects is collected at the same dense grid $\{t_1, \ldots, t_n\}$, and the grid is equally spaced, i.e. $t_{k+1} - t_k = t_k - t_{k-1}$ holds for all k. For example, accelerometers and other wearable devices can record the activity counts at a constantly high frequency (usually summarized during one minute epoch) throughout the day [Gol15; Spi11], and thus gives the form of densely observed functional data. (2) sparse functional data, where the observation grid varies from subject to subject, and the grid per subject usually only consists of a small number of random points. The sparse functional data often refers to the irregularly spaced longitudinal data [MÜL05; Yao05; Ma12].

In this thesis, we focus our work on the dense functional data, and consider the observed functional data of the following form $\{(Y_{ij}, t_j) : i = 1, \ldots, I, j = 1, \ldots, n\}$, where Y_{ij} is the ith subject's measurement observed at point t_j in a compact interval $\mathcal{T}$. In the fMRI data set example, for simplicity's sake, considering only one region of interest and one specific visit, Y_{ij} represents the brain activation measurement observed at the jth time point t_j for the ith subject. Therefore, the actual observed data Y_{ij} is modeled as

$$Y_{ij} = X_i(t_j) + \epsilon_{ij}, \tag{1.1}$$

where $X_i(\cdot)$ is the unknown smooth stochastic process in $\mathcal{L}^2(\mathcal{T})$, i.e. $\mathbb{E}(\int_{t\in\mathcal{T}} X_i^2(t)\mathrm{d}t) < \infty$ with mean function $\mu(t)$ and covariance function $\mathcal{C}(s,t)$, and measurement error ϵ_{ij} is the

independent and identically distributed (i.i.d.) random noise with zero mean and often assumed to have homoscedastic variance σ^2. The sampling schemes have the functional data equipped with natural order in time or location. Although dense functional data has similar form to the multivariate data in the sense that multiple data points using the same sampling design are collected per subject, this ordering nature forbids switching columns for functional data whereas multivariate statistics are free from this restriction. On the other hand, distance between the columns, which is related to smooth and estimate the covariance function, matters for functional data. Moreover, for dense functional data, traditional dimension reduction techniques such as principal component analysis (PCA) is often not applicable due to the high dimension curse [Jol03]. These properties separate functional data analysis from multivariate data analysis, and present special challenges and opportunities.

In FDA, functional principal component analysis (FPCA) has been one of the most important techniques for exploring and interpreting functional data. Extending PCA for multivariate data, FPCA projects the infinite-dimensional curves onto a finite number of principal directions that have the most variability among the curves. It converts the high dimensional functional data to low dimensional uncorrelated random score vectors. The resulting random scores can be fed into regression or classification models for other research purposes. At the same time, FPCA provides a way to smooth out the noisy realizations and recover the latent underlying smooth trajectory $X_i(\cdot)$ based on the discrete observations Y_{ij} for dense or sparse functional data. Modes of variation can also be visualized in standard software through simple calculation on the obtained random scores and eigenfunctions. A lot of efforts has been made for motivating the development of the functional principal component analysis; see [Rao58; CM97; EM96; EM03; Yao05; Di09; Cra09; JW10; Gre11; Xia16; Zho10] for more details.

While a great amount of research work has been done in the framework of univariate functional data, multivariate functional data has been increasingly causing people's attention. As technology proceeds, often more than one functional features can be simultaneously collected from the same observation unit, resulting in the multivariate functional data. Although dimensionality reduction for this type of data can be done through separate FPCA on each functional counterpart or concatenating the multiple functional process as one single long vector following by the standard FPCA [Ram06], the simultaneous variation among the different elements are not directly handled,

which leads to inaccurate estimation of random scores and multicollinearity issues for subsequent analysis based on the obtained scores [HG18; Xia18; MS05; Chi14; LC11]. An incomplete list of work regarding multivariate functional principal component analysis (mFPCA) includes: a normalized mFPCA method with first normalizing the random curves in [Chi14], a mixture density model for clustering multivariate functional data based on the normalized principal component scores in [JP14b], a pointwise univariate PCA in [Ber11], a multivariate FPCA method that is applicable to functional data collected over different dimensional domains based on univariate FPCA in [HG18], a fast algorithm ffor the covariance estimation based on the bivariate penalized spline in [Li18]. We will detail the methodologies regards the multivariate functional analysis later on. In the appendix A.2 of Chapter 2, we propose a new fast covariance estimation algorithm for multivariate dense functional data.

In particular, the blooming of neuroimaging techniques, providing the both functional brain imaging and structural brain imaging as powerful means to enhance the understanding of the relationship between brain inner processing with disease states, psychological process, and other cognitive tasks, has brought in the new chapter of neuroscience. See [BK09] for an extensive review. Functional brain imaging emphasizes the function of brain, which helps reveal when and where brain is functioning. Structural brain imaging, on the other hand, provides anatomical information of brain, and helps visualize the brain activity along neural pathway. FDA has the appealing advantage of exploring these huge amount of information with functional property, and hence is highly demanded for these neuroimaging data analysis [Lin08; Tia10]. There are several modern functional brain imaging techniques including fMRI, electroencephalography (EGG), positron emission tomography (PET), which mainly differ in the ways of measuring neuronal activity. For example, fMRI and PET detect the brain activation through the changes on the oxygenated blood flow, with high spatial resolution and low temporal resolution; EEG measures the brain signals through the electrical activation of neurons with high temporal resolution and low spatial resolution. Modern structural techniques, such as diffusion tensor imaging (DTI), traces the water diffusivity along the white matter tracts [Bas94; Lin08; Has17]. Our work is motivated by the neuroimaging studies including fMRI and DTI.

Recently, there is a vast volume of research studies on clustering methods for functional data. The clustering analysis for functional data aims at finding the homogeneous subgroups among the

various curves. Typical clustering methods for functional data lie in the following four directions [JP14a]: (1) using raw functional data; (2)approximating the curves using basis functions or eigenfunctions and clustering the coefficients; (3)simultaneously performing FPCA with clustering; (4)distance-based clustering methods. For the first approach, the high-dimensional multivariate clustering techniques are often adopted and thus the functional properties (such as continuity, smoothness) of the data are neglected; see [BBS14] for an extensive review. For the second approach, the dimension reduction process using B-splines followed by K-means clustering are considered in [Abr03; PM08]; the multilevel FPCA followed by ANOVA is studied in [SJ12]. For the third approach, a model-based framework is adopted in [JS03; CL07; Hua14]. For the fourth approach, different distance disciplines combined with K-means clustering or Hierarchical clustering are considered in [TK03; FV06; Tok07].

In this dissertation thesis, we focus on the dense functional data, and aim at addressing scientific problems for testing, prediction and classification motivated by neuroimaging studies. Functional data has become more commonly collected in many areas, and provides us with a rich source of information. Therefore, it is of interest to make good use of functional predictors to either predict continuous responses or classify subjects from different groups. For example, in our motivating DTI data set, four diffusivity measurements are collected along 27 brain tracts for two groups of football athletes, with or without concussion disease, and each tract profile has thousands of observations. It is of scientific importance to explore the relationship of brain activation with the injury appearance, and in other words, we could build a model to distinguish two groups of subjects using these multivariate functional predictors. This high dimensional data set, with $n \ll p$, posed challenges in terms of dimension reduction and feature selection. In this thesis, we will develop methods and algorithms to answer these scientific questions arising from functional data in various biostatistical applications.

In the following sections, we will review some common topics in FDA that are closely related to subsequent chapters, and conclude with a summary of the research projects.

1.2 Univariate Functional Data

1.2.1 Functional principal component analysis

FPCA is the fundamental technique in FDA in that it converts the inherently infinite-dimensional functional data to a finite number of uncorrelated vectors, achieving both smoothness and dimension reduction purposes for dense or sparse functional data. In addition, it has been used as building blocks for developing advanced methods in FDA. In this section, we will review FPCA for univariate dense functional data in more details.

We still consider the dense functional data in model (1.1). Mercer's theorem states that spectral decomposition on the covariance function gives

$$C(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t), \tag{1.2}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots 0$ are the non-negative eigenvalues, and $\phi_k(t)$'s are the corresponding orthornormal eigenfunctions. Then using Karhunen-Loéve (KL) expansion, $X_i(t)$ can be represented as

$$X_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik}\phi_{ik}(t) \tag{1.3}$$

where $\xi_{ik} = \int_{\mathcal{T}} \{X_i(t) - \mu(t)\}\phi_k(t)\mathrm{d}t$ is the kth functional principal component (FPC) score of $X_i(t)$. These FPCs ξ_{ik} are random variables with zero mean and variance λ_k, being independent across i and uncorrelated across k. In practice, we often use a truncated number of FPCs to approximate the random process as $X_i(t) = \mu(t) + \sum_{k=1}^{K} \xi_{ik}\phi_k(t)$, where K can be selected using common approaches such as percentage of variance explained (PVE)[Ram06; PM08; Xia16], Akaike information criterion (AIC)[Yao05; Li13], Bayesian Information Criterion(BIC)[Li13], and cross-validation[RS91; HN13].

FPCA further facilitates the development of modes of variation in functional data to understand the fluctuation among these random curves. A lot of applications have demonstrated its usage in illustrating the contribution of each eigenfunction $\phi_k(s)$ in the random process $X(s)$; see more details in [Cas86; JR92; Ram06; PM16]. The kth mode of variation corresponding to

the kth eigenfunction is defined as a set of functions parameterized by $\alpha \in \mathbb{R}$ as

$$\mu(s) \pm \alpha\sqrt{\lambda_k}\phi_k(s). \tag{1.4}$$

Practically for both the KL expansion and modes of variation, the main components in equation (1.3) and equation (1.4) will be substituted by the empirical estimation. Approaches for estimating eigenfunctions and predicting the random scores in FPCA depends on whether the functional data is dense or sparse.

For dense functional data, which is observed on the same dense fine grid for all the subjects, one of the typical approaches is to smooth the covariance operator followed by spectral decomposition. Specifically, first the mean function can be estimated as $\hat{\mu}(t_j) = \frac{1}{I}\sum_{i=1}^{I} Y_{ij}$, and it follows that sample covariance function can be calculated by $\widehat{C}(t_j, t_{j'}) = \frac{1}{I-1}\sum_{i=1}^{I}\{Y_{ij} - \hat{\mu}(t_j)\}\{Y_{ij'} - \hat{\mu}(t_{j'})\}$. Then the sample covariance is smoothed to recover the true surface before the decomposition. A number of relevant research work in this context includes [BR86; SL98; Yao03; ER35; Sil96; KU01; Xia16]. Then the random scores ξ_{ik}'s can be predicted using best linear unbiased predictors (BLUPs) or numerical integration following equation (1.3)[Xia16].

In the situation of sparse functional data, where only a few irregularly spaced observations are recorded per subject, FPCA also serves as an imputation tool by predicting the underlying smooth full trajectory $X(\cdot)$. There are rich literatures on this topic; see more details in [Shi96; SL98; Jam00; RW01; Yao05; PP09; Xia18]. Among them, main approach aims at smoothing the noisy covariance operator using different smoothing techniques such as P-splines, thin-plate regression splines, and local polynomial smoother. Several other methods focus on the rank-restricted models such as reduced rank mixed effect model in [RW01] and restricted maximum likelihood estimator for the FPCs in [PP09]. Following that, the spectral decomposition can be carried out and random scores ξ_{ik}'s can be obtained using BLUPs [Yao05; Xia18].

In addition to the direct work towards FPCA, there has been advanced methods regarding regression and clustering using FPCA as building blocks. Indeed, once the FPCA process is completed, one can use the obtained random scores for other research purposes. For regression purposes, there are typically three different types of functional regression models depending on whether the predictor or response involve the functional parts: (1) scalar on function regression;

(2) function on scalar regression; (3) function on function regression. There are a lot of successful applications. See for example a generalized functional linear model in [MS05], a non-parametric random effect model for predicting the clustering membership in [CL07], a joint modeling and clustering method for the cocaine relapse behavior in [Hua14].

1.2.2 Scalar on function regression

Scalar on function regression models the relationship between the scalar outcomes and the functional predictors. Various methods including linear approaches, nonlinear approaches, and non-parametric approaches have been developed. We refer to [MÜL05; FV06; Mor15; Rei17] for an extensive review. Extending the traditional multivariate linear regression to allow for the functional predictors, functional linear model (FLM) is the commonly used functional regression model to fit this type of regression. Considering the continuous real-valued outcome, FLM with scalar outcomes can be formulated as

$$Y_i = \alpha + \int_{\mathcal{T}} \{X_i(t) - \mu(t)\}\beta(t)\mathrm{d}t + \epsilon_i, \tag{1.5}$$

where Y_i and X_i are the scalar outcome and the observed functional predictor of the ith subject, respectively; α is the scalar intercept, $\beta(t)$ is the coefficient function, and ϵ_i is the i.i.d. random errors with zero mean and constant variance σ^2. Here, different from the multivariate linear regression, FLM uses the integration over the observed domain $\mathcal{T}$ to accommodate the functional predictor, and assumes that both $X_i(\cdot)$ and $\beta(\cdot)$ are the Gaussian process over $\mathcal{T}$.

General approaches to estimate the coefficients in model (1.5) can be splitted into two directions [Rei17]: (1) expand the coefficient $\beta(t)$ using a set of known basis functions, often spline basis, Fourier basis or wavelet basis, then α and $\beta(t)$ can be estimated by optimizing the sum of squared errors with some roughness penalty. Specifically, let $\beta(t) = \mathbf{B}(t)^\mathsf{T}\boldsymbol{\beta}$, where $\mathbf{B}(\cdot) = [B_1(\cdot), \ldots, B_K(\cdot)]^\mathsf{T}$ are the basis functions with corresponding coefficient vector $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]^\mathsf{T} \in \mathbb{R}^K$. The objective function can be derived as $\sum_i [Y_i - \alpha - \int_{\mathcal{T}} \{X_i(t) - \mu(t)\}\mathbf{B}(t)^\mathsf{T}\boldsymbol{\beta}\mathrm{d}t]^2 + \lambda\mathcal{P}(\boldsymbol{\beta})$, where $\mathcal{P}$ is the penalty function on $\boldsymbol{\beta}$. $\beta(t)$ can be estimated once the estimated basis coefficients $\hat{\boldsymbol{\beta}}$ is obtained. (2) expand the coefficient $\beta(t)$ using the orthonormal eigenfunctions derived from the functional predictor $X(\cdot)$ using FPCA. To be more specific,

first, the truncated version of KL expansion gives $X_i(t) = \mu(t) + \sum_{k=1}^{K} \xi_{ik}\phi_k(t)$, then $\beta(t)$ can be projected onto the eigenfunctions $\phi_k(t)$ such that $\beta(t) = \sum_{k=1}^{K} \theta_k\phi_k(t)$. It follows that model (1.5) can be rewritten as

$$Y_i = \alpha + \sum_{k=1}^{K} \xi_{ik}\theta_k, \tag{1.6}$$

which is a linear mixed model and easy to fit. Then $\hat{\beta}(t)$ can be obtained by plugging in $\hat{\theta}_k$.

For the type of scalar response with exponential-family distributions, the functional generalized linear model can be established by replacing the left-hand-side of model (1.5) as the link-transformed expected mean response [Jam02; MS05]. Other nonlinear approaches for scalar on function regression can often be viewed as extensions from (generalized) FLM. Related research work includes but not limited to the multiple-index model in [JS05], functional additive model with the additive structure on the FPCs in [MY08], additive structure on the functional predictor itself by a smooth bivariate function in [Mül13; McL14].

1.3 Repeatedly Observed Functional Data

Recently, functional data with repeatedly observed curves at multiple instances (often visits) on the same unit (often subjects) gives rise to the repeatedly observed functional data, and has attracted a lot of research attentions. Extended from the single-level functional data as is considered in model (1.1), repeatedly observed functional data not only consider the between-curve correlation, but also the within-curve correlation due to the multiple visits made from the same observational unit.

Direct application of FPCA on the multi-level functional data ignores the dependency between curves of the same units. Therefore, a multi-level FPCA is proposed by [Di09] to estimate both within-curve and between-curve covariance structure. Consider the observed multi-level functional data $\{(Y_{ijk}, t_k) : i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, n\}$, where Y_{ijk} is the ith subject's measurement observed over the jth visit at point t_k in a compact interval $\mathcal{T}$. Similarly, the repeated observed curve is modeled as the true underlying smooth process $X_{ij}(\cdot)$ with random errors ϵ_{ijk}. ϵ_{ijk} is independent across i, j and k. First, the random process $X_{ij}(\cdot)$ is decomposed via a hierarchical functional model [Mor03] as: $X_{ij}(t) = \mu(t) + a_j(t) + b_i(t) + v_{ij}(t)$, where $\mu(t)$

is the population mean function, $a_j(t)$ is the visit-specific fixed shift from the population mean function, $b_i(t)$ and $v_{ij}(t)$ are the subject-specific deviation and subject-visit-specific deviation from the visit-specific mean function and subject-specific mean function, respectively. Then subject-specific effect $b_i(t)$ and subject-visit-specific effect $v_{ij}(t)$ can be expanded using the two levels of eigenfunctions via the truncated KL expansion:

$$b_i(t) = \sum_{k=1}^{K_1} \xi_{ik}\phi_k^{(1)}(t), \quad v_{ij}(t) = \sum_{\ell=1}^{K_2} \eta_{ij\ell}\phi_\ell^{(2)}(t), \tag{1.7}$$

where ξ_{ik} and $\eta_{ij\ell}$ are the zero mean random scores associated with the level 1 eigenfunctions $\phi_k^{(1)}(t)$ and level 2 eigenfunctions $\phi_\ell^{(2)}(t)$, respectively. The between- and within- curve covariance functions can be estimated by method of moment estimators for dense functional data, and by bivariate smoother for sparse functional data. FPC scores can be predicted by numerical integration or BLUPs for dense functional data, and by BLUPs for sparse functional data. The number of truncated eigenfunctions K_1 and K_2 again can be selected using AIC, PVE or cross-validataion. Multilevel functional data are also discussed in terms of the wavelet-based multilevel hierarchical model in [Mor03], wavelet-based functional mixed model in [MC06], multilevel functional regression in [Cra09] and among others.

Longitudinal collected functional data naturally the forms the repeatedly observed functional data. Existing studies incorporating the specific visiting time T_{ij} for the ith subject at jth visit allows for another layer of longitudinal dynamics information. For example, [Gre11] extend the classical longitudinal mixed model to the functional context and introduce a functional linear mixed model with a linear structure on the visiting time T_{ij}; [Ger13] extend the work of [Gre11] and propose two new longitudinal FPCA based regression models; [CM12] develop a two-step longitudinal FPCA with time-dependent eigenfunctions which enables the prediction of full trajectory at a future visit; [PS15] propose a model framework for longitudinal sparse design with time-invariant marginal basis functions and time-varying scores, which is computationally faster than [CM12] and also allows for curve reconstruction at any visit time; [Che17] develop the representation framework for longitudinal functional data using marginal FPCA with product FPCA.

1.4 Multivariate Functional Data

With the rapid development of the recording and storage techniques, functional data in the form of multivariate structure becomes more commonly collected recently. For example, in the motivating data set from a fMRI thermal pain study considered in Chapter 2, where brain activation curves from 21 pain-responsive brain regions during the entire experiment time course were recorded for each of 20 participants at multiple repetitions, we essentially have multiple functional predictors and thus are also interested in seeing how these multiple brain regions together can be related to the self-reported pain ratings. Multivariate functional data, which has the multivariate property and is associated with infinite-dimensional curves, presents special challenges of the high dimensionality. Because of this, we need more powerful dimension reduction tools.

Let $\mathbf{X}(t) = [X_1(t), X_2(t), \ldots, X_p(t)]^\mathsf{T} \in \mathbb{R}^p$ $(p > 1)$ denote a $p-$dimensional multivariate functional data, where $X_i(t)(1 \leq i \leq p)$ is the univariate random function in $L_2(\mathcal{T})$. Let $\boldsymbol{\mu}(t) = \mathbb{E}\{\mathbf{X}(t)\} = \{\mu_1(t), \ldots, \mu_p(t)\}^\mathsf{T}$, $\mathcal{C}(s,t) = \text{cov}\{\mathbf{X}(s), \mathbf{X}(t)\} = [C_{j_1 j_2}(s,t)]_{1 \leq j_1, j_2 \leq p} \in \mathbb{R}^{p \times p}$, where $C_{j_1 j_2}(s,t) = \text{cov}\{X_{j_1}(s), X_{j_2}(t)\}$. Multivariate Mercer's theorem admits the spectral decomposition on the covariance function as $\mathcal{C}(s,t) = \sum_{\ell=1}^{\infty} \lambda_\ell \left(\phi_{1\ell}(t), \ldots, \phi_{p\ell}(t)\right)^\mathsf{T} \left(\phi_{1\ell}(t), \ldots, \phi_{p\ell}(t)\right)$, where λ_ℓ is the nonnegative eigenvalue in non-increasing order with corresponding multivariate eigenfunction $\left(\phi_{1\ell}(t), \ldots, \phi_{p\ell}(t)\right)^\mathsf{T}$ such that $\sum_{j=1}^{p} \int \phi_{j\ell_1}(t)\phi_{j\ell_2}(t)\mathrm{d}t = 1_{\{\ell_1 = \ell_2\}}$. Then the multivariate KL expansion represents $\mathbf{X}(t)$ as $\boldsymbol{\mu}(t) + \sum_{\ell=1}^{\infty} \xi_\ell \left(\phi_{1\ell}(t), \ldots, \phi_{p\ell}(t)\right)^\mathsf{T}$, where $\xi_\ell = \sum_{j=1}^{p} \int (X_j(t) - \mu_j(t))\phi_{j\ell}(t)\mathrm{d}t$. Here ξ_ℓ's are uncorrelated random scores with $\mathbb{E}(\xi_\ell) = 0$ and $\text{Var}(\xi_\ell) = \lambda_\ell$. Similar to the univariate FPCA, the truncated number of multivariate FPCs can be determined by PVE or AIC [Chi14; Li18; Hap19].

Traditional FPCA can be carried out for multivariate functional data by conducting univariate FPCA process on each functional variable, or by concatenating multiple functional variables into a single long vector followed by the standard FPCA and restructuring the resulting FPCs back for each functional part as we mentioned earlier. However, for the former case, the resulting FPC scores are in presence of the multicollinearity issues as the joint variation among the functional variables is not directly measured. This also leads to insufficient interpretation of the

data and causes problems for the subsequent regression analysis [MS05; Gol12]. For the latter case, a successful application in the bivariate functional data is discussed in [Ram06], where the author also points out that difference in the variability and magnitude between the two functional elements will affect the FPCA results. Existing approaches for the multivariate FPCA can be generally categorized in the following three groups: (1) covariance operator estimation; (2) reduced rank model; (3) univariate FPCA. For the first group, for example, [Chi14] and [JP14b] investigate the normalized multivariate FPCA methods using the normalized covariance operator, [Li18] propose a fast covariance estimation algorithm using the tensor-product B-spline presentation of the covariance function for multivariate sparse functional data. For the second group, [Zho08] utilize the penalized spline within the reduced-rank model framework for paired longitudinal data extending the work from [Jam00]. For the third group, [Ber11] repeatedly apply the classical multivariate PCA on each observed point to build the FPCs, and [Hap19] propose a multivariate FPCA method for multi-dimensional domain functional data by the relationship between the multivariate FPCA and univariate FPC scores.

1.5 Research Projects

In the subsequent chapters, we develop new functional regression model, testing methods, and clustering algorithm motivated by neuroimaging studies.

In Chapter 2, we propose a new functional mixed model for scalar on function regression to model the relationship between the fMRI time series predictors with the scalar pain ratings reported by the participants in a fMRI thermal pain study. The model extends the standard scalar on function regression for repeated outcomes by corporating subject-specific random functional effects. The new model can be reformulated as a mixed effects model using FPCA which can be easily fit in standard software. Following that, we further propose an equal-variance test to assess the existence of the subject-specific functional random effects within the proposed model framework. The null hypothesis is that the relationship between the fMRI curves with ratings stays the same across subjects, and alternative hypothesis states the existence of the subject-specific functional effect. In this proposed test, testing for the random functional effects can be converted as the zeroness test of a single variance component. In addition, the new

model has been extended to handle the multivariate functional predictors (e.g. multiple fMRI recordings from different brain regions per subject), and thus the test can also be used to detect the subject-specific signal-response relationship across multiple brain regions. We demonstrate that our proposed test has better power and size property compared to the existing methods through a simulation study. The nominal type I error and high power has been demonstrated in the various simulation settings. Data application indicates significant subject-specific effects of the human brain hemodynamics related to pain and provides insights on how the effects might differ across subjects.

In Chapter 3, we propose a combination of new feature construction process and a novel probabilistic subset search (PSS) algorithm for simultaneous flexible feature selection and clustering analysis. We motivate this algorithm by a DTI study to distinguish two groups, with or without sports-related-concussion brain injury. In this study, hundreds of functional predictors in the form of the tract-measurement combinations are provided in the DTI scan for each of 73 football athletes. It is of both interest and importance to figure out the capability of using these brain signals to detect the disease appearance, and more specifically, to find out which tract and measurement make the difference. This data set presents special challenges: (1) there are multivariate functional predictors with each tract profile having thousands of observations; (2) the length of the functional tract profiles are being both subject-specific and tract-specific. Therefore, we first propose to consider the feature extraction equipped with the density function, log quantile density transformation and FPCA to deal with the difficulty of aligning the tract profiles of different lengths and create lower dimensional features. However, this still results in a $n \ll p$ scenario. We then propose a K-means based flexible PSS algorithm to conduct weighted subset search that allows for the homogeneous subgroup clustering structure to identify more relevant features and discard irrelevant features to differentiate two groups. Althrough there are truly two groups, PSS allows for more than two clusters, with potentially multiple clusters corresponding to the subgroups of the true group. The proposed PSS method is applied to the motivating DTI data set. Improved performance in terms of higher out-of-sample classification accuracy in a cross-validation fashion is demonstrated using different starting values compared with the greedy sequential search, sparse K-means clustering and logistic regression with Lasso penalty. A small number of features are identified with relatively high discriminant power, and

a two-cluster subgroup structure for the concussed group confirms the heterogeneity nature of concussion.

CHAPTER

2

A FUNCTIONAL MIXED MODEL FOR SCALAR ON FUNCTION REGRESSION WITH APPLICATION TO A FUNCTIONAL MRI STUDY

2.1 Introduction

Scalar on function regression models [Ram06] are used to relate functional predictors to scalar outcomes and are becoming increasingly popular in statistical applications (e.g., [Gol11], [Mor15] and [Rei17]). These models have also been extended to data with repeated outcomes (e.g., [Gol12] and [Ger13]). However, existing models only model the effects of the functional predictor as fixed and do not allow for random functional effects that are either subject- or outcome-specific.

We are motivated by a functional magnetic resonance imaging (fMRI) study of thermal pain

([Lin12a]). We begin by briefly describing the study, which was performed on 20 participants. A number of stimuli, consisting of thermal stimulations delivered to the participants left forearm, were applied at two different levels (high and low) to each participant. The temperature of these painful (high) and non-painful (low) stimuli were determined using a pain calibration task performed prior to the experiment. After an 18s time period of thermal stimulation (either high or low), a fixation cross was presented for a 14s time period until the words "How painful?" appeared on the screen. After four seconds of silent contemplation, participants rated the overall pain intensity on a visual analog scale (VAS). The ratings took continuous values and were re-scaled within the range of 100 to 600. The experiment concluded with 10s of rest. During the course of the experimental trial, each subject's brain activity was also measured using fMRI. Data was extracted from different known pain-responsive brain regions across the brain. Each time course consisted of 23 equidistant measurements made every 2s, providing a total of 46s of brain activation, ranging from the time of onset of the application of the stimuli to the conclusion of the pain report. The same experiment was conducted multiple times on each participant, with the total number of the repetitions ranging from 39 to 48, thereby giving rise to an unbalanced design. To illustrate the structure of the data, Figure 2.1 shows the fMRI and the pain rating data for two subjects each with three repetitions, mimicking Figure 1 in [Gol12].

