ADVANCED METHODS FOR DISCOVERING GENETIC MARKERS ASSOCIATED WITH HIGH DIMENSIONAL IMAGING DATA

Jingwen Zhang

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2018

Approved by:

Joseph G. Ibrahim

Hongtu Zhu

James S. Marron

Rebecca C. Santelli

Michael I. Love

ABSTRACT

Jingwen Zhang: Advanced Methods for Discovering Genetic Markers
Associated with High Dimensional Imaging Data
(Under the direction of Joseph G. Ibrahim and Hongtu Zhu)

Imaging genetic studies have been widely applied to discover genetic risk factors of inherited neuropsychiatric diseases and neurodevelopmental abnormalities. Despite the notable contribution of genome-wide association studies (GWAS) in neuroimaging research, it has always been difficult to efficiently perform association analysis on imaging phenotypes. There are several challenges arising from this topic, such as the large dimensionality of both imaging data and genetic data, the potential spatial dependency of imaging phenotypes and the computational burden of the GWAS problem. All the aforementioned issues motivate us to investigate new statistical methods in neuroimaging genetic analysis.

In the first project, we develop a hierarchical functional principal regression model (HFPRM) to simultaneously study diffusion tensor bundle statistics on multiple fiber tracts. The model consists of three key components, (i) a varying coefficient model to characterize functional data, (ii) a latent factor model to jointly analyze multiple fiber bundles, and (iii) a multivariate regression model to study the effects of interest using common factors. A hierarchical estimation procedure is proposed and a global statistic is introduced to test hypotheses of interest. Theoretically, the asymptotic distribution of the global test statistic on the common factors has been studied. Simulations are conducted to evaluate the finite sample performance of HFPRM. Finally, we apply our method to a genome-wide association study of a neonate population to explore important genetic architecture in early human brain development.

In the second project, we consider the problem of performing an association test between functional data and scalar variables in a varying coefficient model setting. We propose a functional projection regression model and an associated global testing statistic to aggregate relatively weak signals across the domain of functional data, while reducing the dimension. An optimal functional projection direction is selected to maximize the signal-to-noise ratio with ridge penalty. Theoretically, we examine the asymptotic distribution of the global testing statistic and provide a strategy to adaptively select the tuning parameter. We use simulations to show that the proposed test outperforms existing state-of-the-art methods in functional statistical inference. We also apply the proposed method to a genome-wide association analysis of imaging genetic data in the UK Biobank dataset.

In the third project, the aim is to develop an adaptive projection regression model (APRM) to perform statistical inference on high dimensional imaging responses in the presence of high correlations. We reduce the dimension of the phenotypes through a projection regression model that maximizes the asymptotic signal-to-noise ratio. Independent screening is applied to control noise in non-signal dimensions and a flexible covariance estimation is introduced to account for major dependency within the data. We also implement an adaptive inference procedure to detect signals at multiple levels. Numerical simulations demonstrate that APRM outperforms many state-of-the-art methods in high dimensional inference. Finally, we apply APRM to a genome-wide association analysis of volumetric data on 93 regions of interest in the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset.

To my parents Jinxi Zhang and Yanzhen Ma.
To Lesheng Li.

ACKNOWLEDGEMENTS

First, I would like to thank my advisors and mentors, Dr. Hongtu Zhu and Dr. Joseph Ibrahim, for their guidance and support during my study at UNC. Their expertise, encouragement and enthusiasm guided me during my research and showed me the way to become a scientist. I am grateful for the opportunity to collaborate with Dr. Rebecca Knickmeyer and Dr. Kai Xia. They have taught me valuable experiences in real data analysis and have provided important inspirations to my research. I would also like to thank Dr. Steve Marron and Dr. Michael Love for their insightful comments and valuable advice on this dissertation.

Moreover, I would like to thank my lab mates and my friends, Dr. Chao Huang, Dr. Leo Yufeng Liu, Dr. Zhengwu Zhang, Dr. Tengfei Li, Dr. Baiguo An, Dr. Zhaohua Lu, Dr. Dehan Kong, Dr. Eunjee Lee, Dr. Hojin Yang, Dr. Mihye An, Dr. Wensheng Zhu, Jasmine Yang, Bingxin Zhao and Yue Wang for their kind help during my stay at the UNC biostatistics and imaging analysis lab and the Big-S2 group. I feel lucky to work with and learn from these smart and passionate researchers.

Finally, I would like to thank my parents, Jinxi Zhang and Yanzhen Ma, for their love, understanding and advice. They have supported me wholeheartedly to pursue my dreams and persistently encouraged me not to give up when facing difficulties. I would like to thank my fiance, Lesheng Li, who have always been there for me with love and a listening ear. I could never make it without his company. I also want to acknowledge my best friends, Lili Wei, Zezhe Li, Yujie Xue and Wenhui Gou. Our friendship has helped me through many down moments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Genome-wide association study (GWAS) is a scientific approach to search for common genetic variations associated with a particular trait or a specific disease across the whole genome. It has illuminated enormous biological discoveries in the understanding of human diseases [1, 2, 3]. Particularly, it has been successfully applied to identify genetic risk factors for a number of neuropsychiatric disorders, such as schizophrenia, major depression, autism and ADHD [4, 5, 6]. Despite many valuable findings, there is a substantial gap between the estimated heritability and the proportion of variation explained by significant loci identified in GWAS [7, 8]. Moreover, directly investigating the association between genotypes and diagnosis outcomes is not helpful to reveal the underlying pathways of how genetic factors influencing disease risk.

In recent decades, a number of large-scale neuroimaging cohort studies have been launched, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) [9], the Philadelphia Neurodevelopmental Cohort (PNC) [10], the Pediatric Imaging Neurocognition and Genetics (PING) study [11], the Human Connectome Project (HCP) [12, 13] and many others. These projects have provided a rich source of information to systematically study the structure and function of the human brain and to investigate the influence of genetic, environmental and behavioral factors. Among them, secondary imaging traits serve as essential intermediate phenotypes to study neurological disorders [14]. Biologically, such phenotypes are closer to the primary gene action level, and are expected to have a simpler genetic architecture than clinical and behavior symptoms [15]. Compared with diagnosis outcome, neuroimaging measures are able to quantify disease risk more accurately, in which phenotypic heterogeneity and ambiguity are reduced [8, 16, 17]. Therefore, investigating the association between genetic variants and imaging phenotypes, known as the imaging genetic analysis, becomes increasingly popular. It provides a better understanding of the pathology of inherited neuropsychiatric abnormalities, which should eventually inspire novel therapies in disease prevention, diagnosis and clinical treatment. Moreover, identification of genetic variants associated with imaging traits that are sensitive to disease risk is very likely to improve detection power [18].

Imaging genetic analysis poses four major challenges to current statistical methods. First of all, for complex inherited diseases, each genetic variant may only have small or moderate contribution to imaging traits. The effect might be too weak to be detected in a genome-wide association study (GWAS). A promising approach to overcome such difficulty is to aggregate information from multiple phenotypes and to reduce the search space for genetic markers. Secondly, due to the large number of genetic variants, real-time analysis of GWAS requires extensive computation. A fast, efficient and robust statistical procedure should be proposed to deal with this problem. Moreover, imaging phenotypes also tend to have extremely high dimensionality, and proper dimension reduction techniques should be considered. Finally, the development, functioning and degeneration of brain tissues are not independent in different brain regions and at different time points. Spatial and temporal dependency within imaging data is a critical feature and should be carefully addressed in the statistical model. All these challenges motivate us to develop new statistical methods for imaging genetic analysis. In this dissertation, we consider two types of imaging data, functional imaging data and multivariate imaging data, and propose three novel models to perform association analysis in different scenarios.

Functional data have been commonly observed in neuroimaging studies to characterize the structure and function of the human brain. For example, diffusion properties are measured along neurofiber bundles in diffusion tensor imaging (DTI) to quantify white matter microstructure [19, 20]. In functional magnetic resonance imaging (fMRI), blood-oxygen-level dependent (BOLD) changes are detected across time to characterize brain activity [21]. Moreover, histogram analysis has been widely used in diffusion tensor imaging and magnetic resonance imaging to delineate distributional alterations in brain tissues [22, 23]. In the framework of functional data analysis, we consider two different hypothesis testing problems designed for different scientific aims. In Chapter 3, a hierarchical functional principal regression model (HFPRM) is proposed to jointly analyze diffusion statistics along multiple fiber bundles. In Chapter 4, we study functional phenotypes on a single curve and introduce a powerful test procedure for association analysis.

Brain segmentation classifies imaging voxels into anatomical or functional regions. Imaging measures for different regions or region pairs, such as volume, thickness, surface area and connectivity measures, can be calculated to characterize local brain properties. Compared with voxel-wise measures, region-based statistics have lower dimensions and higher signal-to-noise ratios, and are expected to give more reliable results in association analysis. However, the number of region-based responses typically ranges from hundreds to tens of thousands, which is still high dimensional compared to sample size. Furthermore, brain regions are organized into structural or functional communities. Local measures of the regions within the same community tend to be highly correlated [24]. The potential dependency structure within the phenotypes should be properly adjusted. Therefore, in Chapter 5, we use a high dimensional linear model to characterize region-based data and introduce an efficient global test procedure that allows capture of a flexible covariance-signal structure.

CHAPTER 2: LITERATURE REVIEW

In this chapter, we review existing models and statistical techniques related to our topic. In Section 2.1, we give a brief overview of functional data analysis. In Section 2.2, we introduce some state-of-the-art methods in high dimensional inference. In Section 2.3, we review a popular dimension reduction method in multivariate analysis, the latent factor model.

2.1 Functional Data Analysis

2.1.1 Varying Coefficient Model

Let $y_i(s)$ be a functional outcome of interest and $\boldsymbol{x}_i$ be a $p \times 1$ vector of covariates for subject i, where $i = 1, \cdots, n$, the following varying coefficient model, first introduced in [25], has been widely used to delineate a linear relation between $y_i(s)$ and $\boldsymbol{x}_i$:

$$y_i(s) = \boldsymbol{x}_i^T \boldsymbol{\beta}(s) + e_i(s), \ s \in [0, S]. \tag{2.1}$$

In the above equation, $\boldsymbol{\beta}(s) = (\beta_1(s), \cdots, \beta_j(s), \cdot, \beta_p(s))^T$ is composed of functional coefficients characterizing covariate effects, and $e_i(s)$ represents the error term following $SP\{0, \Sigma_e(s,t)\}$, in which $SP\{\mu(s), \Sigma(s,t)\}$ denotes a stochastic process with mean function $\mu(s)$ and covariance function $\Sigma(s,t)$. The varying coefficient model can be considered as a natural generalization of the linear model to functional data, in which the response $y_i(s)$, the covariate effects $\boldsymbol{\beta}(s)$ and the error term $e_i(s)$ are allowed to change with s over a continuous domain $[0, S]$.

In real data, $y_i(s)$ can only be observed on a set of discrete points in $[0, S]$, which is denoted as $\mathscr{S} = \{s_1, \cdots, s_w, \cdots, s_W\}$. There are two popular methods to estimate $\boldsymbol{\beta}(s)$ from finite samples. One is a spline-based method, in which the functional coefficients $\boldsymbol{\beta}(s)$ are first expanded by a set of basis functions, and penalized regression is then applied to perform curve fitting. A number of regularization methods have been developed within this framework to perform estimation, statistical inference and to address the penalty choices [26, 27, 28]. The other method uses a kernel smoothing technique, in which model (2.1) is fitted by imposing local smoothness. We will introduce this method here in detail, since it is more suitable to deal with the intrinsic local smoothness of the varying coefficient model, and the estimators are easier to compute.

When the functional coefficients $\boldsymbol{\beta}(s)$ in (2.1) have continuous derivations up to order ν, $\boldsymbol{\beta}(s_w)$ can be approximated by Taylor expansion at each observed point s_w as

$$\begin{aligned} \boldsymbol{\beta}(s_w) &\approx \boldsymbol{\beta}(s) + \sum_{k=1}^{\nu} \partial^k \boldsymbol{\beta}(s) \frac{(s_w - s)^k}{k!}, \qquad (2.2) \\ &:= \mathbf{A}(s)\boldsymbol{z}(s_w - s), \qquad (2.3) \end{aligned}$$

where $\boldsymbol{z}(s_w - s) = (1, s_w - s, (s_w - s)^2/2!, \cdots, (s_w - s)^\nu/\nu!)^T$, $\mathbf{A}(s) = (\boldsymbol{\beta}(s), \partial\boldsymbol{\beta}(s), \cdots, \partial^\nu \boldsymbol{\beta}(s))$. Let $K(u)$ be a predetermined smoothing kernel on a closed interval $[-1, 1]$ and $K_{h_1}(u) = {h_1}^{-1} K(u/h_1)$ be the rescaled kernel with bandwidth h_1, $\mathbf{A}(s)$ can be estimated by the minimizer of the weighted least squares (WLS) function

$$\sum_{i=1}^{n} \sum_{w=1}^{W} [y_i(s_w) - x_i^T \mathbf{A}(s)\boldsymbol{z}(s_w - s)] K_{h_1}(s_w - s). \qquad (2.4)$$

In the above varying coefficient model (2.1), the error term $e_i(s)$ incorporates the total variation that can not be explained by $\boldsymbol{x}_i$. To further disentangle the sources of variation in $e_i(s)$, [29] introduced the following varying coefficient model that differentiates between low frequency spatial variation and independent measurement error:

$$y_i(s) = \boldsymbol{x}_i^T \boldsymbol{\beta}(s) + \eta_i(s) + e_i(s), \tag{2.5}$$

where the individual function $\eta_i(s)$ is modeled as a random function following $SP\{0, \Sigma_\eta(s,t)\}$ and $e_i(s)$ follows $SP\{0, \sigma_e^2(s) I\{s=t\}\}$, in which $I(\cdot)$ is the indicator function. It is further assumed that $\eta_i(s)$ and $e_i(s)$ are mutually independent, and that $\Sigma_\eta(s,t)$ has continuous partial derivatives of order ν, i.e., $\Sigma_\eta(s,t) \in C^\nu[0,S]^{\otimes 2}$.

Similar to (2.3), $\eta_i(s)$ can also be approximated by Taylor expansion at point s_w with

$$\eta_i(s_w) \approx \eta_i(s) + \sum_{k=1}^{\nu} \partial^k \eta_i(s) \frac{(s_w - s)^k}{k!} := \mathbf{Q}_i(s)^T \boldsymbol{z}(s_w - s), \tag{2.6}$$

where $\mathbf{Q}_i(s) = (\eta_i(s), \partial\eta_i(s), \cdots, \partial^\nu \eta_i(s))^T$. Subsequently, for a given bandwith h_2, the WLS estimate of $\eta_i(s)$ can be obtained from

$$\widehat{\mathbf{Q}}_i(s) = \operatorname{argmin} \sum_{w=1}^{W} [y_i(s_w) - \boldsymbol{x}_i^T \widehat{\mathbf{A}}(s) \boldsymbol{z}(s_w - s) - \mathbf{Q}_i(s)^T \boldsymbol{z}(s_w - s)] K_{h_2}(s_w - s). \tag{2.7}$$

The asymptotic properties of $\widehat{\boldsymbol{\beta}}(s)$ and $\widehat{\eta}_i(s)$ have been carefully studied in the literature [30, 31, 32, 33, 34, 29, 35]. Estimation consistency has been constructed and convergence rates have been derived under mild conditions, which provides theoretical support for the kernel smoothing method.

In (2.4) and (2.7), the choice of bandwidths (h_1, h_2) controls the trade-off between bias and variance of the estimates. Theoretical bounds for this problem have been investigated by previous works [29, 35]. In practice, the optimal bandwidths are usually determined in a data-driven way, for example, by using generalized cross validation (GCV) [30, 36] or Bayesian approaches [37, 38].

2.1.2 Functional Principal Component Analysis

Given a consistent estimator of an individual function $\eta_i(s)$, a question arises to explore variables contributing to the unspecified variation. For example, in GWAS, scientists are interested in searching for mutations associated with $\eta_i(s)$ from millions of genotyped markers. Further dimension reduction is required on $\eta_i(s)$ in order to give an efficient analysis. Functional principal component analysis, which is considered as a generalization of principal component analysis (PCA) to functional data, has been developed for this purpose. The primary goal of functional PCA is to capture the dominant variation pattern of the infinite-dimensional functional data in a low dimensional space.

In this section, we provide an overview of functional PCA by taking individual functions $\{\eta_i(s)\}_{i=1}^n$ as an example. Let $\{\eta_i(s)\}_{i=1}^n$ be independent and identical copies from a zero-mean, square-integrable stochastic process indexed on a closed and bounded intervel $[0,S]$. When the covariance function $\Sigma_\eta(s,t)$ is continuous on $[0,S]^{\otimes 2}$, $\Sigma_\eta(s,t)$ has the following spectral decomposition from Mercer's Theorem

$$\Sigma_\eta(s,t) = \sum_{l=1}^{+\infty} \tau_l \phi_l(s)\phi_l(t), \tag{2.8}$$

where $\{\tau_l\}_{l=1}^{+\infty}$ are nonnegative eigenvalues in descending order that satisfy $\sum_{l=1}^{\infty} \tau_l < \infty$, and $\{\phi_l(s)\}_{l=1}^{+\infty}$ are the corresponding orthonormal eigenfunctions. Then $\eta_i(s)$ admits a functional principal component decomposition given by the Karhunen-Loeve expansion [39, 40]

$$\eta_i(s) = \sum_{l=1}^{+\infty} \xi_{i,l}\phi_l(s), \tag{2.9}$$

where $\{\xi_{i,l}\}_{l=1}^{+\infty}$ are functional principal component (fPC) scores calculated from

$$\xi_{i,l} = \int_0^S \eta_i(s)\phi_l(s)ds. \tag{2.10}$$

The fPC scores $\{\xi_{i,l}\}_{i=1}^{+\infty}$ are mutually uncorrelated random variables that satisfy $\mathbb{E}\xi_{i,l} = 0$ and $\mathbb{E}\xi_{i,l}^2 = \tau_l$.

As given by (2.9), $\eta_i(s)$ can be equivalently represented by a series of uncorrelated univariate random variables $\{\xi_{i,l}\}_{l=1}^{+\infty}$. When $\sum_{l=1}^{\infty}\tau_l < \infty$, the majority of variation can be captured by a finite number of fPCs, i.e., $\eta_i(s) \approx \sum_{l=1}^{L}\xi_{i,l}\phi_l(s)$ when L is large enough. The extracted features $\{\xi_{i,l}\}_{l=1}^{L}$ can then be studied in univariate or multivariate analysis, such as classification [41, 42], regression [43] and prediction [44]. More importantly, when multivariate functional responses are observed, it is of great interest to study multiple functional features comprehensively in a joint model. The heterogeneity in sample domain $\mathscr{S}$ and the potential inter-correlation among various traits are major difficulties of a joint analysis. Functional PCA has provided a strategy to map heterogeneous features to a common coordinate system, which allows us to merge all features together in a unified model.

2.1.3 Statistical Inference for Functional Data

In the varying coefficient model, we are interested in a global hypothesis testing problem of the following general form:

$$H_0 : \mathbf{C}\boldsymbol{\beta}(s) = \boldsymbol{c}(s),\ \forall s \in [0,S] \text{ v.s. } H_1 : \mathbf{C}\boldsymbol{\beta}(s) \neq \boldsymbol{c}(s),\ \exists s \in [0,S], \tag{2.11}$$

where $\mathbf{C}$ is a $k \times p$ matrix of rank k and $\boldsymbol{c}(s) = (c_1(s), \cdots, c_r(s))^T$ is a $k \times 1$ vector of functions. (2.11) covers a wide range of testing problems in applications, including the global genetic association test across $[0, S]$ as a special case. Many statistics have been proposed to test this problem. Here, we introduce two of them, an integration of local statistics from functional analysis of diffusion tensor tract statistics (FADTTS) [29, 36] and an F-type statistics from linear models of functional responses (FLMtest) [45, 46], as examples.

In FADTTS, a local test statistic at each $s \in [0, S]$ is computed as

$$\begin{aligned} T_n(s) &= [\mathbf{C}\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{c}(s)]^T \{\mathbf{C}[\widehat{\Sigma}_\eta(s,s) \otimes (\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^T)^{-1}]\mathbf{C}^T\}^{-1} [\mathbf{C}\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{c}(s)] \\ &:= \widehat{\boldsymbol{d}}(s)^T \{\mathbf{C}[\widehat{\Sigma}_\eta(s,s) \otimes (\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^T)^{-1}]\mathbf{C}^T\}^{-1} \widehat{\boldsymbol{d}}(s), \end{aligned} \tag{2.12}$$

where $\widehat{\boldsymbol{\beta}}(s)$ and $\widehat{\Sigma}_\eta(s,t)$ are consistent estimators of $\boldsymbol{\beta}(s)$ and $\Sigma_\eta(s,t)$ respectively. Then a global statistic can be calculated by an integration of $T_n(s)$ across $[0, S]$, i.e.,

$$S_n = \int_0^S T_n(s) ds. \tag{2.13}$$

The asymptotic distribution of S_n is difficult to derive and a wild bootstrap procedure has been proposed in [36] to estimate the p-value.

In the FLMtest, an F-type statistic has been generalized to functional data as

$$F_n = \frac{\int_0^S \widehat{\boldsymbol{d}}(s)^T [\mathbf{C}(\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^T)^{-1} \mathbf{C}^T]^{-1} \widehat{\boldsymbol{d}}(s) ds / k}{\int_0^S \sum_{i=1}^n [y_i(s) - \boldsymbol{x}_i^T \widetilde{\boldsymbol{\beta}}(s)]^2 ds / (n-p)}, \tag{2.14}$$

in which $\widetilde{\beta}(s) = (\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^T)^{-1} \sum_{i=1}^n \boldsymbol{x}_i y_i(s)$. The null distribution of F_n can be approximated by an F-distribution $F[\kappa k, \kappa(n-p)]$, where κ is a degree-of-freedom adjustment factor that can be estimated from the covariance function of $y_i(s)$.

The above two statistics are easy to compute, and have been widely used in the global test (2.11). A common feature of FADTTS and FLMtest is, when calculating the global test statistics, an underlying "uniform weight" has been assigned to each point in $[0,S]$. It can be expected that neither of them gives optimal power when the signal under test is heterogeneous across $[0,S]$.

2.2 High Dimensional Inference

To study multivariate imaging traits with dependent structure, we consider a multivariate linear model given by

$$\boldsymbol{y}_i = \mathbf{B}^T \boldsymbol{x}_i + \boldsymbol{e}_i, \tag{2.15}$$

where $i = 1, \cdots, n$ is the subject index, $\boldsymbol{y}_i$ is a $q \times 1$ vector of phenotypes, $\boldsymbol{x}_i = (x_{i,1}, \cdots, x_{i,p})^T$ is a $p \times 1$ vector of covariates, $\mathbf{B} = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_q) = (\boldsymbol{\beta}_1^T, \cdots, \boldsymbol{\beta}_p^T)^T$ is a $p \times q$ matrix of regression coefficients, and $\boldsymbol{e}_i$ is a $q \times 1$ vector of error terms such that $\mathbb{E}(\boldsymbol{e}_i) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{e}_i) = \Sigma_e$. Compared with sample size n, it is assumed that q is large and p is small in the above model. To perform association analysis, we are interested in testing genetic effects on all phenotypes simultaneously, which can be formulated by the following problem in general,

$$H_0 : \mathbf{CB} = \mathbf{C}_0 \text{ v.s. } H_1 : \mathbf{CB} \neq \mathbf{C}_0, \tag{2.16}$$

where $\mathbf{C}$ is a $k \times p$ matrix with rank k and $\mathbf{C}_0$ is a $k \times q$ matrix. For “large q, small n” problem, multivariate test statistics tend to be unreliable. As pointed in [47, 48] and many others, traditional methods suffer from substantial power loss even for the simplest testing questions when $q/n \to \infty$.

Although dimension reduction techniques such as principal component analysis (PCA), canonical correlation analysis (CCA) and partial least square regression (PLSR) can be applied, the solutions bear dramatic deviation from the ground truth due to severe noise contamination. A number of regularization methods have been introduced in high dimensional setting by imposing a sparsity assumption [49, 50, 51, 52], yet most of them did not provide a standard inference procedure. Alternatively, some pooled association tests have been proposed to conduct univariate analysis, and to combine marginal statistics in a global test [53, 54]. Among those tests, we will introduce two state-of-the-art methods in detail. Both of them are computationally efficient and have included an adaptive strategy to detect signal at multiple levels.

2.2.1 Two-Sample Tests for High Dimensional Means with Thresholding

In [55], a two-sample test has been proposed in high dimensional setting to study a special case of (2.16), the equality of two sample means.

Let $\{\boldsymbol{z}_{1,1},\cdots,\boldsymbol{z}_{1,i},\cdots,\boldsymbol{z}_{1,n_1}\}$ and $\{\boldsymbol{z}_{2,1},\cdots,\boldsymbol{z}_{2,i'},\cdots,\boldsymbol{z}_{2,n_2}\}$ be two groups of independent and identically distributed (i.i.d) samples from $\mathrm{RV}(\boldsymbol{\mu}_1,\mathit{\Sigma})$ and $\mathrm{RV}(\boldsymbol{\mu}_2,\mathit{\Sigma})$ respectively, where $\boldsymbol{z}_{m,i}=(z_{m,i}^{(1)},\cdots,z_{m,i}^{(q)})^T$ is a $q\times 1$ vector of random variables with $m=1,2$ and $i=1,\cdots,n_m$, $\boldsymbol{\mu}_1=(\mu_{1,1},\cdots,\mu_{1,q})^T$ and $\boldsymbol{\mu}_2=(\mu_{2,1},\cdots,\mu_{2,q})^T$ are $q\times 1$ vectors denoting multivariate means, $\mathit{\Sigma}$ is a $q\times q$ matrix denoting multivariate covariance, and $\mathrm{RV}(\boldsymbol{\mu},\mathit{\Sigma})$ represents multivariate random variables with mean $\boldsymbol{\mu}$ and covariance $\mathit{\Sigma}$. The primary interest is to test the equality of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, i.e.,

$$H_0:\boldsymbol{\mu}_1=\boldsymbol{\mu}_2 \text{ v.s. } H_1:\boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2. \tag{2.17}$$

For each dimension j in $\{1, \cdots, q\}$, a U-statistic that test marginal hypothesis

$$H_{j,0} : \mu_{1,j} = \mu_{2,j} \text{ v.s. } H_{j,1} : \mu_{1,j} \neq \mu_{2,j}, \tag{2.18}$$

is calculated as

$$T_{n,j} = \frac{1}{n_1(n_1-1)} \sum_{i \neq i'}^{n_1} z_{1,i}^{(j)} z_{1,i'}^{(j)} + \frac{1}{n_2(n_2-1)} \sum_{i \neq i'}^{n_2} z_{2,i}^{(j)} z_{2,i'}^{(j)} - \frac{2}{n_1 n_2} \sum_{i}^{n_1} \sum_{i'}^{n_2} z_{1,i}^{(j)} z_{2,i'}^{(j)}. \tag{2.19}$$

A thresholding test is then proposed as follows:

$$L_{CQT}(\lambda_n) = n \sum_{j=1}^{q} T_{n,j} I\{nT_{n,j} > \lambda_n\}, \tag{2.20}$$

where $n = \frac{n_1 n_2}{n_1 + n_2}$.

When marginal null hypothesis $H_{j,0}$ holds for a large number of dimensions in $\{1, \cdots, q\}$, it has been theoretically demonstrated in [55] that an appropriate λ_n can substantially boost statistical power by reducing the noise level introduced from non-signal dimensions. To deal with the unknown signal density, a multi-level inference procedure is further proposed to adopt different thresholds and a data-driven strategy is introduce to choose λ_n, which is given by

$$\widehat{\lambda}_n = \max_{\lambda} \frac{L_{CQT}(\lambda) - \mathbb{E} L_{CQT}(\lambda)}{\sqrt{\mathbb{E} L_{CQT}^2(\lambda) - \mathbb{E}^2 L_{CQT}(\lambda)}}. \tag{2.21}$$

When the off-diagonal elements of Σ is sparse, along with certain mild conditions, it is proved that $L_{CQT}(\lambda)$ has asymptotic normal distribution. And (2.21) is selecting λ_n that maximizes the asymptotic power.

2.2.2 Adaptive Sum of Powered Score (aSPU) Test

For each $j = 1, \cdots, q$, let U_j be a univariate statistic designed for marginal hypothesis problem

$$H_{j,0} : \mathbf{C}\boldsymbol{b}_j = \mathbf{c}_{0,j} \text{ v.s. } H_{j,1} : \mathbf{C}\boldsymbol{b}_j \neq \mathbf{c}_{0,j}, \tag{2.22}$$

where $\mathbf{C}$ is the $k \times p$ matrix defined by (2.16), $\boldsymbol{b}_j$ is the j-th column of coefficient matrix $\mathbf{B}$ and $\mathbf{c}_{0,j}$ is the j-th row of matrix $\mathbf{C}_0$. In [56], a sum of powered score (SPU) test statistic has been proposed for a given positive integer ν,

$$T(\nu) = \sum_{j=1}^{q} U_j^{\nu}. \tag{2.23}$$

With different choices of power index ν, SPU form a class of statistics that is able to detect flexible signal patterns. When $\nu = 1$, SPU is the analog of burden test that assess the cumulative effect of multiple weak signals. As ν increases, $T(\nu)$ put larger weights on sharp signals. In an extreme case when $\nu \to +\infty$, $T(\nu)$ is equivalent to the supremum statistic, i.e., $\max_{1 \leq j \leq q} U_j$. An optimal choice of ν depends on the underlying signal pattern under test. And a data-driven strategy has been introduced in [56] to determine ν. Specifically, the p-value of each $T(\nu)$, denoted as $P(\nu)$, is estimated from permutation or bootstrap resampling. Then the minimum p-value is used as an adaptive test score, i.e.,

$$T_{\text{aSPU}} = \min_{\nu} P(\nu). \tag{2.24}$$

Intuitively, the above equation is selecting ν that gives the largest statistical power. Since T_{aSPU} is considered as a test statistic rather than a genuine p-value, resampling is also required to approximate its null distribution in order to control type I error. Instead of running double permutation or bootstrap, the p-value of T_{aSPU} can be obtained in the same procedure when calculating $P(\nu)$s.

2.2.3 Projection Regression Model and Heritability Ratio

Most of the existing tests in high dimensional inference are derived from independent responses, such as the two methods mentioned above. The correlation within $\boldsymbol{y}_i$ is either ignored or handled inappropriately. For example, in presence of high correlations, one commonly used strategy is to transform the data by the precision matrix Σ_e^{-1} [55, 57, 58], in which model (2.15) can be written as

$$\boldsymbol{y}_{\Sigma_e^{-1},i} = \Sigma_e^{-1}\boldsymbol{y}_i = \mathbf{B}_{\Sigma_e^{-1}}^T \boldsymbol{x}_i + \boldsymbol{e}_{\Sigma_e^{-1},i}, \tag{2.25}$$

where $\mathbf{B}_{\Sigma_e^{-1}} = \mathbf{B}\Sigma_e^{-1}$ and $\boldsymbol{e}_{\Sigma_e^{-1},i} = \Sigma_e^{-1}\boldsymbol{e}_i$ with $\mathbb{E}(\boldsymbol{e}_{\Sigma_e^{-1},i}) = \mathbf{o}$ and $\mathrm{Cov}(\boldsymbol{e}_{\Sigma_e^{-1},i}) = \Sigma_e^{-1}$. Precision matrix transformation is expected to increase statistical power when the signal is sparse. However, the structure of $\mathbf{B}$ has been ignored in the transformation. In some cases, it may cause severe power loss. A toy example below can clearly demonstrate this problem.

