Statistical Methods for Imaging Genetic Data

Ja-an Lin

A dissertation submitted to the faculty of 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.

Chapel Hill
2013

Approved by:

Dr. Hongtu Zhu

Dr. Joseph G Ibrahim

Dr. Rebecca Santelli

Dr. Wei Sun

Dr. Fei Zou

Abstract

JA-AN LIN: Statistical Methods for Imaging Genetic Data
(Under the direction of Dr. Hongtu Zhu)

More and more large-scale imaging genetic studies are being widely conducted to collect a rich set of imaging, genetic, and clinical data in order to detect susceptibility genes for complexly inherited diseases including common mental disorders (e.g., schizophrenia) and neurodegenerative disorders, among many others. However, the development of statistical and computational methods for the joint analysis of complex imaging phenotypes, genetic data, and clinical data has fallen seriously behind the technological advances. The aim this work is to develop three statistical approaches called Projection Regression Method (PRM) and functional mixed effects model (FMEM) for the joint analysis of high-dimensional imaging data with a set of genetic markers. In PRM, it generalizes a statistical method based on the principal component of heritability for association analysis in genetic studies of complex multivariate phenotypes. The key components of the PRM include an estimation procedure for extracting several principal directions of multivariate phenotypes relating to covariates and a test procedure based on wild-bootstrap method for testing for the association between the weighted multivariate phenotype and explanatory variables. Simulation studies and an imaging genetic dataset are used to examine the finite sample performance of the PRM. In FMEM, to accommodate the correlation structure of the marker set, we model the genetic effects as population-shared random effects with a common variance component (VC), whereas to accommodate spatial feature in imaging data, we spatially model varying associations between imaging measures in a three-dimensional (3D) volume (or 2D surface) with a set of covariates and the genetic random effects. We develop a two-stage estimation

procedure to spatially and adaptively estimate the varying coefficient functions, while preserving its edges among different piecewise-smooth regions. To test hypothesis of interest, we provide two test statistics with well-controlled type I error and better power comparing to traditional voxel-based approach. Simulation studies and a real data analysis of the AlzheimerâĂŹs Disease Neuroimage Initiative (ADNI) show that FMEM significantly outperforms voxel-based approaches in terms of identification of activation regions.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Literature Review

Many studies have been collecting/collected multivariate phenotypes in order to investigate their relationship with some explanatory variables of interest. For example, multivariate imaging phenotypes have been widely collected to characterize brain structures and their functions [Knickmeyer et al., 2008; Lenroot and Giedd, 2006]. Such multivariate imaging phenotypes include diffusion tensor, deformation tensors of deformation field, the hemodynamic response function of functional magnetic resonance images, and the spherical harmonic boundary description of subcortical structures, among many others [Zhu et al., 2007; Styner et al., 2004; Friston, 2007; Huettel et al., 2004; Taylor and Worsley, 2008; Worsley et al., 2004]. Statistical analysis of these multivariate imaging phenotypes with explanatory variables eventually leads to a better understanding of the progression of neuropsychiatric and neurodegenerative diseases or the normal brain development/aging [Styner et al., 2003, 2004; Friston, 2007; Huettel et al., 2004; Taylor and Worsley, 2008; Worsley et al., 2004]. Among the studies, there is a large amount focusing on the whole brain MR images as the multivariate phenotypes and the genomic data as covariate of interest. Due to the features of ultra-high dimension and complex noise structure in both imaging and genetic data, statistical analyses usually take candidate-phenotype/candidate-genotype, candidate-phenotype/genome-wide [Potkin et al., 2009] and phenotype-wide/candidate-genomes [Braskie and et al, 2011].

There are six commonly used approaches to delineate the association between the phenotypes and genotypes.