In previous work, Lindquist [Lin12a] used this data set to study how brain activation affected the pain rating using a scalar on function regression model that treated the continuously observed fMRI data as a functional covariate and the subjective rating as a scalar response. However, they used a population model that did not allow for the subject-specific effect of the fMRI imaging on the pain rating to be appropriately modeled. In this work, we seek to determine whether the fMRI data affects the pain rating in a unified or subject-specific manner. For this purpose, we extend the scalar on function linear regression to a new functional mixed effects model for repeated outcomes, and develop a test to determine if the relation between the brain imaging data (more specifically, fRMI data at one brain region) and the pain rating is subject-specific or not. Suggested by the Associate Editor, we further extend the proposed model and test to simultaneously assess the association between pain rating and fMRI data at multiple brain regions.

Testing for the lack of an effect in a functional predictor, i.e., whether the coefficient function

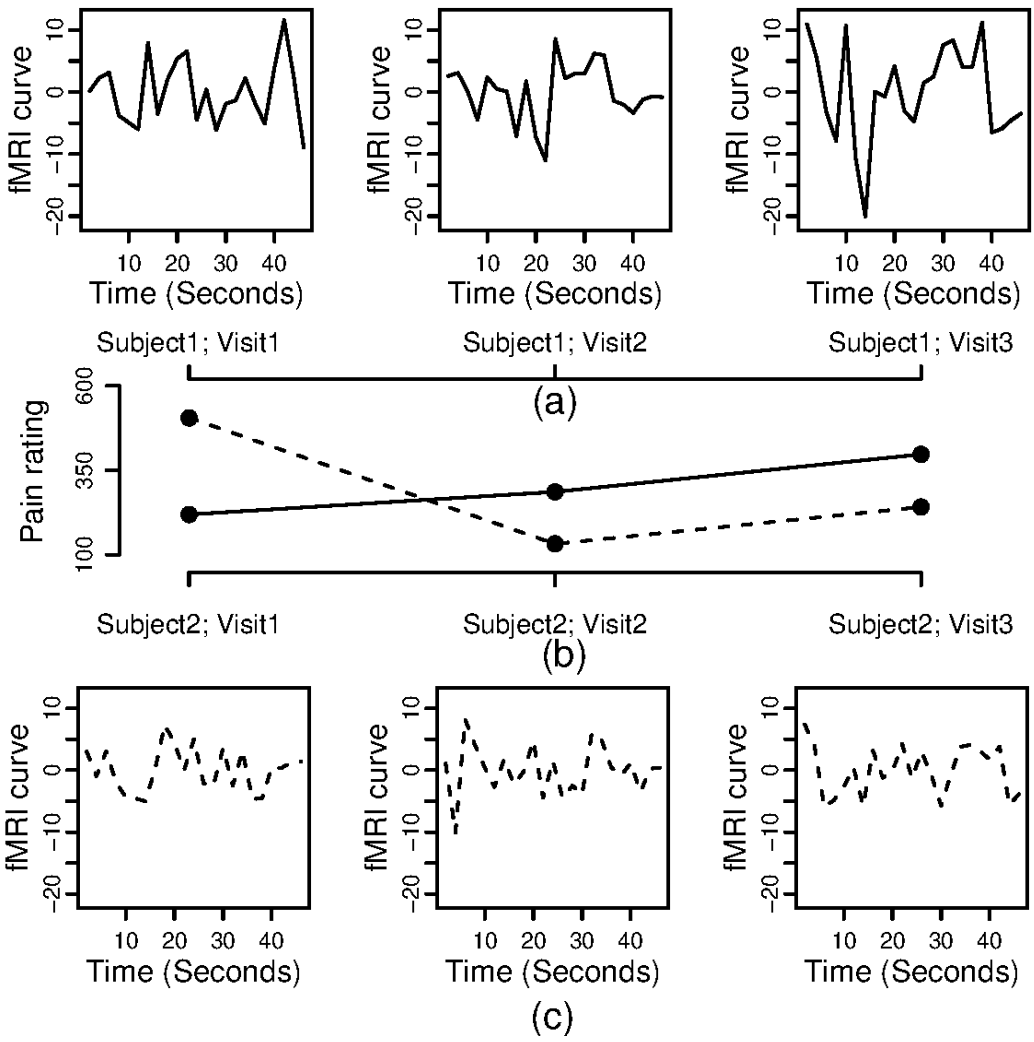

Figure 2.1 Data from the fMRI study: (a) and (c) give the fMRI time series at ROI LAIns for two subjects each with three repetitions; (b) presents the corresponding spaghetti plots of pain ratings.

is exactly zero, has been well developed in the scalar on function regression literature. For example, Cardot et al. [Car03] developed a test using the covariance of the scalar response and the functional predictor. Swihart et al. [Swi14] and McLean et al. [McL15] used the exact likelihood ratio tests of zero variance components ([CR04]). Kong et al. [Kon16] proposed classical Wald, score and F-tests; see also [Su17]. However, these tests all focus on fixed functional effects and hence are not applicable to simultaneously testing a collection of random functional effects. Instead of directly testing if multiple random functional effects are all zero, we propose an equivalent test, which tests if the covariance function of the random functional effects is zero or not. The test can be further formulated as testing whether multiple variance components are zero. Because existing tests for multiple variance components are either computationally intensive or conservative ([Qu13; Dri12; Bae19]), we propose an alternative test which is based on the exact likelihood ratio test of one zero variance component ([CR04]) and can be more powerful for finite sample data.

The remainder of this paper is organized as follows. In Section 2.2, we describe our proposed model along with model estimation and also our test. In Section 2.3, we extend the proposed model to deal with a multivariate functional predictor. In Section 2.4, we assess the numerical performance of our model and test. In Section 2.5, we consider the motivating data application. We conclude the paper with some discussion in Section 2.6.

2.2 Method

2.2.1 Functional mixed model for scalar on function regression with repeated outcomes

We begin by introducing notation. For subject i ($i = 1, 2, \ldots, n = 20$), let Y_{ij} denote the pain rating at the jth repetition with $j = 1, 2, \ldots, n_i$ and n_i denotes the number of repetitions for subject i. Similarly, let Z_{ij} denote the level of stimuli for the jth repetition for subject i, with $Z_{ij} = 1$ representing high and $Z_{ij} = 0$ representing low. We shall first consider the fMRI time series data at one brain region, which corresponds to a univariate functional predictor. In Section 2.3, we shall extend our model to fMRI data at multiple regions, which corresponds to a multivariate functional predictor. Let W_{ijk} denote the observed fMRI data at time $t_k = 2k$

seconds ($k = 1, 2, \ldots, K = 23$), which is assumed to be a noisy observation of the smooth functional data $X_{ij}(t_k)$. Let $\mathcal{T} = [0, 46]$ denote the time course of the experiment.

To model the subject-specific random effect of a functional predictor, we propose a new functional mixed model extending the scalar on function linear regression for repeated outcomes. The proposed model is

$$Y_{ij} = \alpha + \alpha_i + Z_{ij}(\gamma + \gamma_i) + \int_{t \in \mathcal{T}} \{\beta(t) + Z_{ij}\delta(t) + \beta_i(t)\} \{X_{ij}(t) - \mu(t)\} \, \mathrm{d}t + \epsilon_{ij}, \qquad (2.1)$$

where α is the population intercept, α_i is the subject-specific random intercept, γ is the population effect of the covariate Z_{ij}, γ_i is the subject-specific random effect of Z_{ij}, $\mu(\cdot)$ is the mean function of the functional predictor $X_{ij}(t)$, $\beta(\cdot)$ is the population effect of the functional predictor, $\delta(\cdot)$ is the interaction effect of the functional predictor and the scalar covariate, $\beta_i(\cdot)$ is the subject-specific random effect of the functional predictor, and ϵ_{ij} are independently and identically distributed (i.i.d.) random errors with distribution $\mathcal{N}(0, \sigma_\epsilon^2)$. We assume that α_i are i.i.d. with distribution $\mathcal{N}(0, \sigma_\alpha^2)$, γ_i are i.i.d. with distribution $\mathcal{N}(0, \sigma_\gamma^2)$, $\beta_i(\cdot)$ are i.i.d. random functions following a Gaussian process over $\mathcal{T}$ with mean function $\mathbb{E}\{\beta_i(t)\} = 0$ and covariance function $\mathrm{cov}\{\beta_i(s), \beta_i(t)\} = \mathcal{C}(s, t)$, and all random terms are mutually independent across subjects and from each other.

The proposed model is a functional analog to equivalent non-functional multi-subject models commonly used for fMRI data; see, e.g., Lindquist et al. [Lin12b]. The term $\delta(\cdot)$ in the model represents the stimuli-specific difference in response, which is typically the parameter of interest in many situations, and the term $\beta_i(\cdot)$ corresponds to the subject-specific deviation from the population mean, the main interest of this work.

2.2.2 Model for the repeated functional predictor

The repeated functional predictor $X_{ij}(t)$ might be correlated across repetitions, indexed by j. Following Park & Staicu [PS15] and Chen et al. [Che17], we consider a marginal functional principal component model where the functional predictor is projected onto a sequence of orthonormal marginal eigenfunctions and the associated scores are used to model the correlation

between the repeated functions. Specifically, the model takes the form

$$W_{ijk} = X_{ij}(t_k) + e_{ijk}, \quad X_{ij}(t) = \mu(t) + \sum_{\ell \geq 1} \xi_{ij\ell}\phi_l(t), \tag{2.2}$$

where $e_{ijk} \sim \mathcal{N}(0, \sigma_e^2)$ are measurement errors that are independent across i, j and k and are independent from the true random functions X_{ij}, $\xi_{ij\ell}$ are random scores that are independent across i and ℓ and $\phi_\ell(\cdot)$ are orthonormal marginal eigenfunctions, i.e., $\int_{\mathcal{T}} \phi_{\ell_1}(t)\phi_{\ell_2}(t)\mathrm{d}t = 1_{\{\ell_1=\ell_2\}}$. Here $1_{\{.\}}$ is 1 if the statement inside the bracket is true and 0 otherwise. The reason that the functions $\phi_\ell(\cdot)$ are called marginal eigenfunctions and how they can be obtained will be explained soon. The dependence between repeated functions is then modeled via the scores. We use the exchangeable model $\xi_{ij\ell} = \eta_{i\ell} + \zeta_{ij\ell}$, where $\eta_{i\ell} \sim \mathcal{N}(0, \sigma_{0\ell}^2)$ are independent across i and ℓ, and $\zeta_{ij\ell} \sim \mathcal{N}(0, \sigma_{1\ell}^2)$ are independent across i, j and ℓ. The exchangeable model is reasonable for our fMRI data application; however, when the functional predictor is measured repeatedly along a longitudinal time or with a longitudinal covariate T_{ij}, other model specifications such as unspecified or nonparametric covariances as functions of T_{ij} for the scores might be adopted and the proposed methods in the paper are still applicable. The proposed model is similar to the multi-level fPCA in Di et al. [Di09] and if $\sigma_{0\ell}^2 = 0$, then the functional data are independent across repetitions. It follows that marginally X_{ij} are random functions from a Gaussian process with mean function $\mathbb{E}\{X_{ij}(t)\} = \mu(t)$ and covariance function

$$\mathrm{cov}\{X_{ij}(s), X_{ij}(t)\} = \mathcal{K}(s,t) = \sum_{\ell \geq 1} \lambda_\ell \phi_\ell(s)\phi_\ell(t), \tag{2.3}$$

where $\lambda_\ell = \sigma_{0\ell}^2 + \sigma_{1\ell}^2$. Equation (2.3) shows that $\phi_\ell(\cdot)$ are indeed marginal eigenfunctions and can be obtained via the eigendecomposition of the marginal covariance function $\mathcal{K}(\cdot,\cdot)$.

2.2.3 Model estimation

The key is to reformulate model (2.1) into a linear mixed effects model using the marginal functional principal component analysis (fPCA) of the functional predictor X_{ij} descried in Section 2.2.2. For model identifiability, we assume that the coefficient functions $\beta(\cdot)$ and $\delta(\cdot)$ can be represented as linear combinations of the eigenfunctions ϕ_ℓ so that $\beta(t) = \sum_{\ell=1}^{\infty} \theta_\ell \phi_\ell(t)$ and

$\delta(t) = \sum_{\ell=1}^{\infty} \delta_\ell \phi_\ell(t)$, where θ_ℓ and δ_ℓ are associated scalar coefficients to be determined. Similarly, let $\beta_i(\cdot) = \sum_{\ell=1}^{\infty} \theta_{i\ell} \phi_\ell(t)$, where $\theta_{i\ell}$ are independent subject-specific random coefficients with distribution $\mathcal{N}(0, \tau_\ell^2)$. Here the variance components $\tau_\ell^2 \geq 0$ are to be determined as well. Then the induced covariance function $\mathcal{C}(s,t)$ of the random functional effects equals $\sum_{\ell \geq 1} \tau_\ell^2 \phi_\ell(s)\phi_\ell(t)$. It follows that model (2.1) can be rewritten as

$$Y_{ij} = \alpha + \alpha_i + Z_{ij}(\gamma + \gamma_i) + \sum_{\ell=1}^{\infty} \xi_{ij\ell}(\theta_l + Z_{ij}\delta_\ell + \theta_{i\ell}) + \epsilon_{ij}. \tag{2.4}$$

Model (2.4) has infinitely many parameters and hence cannot be fit, a well known problem for scalar on function regression. Following the standard approach, we truncate the number of eigenfunctions for approximating the functional predictor, so that the associated scores and parameters for β and β_i are all finite dimensional. Specifically, let L be the number of eigenfunctions to be selected. Then an approximate and identifiable model is given by

$$Y_{ij} = \alpha + \alpha_i + Z_{ij}(\gamma + \gamma_i) + \sum_{\ell=1}^{L} \xi_{ij\ell}(\theta_\ell + Z_{ij}\delta_\ell + \theta_{i\ell}) + \epsilon_{ij}. \tag{2.5}$$

Conditional on the scores $\xi_{ij\ell}$, model (2.5) is a linear mixed effects model and can be easily fit using standard mixed effects model software.

Equation (2.3) suggests that standard fPCA on X_{ij} ignoring the dependence between repeatedly observed functions can be used to estimate the eigenfunctions ϕ_ℓ. Such an approach was proposed in Park & Staicu [PS15] and Chen et al. [Che17]. The fPCA on X_{ij} can be conducted using a number of methods, e.g., local polynomial methods [Yao05]. We use the fast covariance estimation (FACE) method [Xia16], which is based on penalized splines [EM96] and has been implemented in the R function "fpca.face" in the R package *refund* [Gol16]. Then, we obtain the estimate of the mean function $\hat{\mu}$, estimates of the eigenfunctions, $\hat{\phi}_\ell$, estimates of the eigenvalues, $\hat{\lambda}_\ell$, and the estimate of the error variance $\hat{\sigma}_e^2$. We predict the random scores $\xi_{ij\ell}$ using only the observations $\{W_{ij1}, \ldots, W_{ijK}\}$ and denote the prediction by $\hat{\xi}_{ij\ell}$. While the random scores can also be predicted using all observations from the ith subject, we have found in the simulations that such an approach may give unstable prediction and hence do not use it.

We select the number of eigenfunctions L by percentage of variance explained (PVE);

alternatively one may use AIC on the functional predictor [Li13]. We use a PVE value of 0.95. Denote the selected number by $\hat{L}$. Then a practical model for (2.5) is

$$Y_{ij} = \alpha + \alpha_i + Z_{ij}(\gamma + \gamma_i) + \sum_{\ell=1}^{\hat{L}} \hat{\xi}_{ij\ell}(\theta_\ell + Z_{ij}\delta_\ell + \theta_{i\ell}) + \epsilon_{ij}. \tag{2.6}$$

Denote the corresponding estimates of θ_ℓ and δ_ℓ by $\hat{\theta}_\ell$ and $\hat{\delta}_\ell$, respectively, and the prediction of $\theta_{i\ell}$ by $\hat{\theta}_{i\ell}$. Then, $\hat{\beta}(t) = \sum_{\ell=1}^{\hat{L}} \hat{\beta}_\ell \hat{\phi}_\ell(t)$, $\hat{\delta}(t) = \sum_{\ell=1}^{\hat{L}} \hat{\delta}_\ell \hat{\phi}_\ell(t)$ and $\hat{\beta}_i(t) = \sum_{\ell=1}^{\hat{L}} \hat{\theta}_{i\ell} \hat{\phi}_\ell(t)$. Confidence bands for $\hat{\beta}(\cdot)$ and $\hat{\delta}(\cdot)$ can also be constructed and the details are omitted.

2.2.4 Test of random functional effect

Of interest is to assess if the functional effect is subject-specific or the same across subjects. In other words, if $\beta_i(t) = 0$ for all i and $t \in \mathcal{T}$ in model (2.1) or $\beta_i(t) \neq 0$ for some i at some $t \in \mathcal{T}$. Because β_i are random coefficient functions, the test can be formulated in terms of its covariance function $\mathcal{C}(s,t)$. The null hypothesis is $H_0 : \mathcal{C}(s,t) = 0$ for all $(s,t) \in \mathcal{T}^2$ and the alternative hypothesis is $H_a : \mathcal{C}(s,t) \neq 0$ for some $(s,t) \in \mathcal{T}^2$. Under H_0, $\beta_i(t) = 0$ for all i and $t \in \mathcal{T}$ and model (2.1) reduces to a standard scalar on function linear regression model. Under the truncated model with L functional principal components, $\mathcal{C}(s,t) = \sum_{\ell=1}^{L} \tau_\ell^2 \phi_\ell(s)\phi_\ell(t)$, an equivalent test is $H_0' : \tau_\ell^2 = 0$ for all ℓ against $H_a' : \tau_\ell^2 > 0$ for at least one $\ell \leq L$. Thus, the test of random functional effect reduces to the test of zeroness of multiple variance components.

Several methods have been proposed for simultaneously testing multiple variance components, e.g., a permutation test [Dri12], a score test [Qu13], and recently, an asymptotic likelihood ratio test [Bae19]. The permutation test in [Dri12] is computationally intensive and the asymptotic LRT [Bae19] tends to be conservative in our simulation study. A simple approach is to conduct test of zeroness of each variance component and then use a Bonferroni correction; this test will be referred to as the Bonferroni-corrected test hereafter. Alternatively, following [McL15] which tested the linearity of a bivariate smooth function, we use the working assumption

$$\tau_\ell^2 = \tau^2 \text{ for all } \ell, \tag{2.7}$$

and consider the corresponding test $\tilde{H}_0 : \tau^2 = 0$ against $\tilde{H}_a : \tau^2 \neq 0$. Under H_0, $\tilde{H}_0$ still

holds. This test involves testing a single variance component and will be referred to as the equal-variance test. While H_a is more general than $\tilde{H}_a$, it was noted in [McL15] that the equal-variance test could actually outperform the Bonferroni-corrected test even when the true variance components are not the same, i.e., (2.7) does not hold. We shall conduct extensive simulations to compare the performance of the asymptotic LRT, the Bonferroni-corrected test, and the proposed equal-variance test.

The latter two tests involve testing of zeroness of one variance component and we shall use the exact likelihood ratio test (LRT) in [CR04], which is implemented in the R package RLRsim [Sch08].The advantage of the exact tests is that it is more powerful than asymptotic tests for finite sample data.

A practical issue with the equal-variance test is that standard testing procedures such as the LRT is not directly applicable to model (2.5) because the model has multiple additive random slopes. Therefore, we transform (2.5) into an equivalent mixed effect model under the assumption of (2.7), which has only one random slope term and can therefore easily be tested.

Under assumption (2.7), the random effects and random errors are independent from each other and satisfy the following distributional assumptions:

$$\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2),\ \gamma_i \sim \mathcal{N}(0, \sigma_\gamma^2),\ \theta_{i\ell} \sim \mathcal{N}(0, \tau^2),\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2). \tag{2.8}$$

The goal of the equivalent model formulation is to convert a set of homoscedastic subject-specific random slopes in (2.5) into a simple random slope, so that the test on homoscedastic random slopes can be conducted using standard software.

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ_i})^\mathsf{T} \in \mathbb{R}^{J_i}$, $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iJ_i})^\mathsf{T} \in \mathbb{R}^{J_i}$, $\mathbf{A}_i = (\xi_{ij\ell})_{j\ell} \in \mathbb{R}^{J_i \times L}$, $\mathbf{B}_i = (Z_{ij}\xi_{ij\ell})_{j\ell} \in \mathbb{R}^{J_i \times L}$, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iJ_i})^\mathsf{T} \in \mathbb{R}^{J_i}$. Also let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_L)^\mathsf{T} \in \mathbb{R}^L$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_L)^\mathsf{T} \in \mathbb{R}^L$, and $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iL})^\mathsf{T} \in \mathbb{R}^L$. Then model (2.5) can be written in matrix form as follows:

$$\mathbf{Y}_i = (\alpha + \alpha_i)\mathbf{1}_{J_i} + \mathbf{Z}_i(\gamma + \gamma_i) + \mathbf{A}_i(\boldsymbol{\theta} + \boldsymbol{\theta}_i) + \mathbf{B}_i\boldsymbol{\delta} + \boldsymbol{\epsilon}_i.$$

Let $\boldsymbol{\Delta}_i = \left(\mathbf{1}_{J_i}, \quad \mathbf{Z}_i, \quad \mathbf{A}_i, \quad \mathbf{B}_i\right) \in \mathbb{R}^{J_i \times (2+2L)}$ and $\boldsymbol{\eta} = (\alpha, \gamma, \boldsymbol{\theta}, \boldsymbol{\delta}^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{2+2L}$. It follows that

$$\mathbf{Y}_i = \boldsymbol{\Delta}_i \boldsymbol{\eta} + \alpha_i \mathbf{1}_{J_i} + \gamma_i \mathbf{Z}_i + \mathbf{A}_i \boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i. \tag{2.9}$$

Let $\tilde{J}_i = \max(J_i, L)$. Let $\bar{\mathbf{A}}_i$ be $\mathbf{A}_i$ if $J_i \leq L$ and otherwise $\bar{\mathbf{A}}_i = [\mathbf{A}_i, \mathbf{0}_{J_i \times (J_i - L)}]$. Then $\bar{\mathbf{A}}_i \in \mathbb{R}^{J_i \times \tilde{J}_i}$. Similarly, let $\bar{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i$ if $J_i \leq L$ and otherwise $\bar{\boldsymbol{\theta}}_i = (\boldsymbol{\theta}_i^\mathsf{T}, \boldsymbol{\nu}_i^\mathsf{T})^\mathsf{T}$, where $\boldsymbol{\nu}_i \in \mathbb{R}^{J_i - L}$ is multivariate normal with zero mean and covariance $\tau^2 \mathbf{I}_{J_i - L}$ and independent from all other random terms. The vector $\boldsymbol{\nu}_i$ is used only to simplify the algebraic derivation. Then $\mathbf{A}_i \boldsymbol{\theta}_i = \bar{\mathbf{A}}_i \bar{\boldsymbol{\theta}}_i$ and $\bar{\boldsymbol{\theta}}_i$ are independent and identically distributed multivariate normal with zero mean and covariance $\tau^2 \mathbf{I}_{\tilde{J}_i}$ under the working assumption (2.7). Let $\mathbf{U}_i \mathbf{D}_i^{\frac{1}{2}} \mathbf{V}_i^\mathsf{T}$ be the singular value decomposition of $\bar{\mathbf{A}}_i$, where $\mathbf{U}_i \in \mathbb{R}^{J_i \times J_i}$ and $\mathbf{V}_i \in \mathbb{R}^{\tilde{J}_i \times J_i}$ are orthonormal matrices satisfying $\mathbf{U}_i^\mathsf{T} \mathbf{U}_i = \mathbf{I}_{J_i}$, $\mathbf{V}_i^\mathsf{T} \mathbf{V}_i = \mathbf{I}_{J_i}$, and $\mathbf{D}_i = \mathrm{diag}(d_{i1}, \ldots, d_{iJ_i})$ is a diagonal matrix of the singular values of $\bar{\mathbf{A}}_i$. Let $\tilde{\mathbf{Y}}_i = (\tilde{Y}_{i1}, \ldots, \tilde{Y}_{iJ_i})^\mathsf{T} = \mathbf{U}_i^\mathsf{T} \mathbf{Y}_i \in \mathbb{R}^{J_i}$, $\tilde{\boldsymbol{\theta}}_i = (\tilde{\theta}_{i1}, \ldots, \tilde{\theta}_{iJ_i})^\mathsf{T} = \mathbf{V}_i^\mathsf{T} \bar{\boldsymbol{\theta}}_i$, and $\tilde{\boldsymbol{\epsilon}}_i = (\tilde{\epsilon}_{i1}, \ldots, \tilde{\epsilon}_{iJ_i})^\mathsf{T} = \mathbf{U}_i^\mathsf{T} \boldsymbol{\epsilon}_i$. Then a left multiplication of (2.9) by $\mathbf{U}_i^\mathsf{T}$ gives

$$\tilde{\mathbf{Y}}_i = (\mathbf{U}_i^\mathsf{T} \boldsymbol{\Delta}_i) \boldsymbol{\eta} + (\mathbf{U}_i^\mathsf{T} \mathbf{1}_{J_i}) \alpha_i + (\mathbf{U}_i^\mathsf{T} \mathbf{Z}_i) \gamma_i + \mathbf{D}_i^{\frac{1}{2}} \tilde{\boldsymbol{\theta}}_i + \tilde{\boldsymbol{\epsilon}}_i,$$

or equivalently,

$$\tilde{Y}_{ij} = (\mathbf{U}_{ij}^\mathsf{T} \boldsymbol{\Delta}_i) \boldsymbol{\eta} + (\mathbf{U}_{ij}^\mathsf{T} \mathbf{1}_{J_i}) \alpha_i + (\mathbf{U}_{ij}^\mathsf{T} \mathbf{Z}_i) \gamma_i + \sqrt{d_{ij}} \tilde{\theta}_{ij} + \tilde{\epsilon}_{ij}, \tag{2.10}$$

where $\mathbf{U}_{ij}$ is the jth column of $\mathbf{U}_i$. The specification (2.8) now becomes $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$, $\gamma_i \sim \mathcal{N}(0, \sigma_\gamma^2)$, $\tilde{\theta}_{ij} \sim \mathcal{N}(0, \tau^2)$, $\tilde{\epsilon}_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and the random terms are independent across i and j, and are independent from each other. Model (2.10) can be fit using a standard mixed model, and then the test of $\tau^2 = 0$ can be conducted by the exact LRT [CR04].