Example 2.2.1. *Consider a special case of (2.15) that $q \geq 2$, $\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{e}_i$, in which $\boldsymbol{\mu} = (\mu_1, \mu_2, \boldsymbol{\mu}_0)^T$, where μ_1, μ_2 are scalar values and $\boldsymbol{\mu}_0$ is a $(q-2)\times 1$ vector, and $\Sigma_e = \left(\begin{smallmatrix} \Sigma_{e,1} & 0 \\ 0 & \Sigma_{e,2} \end{smallmatrix}\right)$ with $\Sigma_{e,1} = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$ and $\Sigma_{e,2}$ being arbitrary $(q-2)\times(q-2)$ covariance matrix. We test a simple zero-mean hypothesis problem, i.e., $H_0 : \boldsymbol{\mu} = 0$ v.s. $H_1 : \boldsymbol{\mu} \neq 0$. Particularly, we focus on the signal-to-noise ratios (SNRs) of the first two dimensions. In the original space, SNRs of $y_{i,1}$ and $y_{i,2}$ can be give by*

$$SNR(y_{i,1}) = |\mu_1| \text{ and } SNR(y_{i,2}) = |\mu_2|.$$

By applying precision matrix transformation, the signal-to-noise ratios become

$$SNR(y_{i,1}^*) = \frac{|\mu_1 - \rho\mu_2|}{\sqrt{1-\rho^2}} \text{ and } SNR(y_{i,2}^*) = \frac{|\mu_2 - \rho\mu_1|}{\sqrt{1-\rho^2}},$$

where $y^*_{i,1} = (y_{i,1} - \rho y_{i,2})/(1-\rho^2)$ *and* $y^*_{i,2} = (y_{i,2} - \rho y_{i,1})/(1-\rho^2)$. *Then we consider the following two signal patterns:*

- *(i) when* $\rho \neq 0$, $\mu_1 \neq 0$ *and* $\mu_2 = 0$, *we have* $SNR(y^*_{i,1})/SNR(y_{i,1}) = \sqrt{1-\rho^2}$. *The signal-to-noise ratio increases substantially when the correlation between* $y_{i,1}$ *and* $y_{i,2}$ *is large.*
- *(ii) when* $\rho \to 1$ *and* $\mu_1 = \mu_2 = \mu$, *we have* $SNR(y^*_{i,1})/SNR(y_{i,1}) = SNR(y^*_{i,2})/SNR(y_{i,2})$ $= \sqrt{1-\rho}/\sqrt{1+\rho} \to 0$. *The signal-to-noise ratio is reducing dramatically.*

For most of the existing methods using the precision matrix transformation, it is commonly assumed that both the signals and the off-diagonal elements of Σ_e^{-1} are sparse, and that the active dimensions are randomly distributed in $\{1, \cdots, q\}$. Such assumptions naturally implicate that a signal pattern similiar to case (i) holds with large probability, which is hard to verify in real data. Therefore, the precision matrix transformation does not guarantee to increase power in general. We should seek for other strategies to deal with covariance structure that take into account the signal pattern appropriately in the data.

To properly address the signal-covariance structure in $\boldsymbol{y}_i$, a projection regression model (PRM) on (2.15) has been introduced by [59] as

$$\mathbf{w}^T \boldsymbol{y}_i \triangleq \boldsymbol{y}_{\mathbf{w},i} = \mathbf{B}_{\mathbf{w}}^T \boldsymbol{x}_i + \boldsymbol{e}_{\mathbf{w},i}, \tag{2.26}$$

where $\mathbf{w}$ is a $q \times 1$ vector of linear projection direction, $\mathbf{B_w} = \mathbf{Bw}$ is the transformed regression coefficient matrix and $\boldsymbol{e}_{\mathbf{w},i} = \mathbf{w}^T \boldsymbol{e}_i$ is the tranformed error term with $E(\boldsymbol{e}_{\mathbf{w},i}) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{e}_{\mathbf{w},i}) = \mathbf{w}^T \Sigma_e \mathbf{w}$. To illustrate the key idea of [59], we consider a simplified unit-rank hypothesis question, given as follows:

$$H_0 : \boldsymbol{\beta}_1 = \mathbf{0} \text{ v.s. } H_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}, \tag{2.27}$$

where $\boldsymbol{\beta}_1$ is a $q \times 1$ vector composed of the first row from $\mathbf{B}$. In the projection regression model (2.26), the testing problem in the projected space is

$$H_{w,0} : \boldsymbol{\beta}_{\mathbf{w},1} = 0 \text{ v.s. } H_{w,1} : \boldsymbol{\beta}_{\mathbf{w},1} \neq 0, \tag{2.28}$$

where $\boldsymbol{\beta}_{\mathbf{w},1} = \mathbf{w}^T \boldsymbol{\beta}_1$. Given a projection direction $\mathbf{w}$, univariate test can be directly applied for the above problem.

One remaining question for PRM now is how to determine the projection direction $\mathbf{w}$ that achieves the best statistical power. In [59], a generalized heritability ratio has been introduced for this purpose. Specifically, the signal-to-noise ratio of $\mathbf{w}^T \boldsymbol{y}_i$ can be given as

$$\mathrm{SNR}_i = \frac{|\mathbf{w}^T \boldsymbol{\beta}_1 x_{i,1}|}{\sqrt{\mathbf{w}^T \Sigma_e \mathbf{w}}}. \tag{2.29}$$

A generalized heritability ratio is introduced as the average of SNR_i^2s across all subjects, i.e.,

$$\mathrm{GHR}(\mathbf{w}) = n^{-1} \sum_{i=1}^{n} \mathrm{SNR}_i^2 = \frac{(\mathbf{w}^T \boldsymbol{\beta}_1)^2}{n \mathbf{w}^T \Sigma_e \mathbf{w}} \sum_{i=1}^{n} x_{i,1}^2 \xrightarrow{p} \frac{(\mathbf{w}^T \boldsymbol{\beta}_1)^2 \sigma_{x_1}^2}{\mathbf{w}^T \Sigma_e \mathbf{w}}. \tag{2.30}$$

where $\sigma_{x_1}^2 = \mathbb{E} x_{i,1}^2$. $\mathrm{GHR}(\mathbf{w})$ is expected to dominate the asymptotic power of testing problem (2.27) under H_1. Therefore, an oracle projection direction is proposed as

$$\mathbf{w}^* = \underset{\mathbf{w} \in \mathbb{R}^q}{\operatorname{argmax}} \, \mathrm{GHR}(\mathbf{w}) = \underset{\mathbf{w} \in \mathbb{R}^q}{\operatorname{argmax}} \frac{(\mathbf{w}^T \boldsymbol{\beta}_1)^2}{\mathbf{w}^T \Sigma_e \mathbf{w}} \propto \Sigma_e^{-1} \boldsymbol{\beta}_1. \tag{2.31}$$

To estimate $\mathbf{w}$, we need to obtain valid estimators of Σ_e^{-1} and $\boldsymbol{\beta}_1$, which is a challenging issue in high dimensional setting. To solve this problem, [59] introduced an L_1 penalty to impose sparsity on $\mathbf{w}$. Specifically, let $\widehat{\Sigma}_e$ and $\widehat{\boldsymbol{\beta}}_1$ be two regularized estimators of Σ_e and $\boldsymbol{\beta}_1$, $\widehat{\mathbf{w}}$ is calculated as

$$\widehat{\mathbf{w}} = \operatorname{argmin} \frac{1}{2}\mathbf{w}^T \widehat{\Sigma}_e \mathbf{w} - \mathbf{w}^T \boldsymbol{\beta}_1 + \lambda \|\mathbf{w}\|_1, \tag{2.32}$$

where $\|\cdot\|_1$ denote the L_1 norm. After $\widehat{\mathbf{w}}$ is obtained, standard test statistics, such as the wald test statistic, can be used on $y_{\widehat{\mathbf{w}},i}$ for problem (2.28).

In the presence of high correlations and sparse signals, [59] achieves better statistical power than some regularized methods designed for independent responses. However, there are two major problems. First of all, the generalized heritability ratio defined by (2.30) is problematic. Covariance between $x_{i,1}$ and other covariates has been ignored, which might potentially introduce bias in the test statistics. Also, the proposed method shows no advantage when dealing with weak effect compared with aSPU and CQT. A flexible test procedure should be considered to detect signals at multiple levels.

2.3 Latent Factor Model and Parallel Analysis

2.3.1 Latent Factor Model

Latent factor model provides a useful perspective to understand multivariate responses and time series data. It allows us to characterize the correlation structure of a large number of variables on lower dimensions. For multivariate observations $\boldsymbol{y}_i = (y_{i,1}, \cdots, y_{i,j}, \cdots, y_{i,q})^T$ with zero means, a latent factor model is given by

$$y_{i,j} = \boldsymbol{\lambda}_j^T \boldsymbol{f}_i + u_{i,j}, \text{ for } j = 1, \cdots, q, \tag{2.33}$$

where $\boldsymbol{f}_i$ is an $r \times 1$ vector of common factors, $\boldsymbol{\lambda}_j$ is an $r \times 1$ vector of factor loadings for the j-th response, and $u_{i,j}$ is the error term uncorrelated with $\boldsymbol{f}_i$. Then $\boldsymbol{y}_i$ can be modeled as

$$\boldsymbol{y}_i = \Lambda \boldsymbol{f}_i + \boldsymbol{u}_i, \tag{2.34}$$

where $\Lambda = (\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_q)^T$ and $\boldsymbol{u}_i = (u_{i,1}, \cdots, u_{i,q})^T$. Let Σ_f and Σ_u denote the covariance matrix of $\boldsymbol{f}_i$ and $\boldsymbol{u}_i$ respectively, the covariance of $\boldsymbol{y}_i$ can be written as

$$\Sigma_y = \Lambda \Sigma_f \Lambda^T + \Sigma_u. \tag{2.35}$$

With the above equation, we are decomposing the covariance of $\boldsymbol{y}_i$ into two parts, a common factor component $\Lambda \Sigma_f \Lambda^T$ and an idiosyncratic component Σ_u. Generally, decomposition (2.35) is not identifiable, since only Σ_y can be estimated from observations. However, when $\Lambda \Sigma_f \Lambda^T$ has diverging eigenvalues relative to Σ_u, the common factor component can be recovered from Σ_y using principal component analysis. Let $\{\tau_j\}_{j=1}^r$ be the first r eigenvalues of Σ_y in decreasing order and $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_r)$ be a $q \times r$ matrix composed of the corresponding eigenvectors, let $\{\tau_{f,j}\}_{j=1}^r$ be the eigenvalues of $\Lambda \Sigma_f \Lambda^T$ in decreasing order and let the columns of $\mathbf{V}_f = (\mathbf{v}_{f,1}, \cdots, \mathbf{v}_{f,r})$ be the corresponding eigenvectors, the following conclusion can be proved using the Weyl's Theorem and the $\sin\theta$ Theorem [60]:

Proposition 2.3.1. *Assume the following conditions hold:*

$$\tau_{f,1} \geq \cdots \geq \tau_{f,r} > q\varepsilon_0 \textit{ and } \|\Sigma_u\|_2 < c_1 < +\infty, \tag{2.36}$$

where ε_0, c_1 are fixed positive values. As q is large enough and r is fixed, we have

$$\frac{1}{q}\max_{1\leq j\leq r}|\tau_j - \tau_{f,j}| \leq \|\Sigma_u\|_2/q = O_p(q^{-1}). \tag{2.37}$$

When $\{\tau_{f,j}\}_{j=1}^r$ are distinct eigenvalues that satisfy $\frac{1}{q}\min_{1\leq j\leq r-1}|\tau_{f,j} - \tau_{f,j+1}| > \varepsilon_1 > 0$, we have

$$\max_{1\leq j\leq r}\|\mathbf{v}_j - \mathbf{v}_{f,j}\|_2 = O_p(q^{-1}). \tag{2.38}$$

Under regularized condition that $\Sigma_f = I_r$ and $\Lambda^T\Lambda = diag\{\tau_{f,1}, \cdots, \tau_{f,r}\}$, it can be proved that

$$\|\boldsymbol{f}_i - T^{-1/2}\mathbf{V}^T\boldsymbol{y}_i\|_2 = O_p(q^{-1/2}). \tag{2.39}$$

The above statement implies that the common factors can be accurately recovered as $q \to +\infty$. The assumption that $\{\tau_{f,j}\}_{j=1}^r$ are distinct eigenvalues can be weakened to allow multiplicity greater than one, and $\boldsymbol{f}_i$ can be recovered up to an orthogonal rotation. The key assumption given by (2.36) is known as the pervasiveness assumption in [61]. It requires that the variation in $\mathbf{y}_i$ explained by any non-negligible proportion of common factors should grow at the rate of $O(q)$.

2.3.2 Parallel Analysis

To solve the latent factor model (2.33), determining the number of common factors is a critical issue. Parallel analysis has been considered as a popular method for this purpose using permutation strategy [62]. In the framework of parallel analysis, it is further assumed that individual factors

$\{u_{i,j}\}_{j=1}^{q}$ are mutually uncorrelated. Suppose we have n observations denoted as

$$\mathbf{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_i, \cdots, \boldsymbol{y}_n]^T = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,q} \\ \vdots & \vdots & \vdots & \vdots, \\ y_{n,1} & y_{n,2} & \cdots & y_{n,q} \end{bmatrix}. \tag{2.40}$$

Let $S_n = \frac{1}{n}\mathbf{Y}^T\mathbf{Y}$ be the sample covariance matrix and let $\{\widehat{\tau}_j\}_{j=1}^{\min\{n,q\}}$ be the eigenvalues of S_n in decreasing order. To determine the number of significant factors, we simulate the distribution of eigenvalues under null hypothesis, i.e., Σ_y is diagonal matrix and all correlations equal to 0, as follows:

Algorithm 2.3.1.

(i) For each permutation replicate g, consider $q-1$ random permutations of $\mathscr{N} = \{1, \cdots, n\}$, denoted as $\boldsymbol{\pi}_j^{(g)} = [\pi_j^{(g)}(1), \pi_j^{(g)}(2), \cdots, \pi_j^{(g)}(n)]$ for each $j = 2, \cdots, q$, the permuted variables are generated as

$$\mathbf{Y}_{\pi^{(g)}} = \begin{bmatrix} y_{1,1} & y_{\pi_2^{(g)}(1),2} & \cdots & y_{\pi_q^{(g)}(1),q} \\ \vdots & \vdots & \vdots & \vdots, \\ y_{n,1} & y_{\pi_2^{(g)}(n),2} & \cdots & y_{\pi_q^{(g)}(n),q} \end{bmatrix}. \tag{2.41}$$

(ii) Calculate the eigenvalues of the sample covariance matrix of $\mathbf{Y}_{\pi^{(g)}}$, which are denoted as $\{\widehat{\tau}_{\pi^{(g)},j}\}_{j=1}^{\min\{n,q\}}$.

(iii) Repeat steps (i) and (ii) G times and calculate the p-value for each eigenvalue as

$$p(\widehat{\tau}_j) = \frac{1}{G}\sum_{g=1}^{G} I\{\widehat{\tau}_{\pi^{(g)},j} \geq \widehat{\tau}_j\}. \tag{2.42}$$

(iv) Given a p-value cutoff p_0, the number of significant factors is chose as $\hat{r}(p_0) = \#\{j : 1 \leq j \leq q, \max_{1 \leq j' \leq j} p(\widehat{\tau}_{j'}) \leq p_0\}$, where $\#\{\cdot\}$ denotes the cardinality of a given set.

Extensive numerical experiments have shown that parallel analysis is superior to many standard methods [63, 64], such as Kaiser′s 1-cutoff [65] and Bartlett′s test [66]. When $n, q \to \infty$ and $q/n \to \varepsilon > 0$, there are some theoretical results justifying the consistency of parallel analysis in multiple scenarios [67], including the case when $\Lambda \Sigma_f \Lambda^T$ has spiked eigenvalues.

CHAPTER 3: HIERARCHICAL FUNCTIONAL PRINCIPAL REGRESSION MODEL FOR DIFFUSION TENSOR BUNDLE STATISTICS

3.1 Introduction

Scientifically, investigation in the connectional organization of human brain and its variation across subjects and time is a critical step for the understanding of pathology of many neuro-related disorders [68, 69], such as autism, schizophrenia and bipolar disorder. Diffusion-weighted MRI (dMRI) offers a non-invasive approach to study the underlying tissue structure of brain white matter *in vivo*, including both geometric shape and diffusion properties [70, 71, 72, 73, 74, 75, 76, 77, 78]. Delineating fiber bundle statistics may help identify structural connectivity abnormalities across different spatial and temporal scales in many neuro-related disorders. It could eventually inspire new approaches for disease preventions, diagnoses, and clinical treatments.

Group analysis of fiber bundle statistics poses remarkable computational and mathematical challenges to existing statistical methods. The first challenge is to efficiently and simultaneously study multiple fiber bundles with heterogeneous geometric structures and variation patterns. The second challenge is to correlate fiber bundle statistics with a large number of covariates, such as millions of genetic markers. This challenge is motivated by the demand to carry out a genome-wide association study. Voxel-wise methods [71] and single tract analysis [76, 29, 79] suffer from performing massive multiple comparison adjustments, which would severely reduce the detection power. The third challenge is to properly handle the potential correlation among multiple tracts and to disentangle common variation shared by a large portion of fiber bundles.

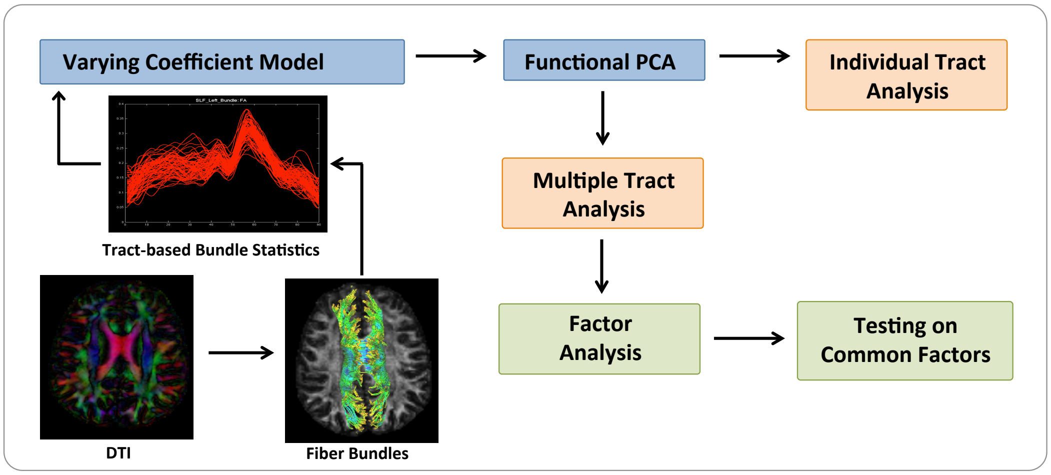

Varying Coefficient Model
Functional PCA
Individual Tract Analysis
Multiple Tract Analysis
Tract-based Bundle Statistics
Factor Analysis
Testing on Common Factors
DTI
Fiber Bundles

Figure 3.1: A schematic overview of the hierarchical functional principal regression model (HFPRM)

In this chapter, we propose a hierarchical functional principal regression model (HFPRM) framework to address the three challenges discussed above. The HFPRM consists of three statistical methods, including a varying coefficient model (VCM), a latent factor analysis (LFA) procedure, and a multivariate regression model (MRM). The path diagram of HFPRM is displayed in Figure 3.1. The VCM not only captures the functional spatial feature of the fiber bundle statistics, but also maps the heterogeneous geometric structure onto a common coordinate system. The LFA is applied to characterize potential inter-tract correlation across multiple fibers. It allows us to explicitly extract common latent features on a lower dimension. The integration of VCM and LFA can dramatically reduce the dimension of fiber bundle statistics of multiple tracts. Finally, by using MRM, we are able to examine the effect of interest and to perform statistical inference on the extracted common features.

In Section 3.2, we introduce the general framework of HFPRM and propose a two-stage estimation-testing procedure to study common features. In Sections 3.4 and 3.5, we use simulation studies and an imaging genetic example to examine the finite sample performance of HFPRM. Section 3.6 concludes with some remarks.

3.2 Methods

3.2.1 Data Structure

In a typical DTI study, we observe functional diffusion properties, such as fractional anisotropty, mean diffusivity and axial diffusivity, along M fiber bundles, some clinical variables as well as genetic variants for n subjects. On the m-th fiber bundle where $m = 1, \cdots, M$, let $s_m \in [0, S_m]$ denotes the arc length of any point relative to a fixed end point, where S_m is the longest tract arc length. For the $i-$th subject where $i = 1, \cdots, n$, functional diffusion property $y_{i,m}(s_m)$ is observed at the point with arc-length s_m. And $\boldsymbol{x}_i$ is a $p \times 1$ vector of variables including demographic variables, clinical biomarkers and genetic variants.

3.2.2 Model Formulation and Problem of Interest

To delineate the association between observed variables $\boldsymbol{x}_i$ and functional response $y_{i,m}(s_m)$ on a specific tract m, the following varying coefficient model has been widely used,

$$y_{i,m}(s_m) = \mu_m(s_m) + \boldsymbol{x}_i^T \boldsymbol{\beta}_m(s_m) + \eta_{i,m}(s_m) + e_{i,m}(s_m), \tag{3.1}$$

where $\mu_m(s_m)$ is the mean function, $\boldsymbol{\beta}_m(s_m) = (\beta_{m,1}(s_m), \cdots, \beta_{m,p}(s_m))^T$ is a $p \times 1$ vector of functions representing covariate effects, $\eta_{i,m}(s_m)$ characterizes spatial variation that cannot be explained by $\boldsymbol{x}_i$ and $e_{i,m}(s_m)$ is the measurement error. Furthermore, $\{\eta_{i,m}(s_m)\}_{i=1}^n$ and $\{e_{i,m}(s_m)\}_{i=1}^n$ are assumed to be i.i.d copies from stochastic processes $\mathrm{SP}\{0, \Sigma_{\eta_m}(s_m, t_m)\}$ and $\mathrm{SP}\{0, \sigma_\varepsilon^2(s) I(s_m = t_m)\}$ respectively, in which $\mathrm{SP}\{\mu(s), \Sigma(s,t)\}$ denotes a stochastic process with mean function $\mu(s)$ and covariance function $\Sigma(s,t)$, and $I(\cdot)$ is the indicator function. It is also assumed that $\eta_{i,m}(s_m)$ and $e_{i,m}(s_m)$ are mutually independent.

In the above varying coefficient model, the primary question we are interested in is to identify genetic variants associated with diffusion properties from all available tracts, which can be formulated as the following hypothesis testing problem in general:

$$H_0 : \mathbf{C}\boldsymbol{\beta}_m(s_m) = \mathbf{0},\ \forall m = 1, \cdots, M \text{ v.s. } H_1 : \mathbf{C}\boldsymbol{\beta}_m(s_m) \neq \mathbf{0},\ \exists m \in \{1, \cdots, M\}, \tag{3.2}$$

where $\mathbf{C}$ is a $k \times p$ matrix with rank k.

Most testing methods in the literatures focus on individual tract, such as FADTTS and FLMtest introduced in Section 2.1.3. A global test is performed using massive multiple comparison adjustment, which tends to be too conservative. More importantly, these methods usually ignore the potential inter-correlations among different tracts. Such correlations can be helpful to increase statistical power in a joint analysis [80]. Also, directly performing statistical inference on the varying coefficient model (3.1) often requires large number of resampling in order to estimate the p-values, which is very time consuming for GWAS problem.

Addressing these issues requires the development of a robust and efficient dimensional reduction and testing framework on functional traits from multiple tracts. However, a joint analysis of multiple tracts is nontrivial. One major difficulty is how to appropriately account for the between-bundle correlations. In addition, the heterogeneity of different fiber bundles in geometric properties, such as length, curvature and sampled grid points, makes it more difficult to include all tracts in a unified model. Therefore, we first perform a dimension reduction procedure on each individual tract and the aim is to extract some key features for further analysis.

3.2.3 Dimension Reduction through Functional Principal Component Analysis (fPCA)

To perform dimension reduction only, we focus on the following varying coefficient model without specifying any fixed effect,

$$y_{i,m}(s_m) = \mu_m(s_m) + \tilde{\eta}_{i,m}(s_m) + e_{i,m}(s_m). \tag{3.3}$$

Compared with (3.1), $\tilde{\eta}_{i,m}(s_m)$ represents spatial variation introduced by both $\boldsymbol{x}_i$ and $\eta_{i,m}(s_m)$, i.e., $\tilde{\eta}_{i,m}(s_m) = \boldsymbol{x}_i^T \boldsymbol{\beta}_m(s_m) + \eta_{i,m}(s_m)$. When $\boldsymbol{x}_i$ are mean-zero random variables with covariance Σ_x and are independent from $\eta_{i,m}(s_m)$ and $e_{i,m}(s_m)$, $\tilde{\eta}_{i,m}(s_m)$ is a sample from stochastic process $\mathrm{SP}\{0, \Sigma_{\tilde{\eta}_m}(s_m,t_m)\}$, in which $\Sigma_{\tilde{\eta}_m}(s_m,t_m)$ is the covariance of $\tilde{\eta}_{im}(s_m)$ given by

$$\Sigma_{\tilde{\eta}_m}(s_m,t_m) = \Sigma_{\eta_m}(s_m,t_m) + \boldsymbol{\beta}_m(s_m)^T \Sigma_x \boldsymbol{\beta}_m(t_m), \tag{3.4}$$

Since $\tilde{\eta}_{i,m}(s_m)$ incorporates all the variations of interest, our primary goal is to extract important features from it. Functional principal component analysis introduced in Section 2.1.2 is adopted here to reduce the dimension of $\tilde{\eta}_{i,m}(s_m)$.

Let $\Sigma_{\tilde{\eta}_m}(s_m,t_m)$ be a continuous covariance function on $[0,S_m]^{\otimes 2}$, Mercer's theorem suggests the following eigen-decomposition:

$$\Sigma_{\tilde{\eta}_m}(s_m,t_m) = \sum_{l=1}^{+\infty} \tau_{m,l} \phi_{m,l}(s_m) \phi_{m,l}(t_m), \tag{3.5}$$

where $\{\phi_{m,l}(s_m)\}_{l=1}^{+\infty}$ are orthonormal eigenfunctions in $L^2[0,S_m]$ that correspond to eigenvalue sequence $\{\tau_{m,l}\}_{l=1}^{+\infty}$ in decreasing order. Given (3.5), $\tilde{\eta}_{i,m}(s_m)$ admits Karhunen-Loeve expansion as

$$\tilde{\eta}_{i,m}(s_m) = \sum_{l=1}^{+\infty} \xi_{i,ml}\phi_{m,l}(s_m), \tag{3.6}$$

where $\xi_{i,ml} = \int_0^{S_m} \tilde{\eta}_{i,m}(s_m)\phi_{m,l}(s_m)ds_m$ is the l-th functional principal component score of subject i on tract m, and $\{\xi_{i,ml}\}_{l=1}^{+\infty}$ are mutually uncorrelated variables with mean zero and variances $\{\tau_{m,l}\}_{l=1}^{+\infty}$.

Through functional PCA, each individual function $\tilde{\eta}_{i,m}(s_m)$ can be equivalently represented by a sequence of functional PC scores $\{\xi_{i,ml}\}_{l=1}^{+\infty}$. When $\sum_{l=1}^{+\infty}\tau_{m,l} < +\infty$ and $\{\tau_{m,l}\}_{l=1}^{+\infty}$ are decreasing quickly, a relatively small number of fPCs would be enough to account for the majority of variation in $\tilde{\eta}_{i,m}(s)$. In other words, we are able to approximate $\tilde{\eta}_{i,m}(s)$ through a finite fPC vector with dimension L_n, i.e., $\boldsymbol{\xi}_{i,m} = (\xi_{i,m1},\ldots,\xi_{i,mL_n})^T \in \mathbb{R}^{L_n}$. For notational simplicity, L_n is assumed to be the same across all M bundles. The choice of L_n depends on both sample size n and eigenvalue sequence $\{\tau_{m,l}\}_{l=1}^{+\infty}$. There are several *ad hoc* procedures to determine L_n. An analog of model selection techniques have been generalized for this purpose, such as Akaike information criterion (AIC), Bayesian information criterion (BIC) [43] and cross-validation (CV). In practice, the percentage of explained variation has been widely used as an appropriate cut-off. For the rest of this section, we assume that the dimension of selected features L_n has been fixed.

Functional PCA not only extracts low dimensional feature from $\tilde{\eta}_{i,m}(s_m)$, but also maps the heterogeneous geometric structure onto a common coordinate system. It allows us to merge all extracted features in a joint model. Specifically, we can get the following multivariate linear model from (3.1):

$$\begin{aligned} \boldsymbol{\xi}_{i,m} &= \boldsymbol{x}_i^T \mathbf{b}_m + \boldsymbol{\delta}_{i,m}, & (3.7) \\ \boldsymbol{\xi}_i &= \boldsymbol{x}_i^T \mathbf{B} + \boldsymbol{\delta}_i, & (3.8) \end{aligned}$$

where $\mathbf{B} = (\mathbf{b}_1, \cdots, \mathbf{b}_M)$ with $\mathbf{b}_m = \int_0^{S_m} \boldsymbol{\beta}_m(s_m)\boldsymbol{\Phi}_m(s_m)ds_m$, $\boldsymbol{\Phi}_m(s_m) = [\phi_{m,1}(s_m), \cdots, \phi_{m,L_n}(s_m)]$, and $\boldsymbol{\delta}_i = (\delta_{i,1}, \cdots, \delta_{i,M})^T$ with $\boldsymbol{\delta}_{i,m} = \int_0^{S_m} \eta_{i,m}(s_m)\boldsymbol{\Phi}_m(s_m)ds_m$.

Correspondingly, the testing question (3.2) lead to the following problem,

$$H_{0,\xi} : \mathbf{CB} = \mathbf{0} \text{ v.s. } H_{1,\xi} : \mathbf{CB} \neq \mathbf{0}. \tag{3.9}$$

When M is large and L_n goes to infinity, direct analysis on (3.8) and (3.9) encounters the challenge of high dimensionality. Therefore, further dimension reduction is required.

3.2.4 Latent Factor Model

Diffusion tensor properties from different tracts are known to have strong correlation that cannot be ignored. To characterize such dependency, we assume a latent factor structure on $\{\tilde{\eta}_{i,m}(s_m)\}_{m=1}^M$, which is given as

$$\tilde{\eta}_{i,m}(s_m) = \boldsymbol{f}_{c,i}^T \gamma_m(s_m) + u_{i,m}(s_m), \tag{3.10}$$

where $\boldsymbol{f}_{c,i}$ is an $r \times 1$ vector of common latent factors that contribute to the variation in multiple tracts, $\gamma_m(s_m) = (\gamma_{m,1}(s_m), \cdots, \gamma_{m,r}(s_m))^T$ is composed of the functional loading coefficients, and $u_{i,m}(s_m)$ represents tract-specific variation of bundle m that is uncorrlated with $\boldsymbol{f}_{c,i}$. It is further assumed that $\{u_{i,m}(s_m)\}_{m=1}^M$ are mutually uncorrelated among different tracts. $y_{i,m}(s_m)$ in (3.3) can then be rewritten as

$$y_{i,m}(s_m) = \mu_m(s_m) + \boldsymbol{f}_{c,i}^T \gamma_m(s_m) + u_{i,m}(s_m) + e_{i,m}(s_m). \tag{3.11}$$

Given the above formulation, the extracted features $\boldsymbol{\xi}_i$ can be also expressed with a latent factor models as

$$\begin{aligned} \boldsymbol{\xi}_{i,m} &= \boldsymbol{\lambda}_m^T \boldsymbol{f}_{c,i} + \boldsymbol{u}_{i,m}, & (3.12) \\ \boldsymbol{\xi}_i &= \Lambda \boldsymbol{f}_{c,i} + \boldsymbol{u}_i, & (3.13) \end{aligned}$$

where $\Lambda = (\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_M)^T$ with $\boldsymbol{\lambda}_m = \int_0^{S_m} \gamma_m(s_m) \boldsymbol{\Phi}_m(s_m) ds_m$ and $\boldsymbol{u}_i = (\boldsymbol{u}_{i,1}, \cdots, \boldsymbol{u}_{i,M})^T$ with $\boldsymbol{u}_{i,m} = \int_0^{S_m} u_{i,m}(s_m) \boldsymbol{\Phi}_m(s_m) ds_m$. Without loss of generality, it is assumed that $\mathrm{Cov}(\boldsymbol{f}_{c,i}) = I_r$. As discussed in Section 2.3, the common factors can be recovered through principal component analysis when M is large enough and a pervasiveness assumption (2.36) is satisfied.

3.2.5 Multivariate Linear Model on Common Factors

Finally, when the common factors $\boldsymbol{f}_{c,i}$ are obtained, the following multivariate linear regression is applied:

$$\boldsymbol{f}_{c,i} = \boldsymbol{x}_i^T \mathbf{B}_c + \boldsymbol{\delta}_{c,i}. \tag{3.14}$$

To study the hypothesis of interest on $\boldsymbol{f}_{c,i}$, we consider testing problem

$$H_{c,0} : \mathbf{C}\mathbf{B}_c = \mathbf{0} \text{ v.s. } H_{c,1} : \mathbf{C}\mathbf{B}_c \neq \mathbf{0}, \tag{3.15}$$

which can be tested by using a wald-type statistic

$$T_n(\boldsymbol{F}_c) \;=\; tr\{\mathbf{F}_c^T H_x^T \mathbf{C}^T[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T]^{-1}\mathbf{C}H_x\mathbf{F}_c\}, \tag{3.16}$$

where $\mathbf{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)^T$, $\mathbf{F}_c = (\boldsymbol{f}_{c,1}, \cdots, \boldsymbol{f}_{c,n})^T$ and $H_x = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.

Let $\mathscr{D}_0 = \{\mathbf{B} : \mathbf{CB} = \mathbf{0}\}$, $\mathscr{D}_1 = \{\mathbf{B} : \mathbf{CB} \neq \mathbf{0}\}$, $\mathscr{D}_{c,0} = \{\mathbf{B} : \mathbf{CB}_c = \mathbf{0}\}$ and $\mathscr{D}_{c,1} = \{\mathbf{B} : \mathbf{CB}_c \neq \mathbf{0}\}$, we have $\mathscr{D}_0 \subset \mathscr{D}_{c,0}$ and $\mathscr{D}_{c,1} \subset \mathscr{D}_1$. It is possible that test on the common factors under H_1 would incur power loss, for example, when $\mathbf{C}\boldsymbol{\beta}_m(s_m) \neq 0$ holds for only a few tracts. In such cases, individual tract analysis can be used to achieve a better detection power. However, when the effects of interest are commonly shared by multiple tracts, the proposed common factor analysis allows us to perform a computationally efficient test on a much lower dimensional space, which would potentially increase statistical power.