1.1.1 Multivariate Linear Model

A standard statistical approach to this problem is to fit a multivariate linear model (MLM) to the multivariate phenotype with the candidate genotypes. The mathematical form of the MLM is usually formulated as

$$\mathbf{Y}_{n\times q} = \mathbf{X}_{n\times p}\mathbf{B}_{p\times q} + E_{n\times q}, \tag{1.1}$$

where n is the total number of subjects, p is the number of covariates including genotypes, q is the dimension of phenotype, each row of the matrices $\mathbf{Y}, \mathbf{X}$ and $\mathbf{E}$ are the phenotype, covariates and random errors for a subject, respectively. With the normality and independence assumption, to test the effect of a single covariate to the whole phenotype, typically Hotelling's T^2 has been used to test hypotheses of interest [Chung et al., 2010; Taylor and Worsley, 2008; Worsley et al., 2004]. Suppose the null hypothesis is formulated as $H_0 : C_{q\times 1}\mathbf{B} = b_0$, where b_0 is a constant. The Hotelling's T^2 statistic is calculated as

$$T^2 = \frac{n-p-q+1}{(n-p)q}(C\hat{\mathbf{B}} - b_0)^T \widehat{Cov(C\mathbf{B})}^{-1}(C\hat{\mathbf{B}} - b_0), \tag{1.2}$$

and it distributes as F-distribution with the numerator degree of freedom q and the denominator degree of freedom $n-p-q+1$. Since MLM involves estimating the covariance matrix of all individual phenotypes, the dimension of the multivariate phenotype is relatively smaller than the sample size.

1.1.2 Component-wise Method

An alternative approach is to fit a marginal linear model and calculate a test statistic for each component of the multivariate phenotype. Then it combines all tests with their associated $p-$values to test an overall hypothesis across all individual phenotypes and adjusting for multiple comparison [Heller et al., 2007; Lazar et al., 2002]. The commonly used adjustment are Bonferroni correction, false discovery rate (FDR) and random field theory (RFT) if the multivariate phenotype is a whole brain image. The Bonferroni correction controls for over all familywise error rate. The FDR controls the expected portion of falsely discovered p-values. Random field theory gives a threshold considering the spatial correlation of the image by involving the imaging smoothness and image size, to cut of p-values.Besides the potential estimation bias due to ignorance of the potential correlation among all individual phenotypes, Bonferroni correction suffers from low statistical power and FDR has inflated type I errors

1.1.3 Principle Component Analysis

Another approach is to directly reduce the dimension of the multivariate phenotype by using dimension reduction techniques, such as principal component analysis (PCA). The procedure of principle component analysis starts with constructing a low dimensional space formed by the eigenvectors of multivariate phenotype and projecting the original data to this low dimensional space. Then the projected pheotype is fitted a MLM with the covariates of interest[Formisano et al., 2008; Teipel et al., 2007; Rowe and Hoffmann, 2006; Kherif et al., 2002]. This method does not properly account for the variation of covariates and their association with the individual phenotypes while projection. Thus it results in low statistical power while examining the phenotype-genotype association.

1.1.4 Partial Least Square Regression

Partial least squares regression (PLSR) is another statistical method that finds a linear regression model by projecting the multivariate phenotype and the explanatory variables to a new and smaller space [Chun and Keles, 2010; Krishnan et al., 2011]. Conceptually, PLSR is trying to find the multidimensional direction in the phenotype space that explains the maximum variance direction in the phenotype space. Suppose we have a MLM model for the multivariate phenotype and covariates of interest with the notations following (1.1). The PLSR is implemented by extracting factors from both $\mathbf{X}$ and $\mathbf{Y}$ such that covariance between the extracted factors is maximized. Pratically, the following underlying model is assumed:

$$\mathbf{X} = TP^T + E \tag{1.3}$$

$$\mathbf{Y} = UQ^T + F, \tag{1.4}$$

where T and U are two $n \times l$ matrices representing the scores similar to the PCs in PCA framework and P and Q are the loadings in the similar context. E and F are unexplained part. PLSR tries to achieve its goal by maximizing the covariance between T and U. However, this method focuses on prediction and classification, instead of investigating the association between the multivariate phenotype and the covariates of interest.