2.3 Extension to Multivariate Functional Predictor

Model (2.1) deals with only fMRI data at one brain region, and it is of interest to consider a model that incorporates fMRI data from multiple regions, i.e., to extend model (2.1) for multivariate functional data. Let $X_{ij}^{(m)}$ denote the mth functional predictor for region $m\,(1 \leq m \leq M)$, where

M is the number of regions to be modeled together. We extend model (2.1) so that

$$Y_{ij} = \alpha + \alpha_i + Z_{ij}(\gamma + \gamma_i) + \sum_{m=1}^{M} \left[\int_{t \in \mathcal{T}} \{\beta_m(t) + Z_{ij}\delta_m(t) + \beta_{im}(t)\} \left\{ X_{ij}^{(m)}(t) - \mu_m(t) \right\} \mathrm{d}t \right] + \epsilon_{ij}, \tag{2.11}$$

where the terms can be similarly interpreted as before. For the repeated multivariate functional predictor, we extend the decomposition model for repeated univariate functional data [PS15; Che17] so that

$$W_{ijk}^{(m)} = X_{ij}^{(m)}(t_k) + e_{ijk}^{(m)}, \quad X_{ij}^{(m)}(t) = \mu_m(t) + \sum_{\ell \geq 1} \xi_{ij\ell}\phi_{m\ell}(t),$$

where $\{\phi_{1\ell}(t), \ldots, \phi_{M\ell}(t)\}^{\mathsf{T}}$ are multivariate eigenfunctions that satisfy $\sum_{m=1}^{M} \int_{\mathcal{T}} \phi_{m\ell_1}(t)\phi_{m\ell_2}(t)\mathrm{d}t = 1_{\{\ell_1=\ell_2\}}$, $\xi_{ij\ell}$ are random scores that are modeled using an exchangeable model as in Section 2.2.3, and $e_{ijk}^{(m)} \sim \mathcal{N}(0, \sigma_{\epsilon k}^2)$ are measurement errors that are independent across i, j, k and m. It follows that $\left\{ X_{ij}^{(1)}, \ldots, X_{ij}^{(M)} \right\}^{\mathsf{T}}$ is marginally following a multivariate Gaussian process with mean function $\mathbb{E}\{X_{ij}^{(m)}(t)\} = \mu_m(t)$ and covariance function

$$\mathrm{cov}\{X_{ij}^{(m_1)}(s), X_{ij}^{(m_2)}(t)\} = \mathcal{K}_{m_1m_2}(s,t) = \sum_{\ell \geq 1} \lambda_\ell \phi_{m_1\ell}(s)\phi_{m_2\ell}(t). \tag{2.12}$$

By letting $\beta_m(t) = \sum_{\ell=1}^{\infty} \theta_\ell \phi_{m\ell}(t)$, $\delta_m(t) = \sum_{\ell=1}^{\infty} \delta_\ell \phi_{m\ell}(t)$ and $\beta_{im}(t) = \sum_{\ell=1}^{\infty} \theta_{i\ell}\phi_{m\ell}(t)$, model (2.11) reduces to (2.4). Equation (2.12) shows that $\{\phi_{1\ell}(t), \ldots, \phi_{M\ell}(t)\}^{\mathsf{T}}$ are indeed marginal multivariate eigenfunctions.

Because of equation (2.12), to estimate the eigenfunctions $\phi_{m\ell}$, we may conduct multivariate fPCA on $\left\{ X_{ij}^{(1)}, \ldots, X_{ij}^{(M)} \right\}^{\mathsf{T}}$, also ignoring the dependence between repeated multivariate functional data. We have extended the fast covariance estimation method [Xia16] to multivariate functional data and developed the corresponding R function, which gives estimate of the mean functions and the multivariate eigenfunctions. See more details in Secion A.2 of Chapter A. Alternatively, one may use the R package *MFPCA* which conducts multivariate fPCA for functions defined on different domains [HG18]. Similar to before, the scores $\xi_{ij\ell}$ are predicted based on the observations at the jth visit for the ith subject.

2.4 A Simulation Study

In this section we conduct simulations to illustrate the performance of the proposed functional mixed model and compare the three tests described in Section 2.2.4 for testing the existence of random subject-specific functional effects. We shall focus on the models with a univariate functional predictor, but a simulation study with multivariate functional predictor is also conducted and the details are reported in Section A.1.2 of Chapter A.

2.4.1 Simulation settings

We let the domain of functional predictors be $\mathcal{T} = [0, 1]$. Each simulated data set consists of I subjects, with each subject having J replicates. Specific values of I and J will be given later. We generate the response Y_{ij} using model (2.5). The model components are specified as $\alpha = 0.5$, $\gamma = 2$, $\alpha_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $\gamma_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $Z_{ij} \overset{i.i.d.}{\sim}$ Bernoulli(0.5), $\theta_\ell = 2$, $\delta_\ell = 2$, $\theta_{i\ell} \overset{i.i.d.}{\sim} \mathcal{N}(0, \tau_\ell^2)$, and $\epsilon_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$. The values of τ_ℓ^2 will be specified later. We let $L = 3$, i.e., the functional predictor X_{ij} has three functional principal components. The functional predictor X_{ij} is generated by model (2.2) with $X_{ij}(t) = \sum_{\ell=1}^{L} \xi_{ij\ell}\phi_\ell(t)$ and $e_{ijk} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_e^2)$. Both independent and correlated functional predictors are considered: (1) independent $X_{ij}(t)$: $\xi_{ij\ell} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_{1\ell}^2)$; (2) correlated $X_{ij}(t)$: $\xi_{ij\ell} = \eta_{i\ell} + \zeta_{ij\ell}$, where $\xi_{i\ell} \sim \mathcal{N}(0, \sigma_{0\ell}^2)$ are independent across i and ℓ, and $\zeta_{ij\ell} \sim \mathcal{N}(0, \sigma_{1\ell}^2)$ are independent across i, j and ℓ. Here $\sigma_{0\ell}^2 = \sigma_{1\ell}^2 = 0.5^\ell, \ell = 1, \ldots, L$, and the eigenfunctions are $\phi_1(t) = \sqrt{2}\sin(2\pi t)$, $\phi_2(t) = \sqrt{2}\cos(4\pi t)$, $\phi_3(t) = \sqrt{2}\sin(4\pi t)$. The noise variance σ_e^2 is chosen so that the signal to noise ratio in the functional data $r = \sigma_e^{-2} \int_\tau \mathcal{K}(t, t)\mathrm{d}t$ equals either 0 or 3. Here $\mathcal{K}(s, t) = \mathrm{cov}\{X_{ij}(s), X_{ij}(t)\}$ is the marginal covariance function. Note that $r = 0$ corresponds to smooth functional data without noises. Finally, the random scores $\theta_{i\ell}$ are generated by $\theta_{i\ell} \sim \mathcal{N}(0, \tau_\ell^2)$ with $\tau_\ell^2 = 2^{1-\ell}\tau^2, \ell = 1, \ldots, L$. The quantity τ^2 measures the level of variation of random subject-specific functional effect and will be specified later.

Given a fixed τ^2, we simulate data using a factorial design with four factors: the number of subjects I, the number of replicates per subject J, the signal to noise ratio r in the functional data, and the independent or correlated functional predictor $X_{ij}(t)$. A total of 24 different model conditions are used: $\{(I, J, r) : I \in \{20, 50, 200\}, J \in \{20, 50\}, r \in \{0, 3\}\}$ with functional predictor being either independent or correlated. Under each model condition, 20000 data sets are

simulated for significance tests, and 1000 data sets are simulated for evaluating model estimation. For tests of subjects-specific random functional effects in the proposed model, simulated data with $\tau^2 = 0$ is used to evaluate the sizes of the tests, and simulated data with multiple values of τ^2 are used to assess the power of tests. The power of the tests will also be assessed in additional settings for generating the random scores to accommodate some realistic situations, e.g., when the random scores corresponding to one of the eigenfunctions are exactly 0; see Section 4.2 for details. For model estimation, we set τ^2 to be either 0.04 or 0.08.

2.4.2 Results on tests

Table A.1 in Section A.1.1 of Chapter A gives the sizes of the asymptotic LRT (denoted asLRT), the Bonferroni-corrected test and the equal-variance test at the 0.05 significance level. Under various model conditions, the asymptotic LRT gives sizes much smaller than 0.05 and hence can be potentially conservative. The other two tests give sizes very close to the 0.05 level for independent functional predictor and then give slightly inflated sizes for correlated functional predictor. The results confirm the validity of the three tests for testing the proposed hypothesis.

Figure 2.2 shows the powers of the three tests as a function of τ^2 for correlated functional predictor. All three tests have increased power when the number of subjects or the number of visits per subject increases. Moreover, they all have higher power when using smooth, i.e, noise-free, functional predictors compared with using noisy functional predictors, as is expected. Under all model conditions, the equal-variance test has higher power than the other two, especially when the number of subjects is small. This agrees with the finding in [McL15], although their settings are different from ours. The asymptotic LRT seems to have the lowest power among the three, showing that it is indeed conservative for finite sample data. It is also interesting to see that increasing the number of visits per subject seems to result in higher power of the tests than instead increasing the number of subjects. Indeed, with 20 subjects and 50 visits per subject, the power curve of the equal-variance test is close to 1 when τ^2 is around 0.05, whereas with 50 subjects and 20 visits per subject, τ^2 has to be 0.06 or larger to reach the same power. The findings remain the same for independent functional predictor and the corresponding power curves are given in Figure A.1 in Section A.1.1 of Chapter A.

In the above simulation, we have considered $\tau^2 = 2^{1-\tau}\tau^2$ for $\ell = 1, \ldots, L = 3$. Now we

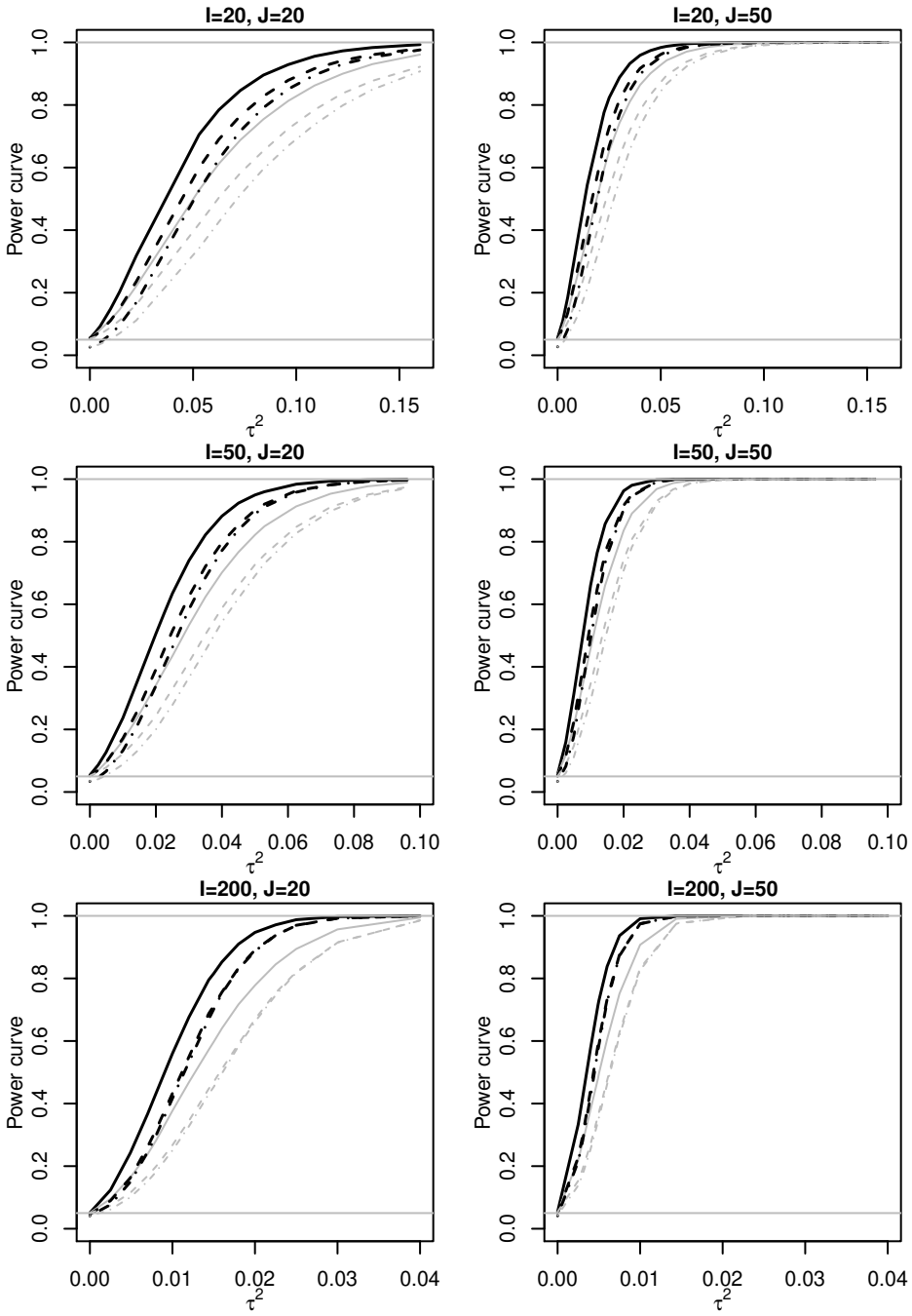

Figure 2.2 Power of three tests at the 5% level for the correlated functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

consider two additional scenarios: scenario 1 with $\tau_1^2 = \tau^2/4$, $\tau_2^2 = \tau^2/2$ and $\tau_3^2 = \tau^2$ and scenario 2 with $\tau_1^2 = \tau^2/2$, $\tau_2^2 = 0$ and $\tau_3^2 = \tau^2$. In scenario 1 the random scores for random functional effects have the smallest variation for the eigenfunction associated with the largest eigenvalue for the functional predictor, while in scenario 2 the random scores corresponding to the second eigenfunction are exactly 0. The power curves for univariate functional predictors are presented in Figure A.7 - Figure A.10 while for multivariate functional predictors they are in Figure A.11-Figure A.14 of the Chapter A. The figures show that the equal-variance test remains the best overall, among the three tests.

To summarize, the simulation study on the tests show that all three tests maintain proper size and have good power. The equal-variance test has the highest power and hence is preferred and will be used in the data application.

2.4.3 Results on estimation

We compare the proposed functional mixed effects model (2.1) (denoted FMM) with the standard scalar on function regression model (denoted FLM), i.e., $\beta_i(t) = 0$ in model (2.1), in terms of both estimation accuracy of the fixed population effects $\beta(t)$ and $\delta(t)$, and out-of-sample prediction accuracy of the response. For the former, we compute mean integrated squared error (MISE) defined as $\int(\beta(t) - \hat{\beta}(t))^2 \mathrm{d}t$ for estimating $\beta(\cdot)$, where $\hat{\beta}(t)$ is the estimate of $\beta(t)$ from either model. The MISE can be similarly computed for $\delta(t)$. For prediction, we use mean squared error (MSE). For each subject in the simulated data, we generate 10 new observations in order to evaluate subject-specific prediction accuracy.

Tables A.2 and A.3 in Section A.1.1 of the Chapter A summarize the results when correlated functional predictor is used. Under each model condition, FMM outperforms FLM with a smaller MSE for predicting the response, and the two methods have comparable performance on estimating the fixed population functional effects with respect to their MISE. Both models have slightly better performance when the functional predictor is smooth without noises. As the sample size increases, both models achieve better performance for fixed effect estimation and response prediction. Increasing τ^2 results in worse prediction result for the response in FLM while slightly deteriorating results for FMM, which indicates the better performance of FMM when there exists strong subject-specific random functional effect of the functional predictor.

Results for the independent functional predictor are shown in Tables A.4 and A.5 in Section A.1.1 of Chapter A, and the findings remain the same.

2.5 Data Application

In this section, we analyze the data from the fMRI study of thermal pain ($n = 20$) described in the Introduction. Recall, fMRI data were extracted from 21 different pain-responsive brain regions. The regions included the anterior insula (AINS), the dorsal anterior cingulate cortex (dACC), thalamus, parahippocampal gyrus (PHG), inferior frontal gyrus (IFG), occipital gyrus, corpus callosum, and the second somatosensory area (SII). These are all brain regions that are often categorized as belonging to the so-called "pain matrix", which is a network of regions thought to generate pain from nociception [Pet02]. The time course extracted from each region consisted of 23 equidistant temporal measurements made every 2s, providing a total of 46s of brain activation, ranging from the time of the application of the heat stimuli to the pain report. We applied the proposed functional mixed model to 6 regions of interest (ROIs), which were found to give statistically significant population effects in [Lin12b]; see Table 2.1 for a list of names of these ROIs.

Table 2.1 Results of FMM and FLM for separate and joint analysis of the fMRI data.

ROI		RMSE	
Separate Analysis	P-value	FLM	FMM
LAIns	0.447	81.31	80.73
RAIns_I	0.035	81.83	79.13
RAIns_II	0.456	81.19	81.19
RThal	0.001	81.13	78.10
RSII_ParOperc	0.404	81.34	81.00
dACC	0.082	81.29	79.18
Joint analysis	0.010	76.35	63.55

We first conduct a joint analysis of all 6 ROIs using the functional mixed model. The residual plot in Figure A.4 in Section A.1.3 of Chapter A indicates that it is reasonable to assume normality of the random errors. The equal-variance test of zeroness of subject-specific functional

effects gives a P-value of 0.010 (Table 2.1), hence favoring the proposed functional mixed model over the standard functional linear mixed model. In addition, the in-sample root mean squared estimation error of the responses for the functional mixed model is 63.55, much smaller than 76.35, the estimation error for the standard functional linear model.

Figure 2.3 plots the estimated subject-specific functional effects $\beta(t) + \delta(t) + \beta_i(t)$ when the hot stimuli is applied. Overall the plots show highly diverse signals at the beginning of the trial, followed by strong positive signal in the middle of the trial, and slightly weaker signal towards the end of trial. The delayed peak occurring in the time period immediately following the conclusion of the thermal stimuli (at time 18s) is consistent with the delayed nature of brain hemodynamics, which peaks roughly 6 seconds after peak neuronal activation, and is consistent with timings of other fMRI experiments [Lin08]. Notably, the secondary peak takes place around the time of the pain reporting (38$\sim$44s), perhaps signaling a contribution of activity during "pain recall".

While the joint analysis and test indicate the existence of subject-specific random functional effects when multiple ROIs are considered together, they cannot assess the existence of subject-specific random effects for each individual ROI. Thus, we next carry out a separate analysis of the data using each ROI as a univariate functional predictor; the results are summarized in Table 2.1. The residual plots in Figure A.5 in Section A.1.3 of Chapter A also indicate it is reasonable to assume normality of random errors. Table 2.1 gives the root MSE (RMSE) of the estimation using FLM and FMM with each ROI. For ROIs right anterior insula (RAIns_I) and right thalamus (RThal), FMM has smaller RMSE than FMM. Among the 6 ROIs, the models with ROIs RAIns_I and RThal give significant subject-specific random functional effect at the 0.05 significance level. For these two ROIs, Figure 2.4 displays the estimated subject-specific functional effects $\beta(t) + \delta(t) + \beta_i(t)$ when the hot stimuli is applied.

Finally, we conduct a 2-cluster analysis of the random functional effects to understand how these effects differ. In each panel of Figure 2.4, the two clusters are denoted by either black solid curves or gray dashed curves. For both ROIs, it appears that subjects mostly differ in the timing of the delayed peak of brain hemodynamics, with one group having peaks around 22s and the other group having a much later peak, e.g., about 24s for ROI RAIns_I. In addition, for ROI RThal, one group of subjects (gray curves) seems to have much pronounced delayed peak as

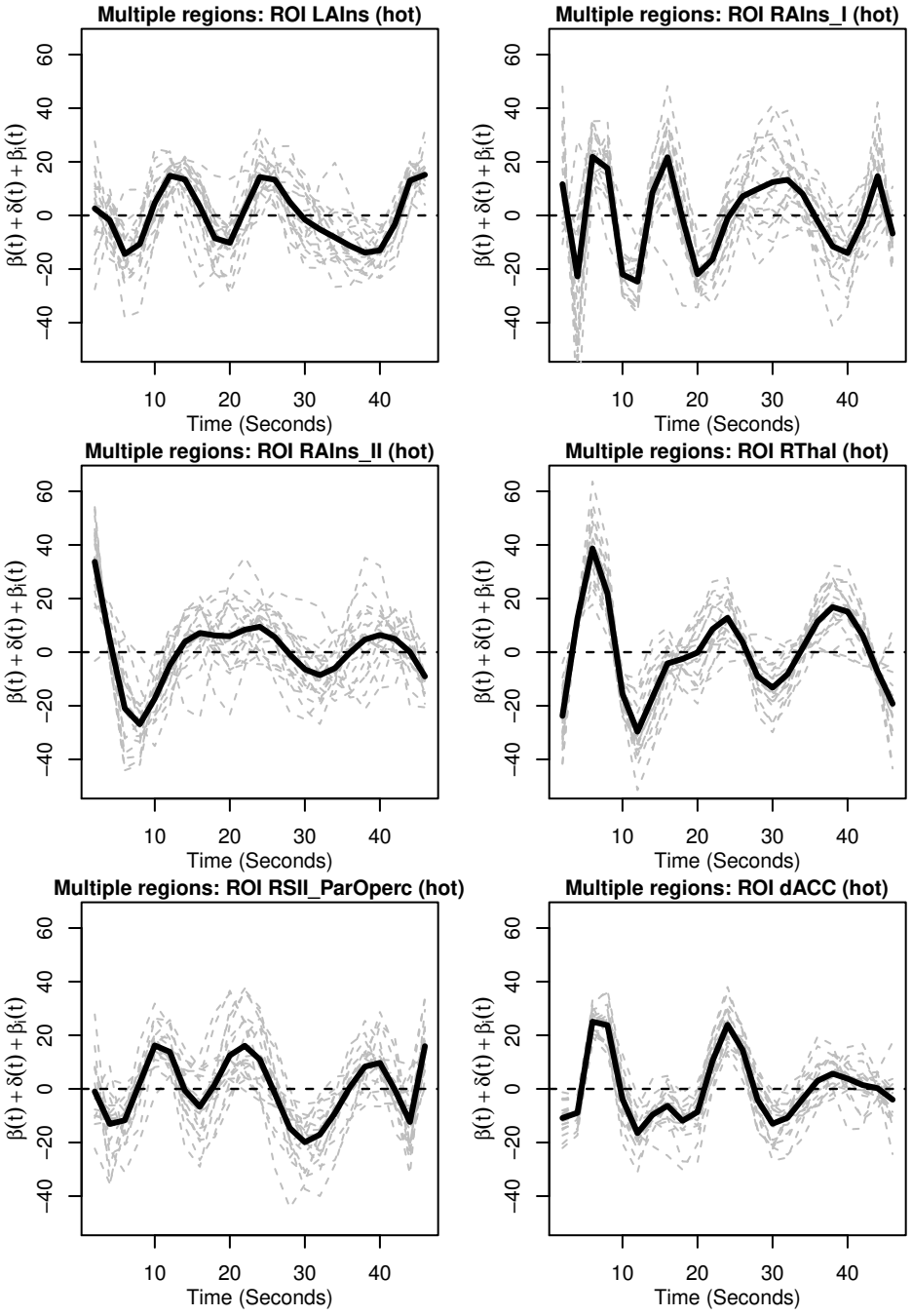

Figure 2.3 Estimated functional effect with a joint analysis of 6 ROIs using FMM. The black solid line is the population functional effect $\beta(t)+\delta(t)$ when the hot stimuli is applied; the gray dashed curves are the subject-specific random functional effect $\beta(t) + \delta(t) + \beta_i(t)$.

well as strong signal during the application of the stimuli; these diverse subject-specific random curves indicate a better fit using the proposed model compared to the fixed population model.

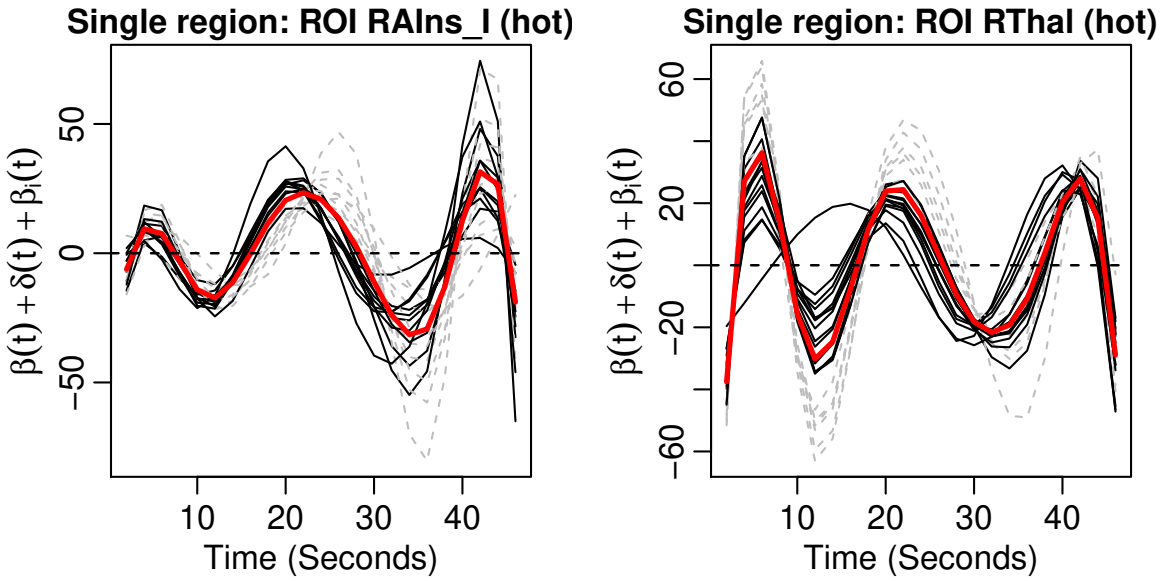

Figure 2.4 Estimated functional effect with a separate analysis of ROI RAIns_I and ROI RThal using FMM. The thick solid line is the population functional effect $\beta(t) + \delta(t)$ when the hot stimuli is applied; the black solid curves are $\beta(t) + \delta(t) + \beta_i(t)$ for one cluster of the subjects and the gray dashed curves are another cluster.

2.6 Discussion

We proposed a functional mixed model to accommodate random functional effects of a univariate or multivariate functional predictor for scalar on function regression, along with a significance test of the random functional effects. Motivated by a fMRI study, we considered subject-specific random effects to assess if the association of the fMRI data with pain rating are subject-specific. We focused on functional data that are observed on a common grid, but the proposed model may be extended to handle sparse functional data. Indeed, the model estimation in Section 2.3 of the marginal decomposition model for repeated functional data can be adapted, e.g., using the FACE method for sparse functional data [Xia18] or for sparse multivariate functional data [Li18]. However, the random score prediction method adopted in the paper might not be optimal. Because of the sparsity of data, the predicted random scores will necessarily be shrunk to zero.

Thus, it remains to be seen how the proposed model and test will perform for sparse functional data.

In the data application, we treated the fMRI data collected at multiple brain regions as multivariate functional data. One may also treat the data as two-way functional data or matrix-variate data as in [Hua17]. An interesting future research direction is to extend the proposed functional mixed model for repeated matrix-variate data.

CHAPTER

3

FLEXIBLE FEATURE SELECTION AND CLUSTER ANALYSIS FOR HETEROGENEOUS DATA WITH APPLICATION TO DIFFUSION TENSOR IMAGING STUDY

3.1 Introduction

In recent decades, the area of structural neuroimaging has seen an explosive growth of the cross-disciplinary research work, and keeps enhancing people's understanding of the brain inner processing related to disease diagnosis([Gol12; Ger13; PS15]). Among different structural brain imaging techniques, Diffusion Weighted Imaging (DWI) is one of the most popular techniques

used in the neuroimaging studies of the sports-related concussions (SRC) ([Mus18; Cub18]).