3.2.6 Estimation and Inference Procedure

In practice, DTI properties are observed on discrete points. For the m-th tract, let $\mathscr{S}_m = \{s_{m,1}, \ldots, s_{m,w}, \ldots, s_{m,W_m}\}$ be the sample grid point of $y_{i,m}(s)$, we first rescale the responses so that $\frac{S_m}{W_m}\sum_{w=1}^{W_m}\sum_{i=1}^{n}[y_{i,m}(s_{m,w}) - \bar{y}_{\cdot,m}(s_{m,w})]^2 = n$, where $\bar{y}_{\cdot,m}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} y_{i,m}(\cdot)$. The following two-stage procedure is then adopted to estimate functional PCA scores $\{\boldsymbol{\xi}_i\}_{i=1}^n$ and common factors $\{\boldsymbol{f}_{c,i}\}_{i=1}^n$:

- Stage I: For each individual tract, $\mu_m(s)$ and $\tilde{\eta}_{i,m}(s)$ are estimated from (3.3) using local polynomial kernel smoothing technique introduced in Section 2.1.1. Functional principal component analysis is then performed to estimate $\boldsymbol{\xi}_{i,m}$.

- Stage II: We merge the estimated fPCs, denoted as $\widehat{\boldsymbol{\xi}}_{i,m}$, from all tracts together and use principal component analysis to estimate common factors $\widehat{\boldsymbol{f}}_{c,i}$. Regression and hypothesis testing can then be performed on $\widehat{\boldsymbol{f}}_{c,i}$.

Details of these two stages are given below.

In Stage I, to estimate the mean curve from model (3.3), we apply the local linear kernel smoothing method given by (2.2) when $\nu = 0$. Specifically, let $h_{1,m}$ be a given bandwidth for tract m, $\mu_m(s_{m,w})$ can be approximated by

$$\mu_m(s_{m,w}) \approx \mu_m(s_m), \text{ as } |s_{m,w} - s_m| \leq h_{1,m}. \tag{3.17}$$

Given a smoothing kernel $K(s)$ on $[-1,1]$, $\widehat{\mu}_m(s_m)$ can be estimated as the minimizers of the following weighted least square function:

$$\sum_{i=1}^{n} \sum_{w=1}^{W_m} [y_{i,m}(s_{m,w}) - \mu_m(s_m)]^2 K_{h_{1,m}}(s_{m,w} - s_m), \tag{3.18}$$

where $K_{h_{1,m}}(s) = \frac{1}{h_{1,m}} K(s/h_{1,m})$. The solution to (4.5) can be explicitly written as

$$\widehat{\mu}_m(s_m) = \frac{\sum_{w=1}^{W_m} \bar{y}_{\cdot,m}(s_{m,w}) K_{h_{1,m}}(s_{m,w} - s_m)}{\sum_{w=1}^{W_m} K_{h_{1,m}}(s_{m,w} - s_m)}, \ \forall s_m \in [0, S_m]. \tag{3.19}$$

The estimated $\widehat{\mu}_m(s_m)$ is a curve with local constant smoothness. More complicated local polynomial structure can be obtained by using higher order expansion if necessary.

Similarly, we can estimate each individual function $\tilde{\eta}_{i,m}(s_m)$ through approximation

$$\tilde{\eta}_{i,m}(s_{m,w}) \approx \tilde{\eta}_{i,m}(s_m), \text{ as } |s_{m,w} - s_m| \leq h_{2,m}, \tag{3.20}$$

where $h_{2,m}$ is a given bandwidth that controls smoothness of the estimated individual functions of tract m. The corresponding weighted least square function and the solution are respectively given as

$$\sum_{w=1}^{W_m} [y_{i,m}(s_{m,w}) - \widehat{\mu}_m(s_{m,w}) - \tilde{\eta}_{i,m}(s_m)]^2 K_{h_{2,m}}(s_{m,w} - s_m), \tag{3.21}$$

and

$$\widehat{\tilde{\eta}}_{i,m}(s_m) = \frac{\sum_{w=1}^{W_m} [y_i(s_{m,w}) - \widehat{\mu}_m(s_{m,w})] K_{h_{2,m}}(s_{m,w} - s_m)}{\sum_{w=1}^{W_m} K_{h_{2,m}}(s_{m,w} - s_m)}, \ \forall s_m \in [0, S_m]. \tag{3.22}$$

For different tracts, the bandwidths are not necessarily equal. We use a leave-one-out cross-validation proposed in [36] to determine the choices of $h_{1,m}$ and $h_{2,m}$. When smoothed individual functions are obtained, we calculate empirical covariance function as $\widehat{\Sigma}_{\tilde{\eta}_m}(s_m, t_m) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\tilde{\eta}}_{i,m}(s_m)\widehat{\tilde{\eta}}_{i,m}(t_m)$. And eigenbases $\{\widehat{\phi}_{m,l}(s_m)\}_{l=1}^{+\infty}$ can be estimated from spectral decomposition

$$\widehat{\Sigma}_{\tilde{\eta}_m}(s_m, t_m) = \sum_{l=1}^{+\infty} \widehat{\tau}_{m,l} \widehat{\phi}_{m,l}(s_m) \widehat{\phi}_{m,l}(t_m). \tag{3.23}$$

Then individual random effect $\widehat{\eta}_{i,m}(s_m)$ is projected onto basis functions $\{\widehat{\phi}_{m,l}(s_m)\}_{l=1}^{+\infty}$ to get functional PC scores, i.e.,

$$\widehat{\xi}_{i,ml} = \frac{S_m}{W_m} \sum_{w=1}^{W_m} \widehat{\eta}_{i,m}(s_{m,w}) \widehat{\phi}_{ml}(s_{m,w}), \tag{3.24}$$

for $l = 1, \cdots, L_n$ and $m = 1, \cdots, M$. Here, L_n is chosen as

$$\widehat{L}_n = \min\{L : 1 \le L \le n, \min_{1 \le m \le M} \frac{\sum_{1 \le l \le L} \widehat{\tau}_{m,l}}{\sum_{1 \le l \le n} \widehat{\tau}_{m,l}}\} > \alpha\} \tag{3.25}$$

where α is a given value close to 1. The above choice of $\widehat{L}_n$ requires us to extract the minimal number of fPCs which include at least $100\alpha\%$ of variation in each tract.

In Stage II, fPC scores from all tracts are merged together as

$$\widehat{\boldsymbol{\xi}}_i = (\widehat{\xi}_{i,11}, \cdots, \widehat{\xi}_{i,1L_n}, \cdots, \widehat{\xi}_{i,m1}, \cdots, \widehat{\xi}_{i,mL_n}, \cdots, \widehat{\xi}_{i,M1}, \cdots, \widehat{\xi}_{i,ML_n})^T. \tag{3.26}$$

Principal component analysis is then applied on the merged features to identify common factors. Let $\{\widehat{\tau}_{\xi,1}, \cdots, \widehat{\tau}_{\xi,r}\}$ be the first r eigenvalues of sample covariance matrix $\widehat{\Sigma}_\xi = \frac{1}{n}\sum_{i=1}^n \widehat{\boldsymbol{\xi}}_i \widehat{\boldsymbol{\xi}}_i^T$ in decreasing order and let $\widehat{\mathbf{v}}_1, \ldots, \widehat{\mathbf{v}}_r$ be the corresponding eigenvectors. The common factors are estimated as

$$\widehat{\boldsymbol{f}}_{c,i} \quad = \quad \widehat{T}^{-1/2} \widehat{\mathbf{V}}_r^T \widehat{\boldsymbol{\xi}}_i, \tag{3.27}$$

where $\widehat{\mathbf{V}}_r = (\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_r)$ and $\widehat{T} = (\widehat{\tau}_{\xi,1}, \cdots, \widehat{\tau}_{\xi,r})$. The dimension of common factors r is determined using parallel analysis. Such a procedure is similar to algorithm 2.3.1, except that the permutation is applied to the indices of tracts, i.e., $\{1, \cdots, M\}$, rather than all dimensions in $\widehat{\boldsymbol{\xi}}_i$.

Finally, we plug in $\widehat{\boldsymbol{F}}_c$ to (3.16) to calculate the test statistic $T_n(\widehat{\boldsymbol{F}}_c)$. It is proved in Section 3.3 that under H_0, $T_n(\widehat{\boldsymbol{F}}_c)$ converges to a mixed χ^2 distribution. The corresponding p-value is then calculated by bootstrap.

3.3 Theoretical Results

In this section, we study the asymptotic distribution of the proposed test statistic $T_n(\widehat{\boldsymbol{F}}_c)$ under both null hypothesis and alternative hypothesis.

3.3.1 Assumptions

Throughout the section, we assume the following assumptions hold. Some of the conditions can be weakened without changing the main conclusion, yet is beyond our interest in this work.

Assumption 3.1. *The arc lengths of M tracts have a universal upper bound, i.e.,* $\max_m S_m < c_1 < +\infty$. *For each tract m, the observed grid point set $\mathscr{S}_m$ is composed of W_m equidistant points in* $[0, S_m]$.

Assumption 3.2. *Smoothing kernel $K(u)$ is a symmetric positive function with continuous first order derivative, i.e., $K(u) \in C^1[-1,1]$. It is further assumed that $K(u)$ and its derivative satisfy* $\sup_{u\in[-1,1]} |\partial K(u)/K(u)| < c_2 < +\infty$.

Assumption 3.3. *The covariates $\{\boldsymbol{x}_i\}_{i=1}^n$ are independent and identically distributed bounded variables that satisfy $\|\boldsymbol{x}_i\|_2 < c_3 < +\infty$ almost surely, with $\mathbb{E}\boldsymbol{x}_i = \mathbf{0}$ and $\mathbb{E}\boldsymbol{x}_i\boldsymbol{x}_i^T = \Sigma_x$, where* $\|\Sigma_x\|_2 + \|\Sigma_x^{-1}\|_2 < c_4 < +\infty$.

Assumption 3.4. *Mean functions $\mu_m(s_m)$ are functions in $C^1[0,S_m]$ with universally bounded first order derivatives, i.e.,* $\max_m \sup_{s_m} |\partial \mu_m(s_m)| < c_5 < +\infty$.

Assumption 3.5. *Fixed effects $\boldsymbol{\beta}_m(s_m)$ are functions in $C^1[0,S_m]$ with universally bounded first order derivatives, i.e.,* $\max_m \sup_{s_m} \|\partial \boldsymbol{\beta}_m(s_m)\|_\infty < c_6 < +\infty$.

Assumption 3.6. *For each m, $\{\eta_{i,m}(s_m)\}_{i=1}^n$ are independently and identically distributed copies from a bounded process. The sample path of each $\eta_{i,m}(s_m)$ has continuous and universally bounded first order derivative on $[0,S_m]$, i.e., $\max_m \sup_{s_m} |\partial \eta_{i,m}(s_m)| < c_7 < +\infty$ holds almost surely. The covariance function $\Sigma_{\eta_m}(s_m,t_m)$ is assumed to belong to $C^1[0,S_m]^{\otimes 2}$ with uniformly bounded first order derivative, i.e., $\max_m \sup_{s_m,t_m} |\partial_{s_m}\Sigma_{\eta_m}(s_m,t_m)| < c_8 < +\infty$. In addition, the eigenfunctions of $\Sigma_{\eta_m}(s_m,t_m)$ are assumed to be universally bounded, i.e.,* $\max_{m,l} \sup_{s_m} |\phi_{m,l}(s_m)| < c_9 < +\infty$.

Assumption 3.7. *For all $m = 1,\cdots,M$, $\{e_{i,m}(s_m)\}_{i=1}^n$ are mutually independent and have a universal bound, i.e., $\max_m \sup_{s_m} |e_{i,m}(s_m)| < c_{10} < +\infty$. In addition, we assume that $\mathbb{E}[e_{i,m}(s_m)|\boldsymbol{x}_i] = 0$ and $\mathbb{E}[e_{i,m}^2(s_m)|\boldsymbol{x}_i] = 0$ hold almost surely.*

Assumption 3.8. *Let* $\{\tau_j\}_{j=1}^{r}$ *be the eigenvalues of an* $r\times r$ *positive definite matrix* $\Omega_\gamma = (\gamma_{j,j'})_{r\times r}$ *in decreasing order, where* $\gamma_{j,j'} = \sum_{m=1}^{M}\int_0^{S_m}\gamma_{m,j}(s_m)\gamma_{m,j'}(s_m)ds_m$ *for* $1\le j,j'\le r$*. It is assumed that* $\frac{\tau_r}{M} > \varepsilon_1 > 0$ *and* $\frac{1}{M}\min_{1\le j\le r-1}\{\tau_j - \tau_{j+1}\} > \varepsilon_2 > 0$.

Assumption 3.9. *The number of grid points* W_m *and the smoothing bandwidths* $h_{1,m}$*,* $h_{2,m}$ *are equal for all* $m = 1,\cdots,M$*, i.e.,* $W_m = W$*,* $h_{1,m} = h_1$ *and* $h_{2,m} = h_2$*. The following conditions are satisfied: (i)* n*,* W *and* $M\to+\infty$ *with* $M/n < c_{11} < +\infty$*; (ii)* $h_1^2\sqrt{\log M}$ *and* $h_2\sqrt{\log M}\to 0$*; (iii)* Wh_1^2*,* $Wh_2^2\to+\infty$*,* $(Wh_2)^{-1}\log M\to+\infty$ *and* $\max\{Wh_1, Wh_2\}/\sqrt{n}\le c_{12} < +\infty$.

Assumption 3.10. *For all* $m = 1,\cdots,M$*, there exists a uniform sequence of functional PCA cutoffs* $\{L_n\}_{n=1}^{+\infty}$ *such that, as* $n\to+\infty$*, we assume that* $\max_m\sum_{l=L_n+1}^{+\infty}\tau_{m,l}\to 0$ *and* $n^{-1}\log L_n\to 0$*. Let* $\omega_0 = \sqrt{\log M}\max\{h_1^2, h_2, (Wh_2)^{-1/2}, (\log M/n)^{-1/2}\}$*, it is required that* $\frac{\omega_0^2}{M}\sum_{m=1}^{M}\sum_{l=1}^{L_n}\tau_{m,l}^{-2}\to 0$.

Assumption 3.11. *Under alternative hypothesis, it is assumed that* $\mathbf{CB}_c = n^{-\frac{1}{2}}\mathbf{C}_0$*, where* $\mathbf{C}_0 = (\mathbf{c}_{0,1},\cdots,\mathbf{c}_{0,r})$ *is a* $k\times r$ *matrix with* $\|\mathbf{C}_0\|_F > \varepsilon_3$.

Assumptions 3.1−3.7 are standard conditions required to obtain consistent estimates of $\mu_m(s_m)$ and $\Sigma_{\tilde{\eta}_m}(s_m,t_m)$ with uniform convergence rates for all tracts. Assumption 3.8 is an analog to the pervasiveness condition (2.36) for model (3.11). It is required to guarantee that the common latent factors can be accurately recovered from the observations as $M\to+\infty$. Assumption 3.9 specifies the rates of h_1, h_2, W and M when $n\to+\infty$. The condition that all tracts have the same bandwidth and sample point size is not necessary, as long as $h_{1,m}$, $h_{2,m}$ and W_m have the same rate with h_1, h_2 and W respectively, and is only required to simplify the proof. Assumption 3.10 specifies the range of L_n, which depends on the decay rate $\{\tau_{m,l}\}_{l=1}^{+\infty}$ for all $m = 1,\cdots,M$. To see this, we consider two decay rates as examples. For an exponential decay rate, i.e., $c_{13}\alpha_1^{-l}\le\tau_{m,l}\le c_{13}\alpha_2^{-l}$ holds for all m with $1 < \alpha_2 < \alpha_1 < +\infty$ and $0 < c_{13} < +\infty$, Assumption 3.10 requires that $L_n\to+\infty$

and $L_n = \log_{\alpha_1}[o(\omega_0^{-2})]$. For a polynomial decay rate, i.e., $c_{14}l^{-r_1} \leq \tau_{m,l} \leq c_{14}l^{-r_2}$ holds for all m with $1 < r_2 < r_1 < +\infty$ and $0 < c_{14} < +\infty$, Assumption 3.10 requires that $L_n \to +\infty$ and $L_n = o[\omega_0^{-2/(r_1+1)}]$. Assumption 3.11 specifies an alternative hypothesis in which the effect of interest is a common effect implied by the latent common factors.

3.3.2 Main Results

The following theorem establishes the asymptotic distribution of the proposed test statistic under both null hypothesis and the alternative hypothesis.

Theorem 3.3.1. *Let $\Sigma_{f|x}$ be the covariance of $\boldsymbol{f}_{c,i}$ conditioned on $\boldsymbol{x}_i$, let $\{\tau_{f|x,j}\}_{j=1}^r$ be the eigenvalues of $\Sigma_{f|x}$ in decreasing order, and let $\{\mathbf{v}_{f|x,j}\}_{j=1}^r$ be the corresponding eigenvectors. When Assumptions 3.1 - 3.10 hold, $T_n(\widehat{\boldsymbol{F}}_c)$ has the following asymptotic distribution under H_0:*

$$T_n(\widehat{\boldsymbol{F}}_c) \xrightarrow{d} \sum_{j=1}^{r} \tau_{f|x,j}\chi_j^2(k),$$

where $\{\chi_j^2(k)\}_{j=1}^r$ denote r independent χ^2 distributions with degree-of-freedom k.

Under the alternative hypothesis specified by Assumption 3.11, $T_n(\widehat{\boldsymbol{F}}_c)$ converges to a non-central mixed χ^2 distribution given by:

$$T_n(\widehat{\boldsymbol{F}}_c) \xrightarrow{d} \sum_{j=1}^{r} \tau_{f|x,j}\chi_j^2(\nu_j,k),$$

where $\{\chi_j^2(\nu_j,k)\}_{j=1}^r$ are r independent non-central χ^2 distributions with non-central parameters $\{\nu_j\}_{j=1}^r$, in which $\nu_j = \mathbf{v}_{f|x,j}^T\mathbf{C}_0^T(\mathbf{C}\Sigma_x^{-1}\mathbf{C}^T)^{-1}\mathbf{C}_0\mathbf{v}_{f|x,j}/\tau_{f|x,j}$, and degree-of-freedom k.

The above theorem also shows that the asymptotic distribution of $T_n(\widehat{\boldsymbol{F}}_c)$ is the same as that of $T_n(\boldsymbol{F}_c)$ under both null hypothesis and alternative hypothesis. It is implied that using the proposed procedure to study testing problem (3.15) is asymptotically equivalent to directly working on latent common factors as if they are observed.

3.4 Simulations

In this section, we apply HFPRM to simulation examples. Simulation data is generated from a clinical study. It is a twin study designed to understand the genetic influence on brain development in early childhood. Detailed description can be found in Section 3.5

3.4.1 Setup

We selected 40 major fiber bundles with fractional anisotropy (FA) value from DTI tractography and extracted 100 uncorrelated subjects from the dataset. The following model was used to generate simulation data, and gestational age (Gage) and gender (G) were included as covariates:

$$y_{i,m}(s_m) = \mu_m(s_m) + c\beta_{m,1}(s_m)\mathrm{Gage}_i + \beta_{m,2}(s_m)\mathrm{G}_i + \eta_{i,m}(s_m) + e_{i,m}(s_m), \tag{3.28}$$

in which model parameters $\beta_{m,2}(s_m), \Sigma_{\eta_m}$ and Σ_{e_m} were estimated from real data, and coefficient $\beta_{m,1}(s_m)$ was rescaled so that all tracts had comparable effect sizes.

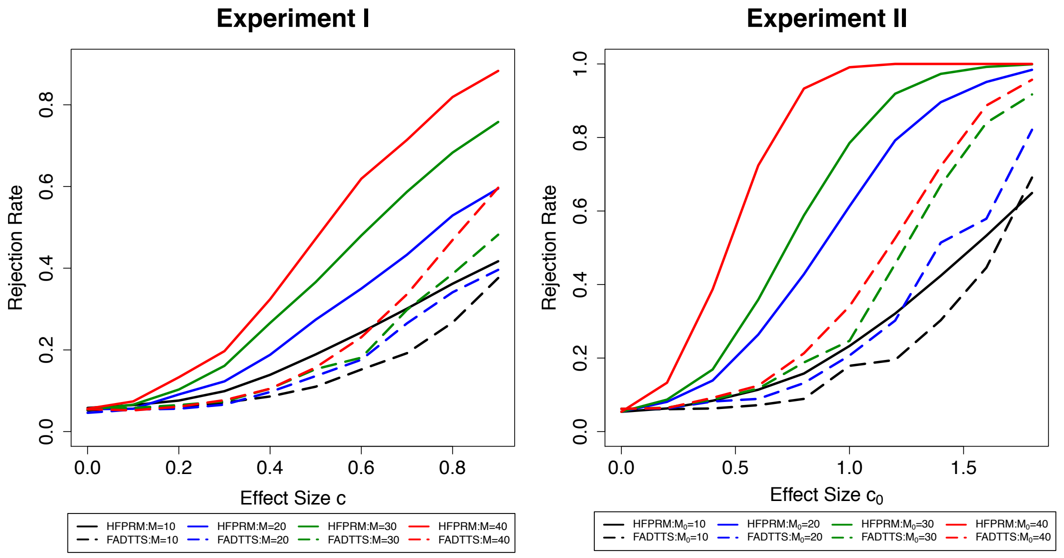
In the first simulation experiment, we aim to examine the influence of total tract number M to the performance of HFPRM when common effect exists. In each simulation run, M tracts were randomly chosen from all 40 bundles and simulation data was generated from model (3.28). M was set to take value from $10, 20, 30$ and 40. The second simulation experiment is to study the performance of HFPRM when different proportions of tracts have real effect. M was set to 40 and M_0 took value from $\{10, 20, 30, 40\}$. In each simulation run, M_0 tracts were randomly selected from all 40 bundles to have $c = c_0$, with $c_0 > 0$ under alternative hypothesis. For other tracts, c was set to 0. In both experiments, we tested the significance of the gestational age effect.

When applying HFPRM, varying coefficient model (3.3) was first fitted to estimate individual functions. Functional principal components were then extracted so that at least 85% of total variation was included for each tract. In factor analysis, parallel analysis was applied and factors with p-values less than 0.05 were selected as common factors. Type I error and statistical power were calculated at significance level $\alpha = 0.05$ based on 1000 simulation replications. FADTTS was also applied on each single tract. The results of multi-tract test with Bonferroni correction were reported as a comparison.

3.4.2 Results

The rejection rate of simulation experiment I is demonstrated in Figure 3.2(a). When $M = 10$, the common factor analysis of HFPRM slightly outperforms single tract analysis with multiple comparison adjustment. When $M = 20, 30$ and 40, HFPRM shows notable power increase compared to single tract analysis when detecting common effect. When M becomes larger, the improvement becomes more substantial. Such results are expected since common effect tends to accumulate in the common factor as M grows.

The rejection rate of simulation experiment II is shown in Figure 3.2(b). When $M_0 = 10$, i.e., 25% of the tracts have real effect, HFPRM and FADTTS have comparable performance. As M_0 increases to $20, 30$ and 40, HFPRM shows notably higher power compared with FADTTS. The power gain becomes larger as M_0 grows. It indicates that the proportion of tracts with true effect is critical to the performance of the common factor analysis in HFPRM. When the proportion is relatively low, HFPRM does not give much power improvement. As the proportion increases, the power gain of HFPRM becomes more substantial.



(a) (b)

Figure 3.2: Simulation results of HFPRM: panels (a) and (b) show the rejection rate of HFPRM and FADTTS in experiment I and experiment II respectively.

3.5 Early Human Brain Development Study

To investigate how genetic factors influence brain structure in prenatal and early postnatal stage, we conducted a genome-wide association study on fiber bundle statistics in a unique cohort of infants from the UNC Early Brain Development Study (UNCEBDS) [81].

3.5.1 Data Acquisition and Preprocessing

MRI scans were acquired either on a 3T Siemens Allegra head-only scanner (N $=$ 566) or on a 3T Siemens TIM Trio 3T scanner (N $=$ 96). For the Allegra model, diffusion weighted images were acquired from 339 subjects by a single shot EPI DTI sequence with the following parameters: TR/ TE $= 5200/73$ ms, voxel resolution $= 2\times2\times2$ mm^3, 6 non-collinear directions with $b = 1000$ s$/$mm^2 and 1 baseline image with $b = 0$. For the remaining subjects scanned on Allegra, DWI was acquired with the following parameters: TR/ TE $= 7680/82$ ms, voxel resolution $= 2\times2\times2$ mm^3, 42 non-collinear directions with diffusion gradients of $b = 1000$ s$/$mm^2 in addition to 7 baseline images. Quality control was applied on raw DWIs using DTIPrep [82], and FSL [83, 84] was performed for skull stripping and brain masking. We used a weighted least squares method [76] to estimate diffusion tensors and followed the UNC-Utah NA-MIC framework [85] to create a study-specific atlas. Subsequently, a total number of 44 fiber tracts listed in Table 3.1 were reconstructed in the atlas space using a streamline algorithm [86]. For each subject, a scalar diffusion property fractional anisotropy (FA) was calculated at each sample point along each tract using neighboring diffusion tensors.

Table 3.1: The UNCEBDS neonate data: A list of fiber tracts in the simulation experiments and the real data analysis

Bundle Group	Tract Segments
Arcuate Fasciculus	right temporo-parietal (ARTP)*, left fronto-temporal (ALFT)*, right fronto-temporal (ARFT)*, left fronto-parietal (ALFP)*, right fronto-parietal (ARFP)*
Corpus Callosum	motor body (CCMB)*, occipital splenium (CCOS)*, parietal body (CCPB)*, premotor body (CCPMB)*, rostrum (CCR)*, temporal tapetum (CCTT), genu (CCG)*
Cingulum	left cingulate gyrus (CLC)*, right cingulate gyrus (CRC)*, right hippocampal (CRH)*
Corticothalamic	left motor (CTLM)*, right motor (CTRM)*, left premotor (CTLPM)*, right premotor (CTRPM)*, left parietal (CTLP)*, right parietal (CTRP)*, left prefrontal (CTLPF)*, right prefrontal (CTRPF)*
CorticoFugal	left motor (CFLM)*, right motor (CFRM)*, left parietal (CFLP)*, right parietal (CFRP), left prefrontal cortex (CFLPFC)*, right prefrontal cortex (CFRPFC)*, left premotor (CFLPM)
Others	left fornix (FL)*, right fornix (FR)*, left inferior fronto-occipital fasciculi (IFOFL), right inferior fronto-occipital fasciculi (IFOFR), left inferior longitudinal fasciculi (ILFL)*, right inferior longitudinal fasciculi (ILFR)*, left medial lemniscus (ML)*, right medial lemniscus (MR)*, left optic (OTL)*, right optic (OTR)*, left superior longitudinal fasciculus (SLFL)*, right superior longitudinal fasciculus (SLFR)*, left uncinate fasciculus (UNCL)*, right uncinate fasciculus (UNCR)*

* marks the 40 selected tracts in the simulation experiments

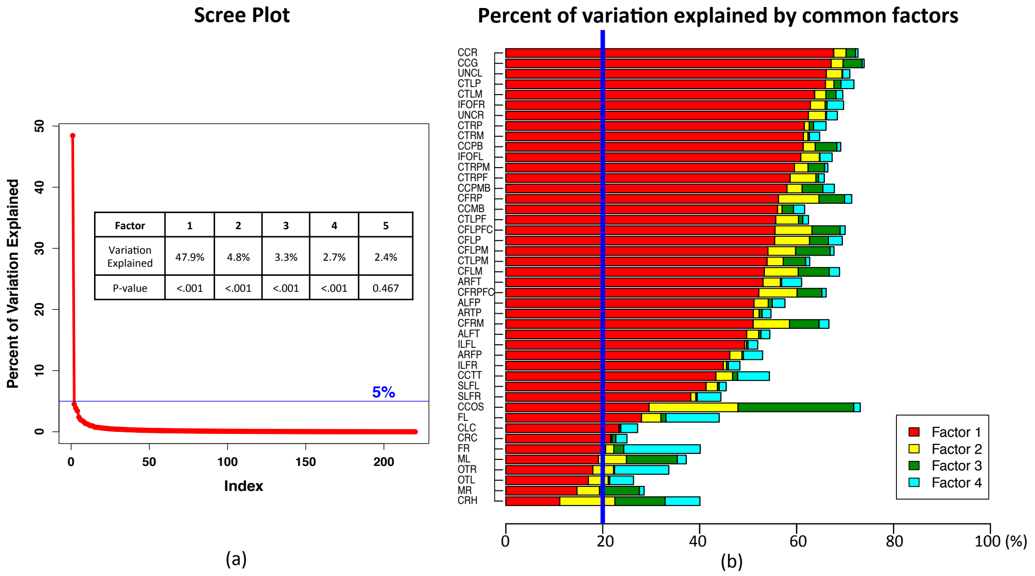
Genotyping of single nucleotide polymorphisms (SNPs) was conducted on Affymetrix Axiom genome-wide LAT Array. Samples with call rates less than 95%, outliers for homozygosity, ancestry outliers and unexpected relatedness were excluded from the study. We also removed genetic markers with Hardy-Weinberg equilibrium p-value less than 10^{-8}, call rate less than 95% and Mendelian error rate larger than 10%. Population stratification was assessed using principal component analysis [87]. Imputation was performed with MaCH-Admix [88] using 1000G reference panel [89]. To evaluate the quality of imputed SNPs, we computed the mean R^2 under varying minor allele frequency (MAF) categories and selected R^2 cutoffs as described in [90]. SNPs with MAF less than 0.01 were excluded from imputed dataset. Eventually, 471 twin subjects (31 MZ pairs, 75 DZ pairs and 260 singletons or unpaired twin subjects) and 8,538,562 genetic markers were retained for further analysis.

3.5.2 Data Analysis

Our primary interest is to perform GWAS on the neonate samples in order to find important genetic variants influencing the development of human brain at early life stage. The fractional anisotropy statistics on 44 major fiber bundles are the primary phenotypes under analysis, since FA value is an important diffusion measure that quantifies the extent of local directional water diffusion and partially reflects the degree of bundle maturation in premature brains and neonatal brains [91]. For a twin study, an ACE model was fitted instead of (3.14) on each common factor to account for correlation within twin subjects. For a twin pair i, a univariate common factor is modeled as

$$\boldsymbol{f}_{c,ij} = \mu_c + \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_{c,x} + \beta_{c,g} g_{ij} + a_{c,ij} + c_{c,i} + e_{c,ij}, \tag{3.29}$$

where $j = 1,2$ represent twin subject indices, $\boldsymbol{x}_{ij}$ are covariates and g_{ij} is additive genetic effect for a specific variant coded as $\{0,1,2\}$. Seven variables were added as covariates in $\boldsymbol{x}_{ij}$, including gestational age at birth, gender, DTI direction type, scanner type and the first three genetic principal component to adjust for population stratification. For variance components, it is assumed that additive genetic variation $a_{c,ij}$ follows normal distribution $N(0,\sigma_a^2)$ with $\text{cor}(a_{c,i1}, a_{c,i2}) = 0.5 + 0.5I_{MZ,i}$, in which $I_{MZ,i} = 1$ if twin pair i are monozygotic, that common environmental variation $c_{c,i}$ follows normal distribution $N(0,\sigma_c^2)$ and that unique environmental variation $e_{c,ij}$ follows normal distribution $N(0,\sigma_e^2)$.



Factor	1	2	3	4	5
Variation Explained	47.9%	4.8%	3.3%	2.7%	2.4%
P-value	<.001	<.001	<.001	<.001	0.467

Figure 3.3: Application of HFPRM to the UNCEBDS neonate data: panel (a) shows the scree plot of the factor analysis, the p-values of the first five factors form the parallel analysis and the percent of variation explained by the significant factors. Panel (b) shows the percent of variation of explained by the significant factors on each individual tract.

3.5.3 Results

In functional PCA, the first 5 functional principal components are extracted for each tract to include more than 70% of variation. As shown in Figure 3.3(a) and (b), factor $1-4$ were significant in parallel analysis (p-value < 0.001). Among these four factors, factor 1 was able to explain 47.9%

of total variation of all tracts and more than 20% of individual variation in most tracts, while each of the factor $2-4$ explained less than 5% of total variation and less than 8% of individual variation in most tracts. This indicates that factor 1 is able to capture common variation shared by most tracts. Therefore, we applied GWAS analysis on factor 1 only.

The results of GWAS are visualized in Figure 3.4. From the Manhattan plot, one genome-wide significant loci with p-value less than 5×10^{-8} was observed on the proteasome inhibitor subunit 1 (PSMF1) gene on chromosome 19. PSMF1 gene is a member in ubiquitin-proteasomal pathway, which is known to play an important role in developmental axonal pruning and synaptic plasticity [92] and is suspected to be a contributor of a wide range of neurophysiological and neuropathological processes [93]. Additionally, 13 locus were observed exceeding the suggestive genome-wide significance threshold (p-value $< 5 \times 10^{-6}$). These top snps are summarized in Table 3.2 and the nearest genes to the variants are presented. We also examined the gene expression level of these top genes in fetal tissue using a publicly available gene expression atlas [94]. Figure 3.5 shows the scaled expression level for 12 out of 14 identified genes with available data. SCAPER, SETBP1, B3GAT1,MAP3K13 genes are predominantly expressed in fetal brain tissues than in other fetal tissues, suggesting active involvement in genesis and differentiation of the central nervous system.