1.1.5 Sparse Reduced Rank Regression

An alternative method to implement the genome-wide/phenotype-wide association study is sparse reduced rank regression method (sRRR) [Vounou et al., 2010]. It is a multivariate modeling of high-dimensional imaging responses which enforces sparsity in the regression coefficients such that the model performs simultaneous genotype and phenotype selection. The reduced-rank regression model (RRR) models the simultaneous

dependence of the phenotype on the genotype such that

$$\mathbf{Y} = \mathbf{XB} = \mathbf{X}WA + E \tag{1.5}$$

where the notation follows (1.1), W is the $p \times r$ matrix of regression coefficients for the p genotypes, and A is the $r \times q$ matrix of regression coefficients for the q phenotype, both of full rank r. The factorization of the regression coefficient matrix $\mathbf{B} = WA$ comes from imposing a reduced rank condition on C, namely that rank(C) is $r \leq min(p, q)$. To find out the regulating genotypes to the phenotype, the sRRR model further involves a L_1 penalty on each column of regression coefficients. Although it may successfully identify the activation voxels, it still does not account for the spatial correlation in phenotypes as well as the LD structure in genotypes only the representing genotype will be selected by L_1 penalty instead of a group of regulation genotypes. The other drawback of this approach is its incapableness of performing hypothesis testing.

1.1.6 Least Square Kernel Machine Method

Another group of approaches use lease square kernel machine to accommodate the LD structure of a group of genes when investigating the relationship between imaging phenotype and a group of genetic information. In Liu et al. [2007], they assume for the i-th subject, the univariate response of interest y_i can be modeled by

$$y_i = x_i^T \beta + h(z_i) + e_i, \tag{1.6}$$

where $x_i$is the p-dimension covariate of interest of the i-th subject; β is p by 1 vector of regression coefficient; z_i is a L by 1 vector of genetic variation for the i-th subject; $h(z_i)$ is an unknown centered smoothed function modeling by a kernel lying in the reproducing kernel space H_K generated by a kernel function $K(\bullet, \bullet)$; and the errors e_i are assumed to be independent and follow $N(0, \sigma^2)$. The regression coefficients β and the function

$h(\bullet)$ are estimated via maximizing the scaled penalized likelihood function

$$J(h,\beta) = -0.5\sum_{i=1}^{n} y_i - x_i\beta - h(z_i) - 0.5\lambda\|h\|^2_{H_K} \tag{1.7}$$

By assuming the function $h(\bullet)$ in the space H_K generating by $K(\bullet,\bullet)$, $h(\bullet)$ can be represented as

$$h(\bullet) = \sum_{i=1}^{n} \alpha_i K(\bullet, z_i), \tag{1.8}$$

where $\alpha = (\alpha_1, \cdots, \alpha_n)^T$ are unknown parameters. The representing form leads the objective function (1.7) becomes

$$J(h,\beta) = -0.5\sum_{i=1}^{n} y_i - x_i\beta - h(z_i) - 0.5\lambda\alpha^T\mathcal{K}\alpha, \tag{1.9}$$

where $\mathcal{K}$ is a n by n matrix with the (i,j)-th element as $\mathcal{K}(z_i, z_j)$. The normal equation to solve the parameters β and α of (1.9) is the same as the linear mixed model

$$y = X\beta + h + e, \tag{1.10}$$

when solving for β and h where β is a p by 1 vector of regression coefficient, h is an n by 1 vector of random effects with distribution $N(0, \tau\mathcal{K})$, and e $\sim$ $N(0, \sigma^2\mathcal{I})$. With this connection between (1.8) and (1.10), it is sufficient to test the integrated effect of genetic variation to the response of interest y by testing the null hypothesis $H_0 : \tau = 0$ using sophisticatedly developed linear mixed model theory. Moreover, this model enables considering the LD structure among a group of genetic variation even when the dimension is high. In Ge and et al [2012], it applies LSKM to structure imaging responses that they repeatedly fit the model (1.6) for each voxel following with testing the null hypothesis $H_0 : \tau = 0$ for overall genetic effect and then apply random field theory (RFT) on p-value map to accommodate imaging spatial structure and multiple comparison adjustment.