In this paper, we are motivated by a DTI study of SRC where the tract-based DTI measurements are recorded for 73 football atheletes with the aim to distinguish two groups with or without SRC. Numerous clustering algorithms have been developed for functional data. Jacques and Preda [JP14a] categorize the clustering methods for functional data into 4 groups: (1) directly working with the discrete realizations from the latent continuous curves; (2) two-step clustering with first conducting dimension reduction and then clustering; (3) clustering while simultaneous dimension reduction; (4) distance-based clustering.

A major challenge presented in clustering the functional data is the high-dimensional realization along the curves. Therefore, dimension reduction techniques are often utilized first to approximate the curve by a finite number of basis expansion coefficients or functional principal component scores. For example, [Abr03] and [SW05] consider the basis expansion coupled with the K-means clustering algorithm([HW79]). [PM08] study the K-means clustering performance for the sparsely and irregularly sampled curves by means of multidimensional scaling on the functional principal component scores. Clustering for sparse functional data is also studied by [JS03] using a model-based clustering method on the functional principal component scores. Other work focusing on the model-based clustering method with functional principal components include [CL07; BJ11; JP13; JP14b].

In our work, the goal is to use four different DTI measurements (Fractional Anisotropy (FA), Mean Diffusivity (MD), Radial Diffusivity (Dr) and Axial Diffusivity (Da)) from the DTI scanning results along 27 brain tracts from 73 players to classify the concussed football players and concussion-free football players. Common approach to the DTI analyses relies on the Tract-Based Spatial Statistics method ([Smi06]). Application of this method relies on the registration of all brain images to a common space and voxel-based analyses. Due to the injury heterogeneity, however, this approach is not optimal as the voxel-based averaging will hide possible group differences existing within a tract, but not perfectly aligned in the voxels. Therefore, we use the data in the subject-specific space, where the number of voxels is subject- and tract-specific. To incorporate all observations in a principled way, we propose to accumulate the voxel-data using density function.

Due to the natural constraint of density space, densities do not live in a vector space

and thus, commonly used Hilbert space based methods of functional data analysis are not directly applicable. Therefore for the density function approach, we consider the log quantile density transformation following [PM16] to map the density function into a linear space using a continuous and invertible function, and then the functional data analysis techniques such as functional PCA can be properly implemented. Following that, we propose a novel flexible feature selection approach based on K-means clustering algorithm for the resulting high-dimensional coefficients matrix. Density of raw data and transformed curves using the metric FA along a tract cst_r are shown in Figure 3.1. This proposed method can achieve two main purposes: first, the homogeneous subgroups are identified based on the K-means algorithm; second, a small subset of features that differentiate the underlying clusters are selected, which facilitates the visualization of the heterogeneity of the brain signals between the two groups. Feature selection within the clustering framework by [WT10] tends to give a poor out-of-sample prediction results in our case.

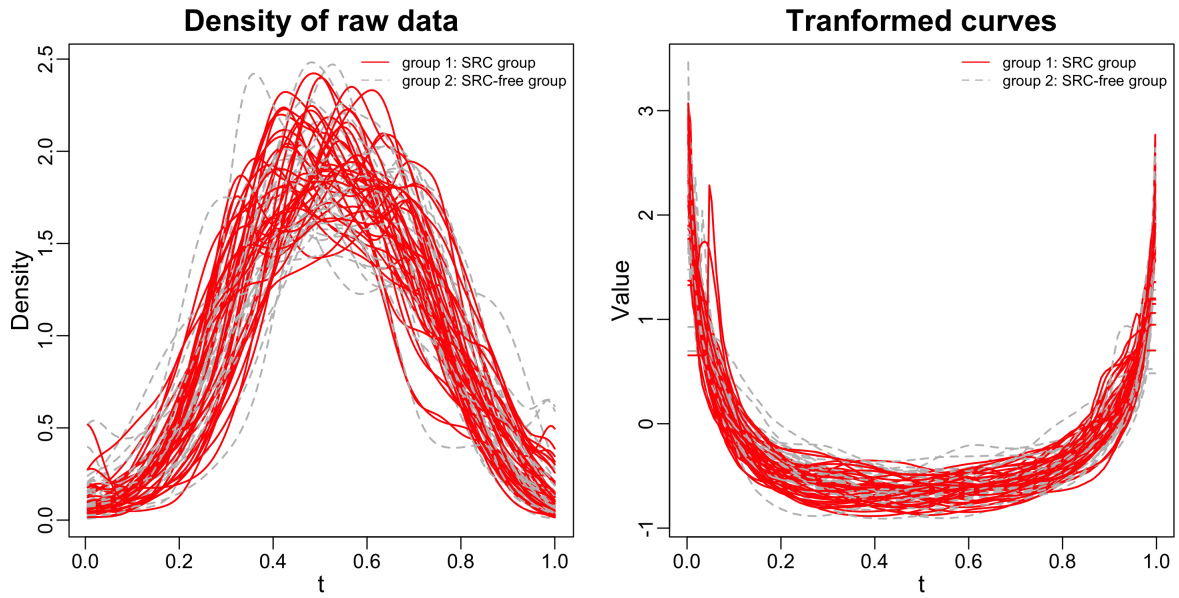

Figure 3.1 Comparison of the raw data density and the log-quantile transformed curves at tract cst_r using measurement FA. The left panel shows the density of the raw data; right panel shows the log-quantile transformed curves. Note that for both the two density calculation, we scaled the data and de-outlier as discussed in Section 3.2.

3.1.1 Data description

We use the data collected in a sub-study of the largest and most comprehensive to-date study of the college athletes' acute concussions: NCAA-DoD Concussion Assessment, Research and Education (CARE) Consortium [Mus18]. We concentrate on the structural brain imaging data obtained from the football players at three US colleges: University of North Carolina (UNC), University of California Los Angeles (UCLA), and Virginia Tech (VT). Data used consists of four DTI measures collected along 27 tracts from 39 concussed and 34 non-concussed football players. Non-concussed group members were matched to the concussed football players via a combination of college (exact), playing position (same group), number of prior concussions and estimated premorbid level of verbal intellectual functioning (i.e., Wechsler Test of Adult Reading ([Wec55]). All participants provided informed consent approved by the guidelines of the institutional review board (IRB) at the three colleges.

Concussed athletes have undergone clinical assessments and multimodal MRI imaging between 24 and 48 hours after the injury. Matched controls were enrolled and scanned within a short time period after their matching with the concussed athletes. In our study, we use the DTI measurements processed by Dr. Wu's lab at Indiana University School of Medicine. Specifically, based on Johns Hopkins white matter anatomic atlas, we use four types of measurements: FA, MD, Dr and Da collected on 27 white matter tracts. The number of voxels in each participant and each tract is different. Thus, in our analyses the data are accumulated within the tracts using density functions.

The reminder of this paper is organized as follows. In Section 3.2, we describe the data preprocessing including density transformation and functional principal component analysis. In Section 3.3, we detail the proposed clustering method with flexible feature selection. In Section 3.4, we illustrate the performance of the proposed method using the motivating DTI data set for the SRC disease. Conclusions and discussions are provided in Section 3.5.

3.2 Data Preprocessing and Feature Extraction

3.2.1 Density covariates

For subject i, denote by $Y_i \in \{0, 1\}$ the group label with 0 indicating the non-concussed group and 1 indicating the concussed group. For the DTI imaging data, denote by D_{ijk} the kth scalar metric along the jth tract for the ith subject. There are $J = 27$ tracts to consider and for each tract there are 4 DTI scalar measurements: FA, MD, Da and Dr. Thus, for each subject, there are in total 108 sample curves. Within each subject, for each tract, the 4 DTI metrics are densely collected and of the same length. However, due to tract variation and subject variation, the observation length of the metrics differ across tracts and subjects. Thus, to deal with the data heterogeneity, instead of working with D_{ijk}, we consider the density of the samples, thus ignoring the order of original observations. In order to confirm the advantage of using density function over the raw data, we also consider an alternative approach of using the raw data. Specifically, we first smooth each curve, and then sample the same number of points to form the tract profiles of the same length. Here we use 200 sampling points for each smoothed curve. The out-of-sample classification result using greedy search is shown in Figure B.2.

3.2.2 Functional principal component analysis for densities

We now consider feature extraction from densities. As densities are a collection of smooth curves, functional principal component analysis (FPCA) can be used for feature extraction via dimension reduction. Because densities do not live in a Hilbert space, standard FPCA is not directly applicable. Following Peterson and Muller [PM16], we first transform densities to a Hilbert space and then apply standard FPCA to the transformed curves. We adopt the log quantile transform and calculate $-\log(f(q(t)))$, where $q(t)$ denotes the quantile function of the normalized curve at $t \times 100$th percentile and $f(\cdot)$ is the density function. We let t be 200 equally spaced time points in $[0.0025, 0.9975]$. In particular, to deal with data heterogeneity and outliers across metrics and tracts, we use the scaled and modified quantile function. Specifically, let $q^*(t)$ be the original quantile function and m and M be the 0.5 and 99.5th percentiles of all

observations across subjects (for each metric and tract), then

$$q(t) = \begin{cases} 0, & \text{for } t \in [0, 0.005], \\ \frac{q^*(t)-m}{M-m}, & \text{for } t \in [0.005, 0.995], \\ 1, & \text{for } t \in [0.995, 1]. \end{cases}$$

Let X_{ijk} be the transformed curves corresponding to D_{ijk}. The functional data model for X_{ijk} is $X_{ijk}(t) = \mu_{jk}(t) + \sum_{\ell=1}^{\infty} \xi_{ijk\ell}\phi_{jk\ell}(t)$, where μ_{jk}s are fixed mean functions, $\phi_{ij\ell}$s are orthonormal eigenfunctions, i.e., $\int_{t=0}^{1} \phi_{ij\ell_1}(t)\phi_{ij\ell_2}(t)dt = I(\ell_1 = \ell_2)$, and $\xi_{ij\ell}$s are random scores that are uncorrelated across i and ℓ.

Practically, truncation is done so that approximately $X_{ijk}(t) = \mu_{jk}(t) + \sum_{\ell=1}^{L_{jk}} \xi_{ijk\ell}\phi_{jk\ell}(t)$. Here L_{jk} is the truncation number for metric k along tract j and can be determined by percentage of variance explained (PVE). We use 0.85 for PVE in the analysis. Then the corresponding predicted scores $\xi_{ijk\ell}$ are the collection of features for subsequent analysis.

One drawback of conducting FPCA on each function is that the extracted features might be correlated across tracts/metrics. An alternative method is multivariate functional principal component analysis (MFPCA), which gives uncorrelated features. However, there is computational and interpretation challenge with MFPCA. First, the dimension is 108 and to conduct MFPCA, we need to smooth 108^2 bivariate functions including self-covariance functions and cross-covariance functions, which is computationally prohibitive. Second and more importantly, the extracted features (scores) from MFPCA involve every functions and as a result, we will be unable to locate the tract/metrics that are useful for differentiating the groups. Therefore, we prefer univariate FPCA over MFPCA for feature extraction.

We have got 73 subjects (39 in the concussed group and 34 in the non-concussed group) with over 400 features.

3.3 Cluster Analysis and Feature Selection

To deal with data heterogeneity, while there are truly only 2 groups of subjects, the cluster analysis should allow for more than 2 clusters, with potentially multiple clusters corresponding

to one group.

Second, because there is a large number of features, we need to conduct a subset search to identify relevant features and discard irrelevant features for differentiating the 2 groups of subjects. Ideally, a small subset of features with high differentiating power is identified. Existing methods for variable selection, e.g., logistic regression with Lasso, does not give satisfactory results. Results of using logistic regression with Lasso is discussed in Section 3.4. One may also use a greedy search, i.e., the features are sequentially selected and added. While the greedy search is straightforward and computationally efficient, it often may not identify a good subset of features and/or may need an unnecessarily large number of features to achieve certain level of classification accuracy.

To alleviate the issue with greedy search and also to allow for flexibility in the number of clusters for grouping the data, we propose a probabilistic subset search (PSS) method. First, we introduce some notation. Let $\mathcal{S} = \{i\}$ be the group of all features. For each feature, we evaluate its marginal relevance for differentiating the 2 groups of subjects by the percentage of correct classification via cross-validated K-means clustering analysis. So for feature $i \in \mathcal{S}$, denote by r_i the corresponding percentage of correct classification. The r_is will be used for the feature sampling method and it is reasonable to sample features more often with a larger r_i; the details are given later.

Now, we introduce the PSS method, which consists of several steps. Suppose we would like to select p features. Let q be a fixed positive integer. The PSS method requires an initial input of p features, denoted by $\mathcal{S}_0$. The initial features may be obtained from, e.g., the greedy search. In Step 1, we randomly select q additional features from the remaining features, i.e., $\mathcal{S}/\mathcal{S}_0$. The probability that a feature $i \in \mathcal{S}/\mathcal{S}_0$ is selected is given by

$$\frac{r_i}{\sum_{i \in \mathcal{S}/\mathcal{S}_0} r_i}.$$

Denote the additional features by $\mathcal{S}_1$. In Step 2, we evaluate the set of combined $(p+q)$ features, $\mathcal{S}_0 \cup \mathcal{S}_1$. Specifically, we conduct an exhaustive search of all possible combinations of p features from $\mathcal{S}_0 \cup \mathcal{S}_1$ to identify the best p features. In this step, we use the K-means method for clustering the training data and evaluate the performance of clusters using the testing data.

We use 5-fold cross validation and to avoid dependence of K-means on the starting values of centers we try different values of centers for initialization. Moreover, we allow the number of clusters to be either of 2, 3, 4 or 5 and identify the best number of clusters to be the one that leads to highest (cross-validated) classification accuracy. In Step 3, we repeat Steps 1 and 2 until reaching convergence. The criteria for convergence is that the algorithm ends if after a large number (denoted by L) of sets of additional features are sampled, the best p features remain always the same. In our analysis, we let $L = 100$, which seems to work well for the DTI data. Algorithms of the PSS method and exhaustive search are summarized in Algorithm 1 and Algorithm 2, respectively.

Algorithm 1 Algorithm for PSS

Input: $\mathbf{X} \in \mathbb{R}^{N\times m}$ (covariates), $\mathbf{Y} \in \mathbb{R}^{N\times 1}$(true label), p, q, n_{cluster}, $\mathbf{S}_{\text{start}} \in \mathbb{R}^{p\times 1}$(initial input features index), $\mathbf{W} \in \mathbb{R}^{m\times 1}$(weight)
Output: $\mathbf{S}_{\text{cur}} \in \mathbb{R}^{p\times 1}$(selected features index), $\epsilon_{\text{cur}} \in \mathbb{R}$(prediction error)
Initialisation :
$L \leftarrow 100$, $\mathbf{S}_{\text{cur}} \leftarrow \mathbf{S}_{\text{start}}$, ind $\leftarrow 1$, count $\leftarrow 0$
Loop until converge:
while count $\leq L$ **do**
  Randomly select q features $\mathbf{S}_q \in \mathbb{R}^{q\times 1}$ based on $\mathbf{W}$
  Find $\mathbf{S}$ with ϵ among $\mathbf{S}_{\text{cur}} \cup \mathbf{S}_q$ by algorithm 2
  if (ind $= 1$ or $\epsilon \leq \epsilon_{\text{cur}}$) **then**
    $\epsilon_{\text{cur}} \leftarrow \epsilon$, $\mathbf{S}_{\text{cur}} \leftarrow \mathbf{S}$, count $\leftarrow 0$
  else
    count$\leftarrow$count$+1$
  end if
  ind $\leftarrow 1$
end while
return $\mathbf{S}_{\text{cur}}, \epsilon_{\text{cur}}$

Compared to the greedy search, the PSS method is slower but could potentially identify a better combination of features in the sense of achieving high classification power with a fewer number of selected features. Indeed, using greedy search, the selected $(p+1)$ features will always contain the p features selected previously but they may not necessarily be the best combination of $(p+1)$ features. The features selected via the PSS method do not necessarily overlap for different values of p and hence have the potential (albeit not guaranteed) of identifying a better

Algorithm 2 Algorithm for exhaustive search

Input: $\mathbf{X} \in \mathbb{R}^{N\times(p+q)}$ (covariates), $\mathbf{Y} \in \mathbb{R}^{N\times 1}$(true label), n_{cluster}, p
Output: $\mathbf{S} \in \mathbb{R}^{p\times 1}$(selected features index), $\epsilon \in \mathbb{R}$(prediction error)
Initialisation:
$\mathbf{C} \in \mathbb{R}^{p\times\binom{p+q}{p}}$, $\boldsymbol{\epsilon} \in \mathbb{R}^{\binom{p+q}{p}\times 1} \leftarrow \mathbf{0}^{\binom{p+q}{p}\times 1}$
Loop:
for i in 1 to $\binom{p+q}{p}$ **do**
  Initialisation: $\boldsymbol{\kappa} \in \mathbb{R}^{\#\{n_{\text{cluster}}\}\times 1} \leftarrow \mathbf{0}^{\#\{n_{\text{cluster}}\}\times 1}$
  for j in n_{cluster} **do**
    $\boldsymbol{\kappa}[j] \leftarrow$ 100 times repeated 5-fold cross-validated prediction error by K-means$\big(\mathbf{X}\big[,\mathbf{C}[,i]\big]$, $K = j\big)$
  end for
  $\boldsymbol{\epsilon}[i] \leftarrow min(\boldsymbol{\kappa})$
end for
$\epsilon \leftarrow min(\boldsymbol{\epsilon})$, $\mathbf{S} \leftarrow \mathbf{C}\big[,\text{argmin}_{i\in\{1,\ldots,\binom{p+q}{p}\}}\,\boldsymbol{\epsilon}[i]\big]$
return $\mathbf{S}, \epsilon$

subset of features. Moreover, the PSS method is better equipped for allowing a flexible number of clusters for grouping the data, again because the PSS method does not impose a restriction as the greedy search does on how the features are searched. Indeed, it may happen that the best 2 features corresponds to 2 clusters while the best 3 features corresponds to 3 clusters. In such case, the greedy search is not as adaptive due to its sequential adding features. But the PSS method allows the features identified in earlier stages to be replaced by a better set of features.

3.4 Results

In this section, we utilize the proposed PSS method to classify the two groups in our motivating DTI data set. Recall that in this DTI study, a number of 73 college football athletes are enrolled to further explore the relationship between the inner working of human brain with the SRC appearance in a group of football players, of which 39 football players are diagnosed with concussion and the rest 34 football players are matched controls. We also compare the performance of the proposed PSS method with the greedy search, the logistic regression with Lasso penalty as well as sparse K-means clustering method. In particular, we use the selected features given by the greedy search as the starting values at each step for the PSS method. As the greedy search assumes the fixed number of clusters, we have different starting values for

PSS method due to different pre-specified number of clusters in greedy search, of which we shall investigate the impact on the classification results. Here, we use the prediction error from 100 bootstrapped 5-fold cross-validation to evaluate the classification accuracy for each method.

Figure 3.2 compares the classification results of greedy search and PSS at the first few feature selection steps. Under each case of the pre-specified number of clusters, namely, different starting features for PSS, one can see that PSS outperforms the greedy search in terms of smaller out-of-sample classification error at each selection step. This confirms the advantage of the proposed PSS method which allows for both flexible selection for the potential multiple clusters and the freedom of replacing the previous selected features. Specifically, PSS method with the initial features from the 2 cluster greedy search provides the out-of-sample misclassification rate as low as 0.22 with only 3 features. While it is shown in Figure B.1 that greedy search can hardly achieve as accurate classification as PSS does unless a larger number of features are selected or a large number of clusters is assumed, which in either case will pose additional challenges on interpreting the results. On the other hand, it is shown that different starting values can end up with different selected features and consequently, different classification performance, implying the PSS method may converge to the local minimum under different sets of initial features.

The out-of-sample classification results for Sparse K-means clustering method and logistic regression with Lasso penalty are illustrated in Figure 3.3 and Figure 3.4, respectively. In Figure 3.3, the performance of Sparse K-means is evaluated at different number of clusters ranging in $\{2, 3, 4, 5\}$. Overall Sparse K-means clustering gives a high out-of-sample classification error (≥ 0.5), which can not compete with the proposed PSS method. The smallest prediction error that Sparse K-means can achieve is 0.5 with 3 pre-specified clusters. Despite its unsatisfactory performance, interestingly, it does show that the prediction accuracy will eventually decrease as the number of clusters increases too much, suggesting that the over-complicated structure overfits the training data and is not generalized to the testing data well. Figure 3.4 displays the 500 times cross-validated classification errors and model complexity for logistic regression with Lasso penalty, showing that the prediction error is 0.53, slightly higher than that of sparse K-means method. Model complexity in 3.4 indicates that most of the time Lasso ends up with zero selected feature and thus classifies all the test observations into one group, which explains why the prediction performance is not good.

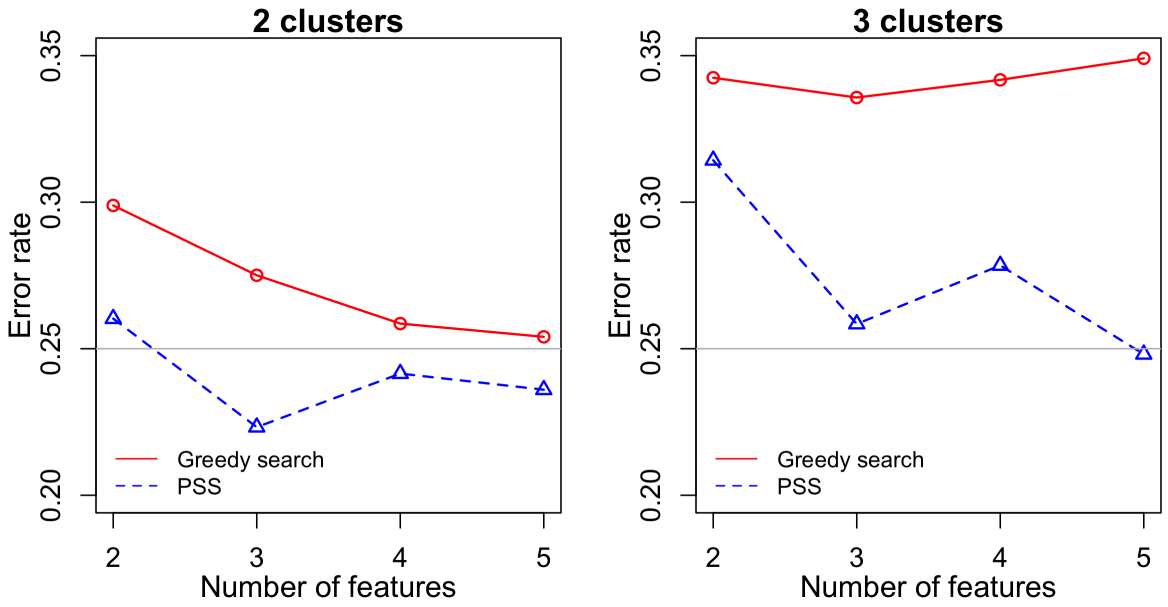

Figure 3.2 PSS method using different starting values vs. greedy search: Out-of-sample classification error by 100 times bootstrapped 5-fold cross-validation at each step of selected features. Solid red lines are for greedy search. Blue dashed lines are for PSS method. Note that in each subfigure, greedy search use the corresponding pre-specified fixed number of clusters shown in the title to select feature at each step; while PSS uses the selected features given by greedy search as starting value at each step.

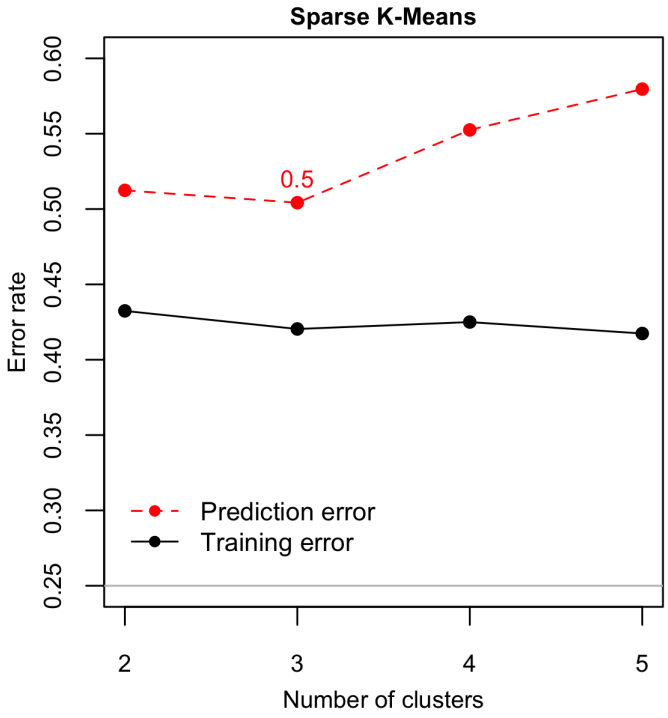

Figure 3.3 Sparse K-means using different number of clusters: Out-of-sample classification error by 100 times bootstrapped 5-fold cross-validation. Solid black line is for training error. Red dashed line is for prediction error.

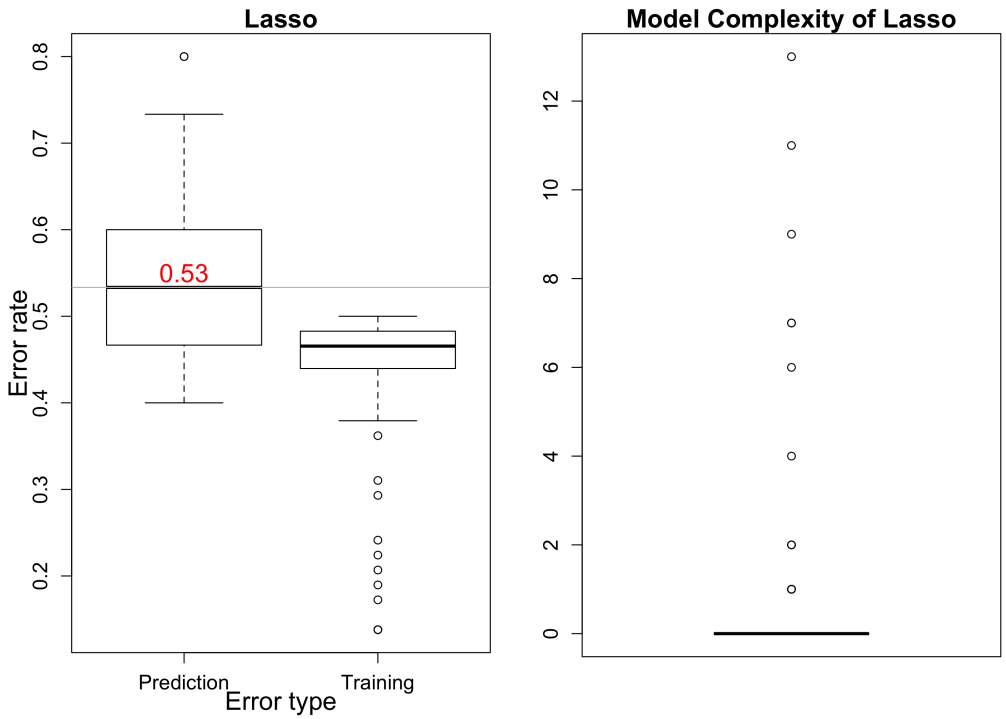

Figure 3.4 Lasso: Out-of-sample classification error by 100 times bootstrapped 5-fold cross-validation. Left panel displays different error types. Right panel displays the model complexity selected by Lasso.