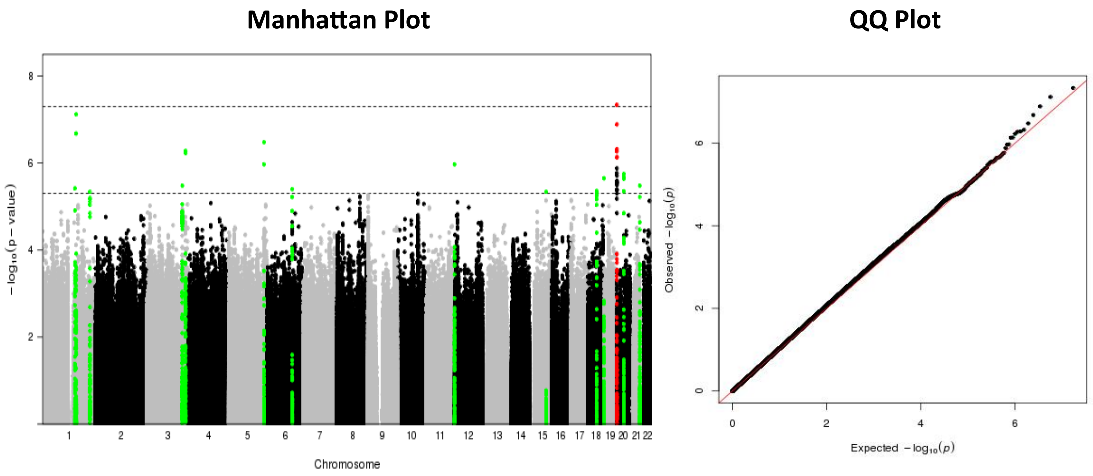


Figure 3.4: Application of HFPRM to the UNCEBDS neonate data: Manhattan plot and QQ plot of the $-\log 10(p\text{-values})$ from GWAS on the common factor.

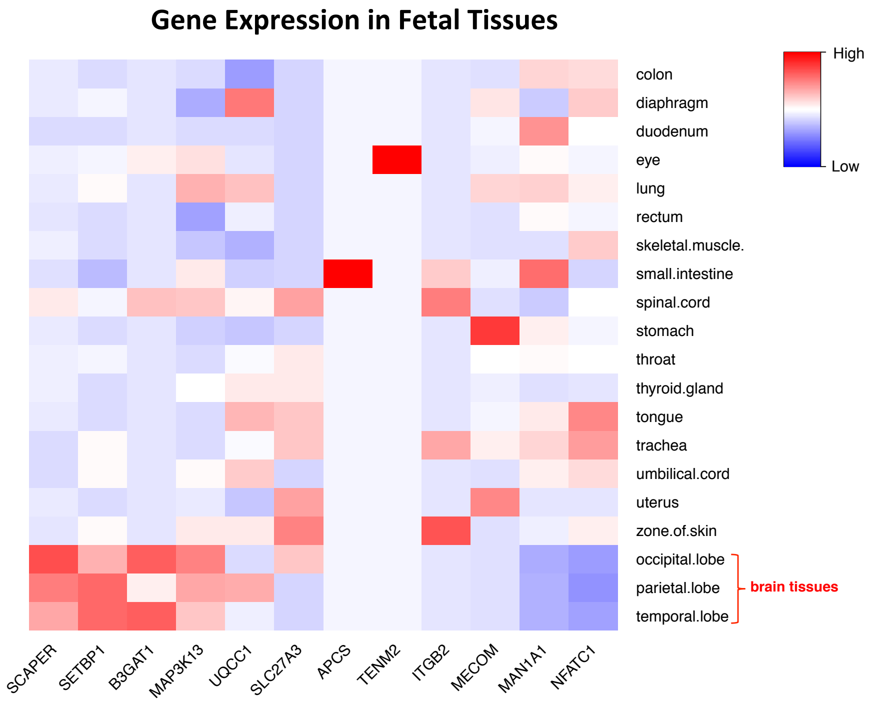


Figure 3.5: Application of HFPRM to the UNCEBDS neonate data: heatmap of relative expression level of the identified genes in fetal tissues. The expression level of SCAPER, SETBP1, B3GAT1 and MAP3K13 in brain tissues is higher than the average expression level in all tissues.

Table 3.2: Application of HFPRM to the LSEBD Neonate Data: Top SNPs from GWAS and their nearest genes

SNP	Chr	*p*-value	Gene
rs6077860	20	4.61E-08	PSMF1
rs1446965	1	7.60E-08	APCS
rs72830077	5	3.32E-07	TENM2
rs6777575	3	5.21E-07	MAP3K13
11:134773378	11	1.08E-06	B3GAT1
rs78070351	20	1.78E-06	UQCC1
rs28627209	18	2.26E-06	NFATC1
rs2216360	3	3.32E-06	MECOM
rs17004715	21	3.34E-06	ITGB2
rs7366960	1	3.84E-06	SLC27A3
rs114172604	6	3.94E-06	MAN1A1
rs11876680	18	4.39E-06	SETBP1
rs79045984	15	4.61E-06	SCAPER
rs2002371	1	4.62E-06	CNIH3

3.6 Conclusion

We developed a Hierarchical Principal Regression Model (HFPRM) to efficiently conduct joint analysis on diffusion tensor bundle statistics from multiple neurofiber tracts. A varying coefficient model was introduced and functional PCA was applied to characterize tract variation. Factor analysis was then adopted to extract common features and a standard multivariate test procedure was applied to study common effect. Simulation results have demonstrated that HFPRM is powerful to detect common effect shared by multiple tracts. Finally, the proposed method has been applied to a genome-wide association study on neonatal diffusion tensor images. We have identified some important genetic architecture related to early human brain development.

CHAPTER 4: A POWERFUL GLOBAL TEST STATISTIC FOR FUNCTIONAL STATISTICAL INFERENCE

4.1 Introduction

Functional regression modeling with a functional response $\mathbf{y} = \{y(s) : s \in \mathscr{S}\}$ and multivariate covariates $\mathbf{x} \in \mathbb{R}^p$ is a popular statistical tool in modern high-dimensional inference, with wide applications in various medical imaging studies [95, 96, 97, 98, 99, 100]. Among them, imaging genetic analysis on functional phenotypes is an important topic [101]. The primary interest is to identify genetic variants $\mathbf{x}$ associated with functional phenotypic variation $\mathbf{y}$ in human brain, which may ultimately lead to discoveries of genes for neuropsychiatric and neurological disorders.

Suppose that we observe functional responses $\mathbf{y}_i(s)$ and a set of clinical variables (e.g., age, genetic markers, and gender) $\boldsymbol{x}_i \in R^p$ for n unrelated subjects. Without loss of generality, we assume $\mathscr{S} = [0, S]$ for a positive scalar S. Throughout this chapter, we consider n independent observations $(\mathbf{y}_i, \boldsymbol{x}_i)$ and a varying coefficient model given by

$$y_i(s) = \boldsymbol{x}_i^T \boldsymbol{\beta}(s) + \eta_i(s) + e_i(s), \tag{4.1}$$

where $\boldsymbol{\beta}(s)$ is a $p \times 1$ vector of functional coefficients, $\eta_i(s)$ is random effect that characterizes subject-specific spatial variation, and $e_i(s)$ represents measurement error. It is assumed that $\eta_i(s)$ and $e_i(s)$ are mutually independent and identical copies of $\mathrm{SP}\{0, \Sigma_\eta(s,t)\}$ and $\mathrm{SP}\{0, \sigma_e^2(s) I(s=t)\}$, respectively, where $\mathrm{SP}(\mu, \Sigma)$ denotes a stochastic process with mean function $\mu(s)$ and covariance function $\Sigma(s,t)$, and $I(\cdot)$ is the indicator function of an event. Many hypothesis testing problems of interest, such as GWAS, can be formulated as a unit-rank global testing problem across $\mathscr{S}$, which is given by

$$H_0 : \mathbf{C}\boldsymbol{\beta}(s) = b_0(s) \ \forall s \in \mathscr{S} \text{ v.s. } H_1 : \mathbf{C}\boldsymbol{\beta}(s) \neq b_0(s) \ \exists s \in \mathscr{S}. \tag{4.2}$$

where $\mathbf{C}$ is a $1 \times p$ vector and $b_0(s)$ is an scalar value. Without loss of generality, we center the covariates, standardize the responses, and assume $b_0(s) = 0$.

The key problem is how to design a powerful global test statistic that can efficiently aggregate weak signals across $\mathscr{S}$, while achieving high statistical power for testing problem (4.2). To the best of our knowledge, such problem has not been fully solved yet. We focus on a specific setting that all components in $\boldsymbol{\beta}(s)$ lie in an infinite-dimensional functional space, but p is relatively small. Existing testing methods fail to detect moderate or weak signals due to two major challenges, (i) infinite-dimensional functional parameters and (ii) complicated covariance structure $\Sigma_\eta(s,t)$. Popular pooled global test statistics are proposed to conduct univariate analysis at each sample grid point of $\mathscr{S}$ and then to combine their results [36, 101]. However, since most of such tests ignore the correlation structure of $\boldsymbol{y}_i(s)$, they may suffer from severe power loss in presence of high correlation. Moreover, testing at each grid point individually in the mass univariate analysis requires a substantial penalty of controlling for multiplicity. The Hotelling's T^2 type test is also not well-defined for our problem of interest, since the sample estimate of Σ_η is not invertible. Although dimension reduction techniques, such as principal component analysis (PCA), are considered to reduce the dimension of functional response, most of the methods ignore the variation of covariates and their associations with responses. Thus, such methods can be sub-optimal for our problem. Finally, some recent developments in regularization methods, such as multiple task learning, do not provide a post-inference tool (e.g., p-values) [102, 103].

The proposed method has three major contributions given as follows:

- A novel functional projection regression model and its associated global test statistic are introduced to aggregate relatively weak signals across $\mathscr{S}$, while reducing the dimension of functional data. An optimal functional projection direction is calculated by maximizing statistical power with ridge penalty.

- The asymptotic distribution of the global test statistic is studied systematically under both null and alternative hypotheses and a data-driven strategy is provided to adaptively select optimal tuning parameter.

- Numerical simulations show that the proposed test outperforms all existing state-of-the-art methods in functional statistical inference.

The rest of the chapter is organized as follows. In Section 4.2, we introduce the functional projection regression model and its associated global test statistic. In Section 4.3, we derive the asymptotic distribution of the test statistic under both null and alternative hypotheses. In Sections 4.4 and 4.5, we use numerical simulations and a real data example to examine the finite sample performance of the proposed test. Section 4.6 concludes with some remarks.

4.2 Method

4.2.1 Functional Projection Regression Model

We propose a functional projection regression model as follows. Specifically, let $\omega(s)$ be a weight function in $\mathscr{L}_2(\mathscr{S})$, we project $y_i(s)$ onto the functional direction $\omega(s)$ such that

$$y_{\omega,i} = \boldsymbol{x}_i^T \boldsymbol{\beta}_\omega + \eta_{\omega,i}, \tag{4.3}$$

in which $\boldsymbol{\beta}_\omega = \int_{\mathscr{S}} \boldsymbol{\beta}(s)\omega(s)ds$, and $\eta_{\omega,i} = \int_{\mathscr{S}} \eta_i(s)\omega(s)ds$. The term associated with $e_i(s)$ would converges to 0 in probability through local kernel smoothing, therefore is asymptotically ignorable. The projected model (4.3) transforms the functional response to a univariate response. Let $\widehat{\boldsymbol{\beta}}_\omega$ and $\widehat{\Sigma}_\eta$ be the estimates of $\boldsymbol{\beta}_\omega$ and Σ_η, respectively. Thus, a standard wald-type statistic can be given by

$$T_n(\omega) = \widehat{\boldsymbol{\beta}}_\omega^T \mathbf{C}^T[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T]^{-1}\mathbf{C}\widehat{\boldsymbol{\beta}}_\omega / \{\iint \widehat{\Sigma}_\eta(s,s')\omega(s)\omega(s')dsds'\}. \tag{4.4}$$

We calculate $\widehat{\boldsymbol{\beta}}_\omega$ and $\widehat{\Sigma}_\eta$ by using local constant kernel smoothing with weighted least square method [29]. Assume functional responses $\{y_i(s)\}_{i=1}^n$ are observed on W discrete sample points $\mathscr{S} = \{s_1, \cdots, s_W\}$ and h_1 is a given smoothing bandwidth, a smooth estimate of $\boldsymbol{\beta}(s)$ can be calculated as

$$\widehat{\boldsymbol{\beta}}_{h_1}(s) = \underset{\boldsymbol{\beta}(s)}{\operatorname{argmin}} \sum_{i=1}^n \sum_{\mathrm{w}=1}^W [y_i(s_\mathrm{w}) - x_i^T\boldsymbol{\beta}(s)]^2 K_{h_1}(s_\mathrm{w} - s), \tag{4.5}$$

Similarly with bandwidth h_2, the random function $\eta_i(s)$ can be estimated by

$$\widehat{\eta}_{i,h_2}(s) = \underset{\eta_i(s)}{\operatorname{argmin}} \sum_{\mathrm{w}=1}^W [y_i(s_\mathrm{w}) - x_i^T\widehat{\boldsymbol{\beta}}(s_\mathrm{w}) - \eta_i(s)]^2 K_{h_2}(s_\mathrm{w} - s). \tag{4.6}$$

With $\{\widehat{\eta}_{i,h_2}(s)\}_{i=1}^n$, we can obtain a consistent estimate of $\Sigma_\eta(s,t)$ as follows,

$$\widehat{\Sigma}_\eta(s,t) = \frac{1}{n}\sum_{i=1}^n \widehat{\eta}_{i,h_2}(s)\widehat{\eta}_{i,h_2}(s'). \tag{4.7}$$

We address the problem of determining $\omega(s)$ in order to achieve optimal power. Specifically, we consider the signal-to-noise ratio of test statistic $T_n(\omega)$, which dominates the asymptotic power, as follows:

$$L_1(\omega) = \frac{\boldsymbol{\beta}_{\omega}^T \mathbf{C}^T [\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T]^{-1}\mathbf{C}\boldsymbol{\beta}_{\omega}}{\iint \Sigma_{\eta}(s,t)\omega(s)\omega(t)dsdt}. \tag{4.8}$$

An optimal projection direction is the maximizer of $L_1(\omega)$. However, when plugging in the estimates of $\boldsymbol{\beta}(s)$ and $\boldsymbol{\Sigma}_{\eta}$, maximizing $L_1(\omega)$ can be an ill-conditioned problem. The eigenvalues of $\widehat{\boldsymbol{\Sigma}}_{\eta}(s,t)$ usually decrease to zero very fast and the maximum value of $L_1(\omega)$ tend to be ∞. To solve this issue, we add a ridge penalty term, which leads

$$\widehat{\omega}_{\lambda}(\cdot) = \underset{\omega(\cdot)}{\text{argmax}} \frac{\widehat{\boldsymbol{\beta}}_{\omega,h_1}^T \mathbf{C}^T [\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T]^{-1}\mathbf{C}\widehat{\boldsymbol{\beta}}_{\omega,h_1}}{\iint \widehat{\Sigma}_{\eta}(s,t)\omega(s)\omega(t)dsdt + \lambda\|\omega(s)\|_2^2}, \tag{4.9}$$

where $\lambda > 0$ is a tuning parameter and $\|\omega(s)\|_2^2 = \int_{\mathscr{S}} \omega^2(s)ds$.

For a given λ, we calculate $\widehat{\omega}_{\lambda}(\cdot)$ as follows. Let $\{\widehat{\tau}_l\}_{l=1}^{+\infty}$ be the eigenvalues of $\widehat{\Sigma}_{\eta}(s,t)$ in a decreasing order and let $\{\widehat{\phi}_l(s)\}_{l=1}^{+\infty}$ be the corresponding eigenfunctions. Assume that the underlying $\omega(s) \in span\{\phi_l(s)\}_{l=1}^{+\infty}$ such that $\omega(s) = \sum_{l=1}^{+\infty} \omega_l \phi_l(s)$, we search $\widehat{\omega}(s)$ in the space spanned by the estimated eigenfunctions $span\{\widehat{\phi}_l(s)\}_{l=1}^{+\infty}$. Then (4.9) can be equivalently formulated as

$$\widehat{\boldsymbol{\omega}}_{\lambda} = (\widehat{\omega}_{1,\lambda}, \cdots, \widehat{\omega}_{l,\lambda}, \cdots) = \underset{\boldsymbol{\omega}}{\text{argmax}} \frac{[\sum_{l=1}^{+\infty} \widehat{d}_{l,h_1}\omega_l]^2}{\sum_{l=1}^{+\infty} \omega_l^2(\widehat{\tau}_l + \lambda)}, \tag{4.10}$$

in which $\widehat{\omega}_{l,\lambda} = \displaystyle\int_0^S \widehat{\omega}_{\lambda}(s)\widehat{\phi}_l(s)ds$ and $\widehat{d}_{l,h_1} = \displaystyle\int_0^S \mathbf{C}\widehat{\boldsymbol{\beta}}_{h_1}(s)\widehat{\phi}_l(s)ds$ are the projections of functional direction $\widehat{\omega}_{\lambda}(s)$ and estimated signal $\mathbf{C}\widehat{\boldsymbol{\beta}}_{h_1}(s)$ on the estimated eigenfunctions $\{\widehat{\phi}_l(s)\}_{l=1}^{\infty}$. The solution to (4.10) can be explicitly expressed as,

$$\widehat{\omega}_{l,\lambda} = \widehat{d}_{l,h_1}/(\widehat{\tau}_l + \lambda). \tag{4.11}$$

Finally, we obtain a global test statistic based on the optimal projection direction $\widehat{\omega}_\lambda(s) = \sum_{l=1}^{+\infty} \widehat{\omega}_{l,\lambda} \widehat{\phi}_l(s)$ as follows:

$$T_n(\widehat{\omega}_\lambda) = \frac{(\sum_{l=1}^{+\infty} \widehat{d}_{l,h_1} \widehat{\omega}_{l,\lambda})^2}{[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T] \sum_{l=1}^{+\infty} \widehat{\tau}_l \widehat{\omega}_{l,\lambda}^2}. \tag{4.12}$$

An unsolved question is how to choose the tuning parameter λ, which will be answered in Section 4.3.

To approximate the distribution of $T_n(\widehat{\omega}_\lambda)$ under null hypothesis, we adopt a wild-boostrap procedure described as follows.

Algorithm 4.2.1.

(i) Fit the varying coefficient model under the null hypothesis and get the estimate of $\widehat{\beta}_0(s)$, $\widehat{\eta}_{i,0}(s)$ and $\widehat{e}_{i,0}(s)$ for $i = 1, \cdots, n$ and $s \in [0, S]$.

(ii) For $g = 1, \cdots, G$, generate independent random numbers $\nu_i^{(g)}$ and $\nu_i^{(g)}(s_w)$ from $N(0,1)$, and the wild bootstrap sample on each grid point can be calculated as

$$\widehat{y}_i^{(g)}(s_w) = \widehat{\beta}_0(s_w)^T \boldsymbol{x}_i + \nu_i^{(g)} \widehat{\eta}_{i,0}(s_w) + \nu_i^{(g)}(s_w) \widehat{e}_{i,0}(s_w). \tag{4.13}$$

(iii) Repeat the test procedure and generate G samples of $T_n(\widehat{\omega}_\lambda^{(g)})$ under the null hypothesis.

(iv) The $p-$value is approximated by $p = G^{-1} \sum_{g=1}^{G} I\{T_n(\widehat{\omega}_\lambda) \geq T_n(\widehat{\omega}_\lambda^{(g)})\}$.

Approximation of the null distribution requires repeated calculations of the estimation-test procedure by G times, and G must be large enough to guarantee approximation accuracy.

4.3 Theoretical Result

In this section, we study the asymptotic distribution of test statistic $T_n(\widehat{\omega}_\lambda)$ and consider the problem of determining the tuning parameter λ for optimally testing (4.2).

4.3.1 Assumptions

Throughout this section, the following assumptions are used to facilitate the technical details. Some of the assumptions might be weakened but the current version simplifies the proof.

Assumption 4.1. *Smoothing kernel $K(u)$ is a symmetric positive function with compact support $[-1,1]$ and upper bound c_1. Moreover, $K(u)$ has continuous first order derivative satisfying $\sup_u |\dot{K}(u)| < c_2 < +\infty$.*

Assumption 4.2. *Variable of interest $\boldsymbol{x}_i$ are identically and independently distributed variables with mean μ_x and positive definite covariance Σ_x, and $\|\boldsymbol{x}_i\|_\infty < c_3 < +\infty$.*

Assumption 4.3. *Sample grid point set $\mathscr{S}$ is composed of M equidistant points on $[0,S]$.*

Assumption 4.4. *Fixed effects $\boldsymbol{\beta}(s)$ are continuous functions in $C^1[0,S]$ with universally bounded first order derivatives, i.e., $\sup_s \|\partial\boldsymbol{\beta}(s)\|_\infty < c_4 < +\infty$.*

Assumption 4.5. *Random functions $\{\eta_i(s)\}_{i=1}^n$ are i.i.d copies from a gaussian process and the sample path has continuous first-order derivative on $[0,S]$. We further assume that $\partial\eta_i(s)$ is also a gaussian process and its covariance function has continuous first-order derivatives, i.e., $\Sigma_{\partial\eta}(s,t) \in C^1[0,S]^{\otimes 2}$.*

Assumption 4.6. *Error terms $\{e_i(s)\}_{i=1}^n$ are i.i.d copies from a universally upper bounded process, i.e., $\sup_s |e_i(s)| < c_7 < +\infty$.*

Assumption 4.7. *Let $\Sigma_\eta(s,t) = \sum_{l=1}^{+\infty} \tau_l \phi_l(s)\phi_l(t)$ be the spectral expansion of $\Sigma_\eta(s,t)$. $\tau_1 > \cdots > \tau_l > \cdots \geq 0$ are eigenvalues with simple multiplicity that satisfy $\min_{j\neq l} |\tau_j - \tau_l|/\tau_l > \varepsilon_0 > 0$. Additionally, we assume that one of the two conditions holds,*

(i) $\{\tau_l\}$ follows polynomial decay rate, i.e., $\tau_l \asymp l^{-r}$ with $r > 1$, it is assumed that $\lambda_n^{1-\frac{1}{r}} Mh_1 \to +\infty$ and $\lambda_n^{3-\frac{1}{r}} \min\{h_1^{-2}, h_2^{-2}, Mh_2, n/\log n\} \to +\infty$.

(ii) $\{\tau_l\}$ follows exponential decay rate, i.e., $\tau_l \asymp \alpha^{-l}$ with $\alpha > 1$, it is assumed that $Mh_1\lambda_n \log \lambda_n^{-1} \to +\infty$ and $\lambda_n^3 \log \lambda_n^{-1} \min\{h_1^{-2}, h_2^{-2}, Mh_2, n/\log n\} \to +\infty$.

Assumption 4.8 (Local Alternative Hypothesis)**.** *A sequence of local alternative hypotheses are defined as, $H_{1n} : \mathbf{C}\beta(s) = n^{-1/2} d_0(s)$, where $d_0(s) \in C^1[0,S] \cap span\{\phi_l(s))\}_{l=1}^{+\infty}$.*

Assumptions 4.1−4.6 are standard conditions in functional data analysis, which are required to guarantee that the estimates of $\beta(s)$ and $\Sigma_\eta(s,t)$ are consistent. Assumption 4.7 is required in order to specify the bound of tuning parameter λ_n according to different decay rates of $\{\tau_l\}$. Here, we only consider distinct eigenvalues. It is assumed that the distances between one eigenvalue and other eigenvalues can not be too large compared to itself. Conclusions for multiplicity greater than one could be reached, yet is beyond the discussion in this work. Assumption 4.8 specifies a sequence of local alternative hypotheses from which we will derive the asymptotic power.

4.3.2 Main Theoretical Results

We present the key results below according to different decay rates of $\{\tau_l\}_{l=1}^{+\infty}$. The proof of the theorem is given in Appendix B.

Theorem 4.3.1. *When Assumptions 4.1 − 4.6 and 4.7(i)/4.7(ii) hold, as $n, M \to +\infty$, for sequence of $\{\lambda_n\}$ satisfying $\lambda_n \downarrow 0$, $T_n(\widehat{\omega}_{\lambda_n})$ has the following asymptotic normality distribution under the* *null hypothesis,*

$$T_n(\widehat{\omega}_{\lambda_n}) \quad \xrightarrow{d} \quad N\{\mu_0, \sigma_0^2\}, \tag{4.14}$$

where $\xrightarrow{d}$ *denotes convergence in distribution and* μ_0 *and* σ_0^2 *are given by,*

$$\mu_0 = \frac{a_1^2}{a_2} \text{ and } \sigma_0^2 = \frac{8a_1^2}{a_2} + \frac{2a_1^4 a_4}{a_2^4} - \frac{8a_1^3 a_3}{a_2^3}, \tag{4.15}$$

in which $a_1, a_2, a_3,$ *and* a_4 *are defined as,*

$$a_1 \quad = \quad \sum_{l=1}^{+\infty} \frac{\tau_l}{\tau_l + \lambda_n},\ a_2 = \sum_{l=1}^{+\infty} (\frac{\tau_l}{\tau_l + \lambda_n})^2,\ a_3 = \sum_{l=1}^{+\infty} (\frac{\tau_l}{\tau_l + \lambda_n})^3,\ a_4 = \sum_{l=1}^{+\infty} (\frac{\tau_l}{\tau_l + \lambda_n})^4.$$

When the local alternative hypothesis holds, let $\delta_{0,l} = \int_0^S d_0(s)\phi_l(s)ds$ *and* $\sigma_c^2 = \mathbf{C}\Sigma_x^{-1}\mathbf{C}^T$*,* $T_n(\widehat{\omega}_{\lambda_n})$ *has the following asymptotic distribution given by*

$$T_n(\widehat{\omega}_{\lambda_n}) \quad \xrightarrow{d} \quad N\{\mu_1, \sigma_1^2\}, \tag{4.16}$$

where μ_1 *and* σ_1^2 *are defined as,*

$$\begin{aligned} \mu_1 &= \frac{(a_1 + d_1)^2}{a_2 + d_2}, \\ \sigma_1^2 &= \frac{8(a_1 + d_1)^2 (a_2 + 2d_2)}{(a_2 + d_2)^2} + \frac{2(a_1 + d_1)^4 (a_4 + 2d_4)}{(a_2 + d_2)^4} - \frac{8(a_1 + d_1)^3 (a_3 + 2d_3)}{(a_2 + d_2)^3}, \end{aligned}$$

and d_1, d_2, d_3, and d_4 are given by

$$d_1 \;=\; \sum_{l=1}^{+\infty} \frac{\delta_{l,0}^2}{\sigma_c^2(\tau_l+\lambda_n)},\; d_2 = \sum_{l=1}^{+\infty} \frac{\tau_l \delta_{l,0}^2}{\sigma_c^2(\tau_l+\lambda_n)^2},\; d_3 = \sum_{l=1}^{+\infty} \frac{\tau_l^2 \delta_{l,0}^2}{\sigma_c^2(\tau_l+\lambda_n)^3},\; d_4 = \sum_{l=1}^{+\infty} \frac{\tau_l^3 \delta_{l,0}^2}{\sigma_c^2(\tau_l+\lambda_n)^4}.$$

Theorem 4.3.1 establishes the asymptotic distribution of the proposed test statistics for fixed λ_n under both null and alternative hypotheses. It also provides a data-driven criterion to select tuning parameter λ_n in order to achieve optimal power. Specifically, we propose to choose λ_n as

$$\widehat{\lambda}_n = \text{argmax}[\widehat{\mu}_1/\widehat{\sigma}_1 - \widehat{\mu}_0/\widehat{\sigma}_0], \tag{4.17}$$

in which $\widehat{\mu}_0, \widehat{\mu}_1, \widehat{\sigma}_0, \widehat{\sigma}_1$ are calculated by plugging in their corresponding estimates.

4.4 Numerical Simulation

4.4.1 Setup

In this section, we use numerical simulation to evaluate the finite-sample performance of the proposed global test statistic. Data was generated from the following varying coefficient model

$$y_i(s) = \beta_0(s) + x_{i,1}\beta_1(s) + \eta_i(s) + e_i(s),\; s \in [0,1],\; i = 1, \cdots n,$$

where $x_{i,1} \sim N(0,1)$. We set $n = 200$ and $S = 1$ and put the number of grid points $M = 100$ even in $[0,1]$. Our primary goal is to test the following hypothesis,

$$H_0 : \beta_1(s) = 0 \;\forall s \in [0,1] \text{ v.s. } H_1 : \beta_1(s) \neq 0 \;\exists s \in [0,1].$$

In this experiment, we simulated two types of signals under alternative hypothesis. In case

I, we chose $\beta_1(s)$ as a relatively homogenous signal that spread along the whole curve. In case II, $\beta_1(s)$ was simulated as a spatially heterogenous function with signal concentrated in a small interval. Other model parameters were estimated from the UK Biobank dataset introduced in Section 4.5. Two types of decay rates of $\{\tau_l\}_{l=1}^{+\infty}$ were considered, including a polynomial decay rate with $\tau_l = l^{-3/2}$ and an exponential decay rate with $\tau_l = 0.75^l$. The signal-to-noise ratios under alternative hypothesis are shown in Figure 4.1(a)-(d) for the two cases and the structure of the covariance functions are presented in Figure 4.1(e)-(f).

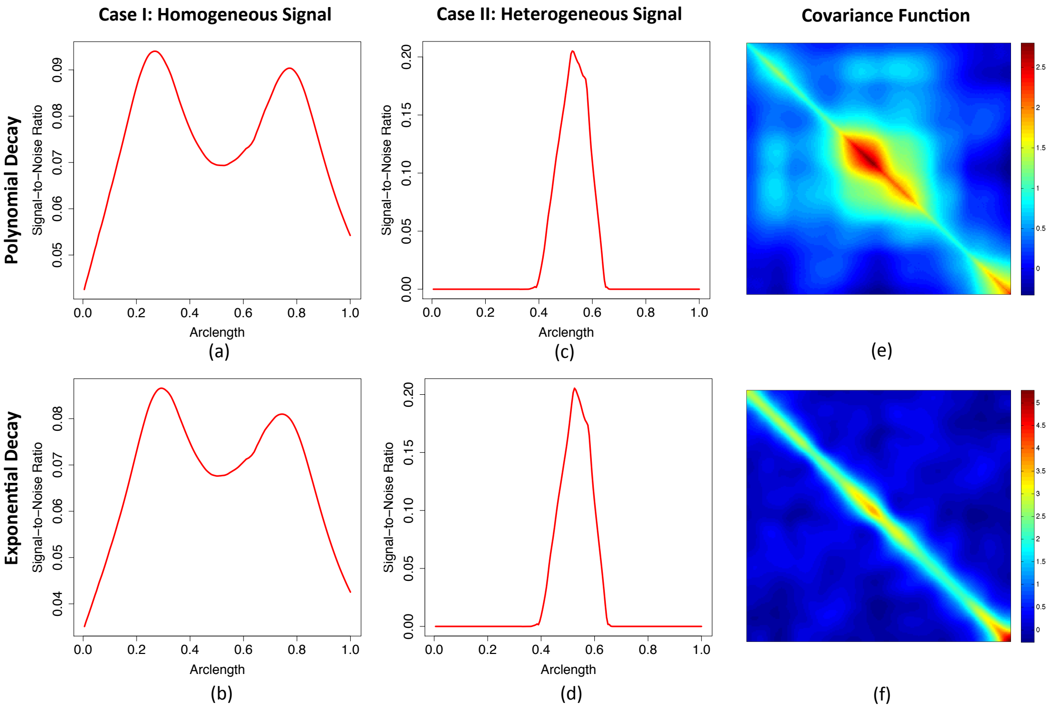


Figure 4.1: Simulation settings for PFGT: panels (a)−(d) demonstrate the signal-to-noise ratios under alternative hypothesis for case I and case II. Panels (e)−(f) visualize the covariance function of simulated responses along the curve.

For the choice of tuning parameter, we considered both fixed quantities where $\log\lambda_n$ takes values from $[-2,0]$ with an equal increment of 0.1 (PFGT-λ_n) and an optimal λ_n selected by (4.17) in each run (PFGT-optimal).

As a comparison, we also applied the two standard methods of FADTTS [36] and FLMtest [46] as reviewed in Section 2.1.3. In each scenario, 1000 simulation replicates were generated to evaluate type I and II error rates respectively. To calculate p-value, $G = 1000$ wild-bootstrap samples were generated in each run.

4.4.2 Results

Simulation results are summarized in Figure 4.2. In both exponential decay case and polynomial decay case, FADTTS controls type I error rates well. Although our global test has slightly inflated false positive rate as λ_n is relatively large, the optimal $\widehat{\lambda}_n$ does not show inflation. For FLMtest, type I error is slightly inflated.