1.2 Introduction to our approach

There is a trend in imaging genetics which utilizes phenotype-wide and genome-wide approaches [Hibar and et al, 2011; Shen et al., 2010; Stein et al., 2010; Ge and et al, 2012]. The standard approach is to repeatedly fit statistical models for each voxel by each genome or gene. Besides computational burden , there are some other limitations. First, the p-values of significant genomes at each nearby voxel may have different order, which makes the biological interpretation complicated. Second, some imaging measures are affected by the interaction between genomes rather than a single genome, which results in missing identification of such genomes. Last, the spatial coherence and spatially contiguous region of activation are not considered in voxel-wise approach thus causing statistical power loss.

Based on the limitations described above, we proposed the following three methods. The first method is called the projection regression model (PRM), which is a new statistical framework, called the projection regression model (PRM). The PRM includes simultaneous selection, estimation, and testing in a general regression setting. We develop an estimation procedure for estimating the optimal weights of the multivariate response in the PRM, while properly accounting for the space of explanatory variables. Particularly, the PRM can accommodate the case that the sample size is relatively smaller than the dimension of the multivariate phenotype. We also propose a test procedure based on a wild-bootstrap method, which leads to a single $p-$value to test for the association between the projected weighted multivariate phenotype and the covariates of interest, such as genetic markers. This test procedure controls the overall type I error, while avoiding the use of an inefficient sample splitting method [Mukhopadhyay et al., 2010; Yang et al., 2010]. Simulation studies are carried out to compare the PRM with several commonly used methods for the multivariate phenotype in terms of both the type I and II error rates. The second proposed method is called functional mixed effects model (FMEM). Specifically, to accommodate the LD structure

of a group of genetic markers, we model the genetic effects as population-shared random effects with a common variance component (VC), whereas to accommodate spatial feature in imaging data, we spatially model varying associations between imaging measures in a three-dimensional (3D) volume (or 2D surface) with a set of covariates and the genetic random effects. Moreover, we assume that the varying associations are piecewise smooth functions with unknown edges and jumps across voxels. We develop a two-stage estimation procedure to spatially and adaptively estimate the varying coefficient functions. Each stage of the estimation procedure includes a multiscale adaptive estimation and testing procedure to independently estimate each varying coefficient function, while preserving its edges among different piecewise-smooth regions. Simulation studies and real data analysis show that FMEM significantly outperforms voxel-wide approaches in terms of both type I and II error rates. However, with wild-bootstrap method in hypothesis testing , the computation is intensive for FMEM. In our third work, we enhance the computational method by revising the weighted likelihood ratio test to the weighted score test which gives comparable statistical power to identify activation regions with more efficient computation. In the weighted likelihood ratio test, the statistical power is enhanced by providing more accurate estimation of VC which results in smaller variance of test statistic. However, the rationale in weighted score test is different that the detail is described in later section.

The rest of this section is organized as follows. Chaper 2 will describe the proposed methods PRM and the result. 2.1 introduces the PRM and its associated estimation and testing procedure. In 2.2, we conduct simulation studies with a known ground truth to examine the finite sample performance of the PRM and several other statistical methods. 2.2 also illustrates an application of PRM in an imaging genetic data set. The detail result of simulation studies and data analysis are given in Appendix. Chapter 3, we describe the proposed model FMEM In detail, the model and its adaptive estimation procedure is in 3.1. The performance of FMEM is evaluated by simulation studies

and real data analysis in 3.2 The analysis result of applying voxel-based method on the same real data is also given. The supporting information of simulation result and real data analysis are in Appendix; detailed derivation of the exact null distribution to test the variance component in FMEM is given in Appendix. The rationale and detailed methodology of weighted score test for FMEM is in 4.1. In 4.2, we evaluate the performance of weighted score test of FMEM by simulation studies and a real data analysis with ADNI. In Chaper 5, we give our discussion and conclusion.