To illustrate the classification results of the PSS method, Figure 3.5 and Figure B.3 compare the true group and the in-sample estimated clustering membership by the PSS method with 3 clusters and initial features from the 2-cluster greedy search at the second and third steps (i.e. $p = 2$ and $p = 3$), respectively. Table 3.1 further summarizes error types and the distribution of the true group among the estimated clustering membership. From table 3.1, the in-sample error of PSS method is 0.22 and 0.23 for $p = 2$ and $p = 3$, respectively. PSS method with $p = 3$ reduces the estimation error compared with that of $p = 2$ by correctly estimating one more subject from the SRC-free group. In addition, PSS method with $p = 2$ and $p = 3$ give the same level of type 2 error, i.e. the error rate of classifying subjects from SRC group as from SRC-free group, which is as low as 0.08. This indicates high estimation accuracy for the SCR group. In both Figure 3.5 and Table 3.1, it can been seen that the two selected features (first principal component of the measurement FA along tract cst_r and first principal component of the measurement Dr along tract slf_r) can well cluster the football players into 3 distinct groups, and the SRC group is identified as having a 2-subgroup cluster structure. The reasonable interpretation can be that

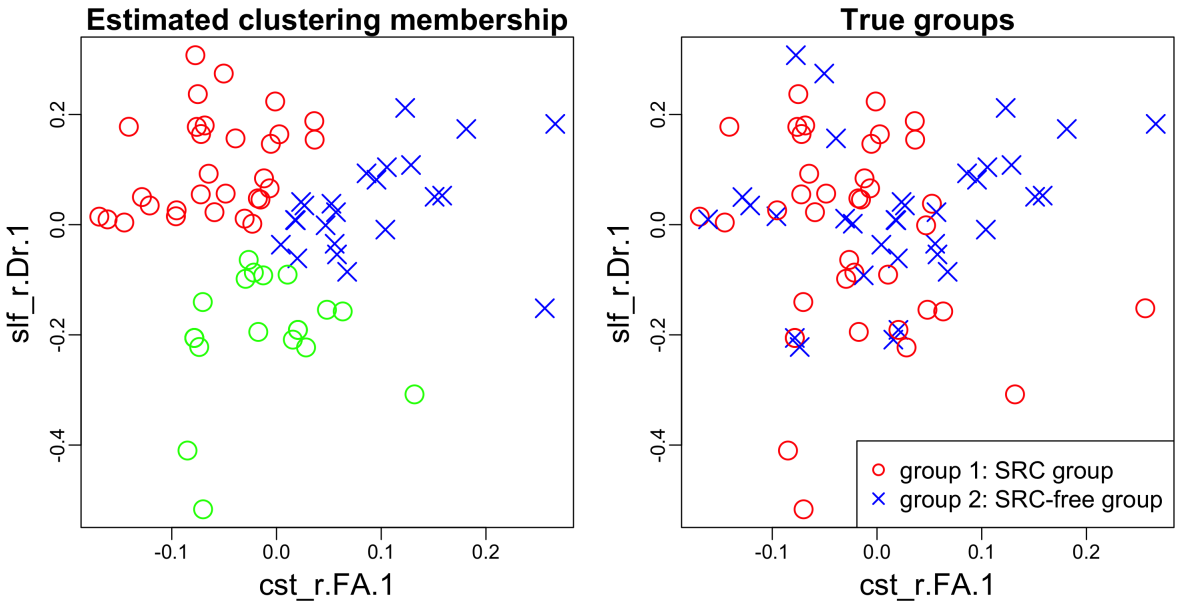

Figure 3.5 $p = 2$: Plots of the in-sample estimated clustering membership vs. the true groups. Left panel shows in-sample clustering result using the first two selected features of the PSS method with starting value from 2-cluster greedy search. Right panel shows the true group of the 73 football players.

the SRC group may have strong and diverse brain injury resulting in heterogeneous clusters whereas the SRC-free group is relatively unified in the brain signals and hence identified as one single cluster. Similar result for $p = 3$ can be found in Figure B.3 and Table 3.1.

Table 3.1 PSS clustering results using the first p selected features and the distribution of the true group in the estimated clustering membership.

	cluster	label	group 1 dist.($N = 39$)	group 2 dist.($N = 34$)	type 1 error	type 2 error	error
$p = 2$	1	1	15	5	0.41 (14/34)	0.08 (3/39)	0.23 (17/73)
	2	1	21	9			
	3	0	3	20			
$p = 3$	1	0	3	21	0.38 (13/34)	0.08 (3/39)	0.22 (16/73)
	2	1	15	5			
	3	1	21	8			

3.5 Discussion

Motivated by a DTI study, we propose a combination of functional data methods and feature selection to distinguish two groups of college football athletes (with or without SRC). The contribution of this work is twofold. First, we propose to consider density function to deal with the difficulty of aligning tract profiles across all the athletes. Second, a flexible probabilistic subset search algorithm is proposed to select features which allows a flexible number of clusters and enables flexible updates during feature selection process. The results of PSS with $p = 2$ and $p = 3$ using the initial features from the 2-cluster greedy search both give insights on the heterogeneous nature of SRC. The proposed PSS method can be used as a feature selection tool for any classification purpose. When selecting the number of clusters, PSS method relies on the out-of-sample classification error, which may be replaced by using criterion such as AIC and BIC.

BIBLIOGRAPHY

[Abr03] Abraham, C. et al. "Unsupervised curve clustering using B-splines". *Scandinavian journal of statistics* **30**.3 (2003), pp. 581–595.

[Bae19] Baey, C. et al. "Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models". *Computational Statistics & Data Analysis* **135** (2019), pp. 107 –122.

[Bas94] Basser, P. J. et al. "MR diffusion tensor spectroscopy and imaging". *Biophysical journal* **66**.1 (1994), pp. 259–267.

[Ber11] Berrendero, J. R. et al. "Principal components for multivariate functional data". *Computational Statistics & Data Analysis* **55**.9 (2011), pp. 2619–2634.

[BR86] Besse, P. & Ramsay, J. O. "Principal components analysis of sampled functions". *Psychometrika* **51**.2 (1986), pp. 285–311.

[BBS14] Bouveyron, C. & Brunet-Saumard, C. "Model-based clustering of high-dimensional data: A review". *Computational Statistics & Data Analysis* **71** (2014), pp. 52–78.

[BJ11] Bouveyron, C. & Jacques, J. "Model-based clustering of time series in group-specific functional subspaces". *Advances in Data Analysis and Classification* **5**.4 (2011), pp. 281–300.

[BK09] Bunge, S. A. & Kahn, I. "Cognition: An overview of neuroimaging techniques" (2009).

[CM97] Capra, W. B. & Müller, H.-G. "An accelerated-time model for response curves". *Journal of the American Statistical Association* **92**.437 (1997), pp. 72–83.

[Car03] Cardot, H. et al. "Testing hypotheses in the functional linear model". *Scandinavian Journal of Statistics* **30**.1 (2003), pp. 241–255.

[Cas86] Castro, P. E. et al. "Principal modes of variation for processes with continuous sample curves". *Technometrics* **28**.4 (1986), pp. 329–337.

[CM12] Chen, K. & Müller, H.-G. "Modeling repeated functional observations". *Journal of the American Statistical Association* **107**.500 (2012), pp. 1599–1609.

[Che17] Chen, K. et al. "Modelling function-valued stochastic processes, with applications to fertility dynamics". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**.1 (2017), pp. 177–196.

[CL07] Chiou, J.-M. & Li, P.-L. "Functional clustering and identifying substructures of longitudinal data". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**.4 (2007), pp. 679–699.

[Chi14] Chiou, J.-M. et al. "Multivariate functional principal component analysis: A normalization approach". *Statistica Sinica* (2014), pp. 1571–1596.

[CR04] Crainiceanu, C. M. & Ruppert, D. "Likelihood ratio tests in linear mixed models with one variance component". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**.1 (2004), pp. 165–185.

[Cra09] Crainiceanu, C. M. et al. "Generalized multilevel functional regression". *Journal of the American Statistical Association* **104**.488 (2009), pp. 1550–1561.

[Cub18] Cubon, V. A. et al. "Preliminary evidence from a prospective DTI study suggests a posterior-to-anterior pattern of recovery in college athletes with sports-related concussion". *Brain and behavior* **8**.12 (2018), e01165.

[Di09] Di, C.-Z. et al. "Multilevel functional principal component analysis". *The annals of applied statistics* **3**.1 (2009), p. 458.

[Dri12] Drikvandi, R. et al. "Testing multiple variance components in linear mixed-effects models". *Biostatistics* **14**.1 (2012), pp. 144–159.

[EM03] Eilers, P. H. & Marx, B. D. "Multivariate calibration with temperature interaction using two-dimensional penalized signal regression". *Chemometrics and intelligent laboratory systems* **66**.2 (2003), pp. 159–174.

[EM96] Eilers, P. & Marx, B. "Flexible smoothing with B-splines and penalties (with Discussion)". *Statist. Sci.* **11** (1996), pp. 89–121.

[ER35] Einstein, A. & Rosen, N. "The particle problem in the general theory of relativity". *Physical Review* **48**.1 (1935), pp. 73–77.

[FV06] Ferraty, F. & Vieu, P. *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media, 2006.

[Ger13] Gertheiss, J. et al. "Longitudinal scalar-on-functions regression with application to tractography data". *Biostatistics* **14**.3 (2013), pp. 447–461.

[Gol11] Goldsmith, J. et al. "Penalized Functional Regression". *Journal of Computational and Graphical Statistics* **20**.4 (2011). PMID: 22368438, pp. 830–851.

[Gol12] Goldsmith, J. et al. "Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements". *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **61**.3 (2012), pp. 453–469.

[Gol15] Goldsmith, J. et al. "Generalized multilevel function-on-scalar regression and principal component analysis". *Biometrics* **71**.2 (2015), pp. 344–353.

[Gol16] Goldsmith, J. et al. "Refund: Regression with functional data". R package. 2016.

[Gre11] Greven, S. et al. "Longitudinal functional principal component analysis". *Recent Advances in Functional Data Analysis and Related Topics*. Springer, 2011, pp. 149–154.

[Hap19] Happ, C. "MFPCA: Multivariate Functional Principal Component Analysis for Data Observed on Different Dimensional Domains". R package. 2019.

[HG18] Happ, C. & Greven, S. "Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains". *Journal of the American Statistical Association* **113**.522 (2018), pp. 649–659.

[HW79] Hartigan, J. A. & Wong, M. A. "Algorithm AS 136: A k-means clustering algorithm". *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **28**.1 (1979), pp. 100–108.

[Has17] Hasenstab, K. et al. "A multi-dimensional functional principal components analysis of EEG data". *Biometrics* **73**.3 (2017), pp. 999–1009.

[HN13] Hosseini-Nasab, M. "Cross-validation approximation in functional linear regression". *Journal of Statistical Computation and Simulation* **83**.8 (2013), pp. 1429–1439.

[Hua14] Huang, H. et al. "Joint modeling and clustering paired generalized longitudinal trajectories with application to cocaine abuse treatment data". *Journal of the American Statistical Association* **109**.508 (2014), pp. 1412–1424.

[Hua17] Huang, L. et al. "Two-way principal component analysis for matrix-variate data, with an application to functional magnetic resonance imaging data". *Biostatistics* **18**.2 (2017), pp. 214–229.

[JP13] Jacques, J. & Preda, C. "Funclust: A curves clustering method using functional random variables density approximation". *Neurocomputing* **112** (2013), pp. 164–171.

[JP14a] Jacques, J. & Preda, C. "Functional data clustering: a survey". *Advances in Data Analysis and Classification* **8**.3 (2014), pp. 231–255.

[JP14b] Jacques, J. & Preda, C. "Model-based clustering for multivariate functional data". *Computational Statistics & Data Analysis* **71** (2014), pp. 92–106.

[Jam02] James, G. M. "Generalized linear models with functional predictors". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**.3 (2002), pp. 411–432.

[JS05] James, G. M. & Silverman, B. W. "Functional adaptive model estimation". *Journal of the American Statistical Association* **100**.470 (2005), pp. 565–576.

[JS03] James, G. M. & Sugar, C. A. "Clustering for sparsely sampled functional data". *Journal of the American Statistical Association* **98**.462 (2003), pp. 397–408.

[Jam00] James, G. M. et al. "Principal component models for sparse functional data". *Biometrika* **87**.3 (2000), pp. 587–602.

[JW10] Jiang, C.-R., Wang, J.-L., et al. "Covariate adjusted functional principal components analysis for longitudinal data". *The Annals of Statistics* **38**.2 (2010), pp. 1194–1226.

[Jol03] Jolliffe, I. "Principal component analysis". *Technometrics* **45**.3 (2003), p. 276.

[JR92] Jones, M. & Rice, J. A. "Displaying the important features of large collections of similar curves". *The American Statistician* **46**.2 (1992), pp. 140–145.

[KU01] Kneip, A. & Utikal, K. J. "Inference for density families using functional principal component analysis". *Journal of the American Statistical Association* **96**.454 (2001), pp. 519–542.

[Kon16] Kong, D. et al. "Classical testing in functional linear models". *Journal of Nonparametric Statistics* **28**.4 (2016), pp. 813–838.

[Li18] Li, C. et al. "Fast Covariance Estimation for Multivariate Sparse Functional Data". *arXiv preprint arXiv:1812.00538* (2018).

[LC11] Li, P.-L. & Chiou, J.-M. "Identifying cluster number for subspace projected functional data clustering". *Computational Statistics & Data Analysis* **55**.6 (2011), pp. 2090–2103.

[Li13] Li, Y. et al. "Selecting the Number of Principal Components in Functional Data". *Journal of the American Statistical Association* **108**.504 (2013), pp. 1284–1294.

[Lin08] Lindquist, M. A. "The statistical analysis of fMRI data". *Statistical science* **23**.4 (2008), pp. 439–464.

[Lin12a] Lindquist, M. A. "Functional causal mediation analysis with an application to brain connectivity". *Journal of the American Statistical Association* **107**.500 (2012), pp. 1297–1309.

[Lin12b] Lindquist, M. A. et al. "Estimating and testing variance components in a multi-level GLM". *Neuroimage* **59**.1 (2012), pp. 490–501.

[Ma12] Ma, S. et al. "A simultaneous confidence band for sparse longitudinal regression". *Statistica Sinica* **22** (2012), p. 95.

[McL14] McLean, M. W. et al. "Functional generalized additive models". *Journal of Computational and Graphical Statistics* **23**.1 (2014), pp. 249–269.

[McL15] McLean, M. W. et al. "Restricted likelihood ratio tests for linearity in scalar-on-function regression". *Statistics and Computing* **25**.5 (2015), pp. 997–1008.

[Mor15] Morris, J. S. "Functional regression". *Annual Review of Statistics and Its Application* **2** (2015), pp. 321–359.

[MC06] Morris, J. S. & Carroll, R. J. "Wavelet-based functional mixed models". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**.2 (2006), pp. 179–199.

[Mor03] Morris, J. S. et al. "Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis". *Journal of the American Statistical Association* **98**.463 (2003), pp. 573–583.

[MÜL05] MÜLLER, H.-G. "Functional modelling and classification of longitudinal data". *Scandinavian Journal of Statistics* **32**.2 (2005), pp. 223–240.

[MS05] Müller, H.-G., Stadtmüller, U., et al. "Generalized functional linear models". *the Annals of Statistics* **33**.2 (2005), pp. 774–805.

[MY08] Müller, H.-G. & Yao, F. "Functional additive models". *Journal of the American Statistical Association* **103**.484 (2008), pp. 1534–1544.

[Mül13] Müller, H.-G. et al. "Continuously additive models for nonlinear functional regression". *Biometrika* **100**.3 (2013), pp. 607–622.

[Mus18] Mustafi, S. M. et al. "Acute white-matter abnormalities in sports-related concussion: a diffusion tensor imaging study from the NCAA-DoD CARE Consortium". *Journal of neurotrauma* **35**.22 (2018), pp. 2653–2664.

[PS15] Park, S. Y. & Staicu, A.-M. "Longitudinal functional data analysis". *Stat* **4**.1 (2015), pp. 212–226.

[PP09] Paul, D., Peng, J., et al. "Consistency of restricted maximum likelihood estimators of principal components". *The Annals of Statistics* **37**.3 (2009), pp. 1229–1271.

[PM08] Peng, J., Müller, H.-G., et al. "Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions". *The Annals of Applied Statistics* **2**.3 (2008), pp. 1056–1077.

[PM16] Petersen, A., Müller, H.-G., et al. "Functional data analysis for density functions by transformation to a Hilbert space". *The Annals of Statistics* **44**.1 (2016), pp. 183–218.

[Pet02] Petrovic, P. et al. *Science* **295**.5560 (2002), pp. 1737–1740.

[Qu13] Qu, L. et al. "Linear score tests for variance components in linear mixed models and applications to genetic association studies". *Biometrics* **69**.4 (2013), pp. 883–892.

[Ram06] Ramsay, J. O. *Functional data analysis*. Wiley Online Library, 2006.

[Rao58] Rao, C. R. "Some statistical methods for comparison of growth curves". *Biometrics* **14**.1 (1958), pp. 1–17.

[Rei17] Reiss, P. T. et al. "Methods for Scalar-on-Function Regression". *International Statistical Review* **85**.2 (2017), pp. 228–249.

[RS91] Rice, J. A. & Silverman, B. W. "Estimating the mean and covariance structure nonparametrically when the data are curves". *Journal of the Royal Statistical Society: Series B (Methodological)* **53**.1 (1991), pp. 233–243.

[RW01] Rice, J. A. & Wu, C. O. "Nonparametric mixed effects models for unequally sampled noisy curves". *Biometrics* **57**.1 (2001), pp. 253–259.

[Sch08] Scheipl, F. et al. "Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models." *Computational Statistics & Data Analysis* **52**.7 (2008), pp. 3283–3299.

[SJ12] Serban, N. & Jiang, H. "Multilevel functional clustering analysis". *Biometrics* **68**.3 (2012), pp. 805–814.

[SW05] Serban, N. & Wasserman, L. "CATS: clustering after transformation and smoothing". *Journal of the American Statistical Association* **100**.471 (2005), pp. 990–999.

[Shi96] Shi, M. et al. "An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves". *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **45**.2 (1996), pp. 151–163.

[Sil96] Silverman, B. W. et al. "Smoothed functional principal components analysis by choice of norm". *The Annals of Statistics* **24**.1 (1996), pp. 1–24.

[Smi06] Smith, S. M. et al. "Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data". *Neuroimage* **31**.4 (2006), pp. 1487–1505.

[Spi11] Spierer, D. K. et al. "A comparison of energy expenditure estimates from the Actiheart and Actical physical activity monitors during low intensity activities, walking, and jogging". *European journal of applied physiology* **111**.4 (2011), pp. 659–667.

[SL98] Staniswalis, J. G. & Lee, J. J. "Nonparametric regression analysis of longitudinal data". *Journal of the American Statistical Association* **93**.444 (1998), pp. 1403–1418.

[Su17] Su, Y.-R. et al. "Hypothesis testing in functional linear models". *Biometrics* **73**.2 (2017), pp. 551–561.

[Swi14] Swihart, B. J. et al. "Restricted Likelihood Ratio Tests for Functional Effects in the Functional Linear Model". *Technometrics* **56**.4 (2014), pp. 483–493.

[TK03] Tarpey, T. & Kinateder, K. K. "Clustering functional data". *Journal of classification* **20**.1 (2003), pp. 093–114.

[Tia10] Tian, T. S. "Functional data analysis in brain imaging studies". *Frontiers in psychology* **1** (2010), p. 35.

[Tok07] Tokushige, S. et al. "Crisp and fuzzy k-means clustering algorithms for multivariate functional data". *Computational Statistics* **22**.1 (2007), pp. 1–16.

[Wan16] Wang, J.-L. et al. "Functional data analysis". *Annual Review of Statistics and Its Application* **3** (2016), pp. 257–295.

[Wec55] Wechsler, D. "Manual for the Wechsler adult intelligence scale." (1955).

[WT10] Witten, D. M. & Tibshirani, R. "A framework for feature selection in clustering". *Journal of the American Statistical Association* **105**.490 (2010), pp. 713–726.

[Xia16] Xiao, L. et al. “Fast covariance estimation for high-dimensional functional data”. *Statistics and computing* **26**.1-2 (2016), pp. 409–421.

[Xia18] Xiao, L. et al. “Fast covariance estimation for sparse functional data”. *Statistics and computing* **28**.3 (2018), pp. 511–522.

[Yao03] Yao, F. et al. “Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate”. *Biometrics* **59**.3 (2003), pp. 676–685.

[Yao05] Yao, F. et al. “Functional data analysis for sparse longitudinal data”. *Journal of the American Statistical Association* **100**.470 (2005), pp. 577–590.

[Zho08] Zhou, L. et al. “Joint modelling of paired sparse functional data using principal components”. *Biometrika* **95**.3 (2008), pp. 601–619.

[Zho10] Zhou, L. et al. “Reduced rank mixed effects models for spatially correlated hierarchical functional data”. *Journal of the American Statistical Association* **105**.489 (2010), pp. 390–400.

APPENDICES

APPENDIX

A

SUPPLEMENTAL MATERIALS FOR CHAPTER 2

A.1 Additional Results

This section summarizes the additional results for the simulation and real data application in Chapter 2. Section A.1.1 presents additional simulation results for the FMM model with a univariate functional predictor. Section A.1.2 presents simulation results (size and power of tests and prediction) for the FMM model with a multivariate functional predictor. Section A.1.3 presents additional results for the fMRI data application. Section A.1.4 presents simulation results on the power of the tests of the FMM model with either a univariate or multivariate functional predictor, under additional model conditions.

A.1.1 Additional simulation results for FMM with a univariate functional predictor

Table A.1 gives the sizes of the asymptotic LRT (denoted asLRT), the Bonferroni-corrected test and the equal-variance test at the 0.05 significance level.

For the independent functional predictor, Figure A.1 gives the power curves of various tests.

For the correlated functional predictor, Table A.2 and Table A.3 give the estimation/prediction errors of FMM and FLM. For the independent functional predictor, Table A.4 and Table A.5 give the estimation/prediction errors of FMM and FLM.

A.1.2 A simulation study for FMM with a multivariate functional predictor

We generate the response Y_{ij} by model (2.5) using the same simulation settings in the main paper. We consider a 3-dimensional functional predictor and the noisy functional predictor $X_{ij}^{(m)}$ is generated by the equation after (2.11) and

$$\begin{aligned}
\{\phi_{11}(t), \phi_{21}(t), \phi_{31}(t)\}^\top &= \{\sqrt{2/3}\sin(2\pi t), \sqrt{2/3}\cos(\pi t), \sqrt{2/3}\sin(\pi t)\}^\top, \\
\{\phi_{12}(t), \phi_{22}(t), \phi_{32}(t)\}^\top &= \{\sqrt{2/3}\cos(4\pi t), \sqrt{2/3}\cos(2\pi t), \sqrt{2/3}\sin(2\pi t)\}^\top, \\
\{\phi_{13}(t), \phi_{23}(t), \phi_{33}(t)\}^\top &= \{\sqrt{2/3}\sin(4\pi t), \sqrt{2/3}\cos(3\pi t), \sqrt{2/3}\sin(3\pi t)\}^\top.
\end{aligned}$$

For the random errors, $e_{ijk}^{(m)} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_e^2)$. The SNR in the noisy observations is defined as $r = \frac{\sum_{k=1}^{3}\lambda_k}{3\sigma_e^2}$. All other specifications are the same as those for a univariate functional predictor in the main paper.

Given a fixed τ^2, we simulate data with 16 different model conditions: $\{(I, J, r) : I \in \{20, 50\}, J \in \{20, 50\}, r \in \{0, 3\}\}$ with the functional predictor being either independent or correlated.

Table A.6 gives the sizes of the three tests, while Figure A.2 and Figure A.3 give the power curves. Table A.7 and Table A.8 give the estimation/prediction errors of the FMM for correlated multivariate functional predictor, while Table A.9 and Table A.10 give the estimation/prediction errors of the FMM for independent multivariate functional predictor.

Table A.1 Sizes of three tests at the 5% level for correlated and independent univariate functional predictor $X_{ij}(t)$. I: number of subjects; J: number of visits per subject; r: noise level in the functional predictor.

Condition	Equal-variance		Bonferroni-corrected		asLRT	
	$r = 0$	$r = 3$	$r = 0$	$r = 3$	$r = 0$	$r = 3$
Correlated functional predictor $X_{ij}(t)$						
$\mathrm{I} = 20, \mathrm{J} = 20$	0.054	0.054	0.055	0.053	0.026	0.027
$\mathrm{I} = 20, \mathrm{J} = 50$	0.055	0.055	0.053	0.051	0.027	0.027
$\mathrm{I} = 50, \mathrm{J} = 20$	0.054	0.052	0.052	0.052	0.034	0.035
$\mathrm{I} = 50, \mathrm{J} = 50$	0.056	0.057	0.051	0.051	0.034	0.034
$\mathrm{I} = 200, \mathrm{J} = 20$	0.051	0.051	0.046	0.050	0.041	0.041
$\mathrm{I} = 200, \mathrm{J} = 50$	0.054	0.053	0.050	0.054	0.041	0.043
Independent functional predictor $X_{ij}(t)$						
$\mathrm{I} = 20, \mathrm{J} = 20$	0.048	0.047	0.048	0.049	0.023	0.024
$\mathrm{I} = 20, \mathrm{J} = 50$	0.052	0.055	0.049	0.051	0.022	0.026
$\mathrm{I} = 50, \mathrm{J} = 20$	0.052	0.053	0.050	0.049	0.030	0.031
$\mathrm{I} = 50, \mathrm{J} = 50$	0.050	0.051	0.048	0.049	0.031	0.032
$\mathrm{I} = 200, \mathrm{J} = 20$	0.051	0.051	0.052	0.051	0.039	0.040
$\mathrm{I} = 200, \mathrm{J} = 50$	0.052	0.050	0.047	0.050	0.037	0.040

Table A.2 Estimation/prediction errors of FMM and FLM across 1000 data sets with a univariate functional predictor under various model conditions.