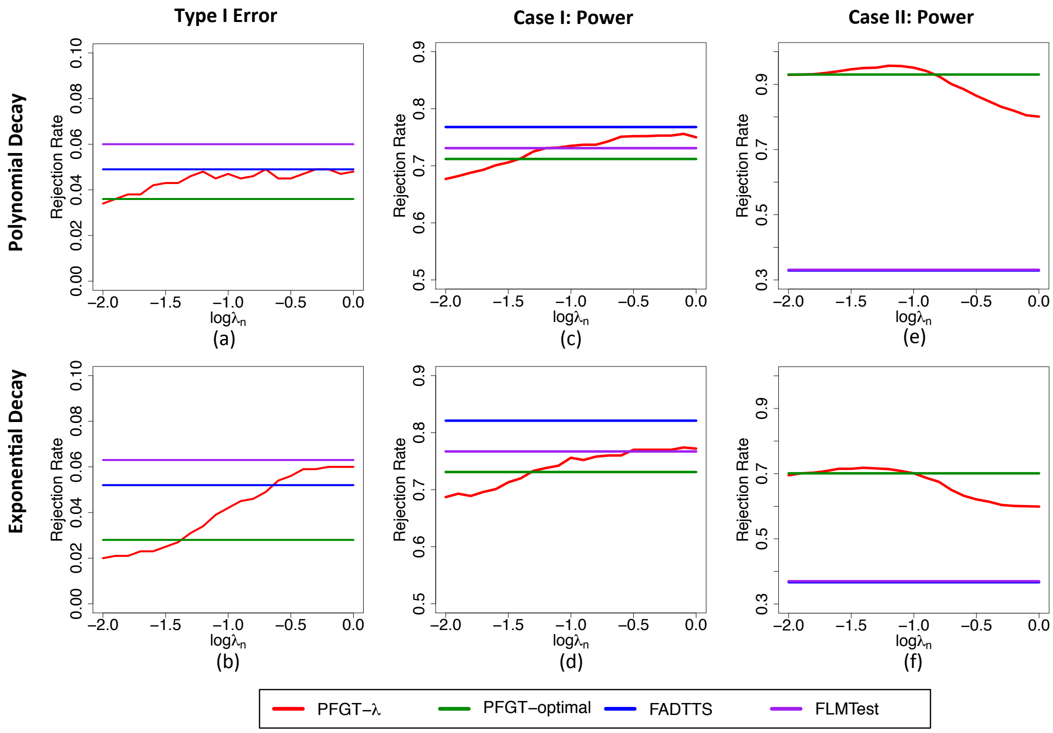


Figure 4.2: Simulation results for PFGT: panels (a)−(b) present the type I error for PFGT-λ_n, PFGT-optimal, FADTTS and FLMTest. Panels (c)−(f) present the power under alternative hypotheses for case I and case II.

Under alternative hypothesis, the proposed method has slightly lower power than FADTTS and

FLMtest in case I, as the signal is relatively homogenous. In case II, our global test substantially outperforms FADTTS and FLMtest for both exponential decay and polynomial decay scenarios. The performance of our data driven strategy for choosing λ_n is comparable to fixed λ_n with the best power.

4.5 Application: the UK Biobank Data Analysis

4.5.1 UK Biobank Study

UK Biobank is a large-scale cohort in the United Kingdom designed to investigate the influences of genetic susceptibility, environmental exposures and lifestyle factors to a wide range of health-related outcomes and disorders in middle aged and elderly population. In this section, we perform GWAS on the functional neuroimaging phenotypes from this study.

Diffusion weighted images (DWI) were acquired for 8751 subjects in total. We ran the TBSS-ENIGMA pipeline [104] on DWIs with the FSL tool set [84] to perform quality control and registration. The ENIGMA skeleton was then projected onto the registered FA images and FA statistics on $26,334$ voxels from 21 regions of interest (ROIs) were obtained. The primary phenotype of interest is the distributional density of voxel-wise FA statistics of the whole brain. As the density function is constrained by the normalization condition, we applied a log quantile density transformation introduced in [105] and took the output as the functional phenotypes for further analysis.

The Affymetrix Axiom platform was used to genotype 8057 subjects from the full population with imaging data, which resulted in a set of $784,256$ single-nucleotide polymorphisms (SNPs). The SNP data were preprocessed by standard quality control steps including dropping any SNP that has more than 5% missing data, imputing the missing values in each SNP with its mode, dropping SNPs with minor allele frequency < 0.01, and screening out SNPs violating the Hardy-Weinberg equilibrium (p-value $< 10^{-6}$). Eventually, $459,588$ SNPs were remained in the dataset for further analysis.

4.5.2 Statistical Analysis and Results

Our problem of interest is to perform GWAS on the log quantile curve of the FA measure. We fitted model (4.1) with covariates including an intercept term, a specific SNP, age, gender, and the top 5 genetic principal components.

To reduce the computational cost of wild bootstrap, we developed an efficient strategy to approximate the p-value of each SNP with different MAFs. In the real data analysis, we considered a pool of SNPs consisting of 7 MAF groups including MAF$\in (0.01,0.03]$, MAF$\in (0.03,0.05]$, MAF$\in (0.05,0.1]$, MAF$\in (0.1,0.2]$, MAF$\in (0.2,0.3]$, MAF$\in (0.3,0.4]$ and MAF$\in (0.4,0.5]$. Each MAF group contains 100 SNPs. For each SNP, we generated 100 wild bootstrap samples. In total, we obtained 10,000 bootstrapped test statistics for each category. Based on the pooled bootstrapped samples, we adopted the method proposed in [106] to approximate the null distribution of the test statistics by a mixed χ^2 distribution of form $a\chi^2(\nu)+b$. Specifically, we matched the first three moments of the bootstrap statistics and those of the mixed chi-square distribution.

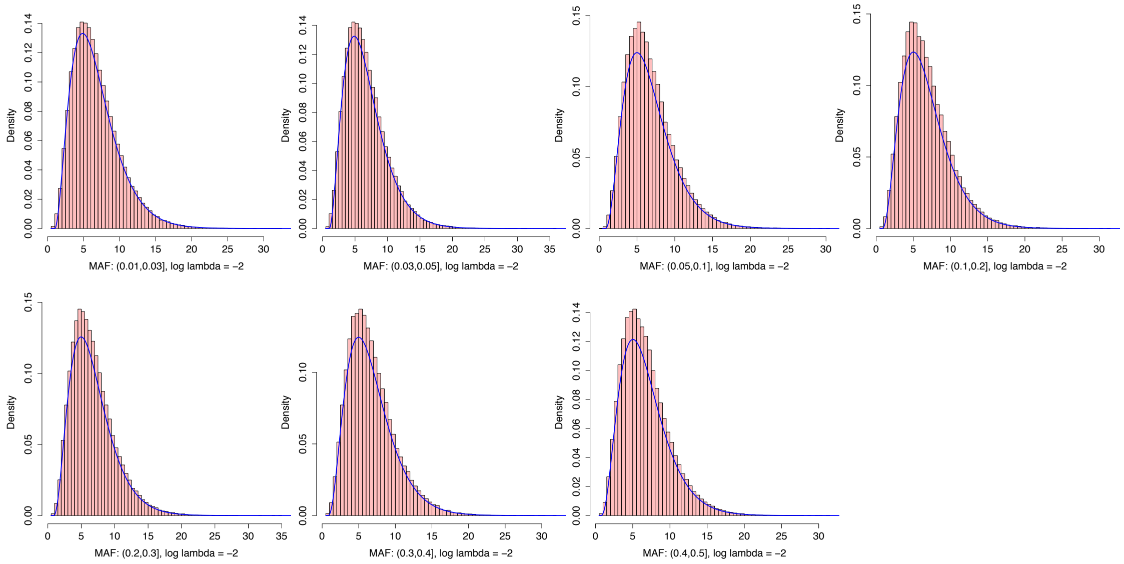


Figure 4.3: Application of PFGT to the UK Biobank data: histograms of wild bootstrap statistics of different MAF intervals when $\lambda_n = 10^{-2}$, along with their density approximations by mixed χ^2 distribution.

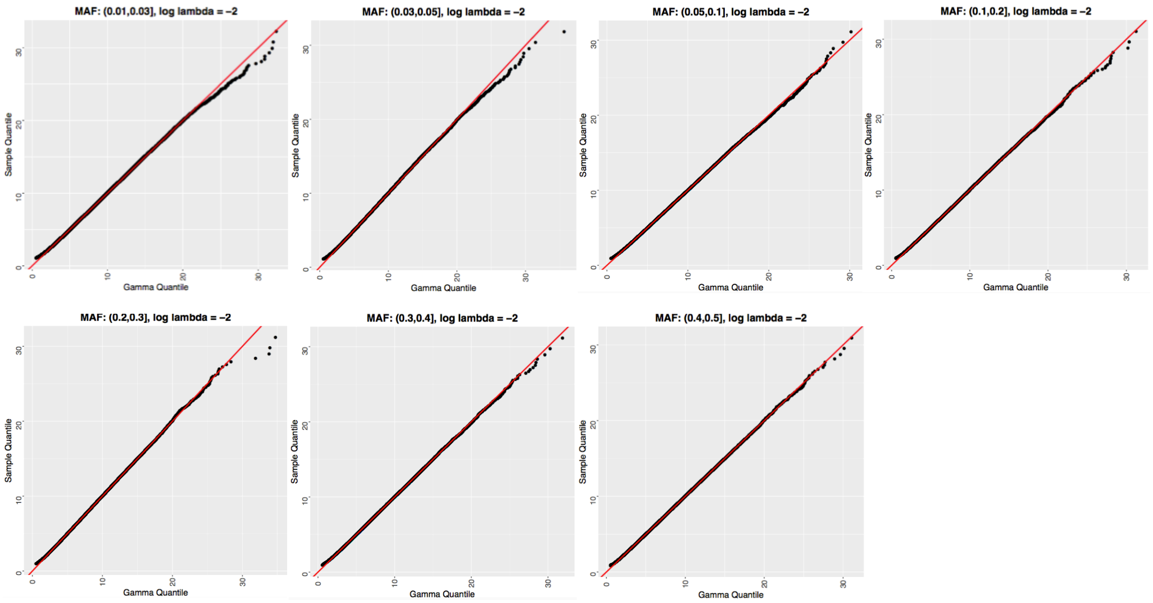


Figure 4.4: Application of PFGT to the UK Biobank data: QQ plots of wild bootstrap statistics of different MAF intervals when $\lambda_n = 10^{-2}$.

The histograms of wild bootstrap statistics with fitted mixed χ^2 distributions and the QQ-plots for $\lambda_n = 10^{-2}$ are shown in Figure 4.3 and Figure 4.4 as an example. The approximation for other λ_n

values shows very similar pattern. It can be seen that the mixed χ^2 approximation works reasonably well for a wide range of MAFs. To obtain a single p-value in GWAS analysis, we first matched each SNP to its closest MAF group in the pool. Then the corresponding p-value is calculated using the approximated distribution of that MAF group with λ_n chosen through (4.17).

We demonstrate the Manhattan plot and the QQ plot of the GWAS results in Figure 4.5. The top 10 loci along with their p-values are summarized in Table 4.1. As can be seen, no genome-wide significant marker (p-value $< 1.08 \times 10^{-7}$) is observed. Additionally, five locus exceed the suggestive genome-wide association threshold (p-value $< 5 \times 10^{-6}$). Among the top locus, CAMK2N1 plays an important role in long-term potentiation, which is a process closely related to learning and memory [107]. ZFP36L1, CEP128, HAS2 and EVI5 are risk genes implicated by certain neurodegenerative diseases [108, 109, 110, 111]. MSI2 gene is known to be related to the proliferation and maintenance of stem cells in the central nervous system [112].

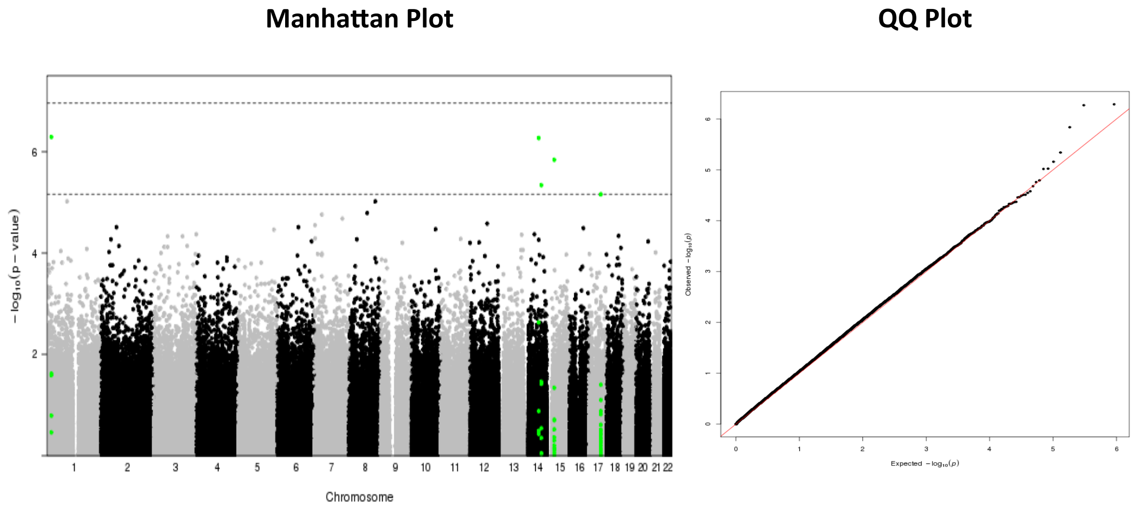


Figure 4.5: Application of PFGT to the UK Biobank data: Manhattan plot and QQ plot of the $-\log_{10} p$-values of 450,899 SNPs.

Table 4.1: Application of PFGT to the UK Biobank data: Top 10 SNPs from GWAS and their nearest genes

SNP	Chr	p-value	Gene
rs6663450	1	5.15E-07	CAMK2N1
rs11158764	14	5.37E-07	ZFP36L1
rs2339157	15	1.45E-06	FMN1
rs143406098	14	3.87E-06	CEP128
rs17821769	17	4.93E-06	MSI2
rs79320696	8	9.47E-06	HAS2
rs72722496	1	9.61E-06	EVI5
rs893282	8	1.61E-05	RALYL
rs73086843	7	1.74E-05	HERPUD2
s55783991	7	2.10E-05	CPA4

4.6 Discussion

We developed a powerful global test statistic for functional responses (PFGTS) to efficiently perform genome-wide association analysis on functional traits. A varying coefficient model was adopted to characterize the spatial smoothness and correlation structure. Then we introduced a functional projection model to reduce dimension. An optimal functional projection direction was selected to maximize the asymptotic signal-to-noise ratio with ridge penalty, which was derived from the hypothesis of interest. The asymptotic distribution of the test statistic was studied systematically and we provided a strategy to adaptively select the optimal tuning parameter to maximize the statistical power. Simulation examples showed that the proposed method outperformed existing state-of-the-art methods in functional data inference. We also applied the method to a genome-wide association analysis of DTI data in UK Biobank dataset.

As a continuation of this work, it is interesting and important to investigate optimal test procedures for other statistical inference problems of parametric and nonparametric models using dimension reduction techiniques and power maximization framework, for example, inference on the transformed measurements [113], test of distributional differences [114], test of independence

[115, 116], test of goodness-of-fit [117] and many others [118, 119].

CHAPTER 5: ADAPTIVE PROJECTION REGRESSION MODEL FOR HIGH DIMENSIONAL DATA WITH DEPENDENT COVARIANCE STRUCTURE

5.1 Introduction

Multivariate responses are frequently acquired in neuroimaging research to characterize brain structure and function [120, 71]. For instance, in a region-of-interest (ROI) analysis, brain measures are calculated for different regions/region pairs, which delineates local brain features/connectivity properties [19, 121]. In this chapter, our primary interest is to identify genetic risk variants associated with multivariate imaging phenotypes. To address the problem in a general framework, we consider a popular multivariate linear model given by

$$y_i = \mathbf{B}^T x_i + e_i, \tag{5.1}$$

where $i = 1, \cdots, n$ is the subject index, $\boldsymbol{y}_i$ is a $q \times 1$ vector of imaging phenotypes, $\boldsymbol{x}_i$ is a $p \times 1$ vector of predictors including genetic markers and other covariates, $\mathbf{B} = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_q)$ is a $p \times q$ matrix of regression coefficients, and e_i is the error term following multivariate normal distribution $N(\mathbf{0}, \Sigma_e)$. Large-scale imaging genetic studies have posed some big data challenges to solve model (5.1). The dimension of imaging phenotype q usually ranges from thousands to millions, and the number of SNPs p in a genomewide study is typically around 6 million. In this chapter, we focus on the case when q is large and p is small compared to sample size n. In other words, we study the effect of a single genetic marker at each time and include only a few number of covariates in x_i in model (5.1). Particularly, we are interested in the following unit-rank hypothesis testing problem which takes the single variant association test as a special case:

$$H_0 : \mathbf{CB} = \mathbf{c}_0 \text{ versus } H_1 : \mathbf{CB} \neq \mathbf{c}_0, \tag{5.2}$$

where $\mathbf{C}$ is an $1 \times p$ matrix and $\mathbf{c}_0 = (c_{1,0}, \cdots, c_{q,0})$ is an $1 \times q$ matrix.

Existing statistical methods for multivariate phenotypes suffer from some major limitations when the number of response variables is large compared with the sample size [122, 123, 53, 124, 125]. In many dimension reduction method, such as the pseudo trait method and envelop methods, (5.1) are limited to relatively small q, i.e., $q \ll n$ [126, 127, 128, 129, 130]. Some recent developments in regularization methods, such as multiple task learning, do not provide a post-inference tool (e.g., p-values) for association analysis [49, 50, 51, 131, 132]. Alternatively, pooled association tests are designed to conduct univariate analysis on each trait and then combine marginal statistics to study global inference problems [53, 54]. Among them, Pan and co-authors have developed a class of sum of powered score (SPU) tests with good finite-sample performance in various settings [56, 133]. Most of these tests are derived under the assumption that the responses are mutually uncorrelated. However, presence of high correlation is a key feature of brain imaging phenotypes, and directly applying such methods would still suffer from power loss in certain cases. Although there are some attempts to account for the correlation structure, such as precision matrix transformation discussed in Section 2.2.3, the methods are not guaranteed to increase power.

Therefore, we develop an adaptive projection regression model (APRM) to perform statistical inference on high dimensional imaging responses with dependent structure. The major contribution of this work is,

- A projection regression model framework is introduced to reduce dimension and a global testing method is proposed to perform statistical inference on a low dimensional space.
- A multi-level test procedure is applied, which allows for flexible signal and covariance structure.
- A data-driven strategy is proposed to choose the tuning parameters for the purpose of maximizing power under alternative hypothesis.

In Section 5.2, we introduce the projection regression model (PRM) framework and propose a novel adaptive procedure (APRM) to extract the most informative directions to test (5.2). In Sections 5.3, we use simulation studies to examine the finite sample performance of APRM and compare it with existing state-of-the-art methods. In Section 5.4, we apply APRM to a GWAS analysis on the ADNI dataset. Section 5.5 concludes with some remarks.

5.2 Adaptive Projection Regression Model

As has been reviewed in Section 2.2.3, the projection regression model introduced in [59] provides a reliable framework to handle signal-covariance structure of multivariate responses y_i. In this section, we generalize the method in [59] and propose an adaptive procedure to detect signals at multiple levels.

5.2.1 Optimal Projection Direction

Let $\mathbf{w}$ be a $q \times 1$ vector denoting a projection direction, the projection regression model is given by

$$\boldsymbol{y}_{\mathbf{w},i} = \mathbf{B}_{\mathbf{w}}^T \boldsymbol{x}_i + \boldsymbol{e}_{\mathbf{w},i}, \tag{5.3}$$

where

$$\boldsymbol{y}_{\mathbf{w},i} = \mathbf{w}^T \boldsymbol{y}_i,\ \mathbf{B}_{\mathbf{w}} = \mathbf{B}\mathbf{w} \text{ and } \boldsymbol{e}_{\mathbf{w},i} = \mathbf{w}^T \boldsymbol{e}_i. \tag{5.4}$$

The key problem is to determine the choice of $\mathbf{w}$. An optimal projection direction is supposed to give the best power under alternative hypothesis. Therefore, we will take a closer look at the influence of projection direction $\mathbf{w}$ to the asymptotic properties of the test statistic.

For a given $\mathbf{w}$, we consider a projected hypothesis testing question

$$H_{w,0} : \mathbf{C}\mathbf{B}_{\mathbf{w}} = \mathbf{c}_0\mathbf{w} \text{ versus } H_{w,1} : \mathbf{C}\mathbf{B}_{\mathbf{w}} \neq \mathbf{c}_0\mathbf{w}. \tag{5.5}$$

Let $\mathbf{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n)^T$ and $\mathbf{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)^T$ be matrices of responses and covariates of all samples respectively, the ordinary least square estimate of $\mathbf{B}_{\mathbf{w}}$ is given by

$$\widehat{\mathbf{B}}_{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\sum_{i=1}^{n}\mathbf{X}^T\mathbf{Y}\mathbf{w}. \tag{5.6}$$

A wald-type statistic for hypothesis testing problem (5.5), denoted by $T_n(\mathbf{w})$ can be given as

$$T_n(\mathbf{w}) = \frac{\mathbf{w}^T\widehat{\boldsymbol{\delta}}^T\Sigma_C^{-1}\widehat{\boldsymbol{\delta}}\mathbf{w}}{\mathbf{w}^T\Sigma_e\mathbf{w}}, \tag{5.7}$$

where $\widehat{\boldsymbol{\delta}} = \mathbf{C}\widehat{B} - \mathbf{c}_0$ and $\Sigma_C = \mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T$. Under alternative hypothesis, let $\boldsymbol{\delta} = \mathbf{C}\mathbf{B} - \mathbf{c}_0$, the signal to noise ratio of $T_n(\mathbf{w})$ can be calculated as

$$\mathrm{SNR}[T_n(\mathbf{w})] = \frac{\mathbb{E}[T_n|(\mathbf{w})|\mathbf{X}]}{\sqrt{\mathrm{Var}[T_n(\mathbf{w})|\mathbf{X}]}} = \frac{1 + (\mathbf{w}^T\boldsymbol{\delta}^T\Sigma_C^{-1}\boldsymbol{\delta}\mathbf{w})/(\mathbf{w}^T\Sigma_e\mathbf{w})^{-1}}{\sqrt{2 + 4(\mathbf{w}^T\boldsymbol{\delta}^T\Sigma_C^{-1}\boldsymbol{\delta}\mathbf{w})/(\mathbf{w}^T\Sigma_e\mathbf{w})^{-1}}}, \tag{5.8}$$

in which $\mathbb{E}[\cdot|\mathbf{X}]$ and $\mathrm{Var}[\cdot|\mathbf{X}]$ denote expectation and variance conditional on $\mathbf{X}$. In equation (5.8), SNR$[T_n(\mathbf{w})]$ increases with the following quantity

$$\frac{\mathbf{w}^T \boldsymbol{\delta}^T \Sigma_C^{-1} \boldsymbol{\delta} \mathbf{w}}{\mathbf{w}^T \Sigma_e \mathbf{w}}. \tag{5.9}$$

Therefore, an optimal projection direction $\mathbf{w}$ for the testing problem (5.2) can be given as

$$\mathbf{w}^* = \max_{\mathbf{w}} \frac{\mathbf{w}^T \boldsymbol{\delta}^T \Sigma_C^{-1} \boldsymbol{\delta} \mathbf{w}}{\mathbf{w}^T \Sigma_e \mathbf{w}}. \tag{5.10}$$

For unit-rank problem, the above equation has a unique solution up to a scalar factor given by

$$\mathbf{w}^* \propto \Sigma_e^{-1} \boldsymbol{\delta}. \tag{5.11}$$

The estimates of $\boldsymbol{\delta}$ and Σ_e are critical to calculating the optimal projection direction. However, consistently estimating regression coefficients and the covariance matrix is a challenging issue in high dimensional setting due to noise contamination. Therefore, we consider the following procedures to obtain reliable estimates of $\boldsymbol{\delta}$ and Σ_e for the testing problem.

5.2.2 Independent Screening Procedure

As discussed in [134, 52, 135, 136] and many others, the ordinary least square estimate $\widehat{\boldsymbol{\delta}}$ contains a lot of noise in high dimensional models. When the true signal is sparse, using $\widehat{\boldsymbol{\delta}}$ directly in the test statistic may cause severe power loss. To deal with this issue, sparse regularization method should applied to remove the non-signal dimensions in $\widehat{\boldsymbol{\delta}}$. Here, we adopt a fast independent screening procedure as introduced in [137] using marginal test statistics. For a given dimension j, where $j = 1, \cdots, q$, a univariate statistic to test marginal hypothesis

$$H_{j,0} : \mathbf{C}\boldsymbol{b}_j = c_{j,0} \text{ v.s. } H_{j,1} : \mathbf{C}\boldsymbol{b}_j \neq c_{j,0} \tag{5.12}$$

is calculated as

$$F_{n,j} = \frac{\hat{\boldsymbol{\delta}}_j^T \Sigma_C^{-1} \hat{\boldsymbol{\delta}}_j}{\hat{\sigma}_{e,jj}^2}, \tag{5.13}$$

where $\hat{\boldsymbol{\delta}}_j = \mathbf{C}\widehat{\boldsymbol{b}}_j - \boldsymbol{c}_{j,0}$ and $\hat{\sigma}_{e,jj}^2$ is the j-th diagonal component of sample covariance matrix $\hat{\Sigma}_e$. For a given threshold λ_n, we select the candidate signal index set as,

$$S_1 = \{j : 1 \leq j \leq q, F_{n,j} > \lambda_n\}, \tag{5.14}$$

and denote the non-signal index set as $S_0 = \{1, \cdots, q\} \backslash S_1$. To select signals in $\boldsymbol{\delta}$ at multiple levels, it is not necessary to try all possible thresholds. Alternatively, we sort the marginal statistics in decreasing order as $\{T_{n,(1)} \geq T_{n,(2)} \geq \cdots \geq T_{n,(q)}\}$, and select the top L dimensions to construct candidate signal set,

$$\widehat{S}_{1,L} = \{j : 1 \leq j \leq q, F_{n,j} \geq T_{n,(L)}\}. \tag{5.15}$$

Then a thresholded estimate of $\boldsymbol{\delta}$ can be given as

$$\widehat{\boldsymbol{\delta}}_L = (\hat{\boldsymbol{\delta}}_{1,L}, \cdots, \hat{\boldsymbol{\delta}}_{j,L}, \cdots, \hat{\boldsymbol{\delta}}_{q,L}) \text{ with } \hat{\boldsymbol{\delta}}_{j,L} = \hat{\boldsymbol{\delta}}_j I\{j \in \widehat{S}_{1,L}\}. \tag{5.16}$$

It should be emphasized that, even though S_0 do not contain useful signal for the testing problem (5.2), we still need to include these dimensions in the optimization equation (5.10) to calculate projection direction $\mathbf{w}^*$. Their contribution can be demonstrated by the following example:

Example 5.2.1. *Let $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \mathbf{0}_{q_0})$ where $\boldsymbol{\delta}_1$ is a $q_1 \times 1$ vector of true signals and $\mathbf{0}_{q_0}$ is a $q_0 \times 1$ zero vector. Let $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_0)$ and $\Sigma_e = \left(\begin{smallmatrix} \Sigma_{11} & \Sigma_{10} \\ \Sigma_{01} & \Sigma_{00} \end{smallmatrix}\right)$ denote the index partition of $\mathbf{w}$ and Σ_e respectively, the optimization equation (5.10) can be rewritten as*

$$w_1^* = \underset{\mathbf{w}_1}{\operatorname{argmax}} \frac{\mathbf{w}_1^T \boldsymbol{\delta}_1^T \Sigma_C^{-1} \boldsymbol{\delta}_1 \mathbf{w}}{\mathbf{w}_1^T \Sigma_{11 \cdot 0} \mathbf{w}_1} \text{ and } \mathbf{w}_0^* = -\Sigma_{00}^{-1} \Sigma_{01} \mathbf{w}_1, \tag{5.17}$$

where $\Sigma_{11\cdot 0} = \Sigma_{11} - \Sigma_{10}\Sigma_{00}^{-1}\Sigma_{01}$ is the conditional covariance matrix.

When the responses from signal set S_1 are highly correlated with responses from non-signal set S_0, $\mathbf{w}_0^*$ helps to reduce the variance of $T_n(\mathbf{w}^*)$.

5.2.3 Block-wise Covariance Estimation

The precision matrix Σ_e^{-1} plays an essential role to reduce noise level in the projection test statistic $T_n(\mathbf{w})$. Consistently estimating the precision matrix in high dimensional space is a difficult problem in general. Many regularization techniques have been proposed for this purpose, which usually require certain structural conditions, such as sparsity assumption or factor model assumption [138, 139, 140, 61], and the methods tend to introduce too much bias in finite sample estimation. Here, we develop a sequential estimation procedure to identify index blocks with limited size. Although an optimal projection may not be obtained, this procedure allows us to capture the major dependent structure in Σ_e for noise shrinkage and to reduce the total number of parameters to be estimated. For a given value B representing the maximum block size allowed, we estimate a block-wise covariance structure as follows:

Algorithm 5.2.1.

(i) Let $S = \{1, \cdots, q\}$ denote the responses index. An initial thresholding [141] is applied to the sample correlation matrix $\hat{R} = (\hat{\rho}_{j,j'})$ to remove noisy terms and weak correlations. Let

$I_0 = (1_{|\hat{\rho}_{j,j'}|>2s\log q/n})$ be the indexing matrix for thresholding, then the post-screening correlation matrix can be calculated as $\tilde{R} = (\tilde{\rho}_{j,j'})_{q\times q} = \hat{R}\cdot I_0$.

(ii) Start from the pair $j, j' \in S = \{1,\cdots,q\}$ with the largest absolute correlation value $|\rho_{j,j'}|$ larger than 0, and take them as the first two elements in index set $I_{(1)}$.

(iii) When the number of elements is not larger than B, find the largest nonnegative $\|\tilde{\rho}_{j,j'}\|$ among all pairs satisfying $j \in S\backslash I_{(1)}$ and $j' \in I_{(1)}$ and include j in $I_{(1)}$.

(iv) Repeat step 3 until the size of $I_{(1)}$ reaches B or no element can be added.

(v) Remove $I_{(1)}$ from S. Then repeat step(ii) - step(iv) to sequentially obtain $I_{(2)}, I_{(3)}, \cdots$ until S becomes empty set.

Finally, we obtain an estimate of covariance matrix with block-diagonal structure as

$$\hat{\Sigma}_{e,B} = (\hat{\sigma}_{e,j,j'}[\sum_k I\{j \in I_{(k)} \cap j' \in I_{(k)}\}])_{q\times q}. \tag{5.18}$$

When B is relatively small, $\hat{\Sigma}_{e,B}^{-1}$ can be used to calculate $\mathbf{w}^*$ in (5.11).

The above algorithm uses marginal correlation to construct covariance blocks. There are other methods designed to estimate block-wise covariance structure more accurately [141, 142, 143], but such improvement is not the major focus of this chapter.

5.2.4 Projected Test Statistics and an Adaptive Inference Procedure

Given the number of selected responses L and maximum block size B, the projection direction is estimated as

$$\widehat{\mathbf{w}}_{L,B} = \hat{\Sigma}_{e,B}^{-1}\widehat{\delta}_L. \tag{5.19}$$

Then a projected test statistic induced by $\widehat{\mathbf{w}}_{L,B}$ can be calculated as

$$T_n(L,B) = T_n(\widehat{\mathbf{w}}_{L,B}) = \frac{\widehat{\boldsymbol{\delta}}_L^T \hat{\Sigma}_{e,B}^{-1} \widehat{\boldsymbol{\delta}}_L}{\sigma_C^2}, \tag{5.20}$$

where σ_C^2 is a scalar value equals to Σ_C.

To detect signal at multiple levels and to allow a flexible covariance structure, different choices of L and B are applied. We then introduce an adaptive framework similar to the aSPU method as reviewed in Section 2.2.2 to select optimal L and B. The specific inference procedure is summarized as follows:

Algorithm 5.2.2.

(i) Marginal test statistics are calculated and sorted in decreasing order as $\{T_{n,(1)} \geq \cdots \geq T_{n,(q)}\}$. Independence screening is applied and the top L responses are selected to form candidate signal set $\widehat{S}_{1,L}$. Then a thresholded estimate $\widehat{\boldsymbol{\delta}}_L = (\hat{\delta}_{1,L}, \cdots, \hat{\delta}_{q,L})$ is calculated with $\hat{\delta}_j I\{j \in \widehat{S}_{1,L}\}$, where $j = 1, \cdots, q$ and $I\{\cdot\}$ is the indicator function.

(ii) A block structure is imposed on covariance matrix Σ_e with maximum block size B, and covariance Σ_e is estimated by algorithm 5.2.1 as $\widehat{\Sigma}_{e,B}$.