Chapter 2

Projection Regression Method

2.1 Methods

Suppose that we observe a $q \times 1$ multivariate phenotype $\mathbf{y}_i = (y_{i1}, \ldots, y_{iq})^T$ and a $p \times 1$ vector of covariates of interest $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T$ for $i = 1, \ldots, N$. We consider a commonly used MLM as follows:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \quad \text{or} \quad \mathbf{y}_i = \mathbf{B}^T\mathbf{x}_i + \mathbf{e}_i, \tag{2.1}$$

where $\mathbf{Y}$ is an $N \times q$ matrix formed by the $q \times 1$ multivariate phenotype of each subject in each row, $\mathbf{X}$ is an $N \times p$ matrix consisting of the $p \times 1$ vector of covariates of each subject in each row, and $\mathbf{B} = (\beta_{jl})$ is a $p \times q$ matrix, in which β_{jl} represents the effect of the $j-$th covariate on the $l-$th response. Moreover, $\mathbf{E}$ is an $N \times q$ matrix representing the random errors and $\mathbf{e}_i^T$ is the $i-$th row of $\mathbf{E}$ with zero mean and covariance matrix V_R. Assuming that $\mathbf{x}_i$ and $\mathbf{e}_i$ are independent, the covariance of $\mathbf{y}_i$ is given by

$$\text{Cov}(\mathbf{y}_i) = V_Q + V_R = \mathbf{B}^T\text{Cov}(\mathbf{x}_i)\mathbf{B} + V_R, \tag{2.2}$$

where V_Q represents the variation coming from the covariates of interest.

Most scientific questions require the comparison across two (or more) diagnostic groups and the association of the genetic marker for each component of $\mathbf{y}_i$. Such ques-

tions can often be formulated as linear hypotheses of $\mathbf{B}$ as follows:

$$H_0 : \mathbf{CB} = \mathbf{B}_0 \quad \text{v.s.} \quad H_1 : \mathbf{CB} \neq \mathbf{B}_0, \tag{2.3}$$

where $\mathbf{C}$ is a $r \times p$ matrix of full row rank and $\mathbf{B}_0$ is a $p \times q$ vector of constants.

We consider a projection of $\mathbf{y}_i$ via a $q \times k$ weight matrix $\mathbf{W}$ and create a $k \times 1$ projection vector $\mathbf{W}^T\mathbf{y}_i$ such that $k << q$. Then, we propose a projection regression model (PRM) given by

$$\mathbf{W}^T\mathbf{y}_i = \beta_{\mathbf{w}}^T\mathbf{x}_i + \varepsilon_i, \tag{2.4}$$

where $\beta_{\mathbf{w}}$ is a $p \times k$ regression coefficient matrix and ε_i is the random vector with $\text{Cov}(\varepsilon_i) = \Sigma_i$. The PRM (2.4) is a heteroscedastic multivariate linear model. When $k = 1$ and $\Sigma_i = \Sigma$ for all i, PRM reduces to the pseudo-phenotype model considered in [Amos et al., 1990, Amos and Laing, 1993, Ott and Rabinowitz, 1999, Lange et al., 2004, Klei et al., 2008]. A direct connection between models (2.1) and (2.4) is that model (2.1) can be rewritten as

$$\mathbf{W}^T\mathbf{y}_i = \beta_{\mathbf{w}}^T\mathbf{x}_i + \varepsilon_i = (\mathbf{BW})^T\mathbf{x}_i + \mathbf{W}^T\mathbf{e}_i. \tag{2.5}$$