Condition	Correlated functional predictor $X_{ij}(t)$, $\tau^2 = 0.04$						
	MISE for $\beta(t)$			MISE for $\delta(t)$		MSE for Y_{ij}	
	r	FLM	FMM	FLM	FMM	FLM	FMM
$\mathrm{I} = 20, \mathrm{J} = 20$	0	0.046	0.046	0.074	0.073	1.289	1.258
	3	0.060	0.061	0.106	0.106	1.788	1.772
$\mathrm{I} = 20, \mathrm{J} = 50$	0	0.024	0.023	0.035	0.034	1.203	1.124
	3	0.032	0.032	0.051	0.050	1.672	1.606
$\mathrm{I} = 50, \mathrm{J} = 20$	0	0.022	0.022	0.034	0.033	1.266	1.223
	3	0.029	0.029	0.049	0.048	1.750	1.720
$\mathrm{I} = 50, \mathrm{J} = 50$	0	0.013	0.013	0.018	0.017	1.201	1.116
	3	0.019	0.019	0.027	0.027	1.664	1.593
$\mathrm{I} = 200, \mathrm{J} = 20$	0	0.010	0.010	0.012	0.012	1.252	1.199
	3	0.015	0.015	0.019	0.018	1.730	1.689
$\mathrm{I} = 200, \mathrm{J} = 50$	0	0.007	0.007	0.008	0.008	1.195	1.108
	3	0.013	0.013	0.014	0.014	1.655	1.580

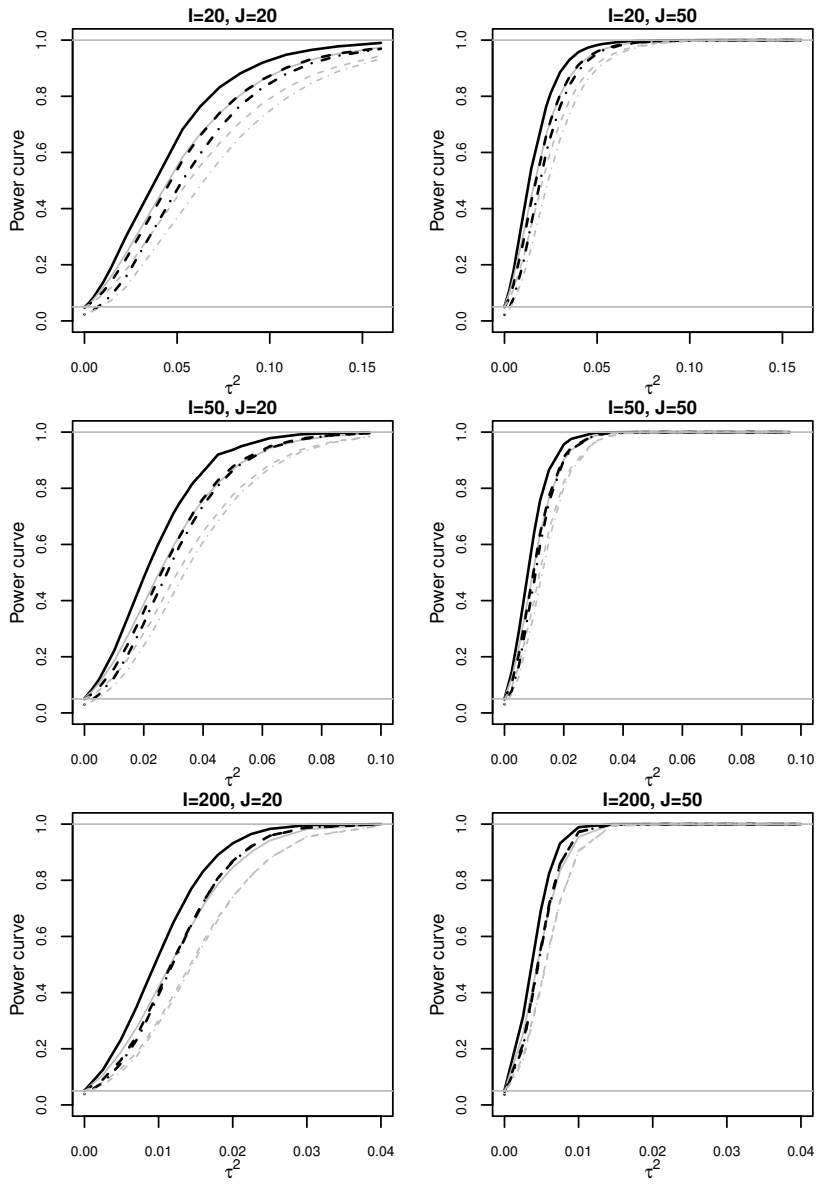

Figure A.1 Power of three tests at the 5% level for the independent univariate functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

Table A.3 Estimation/prediction errors of FMM and FLM across 1000 data sets with a univariate functional predictor under various model conditions.

Condition	Correlated functional predictor $X_{ij}(t)$, $\tau^2 = 0.08$						
	MISE for $\beta(t)$			MISE for $\delta(t)$		MSE for Y_{ij}	
	r	FLM	FMM	FLM	FMM	FLM	FMM
$\mathrm{I} = 20, \mathrm{J} = 20$	0	0.051	0.050	0.078	0.074	1.445	1.314
	3	0.065	0.065	0.109	0.107	1.942	1.842
$\mathrm{I} = 20, \mathrm{J} = 50$	0	0.028	0.027	0.037	0.035	1.356	1.148
	3	0.036	0.035	0.053	0.051	1.825	1.640
$\mathrm{I} = 50, \mathrm{J} = 20$	0	0.024	0.023	0.035	0.034	1.419	1.276
	3	0.031	0.031	0.050	0.049	1.901	1.785
$\mathrm{I} = 50, \mathrm{J} = 50$	0	0.015	0.014	0.018	0.017	1.356	1.140
	3	0.021	0.020	0.028	0.027	1.818	1.626
$\mathrm{I} = 200, \mathrm{J} = 20$	0	0.010	0.010	0.013	0.012	1.409	1.252
	3	0.016	0.016	0.019	0.019	1.885	1.754
$\mathrm{I} = 200, \mathrm{J} = 50$	0	0.008	0.008	0.009	0.008	1.350	1.212
	3	0.013	0.013	0.014	0.014	1.810	1.614

Table A.4 Estimation/prediction errors of FMM and FLM across 1000 data sets with a univariate functional predictor under various model conditions.

Condition	Independent functional predictor $X_{ij}(t)$, $\tau^2 = 0.04$						
	MISE for $\beta(t)$			MISE for $\delta(t)$		MSE for Y_{ij}	
	r	FLM	FMM	FLM	FMM	FLM	FMM
$\mathrm{I} = 20, \mathrm{J} = 20$	0	0.050	0.050	0.087	0.087	1.164	1.153
	3	0.058	0.058	0.110	0.110	1.406	1.399
$\mathrm{I} = 20, \mathrm{J} = 50$	0	0.026	0.025	0.038	0.037	1.090	1.063
	3	0.029	0.029	0.047	0.046	1.317	1.293
$\mathrm{I} = 50, \mathrm{J} = 20$	0	0.023	0.023	0.038	0.038	1.157	1.142
	3	0.027	0.026	0.048	0.047	1.398	1.386
$\mathrm{I} = 50, \mathrm{J} = 50$	0	0.014	0.013	0.019	0.019	1.094	1.067
	3	0.015	0.015	0.022	0.022	1.323	1.298
$\mathrm{I} = 200, \mathrm{J} = 20$	0	0.010	0.010	0.014	0.014	1.150	1.132
	3	0.011	0.011	0.016	0.016	1.387	1.372
$\mathrm{I} = 200, \mathrm{J} = 50$	0	0.008	0.007	0.009	0.008	1.093	1.063
	3	0.008	0.008	0.010	0.009	1.321	1.294

Table A.5 Estimation/prediction errors of FMM and FLM across 1000 data sets with a univariate functional predictor under various model conditions.

Condition	Independent functional predictor $X_{ij}(t)$, $\tau^2 = 0.08$						
	MISE for $\beta(t)$			MISE for $\delta(t)$		MSE for Y_{ij}	
	r	FLM	FMM	FLM	FMM	FLM	FMM
I = 20, J = 20	0	0.056	0.054	0.093	0.089	1.220	1.173
	3	0.064	0.063	0.116	0.113	1.462	1.423
I = 20, J = 50	0	0.031	0.029	0.040	0.037	1.143	1.071
	3	0.034	0.033	0.049	0.047	1.370	1.304
I = 50, J = 20	0	0.026	0.025	0.040	0.039	1.213	1.163
	3	0.029	0.028	0.050	0.048	1.454	1.410
I = 50, J = 50	0	0.015	0.015	0.020	0.019	1.147	1.076
	3	0.016	0.016	0.023	0.022	1.375	1.309
I = 200, J = 20	0	0.011	0.011	0.015	0.014	1.206	1.153
	3	0.012	0.011	0.017	0.016	1.443	1.395
I = 200, J = 50	0	0.008	0.008	0.009	0.009	1.147	1.072
	3	0.008	0.008	0.010	0.009	1.374	1.305

Table A.6 Multivariate functional predictor: Sizes of three tests at the 5% level for correlated and independent functional predictor $X_{ij}(t)$. I: number of subjects; J: number of visits per subject; r: noise level in the functional predictor.

Condition	Equal-variance		Bonferroni-corrected		asLRT	
	$r = 0$	$r = 3$	$r = 0$	$r = 3$	$r = 0$	$r = 3$
Correlated functional predictor $X_{ij}(t)$						
I = 20, J = 20	0.054	0.053	0.054	0.053	0.026	0.026
I = 20, J = 50	0.055	0.054	0.053	0.052	0.025	0.025
I = 50, J = 20	0.054	0.052	0.052	0.051	0.031	0.030
I = 50, J = 50	0.056	0.053	0.051	0.050	0.031	0.031
Independent functional predictor $X_{ij}(t)$						
I = 20, J = 20	0.048	0.047	0.047	0.048	0.024	0.023
I = 20, J = 50	0.050	0.051	0.049	0.049	0.022	0.024
I = 50, J = 20	0.053	0.052	0.049	0.051	0.030	0.031
I = 50, J = 50	0.050	0.049	0.048	0.049	0.031	0.031

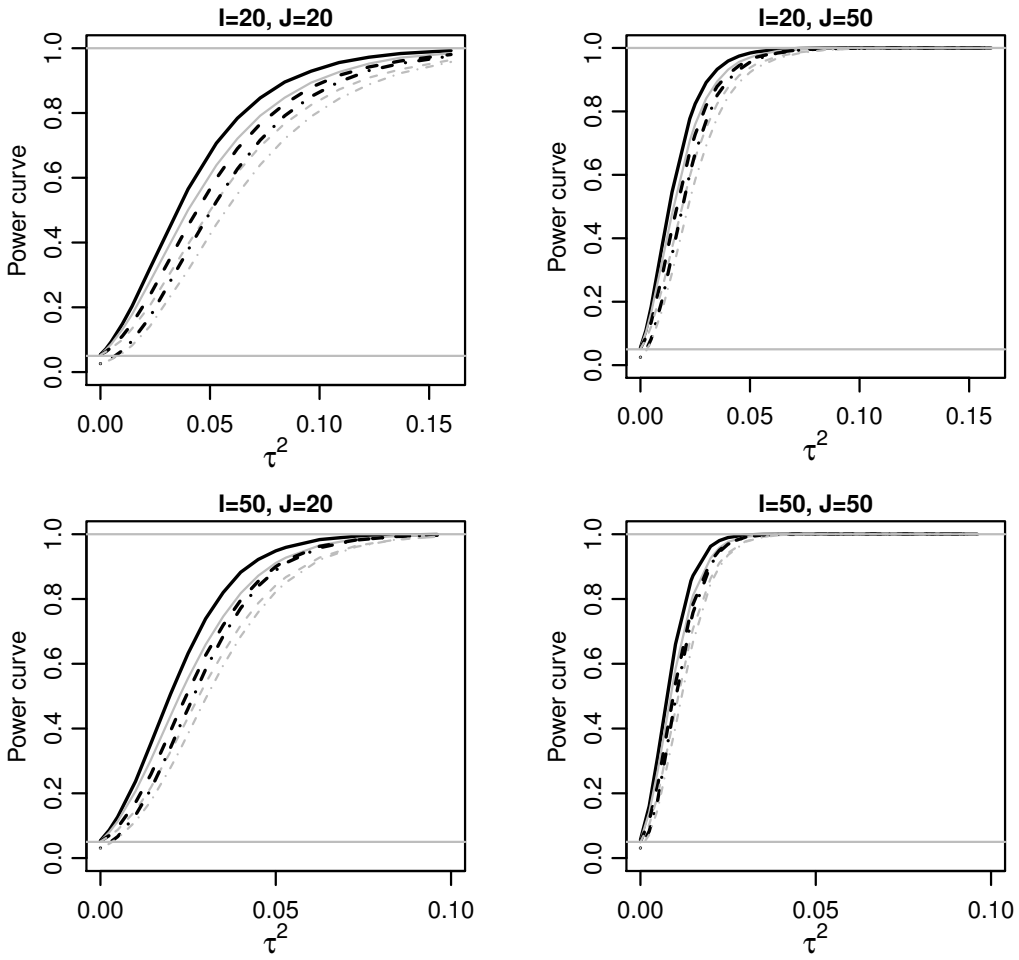

Figure A.2 Multivariate functional predictor: Power of three tests at the 5% level for the correlated functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

Table A.7 Multivariate functional predictor: Estimation/prediction errors of FMM and FLM across 1000 data sets under various model conditions.

Condition	Correlated functional predictor $X_{ij}(t)$, $\tau^2 = 0.04$						
	MISE for $\beta(t)$			MISE for $\delta(t)$		MSE for Y_{ij}	
	r	FLM	FMM	FLM	FMM	FLM	FMM
I = 20, J = 20	0	0.042	0.042	0.071	0.070	1.289	1.258
	3	0.050	0.050	0.085	0.085	1.535	1.510
I = 20, J = 50	0	0.020	0.020	0.032	0.030	1.203	1.124
	3	0.023	0.022	0.037	0.036	1.439	1.365
I = 50, J = 20	0	0.018	0.018	0.030	0.030	1.266	1.223
	3	0.020	0.020	0.034	0.034	1.510	1.472
I = 50, J = 50	0	0.009	0.009	0.014	0.013	1.201	1.116
	3	0.009	0.009	0.015	0.015	1.438	1.359

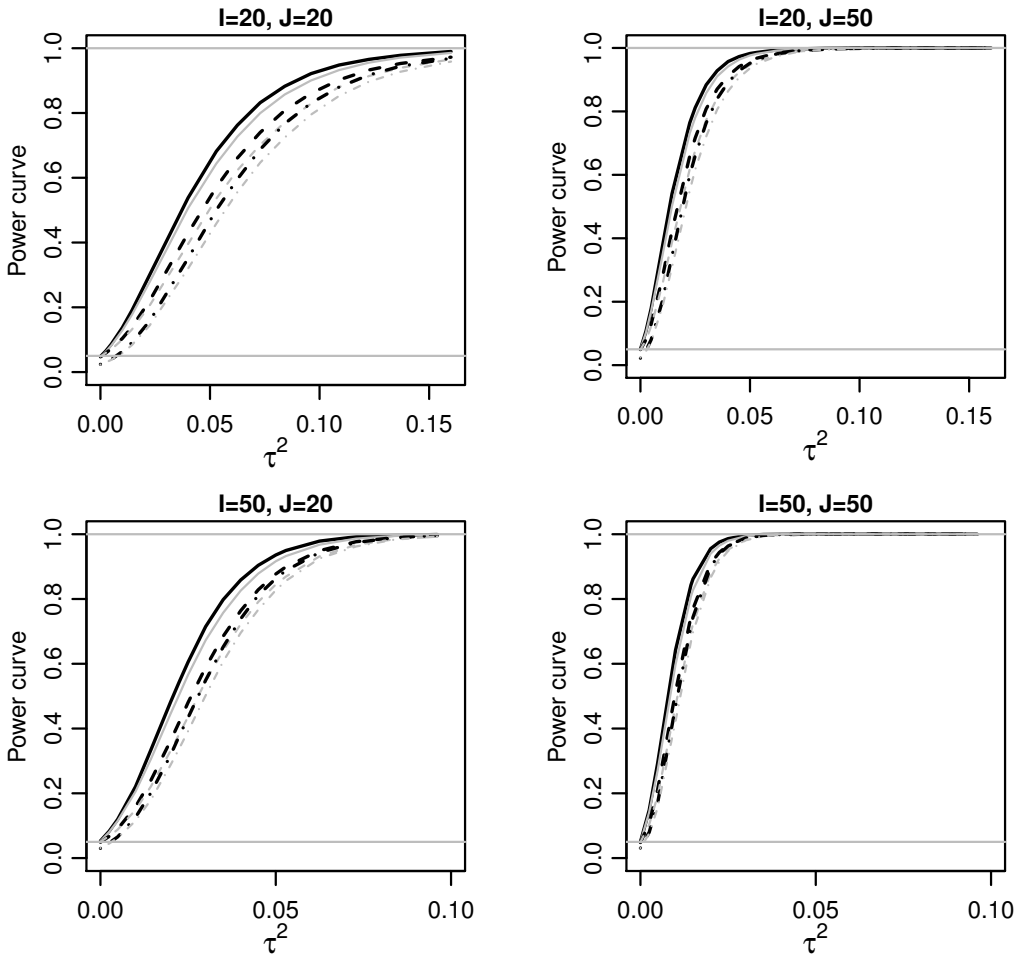

Figure A.3 Multivariate functional predictor: Power of three tests at the 5% level for the independent functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

Table A.8 Multivariate functional predictor: Estimation/prediction errors of FMM and FLM across 1000 data sets under various model conditions.

Condition	Correlated functional predictor $X_{ij}(t)$, $\tau^2 = 0.08$						
	MISE for $\beta(t)$			MISE for $\delta(t)$		MSE for Y_{ij}	
	r	FLM	FMM	FLM	FMM	FLM	FMM
I = 20, J = 20	0	0.047	0.046	0.074	0.071	1.445	1.314
	3	0.055	0.054	0.089	0.086	1.690	1.571
I = 20, J = 50	0	0.024	0.023	0.033	0.031	1.356	1.148
	3	0.027	0.026	0.039	0.036	1.593	1.393
I = 50, J = 20	0	0.020	0.020	0.031	0.030	1.419	1.276
	3	0.022	0.022	0.036	0.034	1.663	1.530
I = 50, J = 50	0	0.011	0.011	0.014	0.013	1.356	1.140
	3	0.011	0.011	0.016	0.015	1.593	1.386

Table A.9 Multivariate functional predictor: Estimation/prediction errors of FMM and FLM across 1000 data sets under various model conditions.

Condition	Independent functional predictor $X_{ij}(t)$, $\tau^2 = 0.04$						
	MISE for $\beta(t)$			MISE for $\delta(t)$		MSE for Y_{ij}	
	r	FLM	FMM	FLM	FMM	FLM	FMM
$I = 20, J = 20$	0	0.045	0.046	0.084	0.084	1.164	1.153
	3	0.054	0.054	0.098	0.098	1.286	1.277
$I = 20, J = 50$	0	0.022	0.021	0.034	0.033	1.090	1.063
	3	0.025	0.025	0.040	0.039	1.207	1.181
$I = 50, J = 20$	0	0.020	0.019	0.034	0.034	1.157	1.142
	3	0.023	0.022	0.040	0.039	1.279	1.266
$I = 50, J = 50$	0	0.010	0.009	0.015	0.015	1.094	1.067
	3	0.011	0.011	0.017	0.017	1.213	1.186

Table A.10 Multivariate functional predictor: Estimation/prediction errors of FMM and FLM across 1000 data sets under various model conditions.

Condition	Independent functional predictor $X_{ij}(t)$, $\tau^2 = 0.08$						
	MISE for $\beta(t)$			MISE for $\delta(t)$		MSE for Y_{ij}	
	r	FLM	FMM	FLM	FMM	FLM	FMM
$I = 20, J = 20$	0	0.052	0.050	0.089	0.086	1.220	1.173
	3	0.060	0.059	0.104	0.100	1.342	1.298
$I = 20, J = 50$	0	0.027	0.025	0.036	0.034	1.143	1.071
	3	0.030	0.029	0.042	0.039	1.260	1.190
$I = 50, J = 20$	0	0.022	0.021	0.037	0.035	1.213	1.163
	3	0.025	0.024	0.042	0.041	1.335	1.288
$I = 50, J = 50$	0	0.011	0.011	0.016	0.015	1.147	1.076
	3	0.013	0.012	0.018	0.017	1.266	1.196

A.1.3 Additional results for the data application

Figure A.4 and Figure A.5 give the residual plots for the FMM model. Figure A.6 gives the violin plots of the mean squared errors (averaged within each subject) for each model fit.

A.1.4 Additional simulation results for the power of the tests

We consider additional simulation scenarios in terms of how the functional effects relate to the marginal eigenfunctions of the repeated functional predictor. In the main manuscript, we have considered $\tau^2 = 2^{1-\tau}\tau^2$ for $\ell = 1, \ldots, L = 3$. Here we consider two additional scenarios. Specifically, we consider: (1) Scenario 1: $\tau_1^2 = \tau^2/4$, $\tau_2^2 = \tau^2/2$ and $\tau_3^2 = \tau^2$ and (2) Scenario 2: $\tau_1^2 = \tau^2/2$, $\tau_2^2 = 0$ and $\tau_3^2 = \tau^2$.The results for univariate functional predictors are presented in Figure A.7 - Figure A.10 while for multivariate functional predictors they are in Figure A.11 - Figure A.14. The figures show that the equal-variance test remains the best overall, among the three tests.

A.2 Fast Covariance Estimation for Multivariate Functional Data

In this section, we detail the multivarite FPCA extended from the FACE method [Xia16] as mentioned in Chapter 2.

A.2.1 Fundamental theory

Assume that $\{\mathbf{X}_i = \big(x_{i\ell}(t), \ldots, x_{ip}(t)\big)^\mathsf{T}, i = 1, \ldots, n\}$ is a set of n independent realizations of random functional process $\mathbf{X}$, which has a p-dimentional multivariate response, $t \in [0, 1]$. Suppose that $\mathbf{X}(t) = \big(x_1(t), \ldots, x_p(t)\big)^\mathsf{T}$ has a p-dimensional smooth mean function $\boldsymbol{\mu}(t)$, that is, $\boldsymbol{\mu}(t) = \mathbb{E}\mathbf{X}(t) = \{\mathbb{E}x_1(t), \ldots, \mathbb{E}x_p(t)\}^\mathsf{T} = \{\mu_1(t), \ldots, \mu_p(t)\}^\mathsf{T}$. Following that, covariance function $\mathbf{K}(s,t) = \mathbb{E}\{(\mathbf{X}(s) - \boldsymbol{\mu}(s))(\mathbf{X}(t) - \boldsymbol{\mu}(t)^\mathsf{T})\} = [K_{j_1 j_2}(s,t)]_{1 \le j_1, j_2 \le p} \in \mathbb{R}^{p \times p}$, and $K_{j_1 j_2}(s,t) = \mathrm{Cov}\{x_{j_1}(s), x_{j_2}(t)\}$.

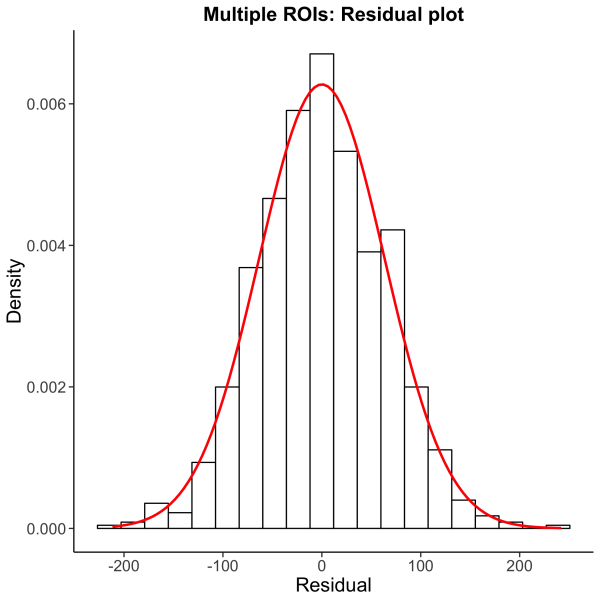

Figure A.4 Residual plot of error term using joint analysis with FMM. Red line denotes the normal density curve fitted over the error.

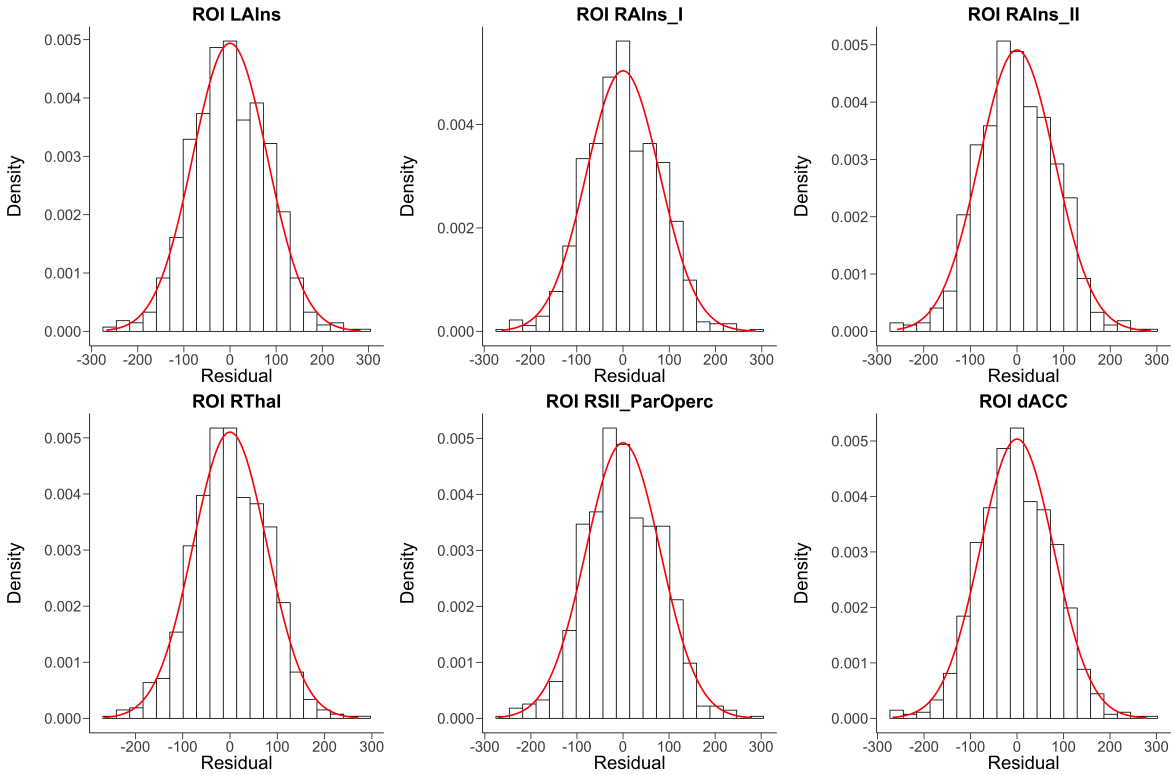

Figure A.5 Residual plot of random errors using univariate analysis with FMM. Red line denotes the normal density curve fitted over the error.

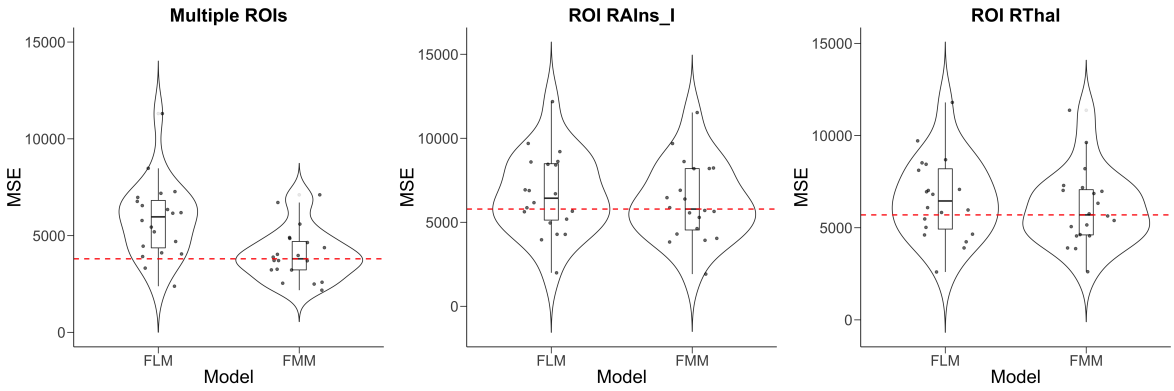

Figure A.6 Violin plots of estimation errors for both joint analysis and separate analysis of ROIs using FMM and FLM. Note that the mean squared error is first averaged within each subject. The red dashed curve is the median of MSE given by FMM.