(iii) For fixed L and B, the projection directions $\{\hat{\mathbf{w}}_{L,B}\}$ is estimated from (5.19) by replacing $\boldsymbol{\delta}$ and Σ_e with $\widehat{\boldsymbol{\delta}}_L$ and $\widehat{\Sigma}_{e,B}$ and the projected test statistic is calculated as $T_n(\hat{\mathbf{w}}_{L,B})$.

(iv) Permutation resampling is performed to obtain G samples under null distribution $\{\boldsymbol{y}_i^{(g)}\}$. Permutation statistics $T_n(\hat{\mathbf{w}}_{L,B}^{(g)})$ are calculated by repeating step(i) and step(iii) while using $\widehat{\Sigma}_{e,B}$ estimated from original samples. The p-value of $T_n(\hat{\mathbf{w}}_{L,B})$ is then given as $\sum_{g=1}^{G} I\{T_n(\hat{\mathbf{w}}_{L,B}^{(g)}) \geq T_n(\hat{\mathbf{w}}_{L,B})\}/G$.

(v) The minimum p-value among all $T_n(\hat{\mathbf{w}}_{L,B})$s is taken as the adaptive global test statistic, i.e.,

$$aT_n = \min_{l,B} \hat{P}[T_n(\hat{\mathbf{w}}_{L,B})], \tag{5.21}$$

where $\hat{P}(\cdot)$ denotes the estimated p-value of a statistic from permutation.

The null distribution of (5.21) can be approximated by the same set of permutation samples directly.

5.3 Simulation Studies

5.3.1 Setup

In this section, we use numerical simulations to evaluate the performance of the proposed method by testing the difference in two sample means in high dimensional setting. Data is generated from model $x_{i1} \sim N(\mathbf{0}, \Sigma_e), i = 1, \cdots, n_1$ and $x_{j2} \sim N(\mu, \Sigma_e), j = 1, \cdots, n_2$ with sample size $n_1 = n_2 = 100$, trait dimension $q = 400$. The hypothesis question is to test $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$. The number of nonzero elements in μ takes value $r = 2, 40, 100$, representing sparse, moderate and dense signal respectively, and the specific location of nonzero entry are generated randomly from discrete uniform distribution on $\{1, \cdots, q\}$. To evaluate APRM for different covariance structures, the following three types of Σ are considered,

- Case 1: (Independent Structure) $\Sigma_e = I_q$.
- Case 2: (Block-wise Compound Structure) $\Sigma_e = (0.6 1_5 1_5^T + 0.4 I_5) \otimes I_{80}$.
- Case 3: (AR-1 Structure) $\Sigma_e = (0.8^{|i-j|})$.

In each setting, we will evaluate the performance of PRM with both fixed block size $B = 1, 2, 5, 10$ and with the adaptive selection strategy (APRM). CQT and aSPU are also examined in our simulation as a comparison. In each scenario, 1000 simulated data sets are generated to evaluate type I and type II error. To calculate p-values, 1000 permutation samples are generated in each run.

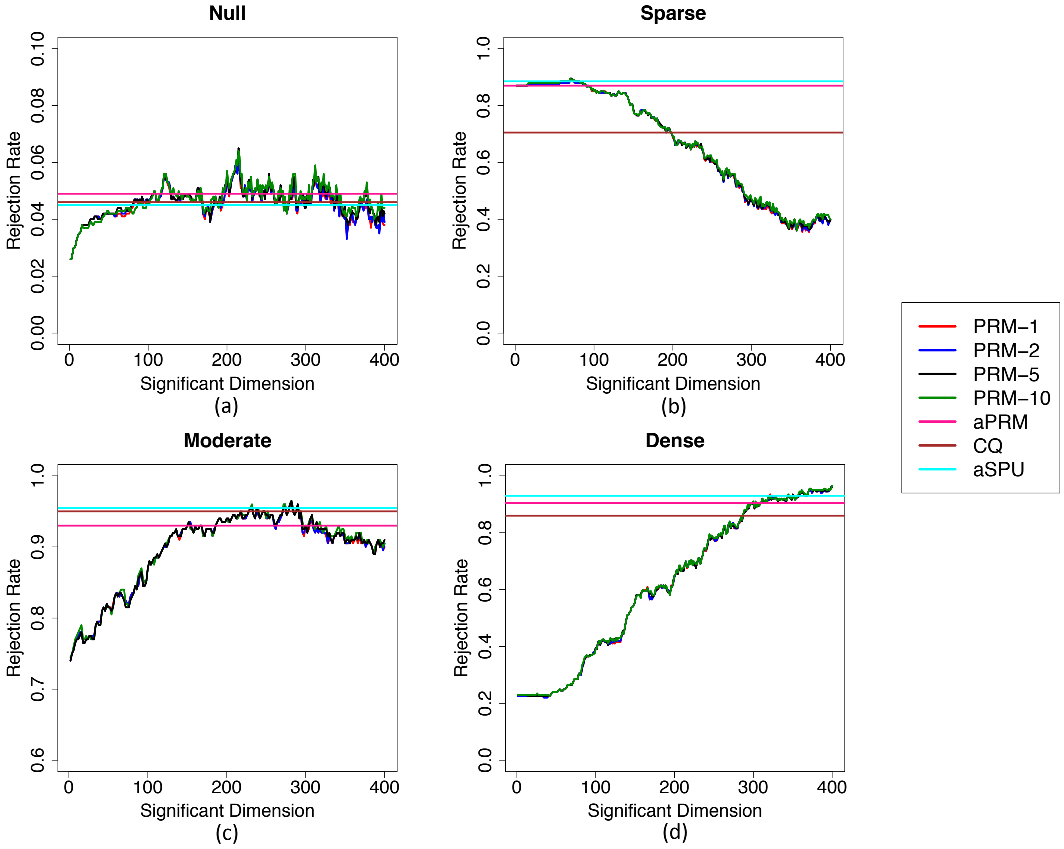


Figure 5.1: Simulation results of APRM for independent structure $\Sigma_e = I_q$: PRM is evaluated under four choices of maximum block size $B = 1, 2, 5, 10$, as well as by adaptive selection strategy (APRM). Results for aSPU and CQT are also presented as comparisons.

5.3.2 Results

Rejection rate of the three cases are given in Figure 5.1 – Figure 5.3. For independent case, all three methods have comparable power, and aSPU performs slightly better than the other two methods. Among all choices of block sizes for PRM, $B = 1$ gives the best performance, which is consistent with the ground truth. Moreover, it can be observed that a large B does not have much influence on the results even when the block size is misspecified. In case 2 and case 3, APRM has the best performance in all scenarios, especially when the signal is non-sparse. The multi-level adaptive strategy (APRM) has slight lower power than largest power achieved, yet still shows substantial power increase compared with aSPU and CQT. This indicate the effectiveness of our methods to detect signal of interest in presence of high correlations.

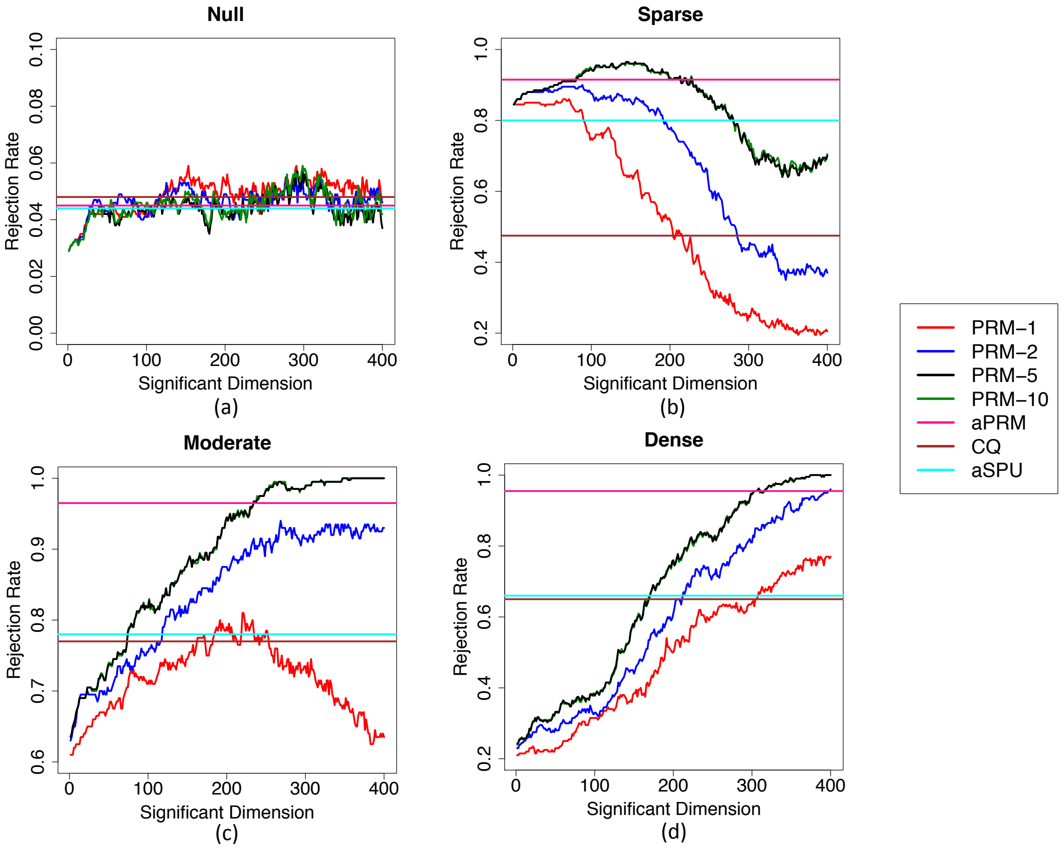


Figure 5.2: Simulation results of APRM for block-wise compound structure $\Sigma_e = (0.6 1_5 1_5^T + 0.4 I_5) \otimes I_{80}$: PRM is evaluated under four choices of maximum block size $B = 1, 2, 5, 10$, as well as by adaptive selection strategy (APRM). Results for aSPU and CQT are also presented as comparisons.

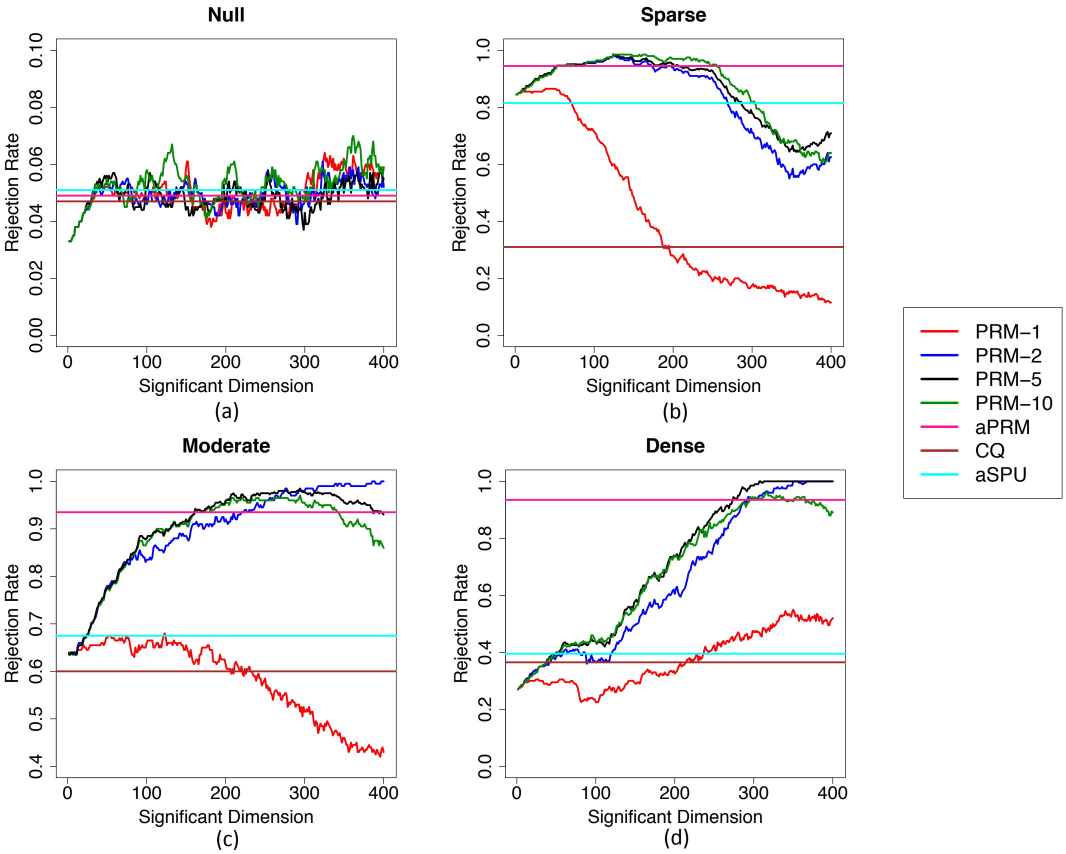


Figure 5.3: Simulation results of APRM for AR-1 structure $\Sigma_e = (0.8^{|i-j|})$: PRM is evaluated under four choices of maximum block size $B = 1, 2, 5, 10$, as well as by adaptive selection strategy (APRM). Results for aSPU and CQT are also presented as comparisons.

5.4 Alzheimer's Disease Neuroimaging Initiative Data Analysis

To illustrate the usefulness of APRM, we considered anatomical MRI data collected at the baseline by the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. Our primary interest is to perform a genome-wide search for snp sets that are significantly associated with the brain volume of 93 regions of interest (ROIs).

5.4.1 Alzheimer's Disease Neuroimaging Initiative

Alzheimer's disease (AD) is a chronic, irreversible neurodegenerative disease that results in fatal deterioration of brain tissues and loss of mental functions. The primary aim of Alzheimer's Disease Neuroimaging Initiative (ADNI) study is to investigate the influence of genetic predispositions and environmental exposures to the development of the disease and to identify biomarkers that can predict the risk and progress of AD. These would eventually inspire novel inventions in disease prevention, diagnosis and effective treatments. For the most up-to-date information, please find at www.adni.loin.usc.edu.

Magnetic resonance imaging (MRI) scans were acquired for 708 subjects (164 patients with Alzheimer's disease, 346 patients with mild cognitive impairment, and 198 normal control subjects) in the ADNI-1 dataset, from a 1.5 TMRI scanners using a sagittal MPRAGE sequence. The image data was processed with standard quality control steps including anterior commissure (AC) and posterior commissure (PC) correction, skull-stripping, cerebellum removal, intensity inhomogeneity correction, registration and segmentation [144]. Then 93 region of interests (ROIs) were labeled using the atlas of the human brain provided by [145] and the volume of each region was computed for each subject.

The Human 610-Quad BeadChip was used to genotype $620,901$ SNPs for 818 subjects. The genetic data was processed by standard quality control procedure using PLINK. Samples with call rates less than 90%, Caucasian ancestry outliers and unexpected relatedness were excluded from the dataset. We also removed genetic markers with Hardy-Weinberg equilibrium p-value less than 10^{-6}, call rate less than 95% and minor allele frequency (MAF) smaller than 5%. Population stratification was assessed using PCA [87]. Eventually, 747 subjects and $501,584$ SNPs remained in the dataset.

5.4.2 Data Analysis and Results

Our primary goal is to perform a genome-wide search for SNPs significantly associated with brain volumn trait. In model (5.1), volumetric measure of 93 ROIs were taken as $\boldsymbol{y}_i$ and several demographic, clinical and genetic variables were added as covariates, including an intercept, age, gender, whole brain volume and the genotype of a bi-allelic SNP coded as $0, 1$ or 2. The first five genetic principal components were also included in $\boldsymbol{x}_i$ to adjust for population stratification. The maximum block size B is allowed to take value from $1, 2, 5, 10, 20$. To accurately estimate the tail distribution of aT_n in (5.21) in a GWAS problem, the bootstrap sample size is typically around 10^8. To reduce computational cost, we gradually increase the number of permutation samples with $G = 10^3, 10^4, 10^5, 10^6, 10^7$, and 10^8. A larger G is adopted only to SNPs with p-values less than $5/G$.

We present the Manhattan plot and the QQ plot of the GWAS results in Figure 5.4. SNP rs2075650 on the chromosome 19 achieved genome-wide significance (p-value $< 1.00 \times 10^{-7}$). It is an intronic variant of the Translocase Of Outer Mitochondrial Membrane 40 (TOMM40) gene, which is a well-known gene associated with Alzheimer's disease [146]. 9 additional loci exceeded the suggestive genome-wide association threshold (p-value $< 5 \times 10^{-6}$), as summarized in Table 5.1. Besides TOMM40, SOCS3 and TMEM106B are also risk genes implicated by Alzheimer's disease [147, 148]. FBXL17, SEMA3D and PLA2G4E are suspected to be associated with other neuropsychiatric disorders as well [149, 150, 151].

Table 5.1: Application of APRM to the ADNI data: Top 10 SNPs from GWAS and their nearest genes

SNP	Chr	p-value	Gene
rs2075650	19	7.10E-08	TOMM40
rs3818698	6	3.80E-07	LOC105378146
rs8074003	17	4.40E-07	SOCS3
rs10058163	5	6.50E-07	FBXL17
rs2178115	7	1.50E-06	SEMA3D
rs10247990	7	3.00E-06	TMEM106B
rs776691	15	8.40E-06	PLA2G4E
rs11941079	4	7.30E-06	DCK
rs34298746	12	4.50E-06	PARP11
rs7001747	8	4.96E-06	KHDRBS3

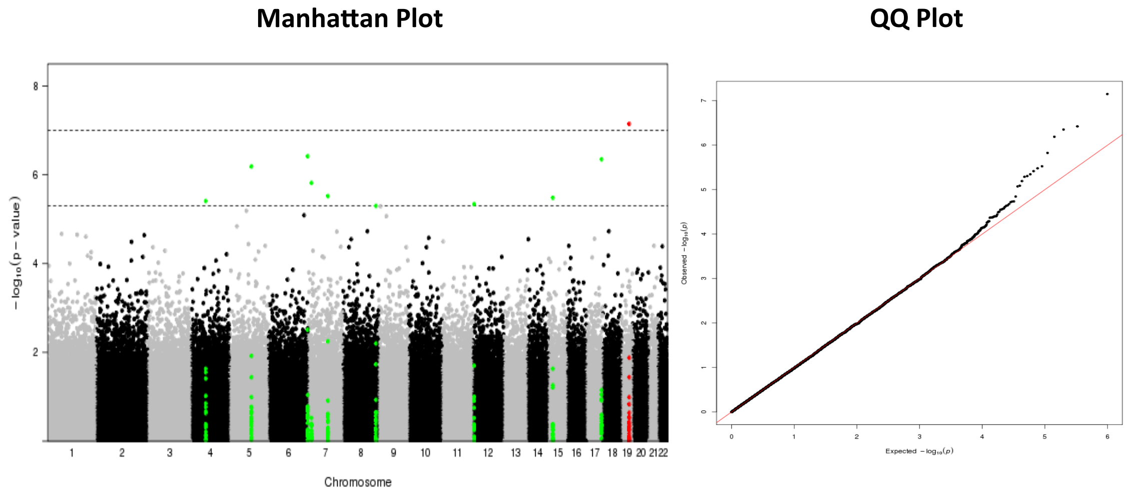


Figure 5.4: Application of APRM to the ADNI data: Manhattan plot and QQ plot of the $-\log 10(p\text{-values})$ of 501,584 SNPs from GWAS.

5.5 Discussion

In this chapter, we developed an adaptive projection regression model (APRM) to perform hypothesis testing on a set of covariates in multivariate regression modeling for a large number of responses with dependent covariance structure. We proposed a dimension reduction strategy by taking advantage of correlations among multivariate responses. A fast and efficient screening procedure base on marginal statistics was first performed to select candidate signal set. Then projection transformation was adopted to maximize asymptotic signal-to-noise ratio. Numerical simulations showed that APRM outperforms many other state-of-the-art methods when dealing with dependent data structure.

APPENDIX A: TECHNICAL DETAILS OF CHAPTER 3

In this chapter, we give the proof to the main theoretical results.

A.1 Lemmas

First, we give several lemmas as the foundation. Positive constants $C_1, C_2, C_3, \cdots$ appeared in the lemmas may vary at each occurrence.

For two positive definite symmetric matrices of size $L \times L$, denoted as Σ and $\widehat{\Sigma}$ respectively, let $\{\tau\}_{l=1}^{L}$ and $\{\widehat{\tau}\}_{l=1}^{L}$ be their eigenvalues in decreasing order, then $\tau_l - \widehat{\tau}_l$ can be bounded by the following lemma:

Lemma A.1.1. *Weyl′s Theorem*

$$\max_{1 \le l \le L} |\tau_l - \widehat{\tau}_l| \le \|\Sigma - \widehat{\Sigma}\|_2. \tag{A.1}$$

In addition, let $\{\mathbf{v}_l\}_{l=1}^{L}$ and $\{\widehat{\mathbf{v}}_l\}_{l=1}^{L}$ be the corresponding eigenvectors, $\mathbf{v}_l - \widehat{\mathbf{v}}_l$ can be bounded by the following $\sin\theta$ theorem [60]:

Lemma A.1.2. *Davis-Kahan′s* $\sin\theta$ *Theorem*

$$\max_{1 \le l \le L} \|\mathbf{v}_l - \widehat{\mathbf{v}}_l\|_2 \le \frac{\sqrt{2}\|\Sigma - \widehat{\Sigma}\|_2}{\min\{|\widehat{\tau}_{l-1} - \tau_l|, |\tau_l - \widehat{\tau}_{l+1}|\}}. \tag{A.2}$$

We also need the following lemmas from [152] to bound the estimation error of $\widehat{\beta}_m(s)$ and $\widehat{\eta}_{i,m}(s)$.

Lemma A.1.3. *For a bounded function class $\mathscr{F}$ composed of functions $f : \mathscr{X} \mapsto [0,1]$, if there* *exist a positive constant K and an integer v such that,*

$$\sup_Q N[\varepsilon,\mathscr{F},L_2(Q)] \le (\frac{K}{\varepsilon})^v, \forall 0<\varepsilon<\varepsilon_0, \tag{A.3}$$

then for every $x > C_1$,

$$P(\sup_{f\in\mathscr{F}} \sqrt{n}|\frac{1}{n}\sum_{i=1}^n f(x_i) - Ef(X)| \ge x) \le C_2 K^{2v} (\frac{x}{\sqrt{v}})^v \exp\{-C_3 x^2\},$$

where C_1, C_2, C_3 are positive constants depending on v and ε_0.

Lemma A.1.4. *Let $\mathscr{F} = \{f : \mathscr{X} \mapsto [0,1]\}$ be a class of bounded functions that satisfy (A.3). Then for every σ with $\sup_{f\in\mathscr{F}} Var[f(X)] \le \sigma^2 < +\infty$ and $x/\sigma > C_1$, we have*

$$P(\sup_{f\in\mathscr{F}} \sqrt{n}|\frac{1}{n}\sum_{i=1}^n f(x_i) - Ef(X)| \ge x) \le C_2 K^{2v} \sigma^{-2v} (\frac{x}{\sigma})^{4v} \exp\{-\frac{C_3 x^2}{2\sigma^2 + (3+x)/\sqrt{n}}\},$$

where C_1, C_2, C_3 are positive constants depending on v and ε_0.

Lemma A.1.5. *For function class $\mathscr{F}$ indexed by set $[0,S]$, i.e., $\mathscr{F} = \{f_s : s \in [0,S]\}$, if there exist a distance measure d on $[0,S]$ and a constant F such that,*

$$|f_s(x) - f_t(x)| \le Fd(s,t),\ \forall s,t \in [0,S],$$

then we have the following uniform bound for the covering number of $\mathscr{F}$,

$$\sup_Q N[\varepsilon,\mathscr{F},L_2(Q)] \le N(\varepsilon/F,[0,S],d).$$

The following lemma establishes a uniform convergence bound for $\widehat{\mu}_m(s_m)$ for all $m = 1, \cdots, M$.

Lemma A.1.6. *When assumptions 3.1 - 3.10 hold, $\Delta\widehat{\mu}_m(s_m) = \widehat{\mu}_m(s_m) - \mu_m(s_m)$ can be bounded by*

$$\begin{aligned}
\max_m \sup_{s_m} |\Delta\widehat{\mu}_m(s_m)| &= O(h_1) + O_p(\sqrt{\frac{logM}{n}}) + O_p(\sqrt{\frac{logh_1^{-1}}{n}}) \\
&\overset{def}{=} O_p(\omega_1).
\end{aligned} \tag{A.4}$$

Proof. Let $\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i$, $\bar{\eta}_{\cdot,m}(s_m) = \frac{1}{n}\sum_{i=1}^n \eta_{i,m}(s_m)$ and $\bar{e}_{\cdot,m}(s_m) = \frac{1}{n}\sum_{i=1}^n e_{i,m}(s_m)$, $\Delta\widehat{\mu}_m(s_m)$ can be decomposed as

$$\begin{aligned}
\Delta\widehat{\mu}_m(s_m) &= \sum_{w=1}^W [\mu_m(s_{m,w}) - \mu_m(s_m)]K_{h_1}(s_{m,w} - s_m) / \sum_{w=1}^W K_{h_1}(s_{m,w} - s_m) \\
&+ \sum_{w=1}^W \bar{\boldsymbol{x}}^T \boldsymbol{\beta}_m(s_{m,w})K_{h_1}(s_{m,w} - s_m) / \sum_{w=1}^W K_{h_1}(s_{m,w} - s_m) \\
&+ \sum_{w=1}^W \bar{\eta}_{\cdot,m}(s_{m,w})K_{h_1}(s_{m,w} - s_m) / \sum_{w=1}^W K_{h_1}(s_{m,w} - s_m) \\
&+ \sum_{w=1}^W \bar{e}_{\cdot,m}(s_{m,w})K_{h_1}(s_{m,w} - s_m) / \sum_{w=1}^W K_{h_1}(s_{m,w} - s_m) \\
&\overset{def}{=} \Delta\widehat{\mu}_m^{(1)}(s_m) + \Delta\widehat{\mu}_m^{(2)}(s_m) + \Delta\widehat{\mu}_m^{(3)}(s_m) + \Delta\widehat{\mu}_m^{(4)}(s_m).
\end{aligned}$$

Then the above four terms can be bounded respectively as follows.

$$\max_m \sup_{s_m} |\Delta\widehat{\mu}_m^{(1)}(s_m)| \leq h_1 \max_m \sup_{s_m} |\partial\mu_m(s_m)|.$$

$$\begin{aligned}\max_m \sup_{s_m} |\Delta \widehat{\mu}_m^{(2)}(s_m)| &= \max_m \sup_{s_m} |\bar{\boldsymbol{x}}^T \sum_{w=1}^W \boldsymbol{\beta}_m(s_{m,w}) K_{h_1}(s_{m,w}-s_m) / \sum_{w=1}^W K_{h_1}(s_{m,w}-s_m)| \\ &\leq \|\bar{\boldsymbol{x}}\|_2 \max_m \sup_{s_m} \| \sum_{w=1}^W \boldsymbol{\beta}_m(s_{m,w}) K_{h_1}(s_{m,w}-s_m) / \sum_{w=1}^W K_{h_1}(s_{m,w}-s_m)\|_2 \\ &= O_p(n^{-1/2}).\end{aligned}$$

To bound $\Delta \widehat{\mu}_m^{(3)}(s_m)$, we introduce a bounded functional class $\mathscr{F} = \{f_{s_m} : s_m \in [0, S_m]\}$ with $f_{s_m} = \sum_{w=1}^W \eta_{i,m}(s_{m,w}) K_{h_1}(s_{m,w}-s_m) / \sum_{w=1}^W K_{h_1}(s_{m,w}-s_m)$. There exists a positive constant C_1 such that

$$\max_m \sup_{s_m} |\frac{\partial f_{s_m}}{\partial s_m}| \leq C_1 h_1^{-1}.$$

Lemma A.1.5 shows that

$$\sup_Q N[\varepsilon, \mathscr{F}, L_2(Q)] \leq \frac{C_2}{\varepsilon h_1}.$$

Then Lemma A.1.3 gives the following bound

$$P(\max_m \sup_{s_m} |\Delta \widehat{\mu}_m^{(3)}(s_m)| \geq x) \leq C_3 M h_1^{-2} \sqrt{n} t \exp\{-C_4 n t^2\},$$

which leads to

$$\max_m \sup_{s_m} |\Delta \widehat{\mu}_m^{(3)}(s_m)| = O_p[\sqrt{\frac{\log h_1^{-1} \vee \log M}{n}}].$$

Similarly for $\Delta \widehat{\mu}_m^{(4)}(s_m)$, we define a bounded functional class as $\mathscr{G} = \{g_{s_m} : s_m \in [0, S_m]\}$ with $g_{s_m} = \sum_{w=1}^W e_{i,m}(s_{m,w}) K_{h_1}(s_{m,w}-s_m) / \sum_{w=1}^W K_{h_1}(s_{m,w}-s_m)$. Then we have

$$\sup_Q N[\varepsilon, \mathscr{G}, L_2(Q)] \leq \frac{C_4}{\varepsilon h_1} \text{ and } \max_m \sup_{s_m} \mathrm{Var}[g_{s_m}] = O[(W h_1)^{-1}].$$

Then Lemma A.1.4 gives the following bound

$$P(\max_m \sup_{s_m} |\Delta\widehat{\mu}_m^{(4)}(s_m)| \geq x) \quad \leq \quad C_5 M h_1^{-2}(Wh_1)^3(\sqrt{n}t)^4 \exp\{-\frac{C_6 n x^2}{2(Wh_1)^{-1}+(3+\sqrt{n}x)/\sqrt{n}}\}.$$

Then we have

$$\max_m \sup_{s_m} |\Delta\widehat{\mu}_m^{(4)}(s_m)| = O_p[\sqrt{\frac{\log M \vee \log Wh_1}{n}}(\sqrt{\frac{1}{Wh_1} \vee \frac{\log M}{n}})].$$

Finally, bound (A.4) can be obtained with some simplification. □

The following lemma establishes a uniform bound for $\widehat{\Sigma}_{\tilde{\eta}_m}(s_m,t_m) - \Sigma_{\tilde{\eta}_m}(s_m,t_m)$ for all $m = 1,\cdots,M$.