Therefore, if $\mathbf{W}$ in (2.4) were known, then one would directly perform an appropriate hypothesis test to address specific research hypotheses as follows:

$$H_{0W} : \mathbf{C}\beta_{\mathbf{w}} = \mathbf{b}_0 \quad \text{v.s.} \quad H_{1W} : \mathbf{C}\beta_{\mathbf{w}} \neq \mathbf{b}_0, \tag{2.6}$$

where $\mathbf{b}_0$ is an $r \times k$ vector of constants. Based on model (2.5), the null hypothesis of (2.6) can be written as $\mathbf{C}\beta_{\mathbf{w}} = \mathbf{CBW} = \mathbf{B}_0\mathbf{W} = \mathbf{b}_0$.

Let $\mathbf{C}_1$ be a $(p - r) \times p$ matrix such that

$$\text{rank}[\mathbf{C}^T\ \mathbf{C}_1^T] = p \ \text{ and } \ \mathbf{C}\mathbf{C}_1^T = \mathbf{0}. \tag{2.7}$$

Let $\mathbf{D} = [\mathbf{C}^T\ \mathbf{C}_1^T]^T$ be a $p \times p$ matrix and $\tilde{\mathbf{x}}_i = (\tilde{\mathbf{x}}_{i1}^T, \tilde{\mathbf{x}}_{i2}^T) = \mathbf{D}^{-T}\mathbf{x}_i$ be a $p \times 1$ vector, where $\tilde{\mathbf{x}}_{i1}$ and $\tilde{\mathbf{x}}_{i2}$ are, respectively, the $r \times 1$ and $(p-r) \times 1$ subvectors of $\tilde{\mathbf{x}}_i$. We define $\tilde{\mathbf{B}} = [\tilde{\mathbf{B}}_1^T \tilde{\mathbf{B}}_2^T]^T$ to be $\tilde{\mathbf{B}} = \mathbf{DB}$ or $\mathbf{B} = \mathbf{D}^{-1}\tilde{\mathbf{B}}$. We consider $\tilde{\mathbf{B}} = [\tilde{\mathbf{B}}_1^T \tilde{\mathbf{B}}_2^T]^T$, where $\tilde{\mathbf{B}}_1$ and $\tilde{\mathbf{B}}_2$ are, respectively, the first r rows and the last $p-r$ rows of $\mathbf{B}$. Therefore, model (2.5) can be rewritten as

$$\mathbf{W}^T\mathbf{y}_i = (\mathbf{D}^{-1}\tilde{\mathbf{B}}\mathbf{W})^T\mathbf{x}_i + \mathbf{W}^T\mathbf{e}_i = \mathbf{W}^T\tilde{\mathbf{B}}_1^T\tilde{\mathbf{x}}_{i1} + \mathbf{W}^T\tilde{\mathbf{B}}_2^T\tilde{\mathbf{x}}_{i2} + \mathbf{W}^T\mathbf{e}_i. \tag{2.8}$$

The next issue is to determine an optimal $q \times k$ matrix $\mathbf{W}$ under some certain criteria. In PCH [Ott and Rabinowitz, 1999; Lange et al., 2004; Klei et al., 2008], the heritability ratio is defined by

$$h(\mathbf{w}) = \frac{\mathbf{w}^T V_Q \mathbf{w}}{\mathbf{w}^T \mathrm{Cov}(\mathbf{y}_i)\mathbf{w}} = \frac{\mathbf{w}^T V_Q \mathbf{w}}{\mathbf{w}^T V_Q \mathbf{w} + \mathbf{w}^T V_R \mathbf{w}}. \tag{2.9}$$

The heritability ratio characterizes the ratio of the variation from the genetic biomarkers $\mathbf{x_i}$ to the total variation of responses $\mathbf{y_i}$. Maximizing $h(\mathbf{w})$ leads to the optimal $\mathbf{W}$.