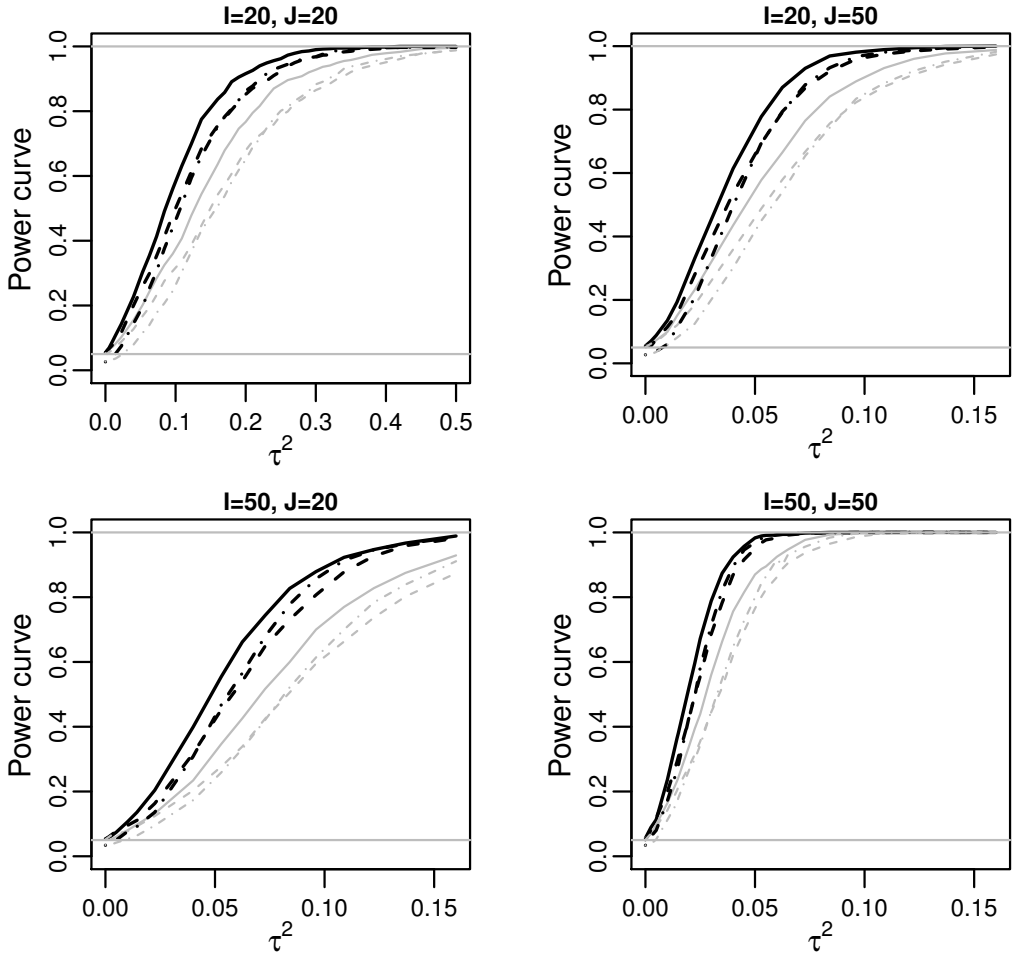

Figure A.7 Scenario 1: Power of three tests at the 5% level for the univariate correlated functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

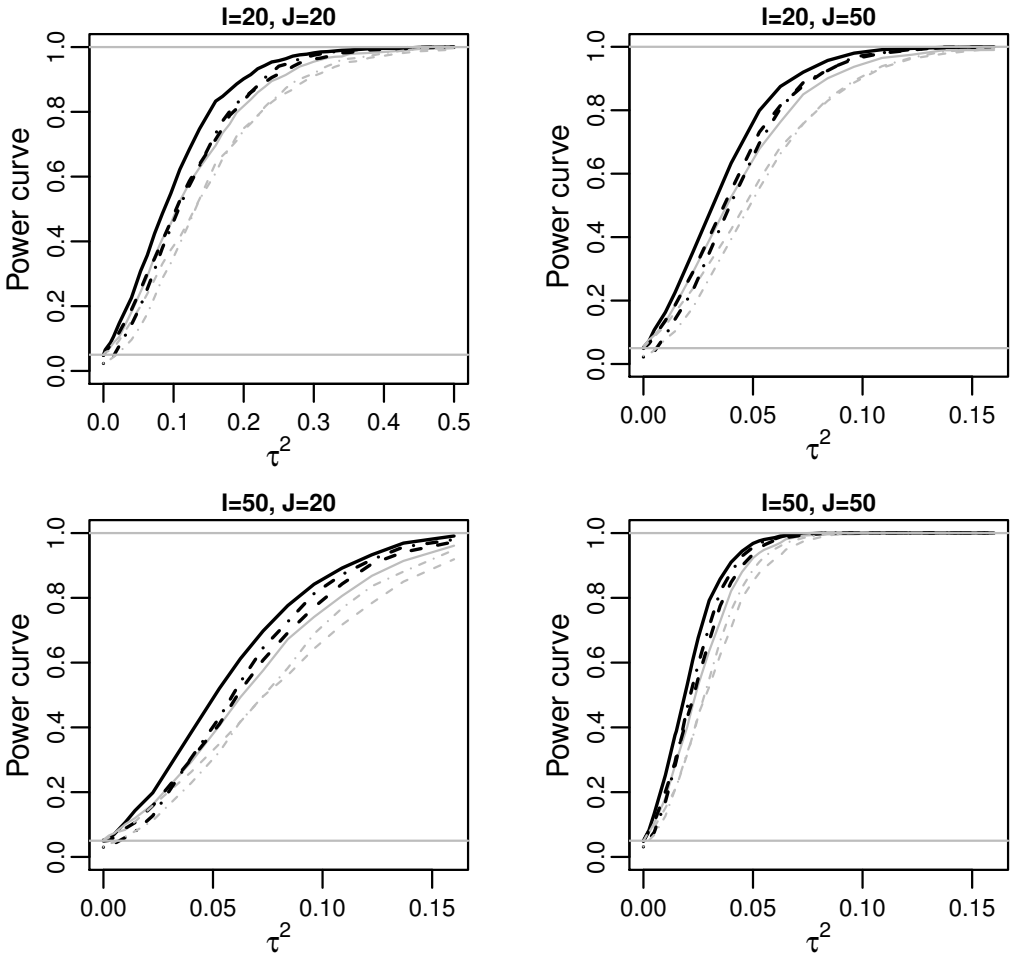

Figure A.8 Scenario 1: Power of three tests at the 5% level for the univariate independent functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

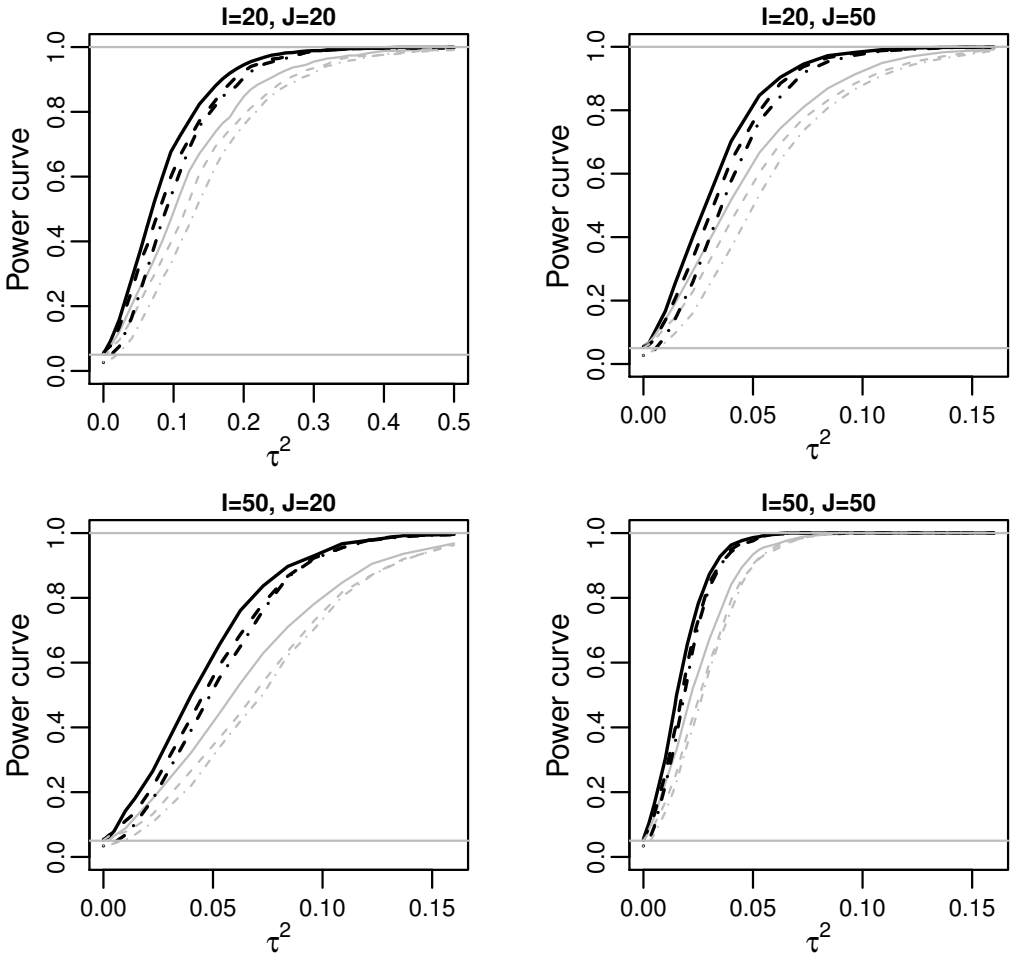

Figure A.9 Scenario 2: Power of three tests at the 5% level for the univariate correlated functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

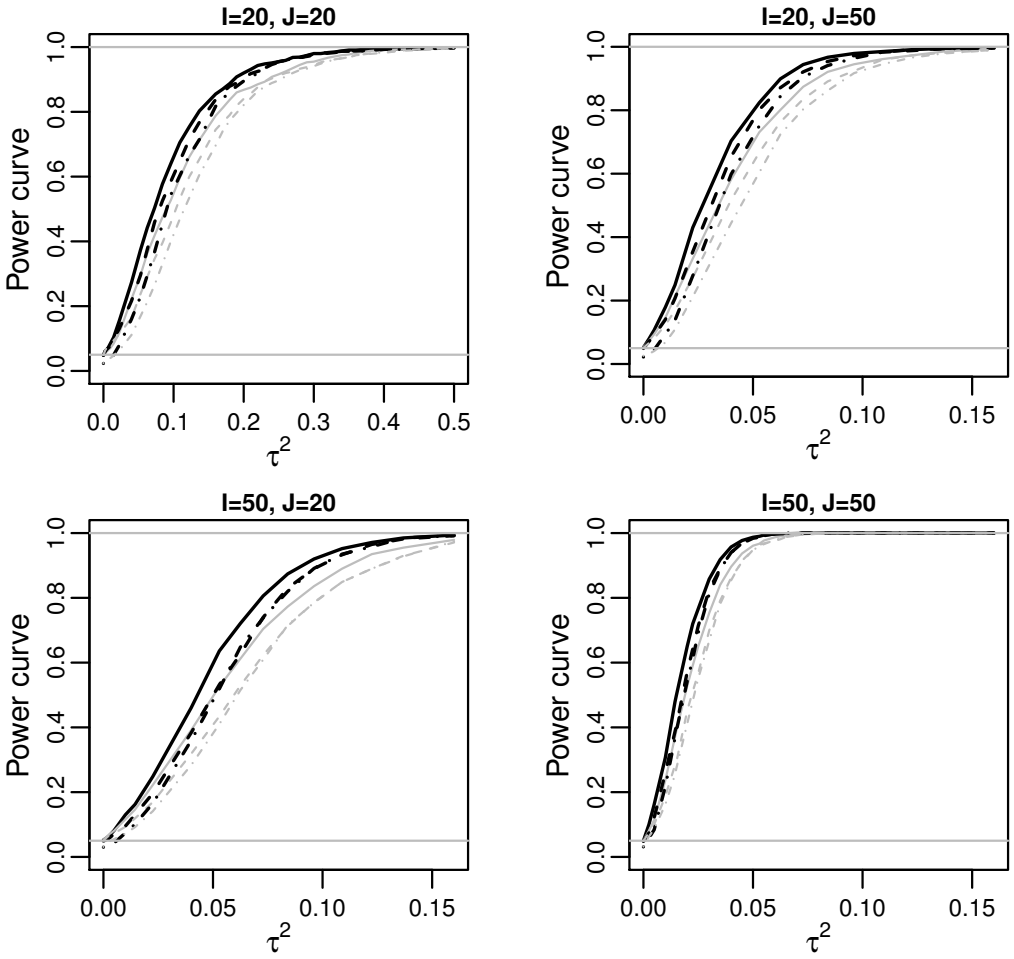

Figure A.10 Scenario 2: Power of three tests at the 5% level for the univariate independent functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

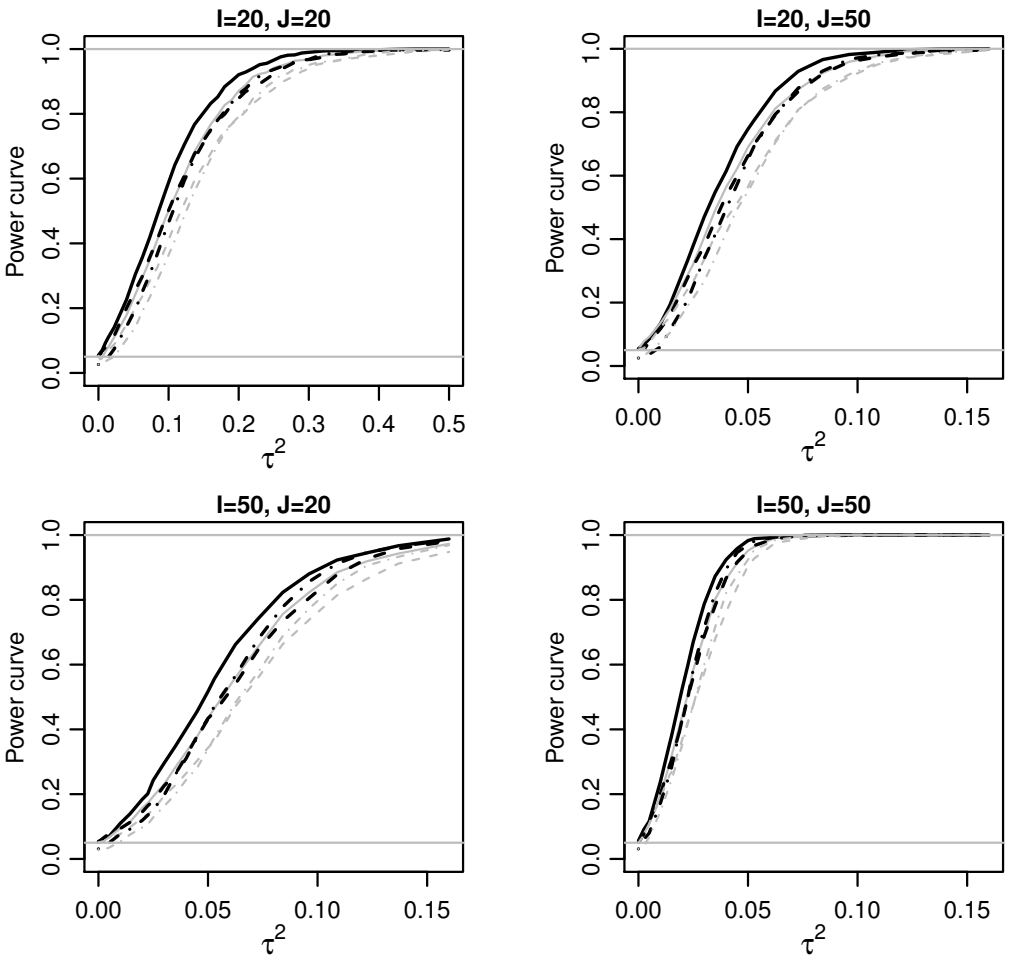

Figure A.11 Multivariate functional predictor with Scenario 1: Power of three tests at the 5% level for the correlated functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

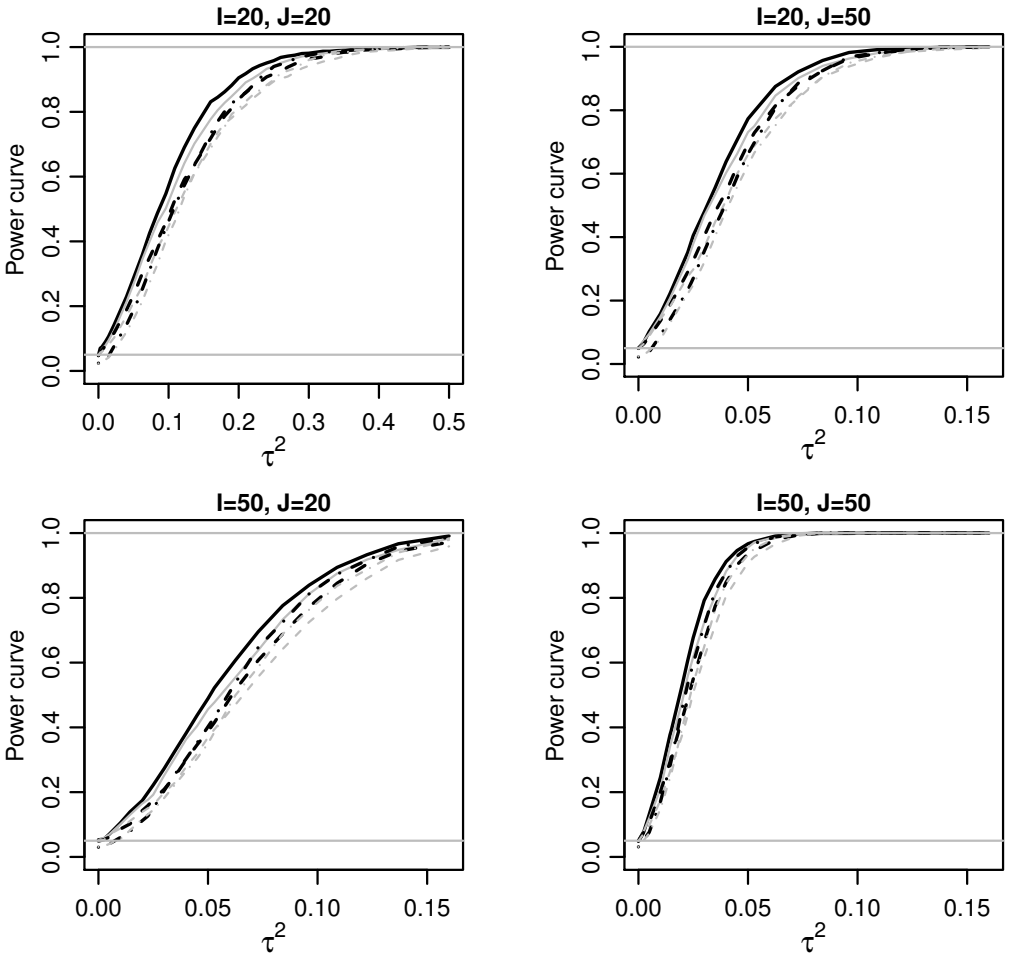

Figure A.12 Multivariate functional predictor with Scenario 1: Power of three tests at the 5% level for the independent functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

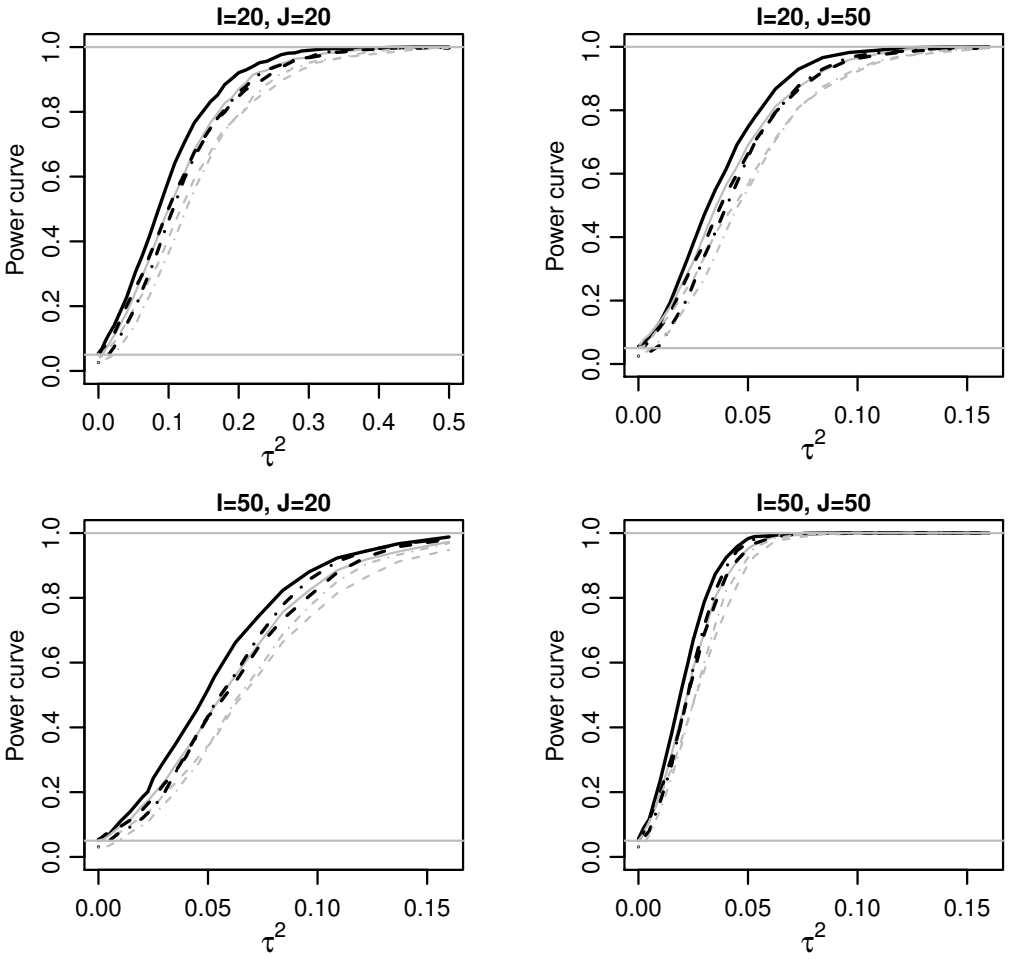

Figure A.13 Multivariate functional predictor with Scenario 2: Power of three tests at the 5% level for the correlated functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

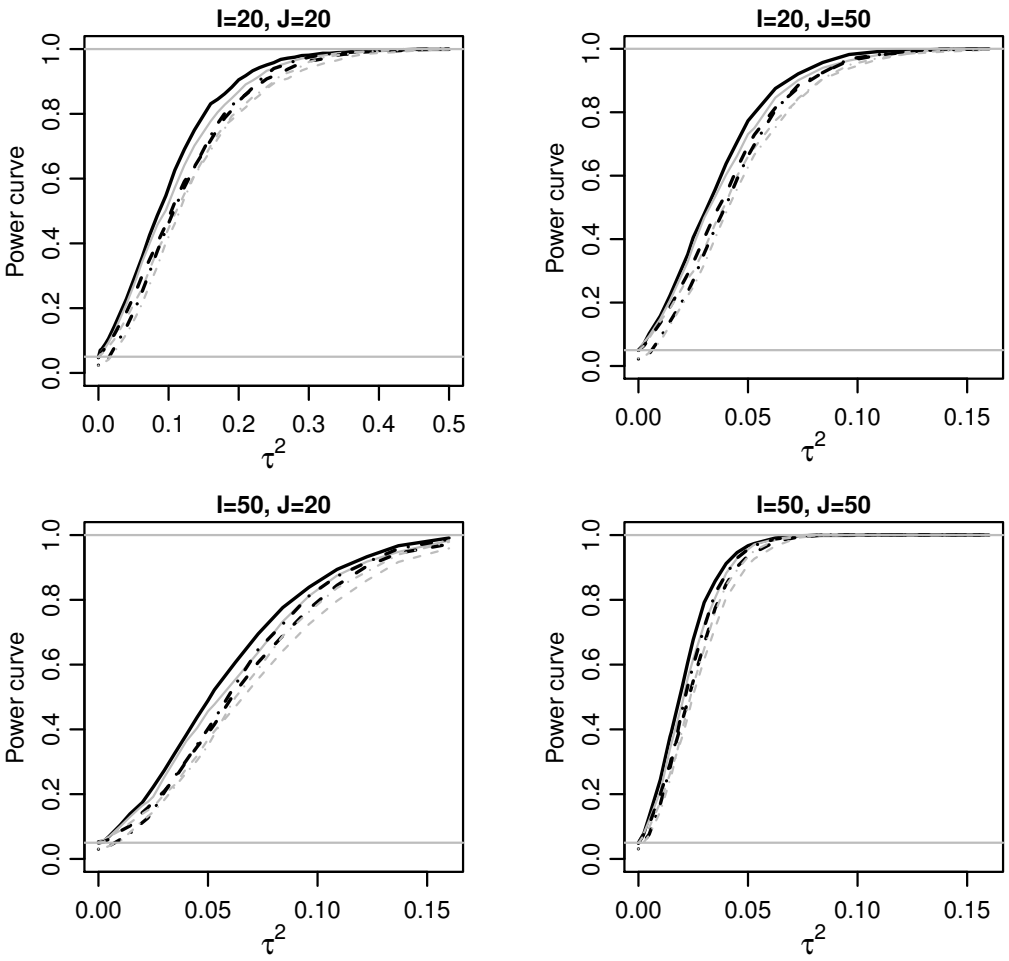

Figure A.14 Multivariate functional predictor with Scenario 2: Power of three tests at the 5% level for the independent functional predictor $X_{ij}(t)$ and as a function of τ^2. Black lines are for smooth functional data, i.e., $r = 0$ while gray lines are for noisy functional data. Solid lines: equal-variance test; dashed lines: Bonferroni-corrected test; dot-dashed lines: asLRT.

By the multivariate Mercer's theorem,

$$\mathbf{K}(s,t) = \sum_{\ell=1}^{\infty} \lambda_\ell \begin{pmatrix} \phi_{1\ell}(s) \\ \vdots \\ \phi_{pl}(s) \end{pmatrix} \begin{pmatrix} \phi_{1\ell}(t), \ldots, \phi_{p\ell}(t) \end{pmatrix}, \tag{A.1}$$

where λ_ℓ is the ℓth largest eigenvalue, $\begin{pmatrix} \phi_{1\ell}(t), \ldots, \phi_{p\ell}(t) \end{pmatrix}^\mathsf{T}$ is the corresponding ℓth multivariate eigenfunction, and $\sum_{j=1}^{p} \int \phi_{j\ell_1}(t)\phi_{j\ell_2}(t)\mathrm{d}t = 1_{\{\ell_1=\ell_2\}}$. Then $K_{j_1j_2}(s,t) = \sum_{\ell=1}^{\infty} \lambda_\ell \phi_{j_1\ell}(s)\phi_{j_2\ell}(t)$. The multivariate Karhunen-Loève expansion gives

$$\mathbf{X}(t) = \boldsymbol{\mu}(t) + \sum_{\ell=1}^{\infty} \xi_\ell \begin{pmatrix} \phi_{1\ell}(t) \\ \vdots \\ \phi_{p\ell}(t) \end{pmatrix}, \tag{A.2}$$

where $\{\xi_\ell, \ell = 1, \ldots\}$ are uncorrelated random scores with $\mathbb{E}(\xi_\ell) = 0$, $\mathrm{Var}(\xi_\ell) = \lambda_\ell$, and $\xi_\ell = \sum_{j=1}^{p} \int (x_j(t) - \mu_j(t))\phi_{j\ell}(t)\mathrm{d}t$.