Lemma A.1.7. *When assumptions 3.1 - 3.10 hold,* $\Delta\Sigma_{\tilde{\eta}_m}(s_m,t_m) = \widehat{\Sigma}_{\tilde{\eta}_m}(s_m,t_m) - \Sigma_{\tilde{\eta}_m}(s_m,t_m)$ *can be bounded as*

$$\begin{aligned}\max_m \sup_{s_m,t_m} |\Delta\Sigma_{\tilde{\eta}_m}(s_m,t_m)| &= O(h_1^2) + O_p(\sqrt{\frac{\log M}{n}}) + O_p(\frac{\log h_1^{-1}}{n}) + O_p(h_2) + O_p(\sqrt{\frac{1}{Wh_2}}) \\ &\overset{def}{=} O_p(\omega_2). \qquad \text{(A.5)}\end{aligned}$$

Proof. $\Delta\Sigma_{\tilde{\eta}_m}(s_m,t_m)$ can be bounded by the following terms:

$$\begin{aligned}\max_m \sup_{s_m,t_m} |\Delta\Sigma_{\tilde{\eta}_m}(s_m,t_m)| &= \max_m \sup_{s_m,t_m} |\frac{1}{n}\sum_{i=1}^n \tilde{\eta}_{i,m}(s_m)\tilde{\eta}_{i,m}(t_m) - \Sigma_{\tilde{\eta}_m}(s_m,t_m)| \\ &+ 2\max_m \sup_{s_m,t_m} |\frac{1}{n}\sum_{i=1}^n [\widehat{\tilde{\eta}}_{i,m}(s_m) - \tilde{\eta}_{i,m}(s_m)]\tilde{\eta}_{i,m}(t_m)| \\ &+ \max_m \sup_{s_m,t_m} |\frac{1}{n}\sum_{i=1}^n [\widehat{\tilde{\eta}}_{i,m}(s_m) - \tilde{\eta}_{i,m}(s_m)][\widehat{\tilde{\eta}}_{i,m}(t_m) - \tilde{\eta}_{i,m}(t_m)]|.\end{aligned}$$

To bound the first term, we consider $\{\tilde{\eta}_m(s_m)\tilde{\eta}_m(t_m)\}$ as a functional class indexed by $(s_m, t_m) \in [0, S_m]^{\otimes 2}$, lemma A.1.3 leads to,

$$\max_m \sup_{s_m, t_m} |\frac{1}{n}\sum_{i=1}^{n} \tilde{\eta}_{i,m}(s_m)\tilde{\eta}_{i,m}(t_m) - \Sigma_{\tilde{\eta}_m}(s_m, t_m)| = O_p(\sqrt{\frac{logM}{n}}).$$

To bound the second term and the third term, we consider the following decomposition of $\widehat{\tilde{\eta}}_{i,m}(s_m) - \tilde{\eta}_{i,m}(s_m)$,

$$\begin{aligned}
\Delta\eta_{i,m}(s_m) &= \widehat{\tilde{\eta}}_{i,m}(s_m) - \tilde{\eta}_{i,m}(s_m) \\
&= \sum_{w=1}^{W}[\mu_m(s_{m,w}) - \widehat{\mu}_m(s_{m,w})]K_{h_2}(s_{m,w} - s_m)/\sum_{w=1}^{W} K_{h_2}(s_{m,w} - s_m) \\
&+ \boldsymbol{x}_i^T \sum_{w=1}^{W}[\boldsymbol{\beta}_m(s_{m,w}) - \boldsymbol{\beta}_m(s_m)]K_{h_2}(s_{m,w} - s_m)/\sum_{w=1}^{W} K_{h_2}(s_{m,w} - s_m) \\
&+ \sum_{w=1}^{W}[\eta_{i,m}(s_{m,w}) - \eta_{i,m}(s_m)]K_{h_2}(s_{m,w} - s_m)/\sum_{w=1}^{W} K_{h_2}(s_{m,w} - s_m) \\
&+ \sum_{w=1}^{W} e_{i,m}(s_{m,w})K_{h_2}(s_{m,w} - s_m)/\sum_{w=1}^{W} K_{h_2}(s_{m,w} - s_m) \\
&\overset{def}{=} \Delta\tilde{\eta}_{i,m}^{(1)}(s_m) + \Delta\tilde{\eta}_{i,m}^{(2)}(s_m) + \Delta\tilde{\eta}_{i,m}^{(3)}(s_m) + \Delta\tilde{\eta}_{i,m}^{(4)}(s_m).
\end{aligned}$$

Then the last two terms can be bounded as

$$\begin{aligned}
\max_m \sup_{s_m, t_m} |\frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}(s_m)\tilde{\eta}_{i,m}(t_m)| &\leq \max_m \sup_{s_m, t_m} |\frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(1)}(s_m)\tilde{\eta}_{i,m}(t_m)| \\
&+ O_p(1)\{\max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(2)2}(s_m) \\
&+ \max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(3)2}(s_m) \max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(4)2}(s_m)]\}^{1/2},
\end{aligned}$$

and

$$\begin{aligned}\max_m \sup_{s_m,t_m} |\frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}(s_m)\Delta\tilde{\eta}_{i,m}(t_m)| &\leq \max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(1)2}(s_m) + \max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(2)2}(s_m) \\ &+ \max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(3)2}(s_m) + \max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(4)2}(s_m).\end{aligned}$$

We then investigate the above terms individually. The term $max_m \sup_{s_m,t_m} |\frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(1)}(s_m)\tilde{\eta}_{i,m}(t_m)|$ can be bounded as,

$$\begin{aligned}&\max_m \sup_{s_m,t_m} |\frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(1)}(s_m)\tilde{\eta}_{i,m}(t_m)| \\ \leq\ &\max_m \sup_{s_m} |\sum_{w=1}^{W} \Delta\mu_m(s_{m,w})K_{h_2}(s_{m,w}-s_m)/\sum_{w=1}^{W} K_{h_2}(s_{m,w}-s_m)| \max_m \sup_{t_m} |\frac{1}{n}\sum_{i=1}^{n} \tilde{\eta}_{i,m}(t_m)| \\ =\ &O_p(\omega_1)O_p(\sqrt{\frac{\log M}{n}}).\end{aligned}$$

The term $\max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(1)2}(s_m)$ can be bounded as

$$\max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(1)2}(s_m) = O_p(\omega_1^2).$$

Terms $\max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(2)2}(s_m)$ and $\max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(3)2}(s_m)$ can be bounded as

$$\begin{aligned}\max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(2)2}(s_m) &= O_p(h_2^2), \\ \max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(3)2}(s_m) &= O_p(h_2^2).\end{aligned}$$

Finally, we consider the term $\max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\eta}_{i,m}^{(4)2}(s_m)$. Let $\mathscr{F} = \{f_{s_m} : s_m \in [0, S_m]\}$ with

$f_{s_m} = \Delta\tilde{\eta}^{(4)2}_{i,m}(s_m)$, we have $\mathbb{E}f_{s_m} \leq O(\frac{1}{Wh_2})$, $\mathbb{E}f^2_{s_m} \leq O(\frac{1}{Wh_2})$ and

$$\sup_Q N[\varepsilon, \mathscr{F}, L_2(Q)] \leq \frac{C_1}{\varepsilon h_1}.$$

Then $\max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^n \Delta\tilde{\eta}^{(4)2}_{i,m}(s_m)$ can be bounded through Lemma A.1.4 as

$$\begin{aligned} \max_m \sup_{s_m} \frac{1}{n}\sum_{i=1}^n \Delta\tilde{\eta}^{(4)2}_{i,m}(s_m) &\leq O_p(\frac{1}{Wh_2}) + O_p(\sqrt{\frac{\log M \vee \log Wh_2}{n}})O_p(\sqrt{\frac{1}{Wh_2}} \vee \sqrt{\frac{\log M}{n}}) \\ &= O_p(\frac{1}{Wh_2}) + O_p(\frac{\log M}{n}). \end{aligned}$$

The bound (A.1.7) can be obtained with some simplification. □

The following lemma establishes a uniform bound for $\widehat{\Sigma}_\xi - \Sigma_\xi$.

Lemma A.1.8. *When assumptions 3.1 - 3.10 hold, $\Delta\Sigma_\xi = \widehat{\Sigma}_\xi - \Sigma_\xi$ can be bounded as*

$$\begin{aligned} \|\Delta\Sigma_\xi\|_2 &= O_p(M\sqrt{\frac{\log ML_n}{n}}) + O_p(Mh_1) + O_p(Mh_2) + O_p(M\sqrt{\frac{\log h_1^{-1}}{n}}) + O_p(\frac{M}{\sqrt{Wh_2}}) \\ &+ O_p(M\omega_2^2)O_p(\frac{1}{M}\sum_{m=1}^M\sum_{l=1}^{L_n}\tau_{m,l}^{-2}) \overset{def}{=} MO_p(\omega_3). \end{aligned} \tag{A.6}$$

Proof. We have the following decomposition for $\|\Delta\Sigma_\xi\|_2$ as

$$\|\Delta\Sigma_\xi\|_2 \leq \|\frac{1}{n}\sum_{i=1}^n \xi_i\xi_i^T - \Sigma_\xi\|_F + \|\frac{1}{n}\sum_{i=1}^n \widehat{\xi}_i\widehat{\xi}_i^T - \frac{1}{n}\sum_{i=1}^n \xi_i\xi_i^T\|_F.$$

The first term can be bounded as

$$\begin{aligned}
\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T-\Sigma_{\xi}\|_F^2 &\leq \sum_{m=1}^{M}\sum_{l=1}^{L_n}\sum_{m'=1}^{M}\sum_{l'=1}^{L_n}|\frac{1}{n}\sum_{i=1}^{n}\xi_{i,ml}\xi_{i,m'l'}-\mathbb{E}\xi_{i,ml}\xi_{i,m'l'}|^2\\
&\leq O_p(\frac{\log ML_n}{n})\sum_{m=1}^{M}\sum_{l=1}^{L_n}\sum_{m'=1}^{M}\sum_{l'=1}^{L_n}(\mathbb{E}\xi_{i,ml}\xi_{i,m'l'})^2.\\
&= O_p(M^2\frac{\log ML_n}{n}).
\end{aligned}$$

Let $\Delta\boldsymbol{\xi}_i=\widehat{\boldsymbol{\xi}}_i-\boldsymbol{\xi}_i$, the second term can be bounded as

$$\|\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{\xi}}_i\widehat{\boldsymbol{\xi}}_i^T-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T\|_F\leq\|\frac{1}{n}\sum_{i=1}^{n}\Delta\boldsymbol{\xi}_i\Delta\boldsymbol{\xi}_i^T\|_F+2\|\frac{1}{n}\sum_{i=1}^{n}\Delta\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T\|_F.$$

We consider a decomposition of $\Delta\xi_{i,ml}=\widehat{\xi}_{i,ml}-\xi_{i,ml}$ as follows:

$$\begin{aligned}
\Delta\xi_{i,ml} &= \int_0^{S_m}\widehat{\tilde{\eta}}_{i,m}(s_m)\widehat{\phi}_{m,l}(s_m)ds_m-\int_0^{S_m}\tilde{\eta}_{i,m}(s_m)\phi_{m,l}(s_m)ds_m\\
&= \int_0^{S_m}[\Delta\tilde{\eta}_{i,m}^{(1)}(s_m)+\Delta\tilde{\eta}_{i,m}^{(2)}(s_m)+\Delta\tilde{\eta}_{i,m}^{(3)}(s_m)+\Delta\tilde{\eta}_{i,m}^{(4)}(s_m)]\widehat{\phi}_{m,l}(s_m)ds_m\\
&+ \int_0^{S_m}\tilde{\eta}_{i,m}(s_m)[\widehat{\phi}_{m,l}(s_m)-\phi_{m,l}(s_m)]ds_m\\
&\overset{def}{=} \Delta\xi_{i,ml}^{(1)}+\Delta\xi_{i,ml}^{(2)}+\Delta\xi_{i,ml}^{(3)}+\Delta\xi_{i,ml}^{(4)}+\Delta\xi_{i,ml}^{(5)}.
\end{aligned}$$

$\|\frac{1}{n}\sum_{i=1}^{n}\Delta\boldsymbol{\xi}_i\Delta\boldsymbol{\xi}_i^T\|_F$ and $\|\frac{1}{n}\sum_{i=1}^{n}\Delta\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T\|_F$ can be decomposed respectively as

$$\|\frac{1}{n}\sum_{i=1}^{n}\Delta\boldsymbol{\xi}_i\Delta\boldsymbol{\xi}_i^T\|_F \leq \sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^2.$$

and

$$\|\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_i\xi_i^T\|_F \;\leq\; \sqrt{\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^2}\sqrt{\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\xi_{i,ml}^2} = O_p(\sqrt{M})\sqrt{\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^2}.$$

Therefore, we only need to bound the term $\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^2$, which is given by

$$\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^2 \;\leq\; 5\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}[\Delta\xi_{i,ml}^{(1)2}+\Delta\xi_{i,ml}^{(2)2}+\Delta\xi_{i,ml}^{(3)2}+\Delta\xi_{i,ml}^{(4)2}+\Delta\xi_{i,ml}^{(5)2}].$$

We then investigate the above terms individually. For $\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(1)2}$, we have

$$\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(1)2} \;\leq\; \frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{M}\int_0^{S_m}\Delta\tilde{\eta}_{i,m}^{(1)2}(s_m)ds_m = O_p(M\omega_1^2).$$

For $\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(2)2}$, we have

$$\begin{aligned}\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(2)2} \;&\leq\; \frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{M}\int_0^{S_m}\Delta\tilde{\eta}_{i,m}^{(2)2}(s_m)ds_m\\ &=\; O_p(Mh_2^2).\end{aligned}$$

For $\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(3)2}$, we have

$$\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(3)2} \;\leq\; \frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{M}\int_0^{S_m}\Delta\tilde{\eta}_{i,m}^{(3)2}(s_m)ds_m = O_p(Mh_2^2).$$

For $\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(4)2}$, we have

$$\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(4)2} \;\leq\; \frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{M}\int_0^{S_m}\Delta\tilde{\eta}_{i,m}^{(4)2}(s_m)ds_m.$$

Since $\max_m \mathbb{E}\Delta\tilde{\eta}_{i,m}^{(4)2}(s_m) = O(\frac{1}{Wh_2})$ and $\max_m \mathbb{E}\Delta\tilde{\eta}_{i,m}^{(4)4}(s_m) \leq O(\frac{1}{Wh_2})$, Bernstein inequality shows that

$$P(\max_m |\frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{M}\int_0^{S_m}\Delta\tilde{\eta}_{i,m}^{(4)2}(s_m)ds_m - \max_m \mathbb{E}\Delta\tilde{\eta}_{i,m}^{(4)2}(s_m)|) \leq 2M\exp\{-\frac{nx^2}{2/(Wh_2)+2C_1t/3}\},$$

which leads to

$$\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(4)2} = O_p(\frac{M\log M}{n}\vee\frac{M}{Wh_2}).$$

Finally, for $\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(5)2}$, we have

$$\begin{aligned}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(5)2} &\leq \frac{1}{n}\sum_{i=1}^{n}\|\tilde{\eta}_{i,m}(s_m)\|_2^2\sum_{l=1}^{L_n}\|\widehat{\phi}_{m,l}(s_m)-\phi_{m,l}(s_m)\|_2^2\\ &\leq O_p(1)\sum_{l=1}^{L_n}\tau_{m,l}^{-2}\|\widehat{\Sigma}_{\tilde{\eta}_m}-\Sigma_{\tilde{\eta}_m}\|_2^2,\end{aligned}$$

which leads to

$$\sum_{m=1}^{M}\sum_{l=1}^{L_n}\frac{1}{n}\sum_{i=1}^{n}\Delta\xi_{i,ml}^{(5)2} = O_p(\omega_2^2)\sum_{m=1}^{M}\sum_{l=1}^{L_n}\tau_{m,l}^{-2}.$$

The bound (A.1.8) can be obtained with some simplification. □

A.2 Proof of Theorem 3.1

Proof. Let $\tau_{f,1},\cdots,\tau_{f,r}$ be the eigenvalues of $\Lambda\Sigma_f\Lambda^T$ in decreasing order, the following inequalities hold through Wely's Theorem,

$$\begin{aligned}\max_{1\leq j\leq r}|\tau_{f,j}-\tau_j| &\leq \|\Lambda\Sigma_f\Lambda^T-\Omega_\gamma\|_2 \leq \|\Lambda\Sigma_f\Lambda^T-\Omega_\gamma\|_F = o(M).\\ \max_j|\widehat{\tau}_{\xi,j}-\tau_{f,j}| &\leq \|\widehat{\Sigma}_\xi-\Lambda\Sigma_f\Lambda^T\|_2 \leq \|\widehat{\Sigma}_\xi-\Sigma_\xi\|_2+\|\Sigma_\xi-\Lambda\Sigma_f\Lambda^T\|_2\\ &= MO_p(\omega_3)+O_p(1) = o(M).\end{aligned}$$

Let $\mathbf{v}_{f,1},\cdots,\mathbf{v}_{f,r}$, we can also prove the following inequality using Davis-Kahan's $\sin\theta$ Theorem

$$\max_{1\leq j\leq r}\|\widehat{\mathbf{v}}_j-\mathbf{v}_{f,j}\|_2\leq O(M^{-1})\|\widehat{\Sigma}_\xi-\Lambda\Sigma_f\boldsymbol{\Lambda}^T\|_2=O_p(\omega_3)+O_p(M^{-1}).$$

To derive the asymptotic distribution, we only need to prove that there exist an $r\times r$ orthogonal matrix $\mathbf{O}$ such that

$$\|\mathbf{C}H_x\widehat{\boldsymbol{F}}_c-\mathbf{C}H_x\boldsymbol{F}_c\mathbf{O}\|_F=o_p(n^{-1/2}).$$

To show this, we consider a decomposition of $\widehat{\boldsymbol{F}}_c-\boldsymbol{F}_c\mathbf{O}$ as

$$\begin{aligned}\widehat{\boldsymbol{F}}_c-\boldsymbol{F}_c\mathbf{O} &= \widehat{\boldsymbol{\xi}}\widehat{\boldsymbol{V}}_r\widehat{T}^{-1/2}-\boldsymbol{\xi}\boldsymbol{V}_{f,r}T_f^{-1/2}\mathbf{O}\\ &= (\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})\widehat{\boldsymbol{V}}_r\widehat{T}^{-1/2}+\boldsymbol{\xi}\widehat{\boldsymbol{V}}_r(\widehat{T}^{-1/2}-T_f^{-1/2})+\boldsymbol{\xi}(\widehat{\boldsymbol{V}}_r-\boldsymbol{V}_{f,r})T_f^{-1/2}.\end{aligned}$$

where $\widehat{\boldsymbol{\xi}}=(\widehat{\xi}_1,\cdots,\widehat{\xi}_n)^T$, $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)^T$, $T_f=\text{diag}\{\tau_{f,1},\cdots,\tau_{f,r}\}$ and $\boldsymbol{V}_{f,r}=(\mathbf{v}_{f,1},\cdots,\mathbf{v}_{f,r})$. Under both null hypothesis and alternative hypothesis (defined by Assumption 3.11), we have the following conclusions,

$$\begin{aligned}\|\mathbf{CH}_x\boldsymbol{\xi}\widehat{\boldsymbol{V}}_r(\widehat{T}^{-1/2}-T_f^{-1/2})\|_F &= O_p(n^{-1/2})\|\widehat{T}^{-1/2}-T_f^{-1/2}\|_F=O_p(n^{-1/2}M^{-1/2}),\\ \|\mathbf{CH}_x\boldsymbol{\xi}(\widehat{\boldsymbol{V}}_r-\boldsymbol{V}_{f,r})T_f^{-1/2}\|_F &= O_p(n^{-1/2})\|\widehat{\boldsymbol{V}}_r-\boldsymbol{V}_{f,r}\|_2O_p(M^{-1/2})\\ &= O_p(n^{-1/2})\max_j|\widehat{\tau}_{\xi,j}-\tau_{f,j}|O_p(1)\\ &= O_p(n^{-1/2})[O_p(\omega_3)+O_p(M^{-1})].\end{aligned}$$

For $\mathbf{CH}_x(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})\widehat{\boldsymbol{V}}_r\widehat{T}^{-1/2}$, using a similar derivation to that in Lemma A.1.8, we can show that

$$\begin{aligned}
\|\mathbf{CH}_x(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})\widehat{\boldsymbol{V}}_r\widehat{T}^{-1/2}\|_F &= \{O_p(\omega_1)+O_p(h_2)+O_p[h_2\sqrt{\log M\log h_2^{-1}}\vee\frac{\log M\log Wh_2}{\sqrt{n}}] \\
&+ O_p[\omega_2\sqrt{\frac{\log M}{M}\sum_{m=1}^{M}\sum_{l=1}^{L_n}\tau_{m,l}^{-2}}] \\
&+ O_p[\sqrt{\frac{\log M\vee\log Wh_2}{Wh_2}}\vee\frac{\log M\vee\log Wh_2}{\sqrt{n}}]\}O_p(n^{-1/2}) \\
&= o_p(1)O_p(n^{-1/2}),
\end{aligned}$$

which finishes the proof. □

APPENDIX B: TECHNICAL DETAILS OF CHAPTER 4

B.1 Proof of Theorem 4.1

$T_n(\widehat{\omega}_\lambda)$ can be explicitly expressed as,

$$T_n(\widehat{\omega}_{\lambda_n}) = \frac{n[\sum_{l=1}^{+\infty} \frac{\widehat{d}_{l,h_1}^2}{\widehat{\tau}_l+\lambda_n}]^2}{\widehat{\sigma}_c^2 \sum_{l=1}^{+\infty} \frac{\widehat{\tau}_l \widehat{d}_{l,h_1}^2}{(\widehat{\tau}_l+\lambda_n)^2}}. \tag{B.1}$$

Let $W_{n,1} = \frac{n}{\widehat{\sigma}_c^2} \sum_{l=1}^{+\infty} \frac{\widehat{d}_{l,h_1}^2}{\widehat{\tau}_l+\lambda_n}$ and $W_{n,2} = \frac{n}{\widehat{\sigma}_c^2} \sum_{l=1}^{+\infty} \frac{\widehat{\tau}_l \widehat{d}_{l,h_1}^2}{(\widehat{\tau}_l+\lambda_n)^2}$, $T_n(\widehat{\omega}_{\lambda_n})$ can be expressed as $T_n(\widehat{\omega}_{\lambda_n}) = \frac{W_{n,1}^2}{W_{n,2}}$. Then we need to study the asymptotic distribution of $(W_{n,1}, W_{n,2})$.

The proof of theorem consists of two parts. In the first part, we prove that

$$W_{n,1} = \sum_{l=1}^{+\infty} \frac{(\sqrt{\tau_l} z_l + \delta_{l,0}/\sigma_c)^2}{\tau_l + \lambda_n} \{1 + o_p(1)\}, \tag{B.2}$$

$$W_{n,2} = \sum_{l=1}^{+\infty} \frac{\tau_l (\sqrt{\tau_l} z_l + \delta_{l,0}/\sigma_c)^2}{(\tau_l + \lambda_n)^2} \{1 + o_p(1)\}, \tag{B.3}$$

where $\{z_l\}_{l=1}^{+\infty}$ are independent variables following $N(0,1)$.

Here we show the derivation of (B.2) in detail. (B.3) can be obtained in a similar way.

Note that $\widehat{d}_{l,0} = \int_0^S \mathbf{C}\widehat{\boldsymbol{\beta}}_{h_1}(s)\widehat{\phi}_l(s)ds$, we first examine the major terms in $\mathbf{C}\widehat{\boldsymbol{\beta}}_{h_1}(s)$. For arbitrary point s in $[0,S]$, let $\boldsymbol{\eta}(s) = [\eta_1(s), \cdots, \eta_n(s)]^T$ and $\mathbf{E}(s) = [e_1(s), \cdots, e_n(s)]^T$, $\mathbf{C}\widehat{\boldsymbol{\beta}}_{h_1}(s)$ can be expressed as,

$$\begin{aligned}
\mathbf{C}\widehat{\boldsymbol{\beta}}_{h_1}(s) &= \mathbf{C}\boldsymbol{\beta}(s)+\frac{\sum_{m=1}^{M}K_{h_1}(s_m-s)[\boldsymbol{\beta}(s_m)-\boldsymbol{\beta}(s)]}{\sum_{m=1}^{M}K_{h_1}(s_m-s)} \\
&+ \mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\eta}(s)+\frac{\sum_{m=1}^{M}K_{h_1}(s_m-s)\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T[\boldsymbol{\eta}(s_m)-\boldsymbol{\eta}(s)]}{\sum_{m=1}^{M}K_{h_1}(s_m-s)} \\
&+ \frac{\sum_{m=1}^{M}K_{h_1}(s_m-s)\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{E}(s)}{\sum_{m=1}^{M}K_{h_1}(s_m-s)} \\
&= \mathbf{C}\boldsymbol{\beta}(s)+\Delta\widehat{\boldsymbol{\beta}}_{h_1,1}(s)+\Delta\widehat{\boldsymbol{\beta}}_{h_1,2}(s)+\Delta\widehat{\boldsymbol{\beta}}_{h_1,3}(s)+\Delta\widehat{\boldsymbol{\beta}}_{h_1,4}(s).
\end{aligned}$$

Let $d_{l,0}=\int_0^S[\mathbf{C}\boldsymbol{\beta}(s)+\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\eta}(s)]\phi_l(s)ds$, we have

$$\frac{n}{\widehat{\sigma}_c^2}\sum_{l=1}^{+\infty}\frac{d_{l,0}^2}{\tau_l+\lambda_n}\xrightarrow{d}\sum_{l=1}^{+\infty}\frac{(\sqrt{\tau_l}z_l+\delta_{l,0}/\sigma_c)^2}{\tau_l+\lambda_n}\asymp\begin{cases}O_p(log\lambda_n^{-1}), \ \tau_l\asymp\alpha^{-l}, \\ O_p(\lambda_n^{-\frac{1}{r}}), \ \tau_l\asymp l^{-r}.\end{cases} \tag{B.4}$$

Then we only need to show that $\frac{n}{\widehat{\sigma}_c^2}\sum_{l=1}^{+\infty}\frac{\widehat{d}_{l,0}^2-d_{l,0}^2}{\widehat{\tau}_l+\lambda_n}$ and $\frac{n}{\widehat{\sigma}_c^2}\sum_{l=1}^{+\infty}[\frac{d_{l,0}^2}{\widehat{\tau}_l+\lambda_n}-\frac{d_{l,0}^2}{\tau_l+\lambda_n}]$ are ignorable compared to $\frac{n}{\widehat{\sigma}_c^2}\sum_{l=1}^{+\infty}\frac{d_{l,0}^2}{\tau_l+\lambda_n}$. The second term can be bounded by,

$$\begin{aligned}
|\frac{n}{\widehat{\sigma}_c^2}\sum_{l=1}^{+\infty}[\frac{d_{l,0}^2}{\widehat{\tau}_l+\lambda_n}-\frac{d_{l,0}^2}{\tau_l+\lambda_n}]| &= |\frac{n}{\widehat{\sigma}_c^2}\sum_{l=1}^{+\infty}d_{l,0}^2\frac{\tau_l-\widehat{\tau}_l}{(\tau_l+\lambda_n)(\widehat{\tau}_l+\lambda_n)}| \\
&\leq \frac{n}{\widehat{\sigma}_c^2}\sum_{l=1}^{+\infty}\frac{d_{l,0}^2}{\tau_l+\lambda_n}\max_l|\frac{\tau_l-\widehat{\tau}_l}{\lambda_n}| \\
&\leq \frac{n}{\widehat{\sigma}_c^2}\sum_{l=1}^{+\infty}\frac{d_{l,0}^2}{\tau_l+\lambda_n}\frac{||\widehat{\Sigma}_\eta-\Sigma_\eta||_2}{\lambda_n}.
\end{aligned}$$

Following a similar derivation of Theorem 3(i) in [36], we can show that $||\widehat{\Sigma}_\eta - \Sigma_\eta||_2 = O_p[(Mh_2)^{-\frac{1}{2}} + h_1 + h_2 + (logn/n)^{\frac{1}{2}}]$, which gives

$\frac{n}{\widehat{\sigma}_c^2}\sum_{l=1}^{+\infty}[\frac{d_{l,0}^2}{\widehat{\tau}_l+\lambda_n}-\frac{d_{l,0}^2}{\tau_l+\lambda_n}]=o_p(\frac{n}{\widehat{\sigma}_c^2}\sum_{l=1}^{+\infty}\frac{d_{l,0}^2}{\tau_l+\lambda_n})$.

For the first term, we consider a decomposition of $\widehat{d}_{l,0}-d_{l,0}$,

$$\begin{aligned}\widehat{d}_{l,0}-d_{l,0} &= \int_0^S \mathbf{C}\widehat{\boldsymbol{\beta}}_{h_1}(s)[\widehat{\phi}_l(s)-\phi_l(s)]ds+\int_0^S \mathbf{C}\Delta\widehat{\boldsymbol{\beta}}_{h_1,1}(s)\phi_l(s)ds \\ &+ \int_0^S \mathbf{C}\Delta\widehat{\boldsymbol{\beta}}_{h_1,3}(s)\phi_l(s)ds+\int_0^S \mathbf{C}\Delta\widehat{\boldsymbol{\beta}}_{h_1,4}(s)\phi_l(s)ds \\ &= \widehat{d}_{l,1}+\widehat{d}_{l,2}+\widehat{d}_{l,3}+\widehat{d}_{l,4}.\end{aligned}$$

Then we only need to show $\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{\widehat{d}_{l,r}^2}{\widehat{\tau}_l+\lambda_n}=o_p\{\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{d_{l,0}^2}{\tau_l+\lambda_n}\}$ for $r=1,2,3,4$.

As $\boldsymbol{\beta}(s)$ are composed of functions with bounded first-order derivatives, $\|\mathbf{C}\Delta\widehat{\boldsymbol{\beta}}_{h_1,1}(s)\|_2^2=O(n^{-1}h_1^2)$. Then we have,

$$\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{\widehat{d}_{l,2}^2}{\widehat{\tau}_l+\lambda_n}=O_p(h_1^2\lambda_n^{-1}).$$

For $\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{\widehat{d}_{l,3}^2}{\widehat{\tau}_l+\lambda_n}$, following Proposition A.2.7 in [152], it can be shown that

$$\sup_s|\mathbf{C}\Delta\widehat{\boldsymbol{\beta}}_{h_1,3}(s)|=\sup_s|\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\partial\boldsymbol{\eta}(s)|h_1=O_p(n^{-1/2}h_1).$$

Then $\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{\widehat{d}_{l,3}^2}{\widehat{\tau}_l+\lambda_n}=O_p(h_1^2\lambda_n^{-1})$. Following Theorem 2.14.16 in [152], we can also show that,

$$\sup_s|\mathbf{C}\Delta\widehat{\boldsymbol{\beta}}_{h_1,4}(s)|=O_p(\sqrt{\frac{1}{nMh_1}})\text{ and }\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{\widehat{d}_{l,3}^2}{\widehat{\tau}_l+\lambda_n}=O_p[(\lambda_nMh_1)^{-1}].$$

Therefore, $\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{\widehat{d}_{l,r}^2}{\widehat{\tau}_l+\lambda_n}=o_p\{\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{d_{l,0}^2}{\tau_l+\lambda_n}\}$ for $r=2,3,4$.

We decompose $\widehat{d}_{l,1}$ as,

$$\begin{aligned}
\widehat{d}_{l,1} &= \int_0^S [\mathbf{C}\boldsymbol{\beta}(s) + \mathbf{C}\Delta\boldsymbol{\beta}_{h_1,2}(s)][\widehat{\phi}_l(s) - \phi_l(s)]ds + \int_0^S \mathbf{C}\Delta\boldsymbol{\beta}_{h_1,1}(s)[\widehat{\phi}_l(s) - \phi_l(s)]ds \\
&+ \int_0^S \mathbf{C}\Delta\boldsymbol{\beta}_{h_1,3}(s)[\widehat{\phi}_l(s) - \phi_l(s)]ds + \int_0^S \mathbf{C}\Delta\boldsymbol{\beta}_{h_1,4}(s)[\widehat{\phi}_l(s) - \phi_l(s)]ds \\
&= \widetilde{d}_{l,1} + \widetilde{d}_{l,2} + \widetilde{d}_{l,3} + \widetilde{d}_{l,4}.
\end{aligned}$$

Similar to previous derivations, it can be shown that $\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{\widetilde{d}_{l,r}^2}{\widehat{\tau}_l+\lambda_n} = o_p\{\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{d_{l,0}^2}{\tau_l+\lambda_n}\}$ for $r = 2,3,4$. For $\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{\widetilde{d}_{l,1}^2}{\widehat{\tau}_l+\lambda_n}$, we consider the summation of the first L_n terms first, where L_n satisfy $\tau_{L_n} \asymp \lambda_n$. To quantify $\widehat{\phi}_l(s) - \phi_l(s)$, we use the L^2 expansion in [35],

$$\widehat{\phi}_l(s) - \phi_l(s) = \sum_{j\neq l}\frac{\langle\Delta\phi_j, \phi_j\rangle\phi_j}{\tau_l - \tau_j} + O(\|\widehat{\Sigma}_\eta - \Sigma_\eta\|_2^2), \tag{B.5}$$

where $\Delta\phi_j(s) = \iint [\widehat{\Sigma}_\eta(s,t) - \Sigma_\eta(s,t)]\phi_j(t)dt$. Since $\widetilde{d}_{l,1}^2 \leq \|\mathbf{C}\boldsymbol{\beta}(s) + \mathbf{C}\Delta\boldsymbol{\beta}_{h_1,2}\|_2^2\|\widehat{\phi}_l(s) - \phi_l(s)\|_2^2$, it follows that,

$$\begin{aligned}
\frac{n}{\sigma_c^2}\sum_{l=1}^{L_n}\frac{\widetilde{d}_{l,1}^2}{\widehat{\tau}_l + \lambda_n} &\leq O_p(1)O(\sum_{l=1}^{L_n}\frac{\|\Delta\phi_j\|_2^2/\tau_l^2 + \|\widehat{\Sigma}_\eta - \Sigma_\eta\|_2^2}{\widehat{\tau}_l + \lambda_n}) \\
&= O_p(1)[O(\sum_{l=1}^{L_n}\frac{\|\Delta\phi_j\|_2^2}{\tau_l^2(\widehat{\tau}_l + \lambda_n)}) + O(\sum_{l=1}^{L_n}\frac{\|\widehat{\Sigma}_\eta - \Sigma_\eta\|_2^2}{\widehat{\tau}_l + \lambda_n})] \\
&= O_p(1)[O(\frac{1}{\lambda_n^3}\sum_{l=1}^{L_n}\|\Delta\phi_j\|_2^2) + o_p(1)] = O(\frac{\|\widehat{\Sigma}_\eta - \Sigma_\eta\|_2^2}{\lambda_n^3}) = o_p(\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{d_{l,0}^2}{\tau_l + \lambda_n}).
\end{aligned}$$

$\frac{n}{\sigma_c^2}\sum_{l=L_n+1}^{+\infty}\frac{\widetilde{d}_{l,1}^2}{\widehat{\tau}_l+\lambda_n}$ can be bounded by,

$$\begin{aligned}
\frac{n}{\sigma_c^2}\sum_{l=L_n+1}^{+\infty}\frac{\widetilde{d}_{l,1}^2}{\widehat{\tau}_l+\lambda_n} &= \frac{n}{\lambda_n\sigma_c^2}\sum_{l=L_n+1}^{+\infty}\widetilde{d}_{l,1}^2 \leq \frac{n}{\lambda_n\sigma_c^2}\sum_{l=L_n+1}^{+\infty}[d_{l,0}+\int_0^S(\mathbf{C}\beta(s)+\mathbf{C}\Delta\beta_{h_1,2}(s))\widehat{\phi}_l(s)ds]^2 \\
&\leq \frac{2n}{\lambda_n\sigma_c^2}\sum_{l=L_n+1}^{+\infty}d_{l,0}^2+\frac{2n}{\lambda_n\sigma_c^2}\sum_{l=L_n+1}^{+\infty}[\int_0^S(\mathbf{C}\beta(s)+\mathbf{C}\Delta\beta_{h_1,2}(s))\widehat{\phi}_l(s)ds]^2 \\
&\leq \frac{4n}{\lambda_n\sigma_c^2}\sum_{l=L_n+1}^{+\infty}d_{l,0}^2+\frac{2n}{\lambda_n\sigma_c^2}\sum_{l=1}^{L_n}\widetilde{d}_{l,1}^2+\frac{4n}{\lambda_n\sigma_c^2}\sum_{l=1}^{L_n}\widetilde{d}_{l,1}d_{l,0} \\
&\leq O(\frac{4n}{\lambda_n\sigma_c^2}\sum_{l=L_n+1}^{+\infty}d_{l,0}^2)+O(\frac{2n}{\lambda_n\sigma_c^2}\sum_{l=1}^{L_n}\widetilde{d}_{l,1}^2) \\
&= O(\frac{4n}{\sigma_c^2}\sum_{l=L_n+1}^{+\infty}\frac{d_{l,0}^2}{\tau_l+\lambda_n})+O(\frac{\|\widehat{\Sigma}_\eta-\Sigma_\eta\|_2^2}{\lambda_n^3})=o_p(\frac{n}{\sigma_c^2}\sum_{l=1}^{+\infty}\frac{d_{l,0}^2}{\tau_l+\lambda_n}).
\end{aligned}$$

Then the proof of the first part is finished.

In the second part, we calculate the asymptotic distribution of $T_n(\widehat{\omega}_{\lambda_n})$.

With (B.2) and (B.3), as $\lambda_n\to 0$, the following distribution can be achieved under null hypothesis [153],

$$\begin{pmatrix} W_{n,1} \\ W_{n,2}\end{pmatrix}\xrightarrow{d}N\{\begin{pmatrix} a_1 \\ a_2\end{pmatrix},2\begin{pmatrix} a_2 & a_3 \\ a_3 & a_4\end{pmatrix}\}.$$

Using delta method, the asymptotic distribution of $T_n(\widehat{\omega}_{\lambda_n})=W_{n,1}^2/W_{n,2}$ is given by,

$$T_n(\widehat{\omega}_{\lambda_n})\xrightarrow{d}N\{\frac{a_1^2}{a_2},\frac{8a_1^2}{a_2}+\frac{2a_1^4a_4}{a_2^4}-\frac{4a_1^3a_3}{a_2^3}\}.$$

Under alternative hypothesis, as $\lambda_n \to 0$, the following distribution can be achieved [153],

$$\begin{pmatrix} W_{n,1} \\ \\ W_{n,2} \end{pmatrix} \xrightarrow{d} N\{\begin{pmatrix} a_1 + d_1 \\ \\ a_2 + d_2 \end{pmatrix}, 2\begin{pmatrix} a_2 + 2d_2 & a_3 + 2d_3 \\ \\ a_3 + 2d_3 & a_4 + 2d_4 \end{pmatrix}\}.$$

Using delta method, the asymptotic distribution of $T_n(\widehat{\omega}_{\lambda_n}) = W_{n,1}^2/W_{n,2}$ is given by,

$$T_n(\widehat{\omega}_{\lambda_n}) \xrightarrow{d} N\{\frac{(a_1+d_1)^2}{a_2+d_2}, \frac{8(a_1+d_1)^2(a_2+2d_2)}{(a_2+d_2)^2} + \frac{2(a_1+d_1)^4(a_4+2d_4)}{(a_2+d_2)^4} - \frac{4(a_1+d_1)^3(a_3+2d_3)}{(a_2+d_2)^3}\}.$$

REFERENCES

[1] Peter M Visscher, Matthew A Brown, Mark I McCarthy, and Jian Yang. Five years of gwas discovery. *The American Journal of Human Genetics*, 90(1):7–24, 2012.

[2] Peter M Visscher, Naomi R Wray, Qian Zhang, Pamela Sklar, Mark I McCarthy, Matthew A Brown, and Jian Yang. 10 years of gwas discovery: biology, function, and translation. *The American Journal of Human Genetics*, 101(1):5–22, 2017.

[3] Barbara E Stranger, Eli A Stahl, and Towfique Raj. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*, 2010.

[4] Patrick F Sullivan. The psychiatric gwas consortium: big science comes to psychiatry. *Neuron*, 68(2):182–186, 2010.

[5] Evangelia Stergiakouli, Marian Hamshere, Peter Holmans, Kate Langley, Irina Zaharieva, deCODE Genetics, Psychiatric GWAS Consortium: ADHD Subgroup, Ziarah Hawi, Lindsey Kent, Michael Gill, et al. Investigating the contribution of common genetic variants to the risk and pathogenesis of adhd. *American Journal of Psychiatry*, 169(2):186–194, 2012.

[6] Elise B Robinson, Beate St Pourcain, Verneri Anttila, Jack A Kosmicki, Brendan Bulik-Sullivan, Jakob Grove, Julian Maller, Kaitlin E Samocha, Stephan J Sanders, Stephan Ripke, et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nature genetics*, 48(5):552, 2016.

[7] Elliot S Gershon, Ney Alliey-Rodriguez, and Chunyu Liu. After gwas: searching for genetic risk for schizophrenia and bipolar disorder. *American Journal of Psychiatry*, 168(3):253–256, 2011.

[8] Andreas Meyer-Lindenberg and Daniel R Weinberger. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *Nature Reviews Neuroscience*, 7(10):818, 2006.

[9] Clifford R Jack, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.

[10] Theodore D Satterthwaite, Mark A Elliott, Kosha Ruparel, James Loughead, Karthik Prabhakaran, Monica E Calkins, Ryan Hopson, Chad Jackson, Jack Keefe, Marisa Riley, et al. Neuroimaging of the philadelphia neurodevelopmental cohort. *Neuroimage*, 86:544–553, 2014.

[11] Terry L Jernigan, Timothy T Brown, Donald J Hagler, Natacha Akshoomoff, Hauke Bartsch, Erik Newman, Wesley K Thompson, Cinnamon S Bloss, Sarah S Murray, Nicholas Schork, et al. The pediatric imaging, neurocognition, and genetics (ping) data repository. *Neuroimage*, 124:1149–1154, 2016.

[12] David C Van Essen, Kamil Ugurbil, E Auerbach, D Barch, TEJ Behrens, R Bucholz, A Chang, Liyong Chen, Maurizio Corbetta, Sandra W Curtiss, et al. The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231, 2012.

[13] Stamatios N Sotiropoulos, Saad Jbabdi, Junqian Xu, Jesper L Andersson, Steen Moeller, Edward J Auerbach, Matthew F Glasser, Moises Hernandez, Guillermo Sapiro, Mark Jenkinson, et al. Advances in diffusion mri acquisition and processing in the human connectome project. *Neuroimage*, 80:125–143, 2013.

[14] Roberta Rasetti and Daniel R Weinberger. Intermediate phenotypes in psychiatric disorders. *Current opinion in genetics & development*, 21(3):340–348, 2011.

[15] Irving I Gottesman and Todd D Gould. The endophenotype concept in psychiatry: etymology and strategic intentions. *American Journal of Psychiatry*, 160(4):636–645, 2003.

[16] Gilbert A Preston and Daniel R Weinberger. Intermediate phenotypes in schizophrenia: a selective review. *Dialogues in clinical neuroscience*, 7(2):165, 2005.

[17] Elena I Ivleva, David W Morris, Amanda F Moates, Trisha Suppes, Gunvant K Thaker, and Carol A Tamminga. Genetics and intermediate phenotypes of the schizophrenia-bipolar disorder boundary. *Neuroscience & Biobehavioral Reviews*, 34(6):897–921, 2010.

[18] Kristin L Bigos and Daniel R Weinberger. Imaging genetics'days of future past. *Neuroimage*, 53(3):804–809, 2010.

[19] Andrew L Alexander, Jee Eun Lee, Mariana Lazar, and Aaron S Field. Diffusion tensor imaging of the brain. *Neurotherapeutics*, 4(3):316–329, 2007.

[20] John H Gilmore, Chaeryon Kang, Dianne D Evans, Honor M Wolfe, J Keith Smith, Jeffrey A Lieberman, Weili Lin, Robert M Hamer, Martin Styner, and Guido Gerig. Prenatal and neonatal brain structure and white matter maturation in children at high risk for schizophrenia. *American Journal of Psychiatry*, 2010.

[21] Scott A Huettel, Allen W Song, Gregory McCarthy, et al. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, MA, 2004.

[22] R Della Nave, S Foresti, A Pratesi, A Ginestroni, M Inzitari, E Salvadori, M Giannelli, S Diciotti, D Inzitari, and M Mascalchi. Whole-brain histogram and voxel-based analyses of diffusion tensor imaging in patients with leukoaraiosis: correlation with motor and cognitive impairment. *American journal of neuroradiology*, 28(7):1313–1319, 2007.

[23] Marco Rovaris, Giuseppe Iannucci, Mara Cercignani, Maria Pia Sormani, Nicola De Stefano, Simonetta Gerevini, Giancarlo Comi, and Massimo Filippi. Age-related changes in conventional, magnetization transfer, and diffusion-tensor mr imaging findings: Study with whole-brain tissue histogram analysis1. *Radiology*, 227(3):731–738, 2003.

[24] Aaron Alexander-Bloch, Jay N Giedd, and Ed Bullmore. Imaging structural co-variance between human brain regions. *Nature Reviews Neuroscience*, 14(5):322, 2013.

[25] Trevor Hastie and Robert Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 757–796, 1993.

[26] James O Ramsay. *Functional data analysis*. Wiley Online Library, 2006.

[27] Hervé Cardot, Frédéric Ferraty, and Pascal Sarda. Spline estimators for the functional linear model. *Statistica Sinica*, pages 571–591, 2003.

[28] Chin-Tsang Chiang, John A Rice, and Colin O Wu. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *Journal of the American Statistical Association*, 96(454):605–619, 2001.

[29] Hongtu Zhu, Linglong Kong, Runze Li, Martin Styner, Guido Gerig, Weili Lin, and John H Gilmore. Fadtts: functional analysis of diffusion tensor tract statistics. *NeuroImage*, 56(3):1412–1425, 2011.

[30] Jin-Ting Zhang, Jianwei Chen, et al. Statistical inferences for functional data. *The Annals of Statistics*, 35(3):1052–1079, 2007.

[31] Uwe Einmahl and David M Mason. An empirical process approach to the uniform consistency of kernel-type function estimators. *Journal of Theoretical Probability*, 13(1):1–37, 2000.

[32] Jianqing Fan and Wenyang Zhang. Statistical estimation in varying coefficient models. *Annals of Statistics*, pages 1491–1518, 1999.

[33] Colin O Wu and Chin-Tsang Chiang. Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statistica Sinica*, pages 433–456, 2000.

[34] Yehua Li and Tailen Hsing. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, pages 3321–3351, 2010.

[35] Peter Hall, Hans-Georg Müller, and Jane-Ling Wang. Properties of principal component methods for functional and longitudinal data analysis. *The annals of statistics*, pages 1493–1517, 2006.

[36] Hongtu Zhu, Runze Li, and Linglong Kong. Multivariate varying coefficient model for functional responses. *Annals of statistics*, 40(5):2634, 2012.

[37] Mark J Brewer. A bayesian model for local smoothing in kernel density estimation. *Statistics and Computing*, 10(4):299–309, 2000.

[38] Xibin Zhang, Maxwell L King, and Rob J Hyndman. A bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics & Data Analysis*, 50(11):3009–3031, 2006.

[39] Kari Karhunen. Zur spektraltheorie stochastischer prozesse. 1946.

[40] Michel Loève. Fonctions aléatoires à décomposition orthogonale exponentielle. *La Revue Scientifique*, 84:159–162, 1946.

[41] Xiaoyan Leng and Hans-Georg Müller. Classification using functional data analysis for temporal gene expression data. *Bioinformatics*, 22(1):68–76, 2006.

[42] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Mueller. Review of functional data analysis. *arXiv preprint arXiv:1507.05135*, 2015.

[43] Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

[44] Jeng-Min Chiou and Hans-Georg Müller. Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting. *Journal of the American Statistical Association*, 104(486):572–585, 2009.

[45] Qing Shen and Julian Faraway. An f test for linear models with functional responses. *Statistica Sinica*, pages 1239–1257, 2004.

[46] Jin-Ting Zhang. Statistical inferences for linear models with functional responses. *Statistica Sinica*, pages 1431–1451, 2011.

[47] Jianqing Fan. Test of significance based on wavelet thresholding and neyman's truncation. *Journal of the American Statistical Association*, 91(434):674–688, 1996.

[48] Zhidong Bai and Hewa Saranadasa. Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, pages 311–329, 1996.

[49] Jianxin Yin and Hongzhe Li. A sparse conditional gaussian graphical model for analysis of genetical genomics data. *The annals of applied statistics*, 5(4):2630, 2011.

[50] M. Vounou, E. Janousova, R. Wolz, J.L. Stein, P.M. Thompson, D. Rueckert, G. Montana, and ADNI. Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in alzheimer's disease. *Neuroimage*, 60:700–716, 2012.

[51] J. Peng, J. Zhu, A. Bergamaschi, W. Han, D. Noh, J. R. Pollack, and P. Wang. Regularized multivariate regression for identifying master predictors with application to integrative genomics study of breast cancer. *Annals of Applied Statistics*, 4:53–77, 2010.

[52] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[53] Qiong Yang and Yuanjia Wang. Methods for analyzing multivariate phenotypes in genetic association studies. *Journal of probability and statistics*, 2012, 2012.

[54] Yiwei Zhang, Zhiyuan Xu, Xiaotong Shen, Wei Pan, Alzheimer's Disease Neuroimaging Initiative, et al. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *NeuroImage*, 96:309–325, 2014.

[55] Song Xi Chen, Jun Li, and Ping-Shou Zhong. Two-sample tests for high dimensional means with thresholding and data transformation. *arXiv preprint arXiv:1410.2848*, 2014.

[56] Wei Pan, Junghi Kim, Yiwei Zhang, Xiaotong Shen, and Peng Wei. A powerful and adaptive association test for rare variants. *Genetics*, 197(4):1081–1095, 2014.

[57] Jun Li and Ping-Shou Zhong. A rate optimal procedure for sparse signal recovery under dependence. *arXiv preprint arXiv:1410.2839*, 2014.

[58] Yingying Fan, Jiashun Jin, Zhigang Yao, et al. Optimal classification in sparse gaussian graphic model. *The Annals of Statistics*, 41(5):2537–2571, 2013.

[59] Qiang Sun, Hongtu Zhu, Yufeng Liu, and Joseph G Ibrahim. Sprem: sparse projection regression model for high-dimensional linear regression. *Journal of the American Statistical Association*, 110(509):289–302, 2015.

[60] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[61] Jianqing Fan, Yuan Liao, and Martina Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680, 2013.

[62] Andreas Buja and Nermin Eyuboglu. Remarks on parallel analysis. *Multivariate behavioral research*, 27(4):509–540, 1992.

[63] William R Zwick and Wayne F Velicer. Comparison of five rules for determining the number of components to retain. *Psychological bulletin*, 99(3):432, 1986.

[64] Pedro R Peres-Neto, Donald A Jackson, and Keith M Somers. How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, 49(4):974–997, 2005.

[65] Henry F Kaiser. The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1):141–151, 1960.

[66] Maurice S Bartlett. Tests of significance in factor analysis. *British Journal of statistical psychology*, 3(2):77–85, 1950.

[67] Edgar Dobriban. Factor selection by permutation. *arXiv preprint arXiv:1710.00479*, 2017.

[68] Emily L Dennis and Paul M Thompson. Typical and atypical brain development: a review of neuroimaging studies. *Dialogues in clinical neuroscience*, 15(3):359, 2013.

[69] S Mueller, D Keeser, MF Reiser, S Teipel, and T Meindl. Functional and structural mr imaging in neuropsychiatric disorders, part 2: application in schizophrenia and autism. *American journal of neuroradiology*, 2011.

[70] Van J Wedeen, Douglas L Rosene, Ruopeng Wang, Guangping Dai, Farzad Mortazavi, Patric Hagmann, Jon H Kaas, and Wen-Yih I Tseng. The geometric structure of the brain fiber pathways. *Science*, 335(6076):1628–1634, 2012.

[71] Stephen M Smith, Mark Jenkinson, Heidi Johansen-Berg, Daniel Rueckert, Thomas E Nichols, Clare E Mackay, Kate E Watkins, Olga Ciccarelli, M Zaheer Cader, Paul M Matthews, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31(4):1487–1505, 2006.

[72] Michael Bach, Frederik B Laun, Alexander Leemans, Chantal MW Tax, Geert J Biessels, Bram Stieltjes, and Klaus H Maier-Hein. Methodological considerations on tract-based spatial statistics (tbss). *Neuroimage*, 100:358–369, 2014.

[73] Eleftherios Garyfallidis, Omar Ocegueda, Demian Wassermann, and Maxime Descoteaux. Robust and efficient linear registration of white-matter fascicles in the space of streamlines. *NeuroImage*, 117:124–140, 2015.

[74] Yan Jin, Yonggang Shi, Liang Zhan, Boris A Gutman, Greig I de Zubicaray, Katie L McMahon, Margaret J Wright, Arthur W Toga, and Paul M Thompson. Automatic clustering of white matter fibers in brain diffusion mri with an application to genetics. *NeuroImage*, 100:75–90, 2014.

[75] P. Guevara, C. Poupon, D. Rivière, Y. Cointepas, M. Descoteaux, B. Thirion, and J. Mangin. Robust clustering of massive tractography datasets. *NeuroImage*, 54(3):1975–1993, 2011.

[76] C. B. Goodlett, P. T. Fletcher, J. H. Gilmore, and G. Gerig. Group analysis of DTI fiber tract statistics with application to neurodevelopment. *NeuroImage*, 45:S133–S142, 2009.

[77] L. J. O'Donnell, C. F. Westin, and A. J. Golby. Tract-based morphometry for white matter group analysis. *NeuroImage*, 45:832–844, 2009.

[78] P. A. Yushkevich, H. Zhang, T. J. Simon, and J. C. Gee. Structure-specific statistical mapping of white matter tracts. *NeuroImage*, 41:448–461, 2008.

[79] Ying Yuan, John H Gilmore, Xiujuan Geng, Styner Martin, Kehui Chen, Jane-ling Wang, and Hongtu Zhu. Fmem: Functional mixed effects modeling for the analysis of longitudinal white matter tract data. *NeuroImage*, 84:753–764, 2014.

[80] Heather F Porter and Paul F O'Reilly. Multivariate simulation framework reveals performance of multi-trait gwas methods. *Scientific reports*, 7:38837, 2017.

[81] John H Gilmore, James Eric Schmitt, Rebecca C Knickmeyer, Jeffrey K Smith, Weili Lin, Martin Styner, Guido Gerig, and Michael C Neale. Genetic and environmental contributions to neonatal brain structure: a twin study. *Human brain mapping*, 31(8):1174–1182, 2010.

[82] Ipek Oguz, Mahshid Farzinfar, Joy Matsui, Francois Budin, Zhexing Liu, Guido Gerig, Hans J Johnson, and Martin Andreas Styner. Dtiprep: quality control of diffusion-weighted images. *Frontiers in neuroinformatics*, 8:4, 2014.

[83] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004.

[84] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.

[85] Audrey Rose Verde, Francois Budin, Jean-Baptiste Berger, Aditya Gupta, Mahshid Farzinfar, Adrien Kaiser, Mihye Ahn, Hans J Johnson, Joy Matsui, Heather C Hazlett, et al. Unc-utah na-mic framework for dti fiber tract analysis. *Frontiers in neuroinformatics*, 7:51, 2014.

[86] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341, 2012.

[87] Alkes L Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and David Reich. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 38(8):904–909, 2006.

[88] Eric Yi Liu, Mingyao Li, Wei Wang, and Yun Li. Mach-admix: genotype imputation for admixed populations. *Genetic epidemiology*, 37(1):25–37, 2013.

[89] 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56–65, 2012.

[90] Eric Yi Liu, Steven Buyske, Aaron K Aragaki, Ulrike Peters, Eric Boerwinkle, Chris Carlson, Cara Carty, Dana C Crawford, Jeff Haessler, Lucia A Hindorff, et al. Genotype imputation of metabochipsnps using a study-specific reference panel of 4,000 haplotypes in african americans from the women's health initiative. *Genetic epidemiology*, 36(2):107–117, 2012.

[91] Jessica Dubois, Lucie Hertz-Pannier, Ghislaine Dehaene-Lambertz, Y Cointepas, and D Le Bihan. Assessment of the early organization and maturation of infants' cerebral white matter fiber bundles: a feasibility study using quantitative diffusion tensor imaging and tractography. *Neuroimage*, 30(4):1121–1132, 2006.

[92] Ashok N Hegde and Sudarshan C Upadhya. Role of ubiquitin–proteasome-mediated proteolysis in nervous system disease. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1809(2):128–140, 2011.

[93] Christoph S Clemen, Marija Marko, Karl-Heinz Strucksberg, Juliane Behrens, Ilka Wittig, Linda Gärtner, Lilli Winter, Frederic Chevessier, Jan Matthias, Matthias Türk, et al. Vcp and psmf1: antagonistic regulators of proteasome activity. *Biochemical and biophysical research communications*, 463(4):1210–1217, 2015.

[94] Misha Kapushesky, Tomasz Adamusiak, Tony Burdett, Aedin Culhane, Anna Farne, Alexey Filippov, Ele Holloway, Andrey Klebanov, Nataliya Kryvych, Natalja Kurbatova, et al. Gene expression atlas update-a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic acids research*, 40(D1):D1077–D1081, 2011.

[95] U. Grenander and M. I. Miller. *Pattern Theory From Representation to Inference*. Oxford University Press, 2007.

[96] M. I. Miller and A. Qiu. The emerging discipline of computational functional anatomy. *NeuroImage*, 45:S16–S39, 2009.

[97] David George Kendall, Dennis Barden, Thomas K Carne, and Huiling Le. *Shape and shape theory*, volume 500. John Wiley & Sons, 2009.

[98] Anuj Srivastava and Eric P Klassen. *Functional and shape data analysis*. Springer, 2016.

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

[100] S. M. Smith and T. E. Nichols. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44:83–98, 2009.

[101] Chao Huang, Paul Thompson, Yalin Wang, Yang Yu, Jingwen Zhang, Dehan Kong, Rivka R Colen, Rebecca C Knickmeyer, Hongtu Zhu, Alzheimer's Disease Neuroimaging Initiative, et al. Fgwas: Functional genome wide association analysis. *NeuroImage*, 159:107–121, 2017.

[102] Sun L, Ji S, and Ye J. *Multi-label dimensionality reduction*. Chapman and Hall/CRC, 2016.

[103] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd)*. Springer, Hoboken, New Jersey., 2009.

[104] MAB McMahon and PM Thompson. Enhancing neuro imaging genetics through meta analysis: global collaborations in psychiatry by the enigma consortium. *European Neuropsychopharmacology*, 27:S715, 2017.

[105] Alexander Petersen, Hans-Georg Müller, et al. Functional data analysis for density functions by transformation to a hilbert space. *The Annals of Statistics*, 44(1):183–218, 2016.

[106] Jin-Ting Zhang. Approximate and asymptotic distributions of chi-squared–type mixtures with applications. *Journal of the American Statistical Association*, 100(469):273–285, 2005.

[107] John Lisman, Howard Schulman, and Hollis Cline. The molecular basis of camkii function in synaptic and behavioural memory. *Nature Reviews Neuroscience*, 3(3):175, 2002.

[108] Yi Yuan, Bei-sha Tang, Ri-li Yu, Kai Li, Zhan-yun Lv, Xin-xiang Yan, and Ji-feng Guo. Marginal association between snp rs2046571 of the has2 gene and parkinson's disease in the chinese female population. *Neuroscience letters*, 552:58–61, 2013.

[109] S Perga, F Montarolo, S Martire, P Berchialla, S Malucchi, and A Bertolotto. Anti-inflammatory genes associated with multiple sclerosis: a gene expression study. *Journal of neuroimmunology*, 279:75–78, 2015.

[110] Ellen M Mowry, Robert F Carey, Maria R Blasco, Jean Pelletier, Pierre Duquette, Pablo Villoslada, Irina Malikova, Elaine Roger, R Phillip Kinkel, Jamie McDonald, et al. Multiple sclerosis susceptibility genes: associations with relapse severity and recovery. *PloS one*, 8(10):e75416, 2013.

[111] H Le-Niculescu, SD Patel, M Bhat, R Kuczenski, SV Faraone, MT Tsuang, FJ McMahon, NJ Schork, JI Nurnberger Jr, and AB Niculescu Iii. Convergent functional genomics of genome-wide association data for bipolar disorder: Comprehensive identification of candidate genes, pathways and mechanisms. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 150(2):155–181, 2009.

[112] Shin-ichi Sakakibara, Yuki Nakamura, Hitoshi Satoh, and Hideyuki Okano. Rna-binding protein musashi2: developmentally regulated expression in neural precursor cells and subpopulations of neurons in mammalian cns. *Journal of Neuroscience*, 21(20):8091–8107, 2001.

[113] Hao Zhou, Vamsi K Ithapu, Sathya Narayanan Ravi, Vikas Singh, Grace Wahba, and Sterling C Johnson. Hypothesis testing in unsupervised domain adaptation with applications in alzheimer's disease. In *Advances in neural information processing systems*, pages 2496–2504, 2016.

[114] Wittawat Jitkrittum, Zoltán Szabó, Kacper P Chwialkowski, and Arthur Gretton. Interpretable distribution features with maximum testing power. In *Advances in Neural Information Processing Systems*, pages 181–189, 2016.

[115] Ruth Heller and Yair Heller. Multivariate tests of association based on univariate tests. In *Advances in Neural Information Processing Systems*, pages 208–216, 2016.

[116] Rajat Sen, Ananda Theertha Suresh, Karthikeyan Shanmugam, Alexandros G Dimakis, and Sanjay Shakkottai. Model-powered conditional independence test. In *Advances in Neural Information Processing Systems*, pages 2955–2965, 2017.

[117] Wittawat Jitkrittum, Wenkai Xu, Zoltán Szabó, Kenji Fukumizu, and Arthur Gretton. A linear-time kernel goodness-of-fit test. In *Advances in Neural Information Processing Systems*, pages 261–270, 2017.

[118] Jeremiah Liu and Brent Coull. Robust hypothesis test for nonlinear effect with gaussian processes. In *Advances in Neural Information Processing Systems*, pages 795–803, 2017.

[119] Fabio Cecchi and Nidhi Hegde. Adaptive active hypothesis testing under limited information. In *Advances in Neural Information Processing Systems*, pages 4038–4046, 2017.

[120] Zhengwu Zhang, Maxime Descoteaux, Jingwen Zhang, Gabriel Girard, Maxime Chamberland, David Dunson, Anuj Srivastava, and Hongtu Zhu. Mapping population-based structural connectomes. *NeuroImage*, 172:130–145, 2018.

[121] Jee Eun Lee, Moo K Chung, Mariana Lazar, Molly B DuBray, Jinsuh Kim, Erin D Bigler, Janet E Lainhart, and Andrew L Alexander. A study of diffusion tensor imaging by tissue-specific, smoothing-compensated voxel-based analysis. *Neuroimage*, 44(3):870–883, 2009.

[122] Matthew Stephens. A unified framework for association analysis with multiple related phenotypes. *PloS one*, 8(7):e65245, 2013.

[123] Debashree Ray, James S Pankow, and Saonli Basu. Usat: A unified score-based association test for multiple phenotype-genotype analysis. *Genetic epidemiology*, 40(1):20–34, 2016.

[124] Xiaobo Guo, Zhifa Liu, Xueqin Wang, and Heping Zhang. Genetic association test for multiple traits at gene level. *Genetic epidemiology*, 37(1):122–129, 2013.

[125] Heping Zhang, Ching-Ti Liu, and Xueqin Wang. An association test for multiple traits based on the generalized kendall's tau. *Journal of the American Statistical Association*, 105(490):473–481, 2010.

[126] J&uuml rg Ott and Daniel Rabinowitz. A principal-components approach based on heritability for combining phenotype information. *Human heredity*, 49(2):106–111, 1999.

[127] Lambertus Klei, Diana Luca, B Devlin, and Kathryn Roeder. Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genetic epidemiology*, 32(1):9–19, 2008.

[128] R. D. Cook, B. Li, and F. Chiaromonte. Envelope models for parsimonious and efficient multivariate linear regression. *Statist. Sinica*, 20:927–1010, 2010.

[129] R. D. Cook, I. S. Helland, and Z. Su. Envelopes and partial least squares regression. *Journal of the Royal Statistical Society, Series B, Statistical Methodology, To appear*, 2013.

[130] Hugues Aschard, Bjarni J Vilhjálmsson, Nicolas Greliche, Pierre-Emmanuel Morange, David-Alexandre Trégouët, and Peter Kraft. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *The American Journal of Human Genetics*, 94(5):662–676, 2014.

[131] H. Chun and S. Keles. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 72:3–25, 2010.

[132] Shuo Xiang, Tao Yang, and Jieping Ye. Simultaneous feature and feature group selection through hard thresholding. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 532–541. ACM, 2014.

[133] Junghi Kim, Yun Bai, and Wei Pan. An adaptive association test for multiple phenotypes with gwas summary statistics. *Genetic epidemiology*, 39(8):651–663, 2015.

[134] David Donoho and Jiashun Jin. Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. *Proceedings of the National Academy of Sciences*, 105(39):14790–14795, 2008.

[135] Jianqing Fan, Fang Han, and Han Liu. Challenges of big data analysis. *National science review*, 1(2):293–314, 2014.

[136] Jianqing Fan, Jinchi Lv, and Lei Qi. Sparse high-dimensional models in economics. 2011.

[137] Jianqing Fan and Jinchi Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):849–911, 2008.

[138] Peter J Bickel and Elizaveta Levina. Covariance regularization by thresholding. *The Annals of Statistics*, pages 2577–2604, 2008.

[139] Peter J Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, pages 199–227, 2008.

[140] Jianqing Fan, Yingying Fan, and Jinchi Lv. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197, 2008.

[141] Shikai Luo, Rui Song, and Daniela Witten. Sure screening for gaussian graphical models. *arXiv preprint arXiv:1407.7819*, 2014.

[142] Michael T Brannick and Paul E Spector. Estimation problems in the block-diagonal model of the multitrait-multimethod matrix. *Applied Psychological Measurement*, 14(4):325–339, 1990.

[143] T Tony Cai, Ming Yuan, et al. Adaptive covariance matrix estimation through block thresholding. *The Annals of Statistics*, 40(4):2014–2042, 2012.

[144] Dinggang Shen and Christos Davatzikos. Measuring temporal morphological changes robustly in brain mr images via 4-dimensional template warping. *NeuroImage*, 21(4):1508–1517, 2004.

[145] Noor Jehan Kabani, Alan C Evans, David J MacDonald, and Colin J Holmes. 3d anatomical atlas of the human brain. *NeuroImage*, 7:S717, 1998.

[146] Steven G Potkin, Guia Guffanti, Anita Lakatos, Jessica A Turner, Frithjof Kruggel, James H Fallon, Andrew J Saykin, Alessandro Orro, Sara Lupoli, Erika Salvi, et al. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for alzheimer's disease. *PloS one*, 4(8):e6501, 2009.

[147] DG Walker, AM Whetzel, and L-F Lue. Expression of suppressor of cytokine signaling genes in human elderly and alzheimer's disease brains and human microglia. *Neuroscience*, 302:121–137, 2015.

[148] Nicola J Rutherford, Minerva M Carrasquillo, Ma Li, Gina Bisceglio, Joshua Menke, Keith A Josephs, Joseph E Parisi, Ronald C Petersen, Neill R Graff-Radford, Steven G Younkin, et al. Tmem106b risk variant is implicated in the pathologic presentation of alzheimer disease. *Neurology*, pages WNL–0b013e318264e3ac, 2012.

[149] SJ Winham, AB Cuellar-Barboza, A Oliveros, SL McElroy, S Crow, C Colby, DS Choi, M Chauhan, M Frye, and JM Biernacka. Genome-wide association study of bipolar disorder accounting for effect of body mass index identifies a new risk allele in tcf7l2. *Molecular psychiatry*, 19(9):1010, 2014.

[150] Sheila P Gregório, Paulo C Sallet, Kim-Anh Do, E Lin, Wagner F Gattaz, and Emmanuel Dias-Neto. Polymorphisms in genes involved in neurodevelopment may be associated with altered brain morphology in schizophrenia: preliminary evidence. *Psychiatry research*, 165(1-2):1–9, 2009.

[151] PH Kuo, LC Chuang, JR Liu, CM Liu, MC Huang, SK Lin, H Sunny Sun, MH Hsieh, H Hung, and RB Lu. Identification of novel loci for bipolar i disorder in a multi-stage genome-wide association study. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 51:58–64, 2014.

[152] Aad W Van Der Vaart and Jon A Wellner. Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer, 1996.

[153] Hanxiang Peng and Anton Schick. Asymptotic normality of quadratic forms with random vectors of increasing dimension. *Journal of Multivariate Analysis*, 164:22–39, 2018.