A.2.2 Data structure

With a certain amount of measurement error, the observed data of ith subject's jth response at sampling time point $t_k \in \{t_1, \ldots, t_m\}$ has the following form:

$$y_{ijk} = x_{ij}(t_k) + \epsilon_{ijk} = \mu_j(t_k) + \sum_{\ell=1}^{\infty} \xi_{i\ell}\phi_{jl}(t_k) + \epsilon_{ijk}, \tag{A.3}$$

where ϵ_{ijk} are random errors with zero means and variance σ_j^2, and independent across i, j and k, $i = 1, \ldots, n$, $j = 1, \ldots, p$, $k = 1, \ldots, m$. Let $\mathbf{Y}_{ij} = \begin{pmatrix} y_{ij1}, \ldots, y_{ijm} \end{pmatrix}^\mathsf{T}$, and $\mathbf{Y}_j = \begin{bmatrix} \mathbf{Y}_{1j}, \ldots, \mathbf{Y}_{nj} \end{bmatrix}$ is a $m \times n$ dimensional matrix. Not losing generality, we suppose that $\mathbf{Y}_j$ has been centered across subjects. Then cross sample covariance can be computed as $\widehat{\mathbf{K}}_{j_1j_2} = \frac{1}{n}\mathbf{Y}_{j_1}\mathbf{Y}_{j_2}^\mathsf{T} \in \mathbb{R}^{m\times m}$ with the realization at sampling point (t_1, t_2), $\widehat{\mathbf{K}}_{j_1j_2}(t_1, t_2) = \frac{1}{n}\sum_{i=1}^{n} y_{ij_1t_1}y_{ij_2t_2} (1 \le t_1, t_2 \le m)$,

and sample covariance $\widehat{\mathbf{K}}$ is

$$\begin{pmatrix} \widehat{\mathbf{K}}_{11} & \cdots & \widehat{\mathbf{K}}_{1j} & \cdots & \widehat{\mathbf{K}}_{1p} \\ \vdots & \ddots & & \vdots & \vdots \\ \widehat{\mathbf{K}}_{j1} & & \widehat{\mathbf{K}}_{jj} & & \widehat{\mathbf{K}}_{jp} \\ \vdots & & & \ddots & \vdots \\ \widehat{\mathbf{K}}_{p1} & \cdots & \widehat{\mathbf{K}}_{pj} & \cdots & \widehat{\mathbf{K}}_{pp} \end{pmatrix}_{mp \times mp} .$$

A.2.3 Covariance estimation

Our interest lies in the estimation of covariance function $K_{j_1 j_2}(s,t)$. To do this, we start with smoothing the sample covariance matrix $\widehat{\mathbf{K}}$. In order to obtain the smoothed covariance matrix $\widetilde{\mathbf{K}}$, we smooth each $\widehat{\mathbf{K}}_{j_1 j_2}$.

When $j_1 = j_2 = j$, we adopt the FACE estimator in [Xia16] to smooth the auto-covariance matrix $\widehat{\mathbf{K}}_{jj}(j = 1, \ldots, p)$. It follows that smoothed auto-covariance matrix $\widetilde{\mathbf{K}}_{jj}$ is given by a sandwich form $\mathbf{S}_j \widehat{\mathbf{K}}_{jj} \mathbf{S}_j$, where $\mathbf{S}_j$ is a $m \times m$ - dimensional symmetric smoother matrix. Penalized B-splines are used to construct $\mathbf{S}_j$, and thus $\mathbf{S}_j = \mathbf{B}(\mathbf{B}^\mathsf{T}\mathbf{B} + \lambda_j \mathbf{P})^{-1}\mathbf{B}^\mathsf{T}$, where $\mathbf{B} = \big(\mathbf{B}_1(\cdot), \ldots, \mathbf{B}_c(\cdot)\big)$ is the $m \times c$ design matrix, $\mathbf{P}$ is a $c \times c$ symmetric penalty matrix, λ_j is the smoothing parameter, c is the number of interior knots plus the order (degree plus 1) of B-splines. When $j_1 \neq j_2$, cross-covariance matrix $\widehat{\mathbf{K}}_{j_1 j_2}$, similarly, is smoothed by $\widetilde{\mathbf{K}}_{j_1 j_2} = \mathbf{S}_{j_1} \widehat{\mathbf{K}}_{j_1 j_2} \mathbf{S}_{j_2}$, where $\mathbf{S}_{j_1}$ and $\mathbf{S}_{j_2}$ are the corresponding smoother matrix of $\widehat{\mathbf{K}}_{j_1 j_1}$ and $\widehat{\mathbf{K}}_{j_2 j_2}$. After smoothing each covariance matrix $\widehat{\mathbf{K}}_{j_1 j_2}$, we pool the estimates together and use the proposition 1 from [Li18] to obtain the decomposition from $\widetilde{\mathbf{K}}$.

To prove that we have the proper condition, that is, $K_{j_1 j_2}(s,t)$ can be represented as $\mathbf{B}(s)^\mathsf{T} \mathbf{\Theta}_{j_1 j_2} \mathbf{B}(t)$, for using the Proposition 1, we give some details as follows.

$$
\begin{aligned}
\widetilde{\mathbf{K}}_{jj} &= \mathbf{S}_j \widehat{\mathbf{K}}_{jj} \mathbf{S}_j \\
&= \mathbf{B} \underbrace{(\mathbf{B}^\mathsf{T}\mathbf{B} + \lambda_j \mathbf{P})^{-1} \mathbf{B}^\mathsf{T} \widehat{\mathbf{K}}_{jj} \mathbf{B} (\mathbf{B}^\mathsf{T}\mathbf{B} + \lambda_j \mathbf{P})^{-1}}_{\widehat{\mathbf{\Theta}}_{jj} \in \mathbb{R}^{c \times c}} \mathbf{B}^\mathsf{T} \\
&= \begin{pmatrix} B_1(t_1) & B_2(t_1) & \cdots & B_c(t_1) \\ \vdots & \vdots & \cdots & \vdots \\ B_1(t_m) & B_2(t_m) & \cdots & B_c(t_m) \end{pmatrix} \widehat{\mathbf{\Theta}}_{jj} \begin{pmatrix} B_1(t_1) & B_2(t_1) & \cdots & B_c(t_1) \\ \vdots & \vdots & \cdots & \vdots \\ B_1(t_m) & B_2(t_m) & \cdots & B_c(t_m) \end{pmatrix}^\mathsf{T}
\end{aligned} \tag{A.4}
$$

Therefore, $\widetilde{K}_{jj}(s,t) = \{B_1(s), \ldots, B_c(s)\} \widehat{\mathbf{\Theta}}_{jj} \{B_1(t), \ldots, B_c(t)\}^\mathsf{T}$. Let $\mathbf{B}(s)$ denote $\{B_1(s), \ldots, B_c(s)\}^\mathsf{T}$, then $\widetilde{K}_{jj}(s,t) = \mathbf{B}(s)^\mathsf{T} \widehat{\mathbf{\Theta}}_{jj} \mathbf{B}(t)$. For cross-covariance matrix,

$$
\begin{aligned}
\widetilde{\mathbf{K}}_{j_1 j_2} &= \mathbf{S}_{j_1} \widehat{\mathbf{K}}_{j_1 j_2} \mathbf{S}_{j_2} \\
&= \mathbf{B} \underbrace{(\mathbf{B}^\mathsf{T}\mathbf{B} + \lambda_{j_1} \mathbf{P})^{-1} \mathbf{B}^\mathsf{T} \widehat{\mathbf{K}}_{j_1 j_2} \mathbf{B} (\mathbf{B}^\mathsf{T}\mathbf{B} + \lambda_{j_2} \mathbf{P})^{-1}}_{\widehat{\mathbf{\Theta}}_{j_1 j_2} \in \mathbb{R}^{c \times c}} \mathbf{B}^\mathsf{T} \\
&= \begin{pmatrix} B_1(t_1) & B_2(t_1) & \cdots & B_c(t_1) \\ \vdots & \vdots & \cdots & \vdots \\ B_1(t_m) & B_2(t_m) & \cdots & B_c(t_m) \end{pmatrix} \widehat{\mathbf{\Theta}}_{j_1 j_2} \begin{pmatrix} B_1(t_1) & B_2(t_1) & \cdots & B_c(t_1) \\ \vdots & \vdots & \cdots & \vdots \\ B_1(t_m) & B_2(t_m) & \cdots & B_c(t_m) \end{pmatrix}^\mathsf{T},
\end{aligned} \tag{A.5}
$$

where

$$
\begin{aligned}
\widehat{\mathbf{\Theta}}_{j_1 j_2} &= (\mathbf{B}^\mathsf{T}\mathbf{B} + \lambda_{j_1} \mathbf{P})^{-1} \mathbf{B}^\mathsf{T} \widehat{\mathbf{K}}_{j_1 j_2} \mathbf{B} (\mathbf{B}^\mathsf{T}\mathbf{B} + \lambda_{j_2} \mathbf{P})^{-1} \\
&= (\mathbf{B}^\mathsf{T}\mathbf{B} + \lambda_{j_1} \mathbf{P})^{-1} \mathbf{B}^\mathsf{T} (n^{-1} \mathbf{Y}_{j_1} \mathbf{Y}_{j_2}^\mathsf{T}) \mathbf{B} (\mathbf{B}^\mathsf{T}\mathbf{B} + \lambda_{j_2} \mathbf{P})^{-1} \\
&= \mathbf{F}_{j_1} \mathbf{F}_{j_2}^\mathsf{T} / n, \\
\mathbf{F}_j &= (\mathbf{B}^\mathsf{T}\mathbf{B} + \lambda_j \mathbf{P})^{-1} \mathbf{B}^\mathsf{T} \mathbf{Y}_j \in \mathbb{R}^{c \times n}
\end{aligned} \tag{A.6}
$$

Hence $\widetilde{K}_{j_1 j_2}(s,t) = \mathbf{B}(s)^\mathsf{T} \widehat{\mathbf{\Theta}}_{j_1 j_2} \mathbf{B}(t)$, and the condition for proposition 1 is met. Following Proposition 1, once we have the estimate of the coefficient matrix $\widehat{\Theta}_{j_1 j_2}$, let $\mathbf{G} = \int \mathbf{B}(t) \mathbf{B}(t)^\mathsf{T} \in \mathbb{R}^{c \times c}$, spectral decompositon of $\left[\mathbf{G}^{\frac{1}{2}} \widehat{\mathbf{\Theta}}_{j_1 j_2} \mathbf{G}^{\frac{1}{2}} \right]_{1 \le j_1, j_2 \le p} \in \mathbb{R}^{pc \times pc}$, $\sum_{l=1}^{\infty} d_l \boldsymbol{\mu}_l \boldsymbol{\mu}_l^\mathsf{T}$, where d_l is the lth largest eigenvalue, and $\boldsymbol{\mu}_l = \left\{ \boldsymbol{\mu}_{1l}^\mathsf{T}, \ldots, \boldsymbol{\mu}_{pl}^\mathsf{T} \right\}^\mathsf{T} \in \mathbb{R}^{pc}$ is the corresponding eigenvector with

$\boldsymbol{\mu}_{jl} \in \mathbb{R}^c$. Then $\phi_{jl}(t) = \mathbf{B}(t)^\mathsf{T}\mathbf{G}^{-\frac{1}{2}}\boldsymbol{\mu}_{jl}$.

Note that $\mathbf{F}_j$ only depends on λ_j. To calculate $\mathbf{F}_j$, we start with the decomposition $(\mathbf{B}^\mathsf{T}\mathbf{B})^{-1/2}\mathbf{P}(\mathbf{B}^\mathsf{T}\mathbf{B})^{-1/2} = \mathbf{U}\text{diag}(s)\mathbf{U}^\mathsf{T}$, where $\mathbf{U}$ is the eigenvector matrix and s is the eigenvalues vector. Let $\mathbf{A} = (\mathbf{B}^\mathsf{T}\mathbf{B})^{-1/2}\mathbf{U}$. Then $(\mathbf{B}^\mathsf{T}\mathbf{B} + \lambda_j\mathbf{P})^{-1} = \mathbf{A}\{\mathbf{I}_c + \lambda_j\text{diag}(s)\}^{-1}\mathbf{A}^\mathsf{T}$. Let $\widetilde{\mathbf{Y}}_j = \mathbf{B}^\mathsf{T}\mathbf{Y}_j$. It follows that $\mathbf{F}_j = \mathbf{A}\{\mathbf{I}_c+\lambda_j\text{diag}(s)\}^{-1}\mathbf{A}^\mathsf{T}\widetilde{\mathbf{Y}}_j$. Furthermore, let $\widetilde{\mathbf{F}}_j = \mathbf{G}^{\frac{1}{2}}\mathbf{F}_j$, then $\mathbf{G}^{\frac{1}{2}}\widehat{\boldsymbol{\Theta}}_{j_1j_2}\mathbf{G}^{\frac{1}{2}} = \widetilde{\mathbf{F}}_{j_1}\widetilde{\mathbf{F}}_{j_2}^\mathsf{T}/n$. Hence, $\left[\mathbf{G}^{\frac{1}{2}}\widehat{\boldsymbol{\Theta}}_{j_1j_2}\mathbf{G}^{\frac{1}{2}}\right]_{1\le j_1,j_2\le p} = \widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^\mathsf{T}/n$, where $\widetilde{\mathbf{F}} = \left(\widetilde{\mathbf{F}}_1, \ldots, \widetilde{\mathbf{F}}_p\right)^\mathsf{T}$.

The algorithm for the fast covariance estimation for multivariate functional data is as follows:

Step 1: Obtain the decomposition $(\mathbf{B}^\mathsf{T}\mathbf{B})^{-1/2}\mathbf{P}(\mathbf{B}^\mathsf{T}\mathbf{B})^{-1/2} = \mathbf{U}\text{diag}(s)\mathbf{U}^\mathsf{T}$;

Step 2: Specify $\mathbf{S}_j = \mathbf{A}_s\boldsymbol{\Sigma}_s\mathbf{A}_s^\mathsf{T}$ with $\boldsymbol{\Sigma}_s = \{\mathbf{I}_c + \lambda_j\text{diag}(s)\}^{-1}$ and $\mathbf{A}_s = \mathbf{B}(\mathbf{B}^\mathsf{T}\mathbf{B})^{-1/2}\mathbf{U}$ by calculating and storing s;

Step 3: Calculate and store $\widetilde{\mathbf{Y}}_{jS} = \mathbf{A}_s^\mathsf{T}\mathbf{Y}_j$;

Step 4: Select λ_j by minimizing the pooled generalized cross validation (PGCV), i.e., $\sum_{i=1}^n ||\mathbf{Y}_{ij} - \mathbf{S}_j\mathbf{Y}||^2/\{1 - \text{tr}(\mathbf{S}_j)/m\}^2$ using Proposition 2 in [Xia16];

Step 5: Calculate $\widetilde{\mathbf{Y}}_j = \mathbf{B}^\mathsf{T}\mathbf{Y}_j$;

Step 6: Calculate $\mathbf{H}_j = \mathbf{A}\{\mathbf{I}_c + \lambda_j\text{diag}(s)\}^{-1}\mathbf{A}^\mathsf{T}$;

Step 7: Calculate $\mathbf{F}_j = \mathbf{H}_j\widetilde{\mathbf{Y}}_j$;

Step 8: Calculate $\widetilde{\mathbf{F}}_j = \mathbf{G}^{\frac{1}{2}}\mathbf{F}_j$;

Step 9: Eigen decomposition $\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^\mathsf{T}/n = \sum_{l=1}^\infty d_l\boldsymbol{\mu}_l\boldsymbol{\mu}_l^\mathsf{T}$;

Step 10: Calculate $\phi_{jl}(t) = \mathbf{B}(t)^\mathsf{T}\mathbf{G}^{-\frac{1}{2}}\boldsymbol{\mu}_{jl}$.

APPENDIX

B

SUPPLEMENTAL MATERIALS FOR CHAPTER 3

B.1 Results for Greedy Search

This section reports the out-of-sample classification results using greedy search based on the transformed data and raw data as discussed in Chapter 3. Figure B.1 and Figure B.2 give the cross-validated out-of-sample classification error with different pre-specified number of clusters using log quantile transformed curves and raw data, respectively. It shows that classification performance using transformed data gives better prediction performance than that of raw data.

B.2 Additional Results for PSS

Figure B.3 compares the true group and the in-sample estimated clustering membership by the PSS method with 3 clusters and initial features from the 2-cluster greedy search at the third step (i.e. $p = 3$).

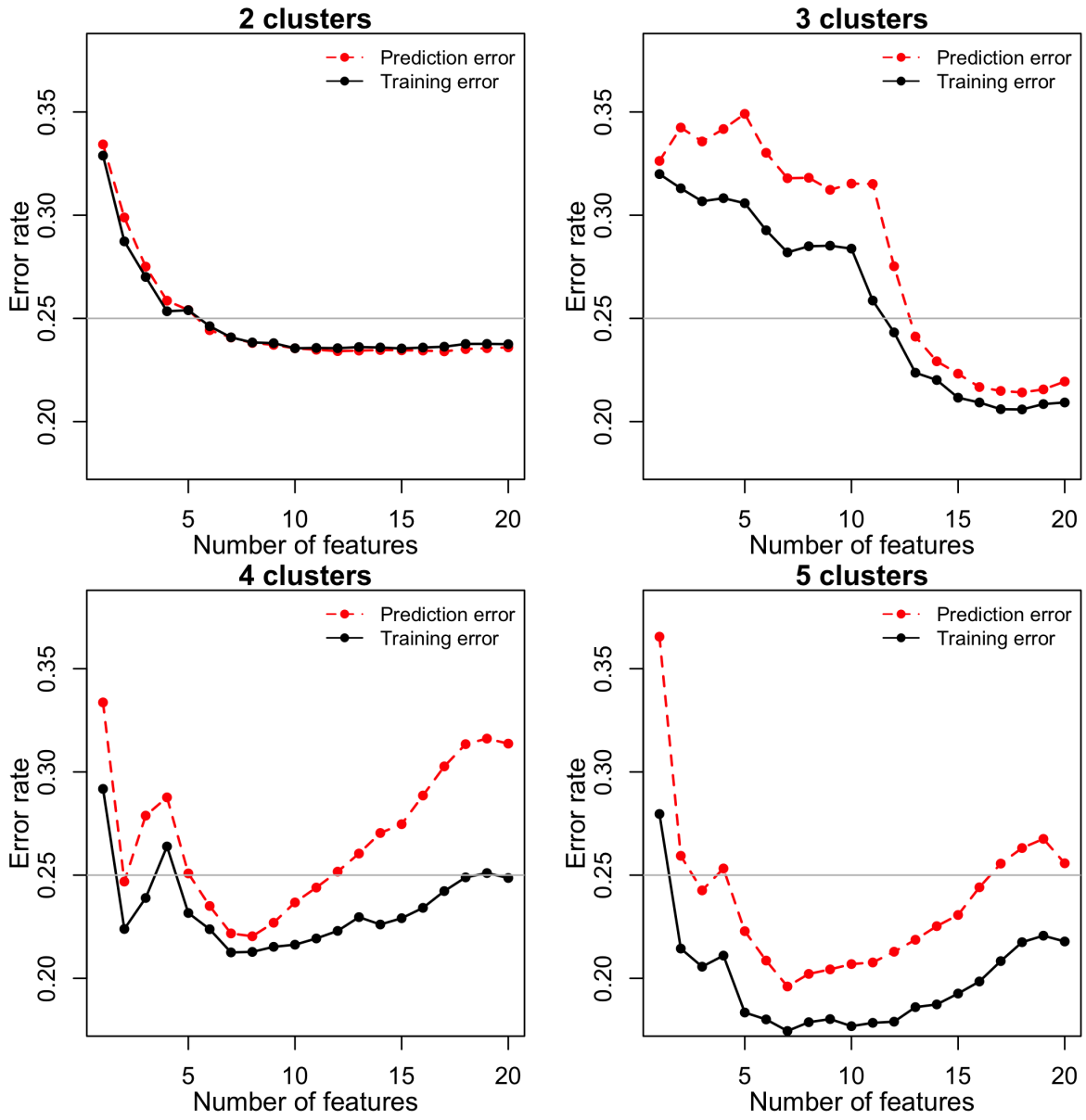

Figure B.1 Greedy search based on transformed curves: Out-of-sample classification error by 100 bootstrapped 5-fold cross-validation at each step of selected features. Solid black lines are for training error. Red dashed lines are for prediction error. Note that in each subfigure, greedy search use the corresponding pre-specified fixed number of clusters shown in the title to select feature at each step.

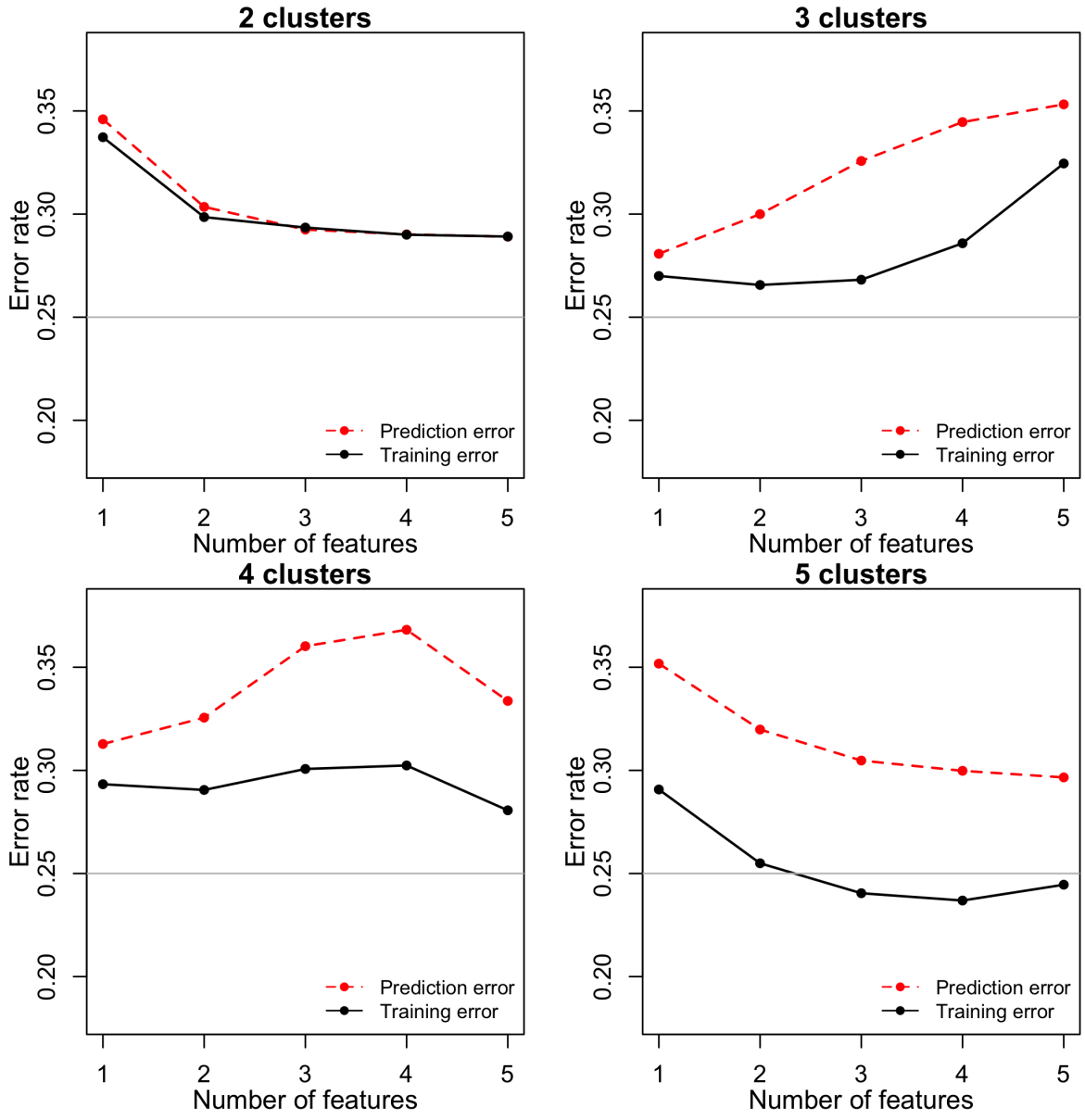

Figure B.2 Greedy search based on raw data using different number of clusters: Out-of-sample classification error by 100 bootstrapped 5-fold cross-validation at each step of selected features. Solid black lines are for training error. Red dashed lines are for prediction error. Note that in each subfigure, greedy search use the corresponding pre-specified fixed number of clusters shown in the title to select feature at each step.

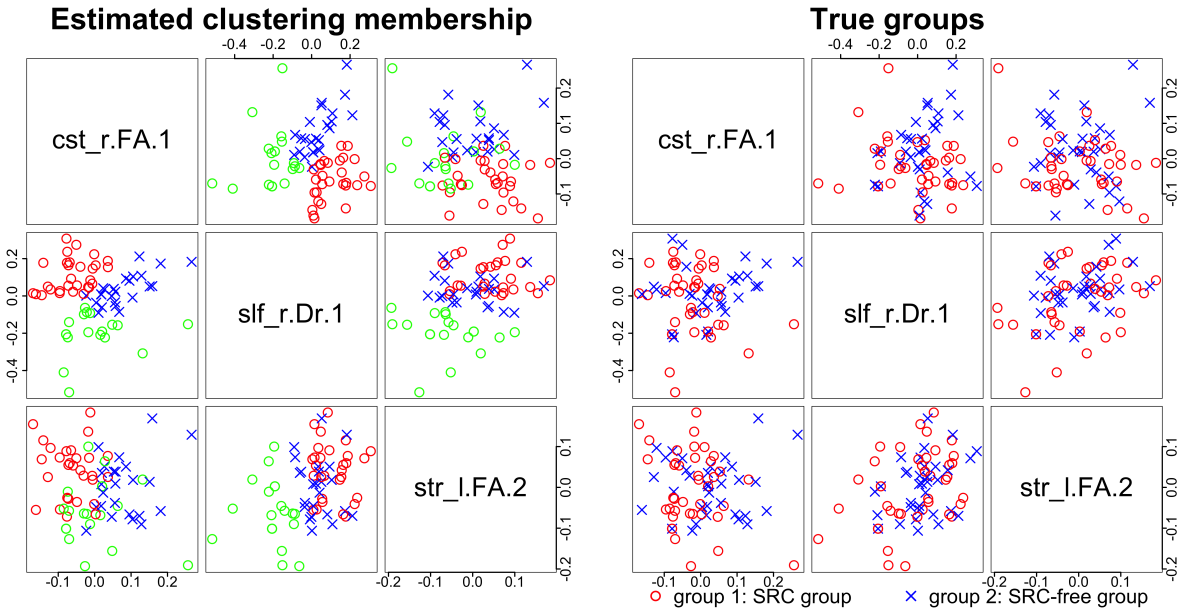

Figure B.3 $p = 3$: Plots of the in-sample estimated clustering membership vs. the true groups. Left panel shows in-sample clustering result using the first three selected features of the PSS method with starting value from 2-cluster greedy search. Right panel shows the true group of the 73 football players.