Instead of directly using the heritability ratio $h(\mathbf{w})$, we consider a generalized 'heritability' ratio $H(\mathbf{w})$ for a given $q \times 1$ vector $\mathbf{w}$ as follows:

$$H(\mathbf{w}) = \frac{\mathbf{w}^T \tilde{\mathbf{B}}_1^T \mathrm{Cov}(\tilde{\mathbf{x}}_{i1}) \tilde{\mathbf{B}}_1 \mathbf{w}}{\mathbf{w}^T V_R \mathbf{w}}. \tag{2.10}$$

The $H(\mathbf{w})$ can be interpreted as the ratio of the variance of $\mathbf{w}^T\tilde{\mathbf{B}}_1^T\tilde{\mathbf{x}}_{i1}$ relative to that of $\mathbf{w}^T\mathbf{e}_i$ under the null hypothesis. We require that the optimal $\mathbf{W}$ enhances the power of detecting the association between $\mathbf{W}^T\mathbf{y}_i$ and $\mathbf{x}_i$ for the null hypothesis (2.6). Thus, we need to find a $\mathbf{W}$ to project the data into a space containing the most information on the null hypothesis of (2.3). Let $\Sigma_X = \mathrm{Cov}(\mathbf{x})$. It can be shown that $\tilde{H}(\mathbf{w})$ reduces to

$$\tilde{H}(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{B}^T\mathbf{C}^T(\mathbf{D}^{-T}\Sigma_X\mathbf{D}^{-1})_{(1,1)}\mathbf{CBw}}{\mathbf{w}^T V_R \mathbf{w}}, \tag{2.11}$$

where $(\mathbf{D}^{-T}\Sigma_X\mathbf{D}^{-1})_{(1,1)}$ is the upper $r \times r$ submatrix of $\mathbf{D}^{-T}\Sigma_X\mathbf{D}^{-1}$. When $\mathbf{C} = [\mathbf{I}_r\, \mathbf{0}]$, $\tilde{H}(\mathbf{w})$ reduces to the ratio of $\mathbf{w}^T\mathbf{B}_1^T(\Sigma_X)_{(1,1)}\mathbf{B}_1\mathbf{w}$ to $\mathbf{w}^T V_R\mathbf{w}$, in which $(\Sigma_X)_{(1,1)}$ is the upper $r \times r$ submatrix of Σ_X.

When V_R is positive definite, maximizing (2.11) is equivalent to maximizing

$$\tilde{H}(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{L}\mathbf{L}^{-1}\mathbf{B}^T\mathbf{C}^T(\mathbf{D}^{-T}\Sigma_X\mathbf{D}^{-1})_{(1,1)}\mathbf{C}\mathbf{B}(\mathbf{L}^{-1})^T\mathbf{L}^T\mathbf{w}}{\mathbf{w}^T\mathbf{L}\mathbf{L}^T\mathbf{w}}, \tag{2.12}$$

where $\mathbf{L}$ is the lower triangular matrix obtained from the Cholesky decomposition of $V_R = \mathbf{L}\mathbf{L}^T$. Letting $V_{C,X} = \mathbf{L}^{-1}\mathbf{B}^T\mathbf{C}^T(\mathbf{D}^{-T}\Sigma_X\mathbf{D}^{-1})_{(1,1)}\mathbf{C}\mathbf{B}(\mathbf{L}^{-1})^T$. Let $\mathbf{v}$ be the eigenvector corresponding to the largest eigenvalue of the matrix $V_{C,X}$, then (2.11) is maximized when $\mathbf{L}^T\hat{\mathbf{w}}$ equals $\mathbf{v}$. Hence, (2.12) is maximized when $\hat{\mathbf{w}}$ equals $\mathbf{L}^{-T}\mathbf{v}$. If q is relatively small compared to N, based on (2.11), we take the $q \times k$ matrix $\mathbf{W}$ in (2.4) by choosing the largest k sparse eigenvectors of $V_{C,X}$ using PCA. However, when q is relatively large compared to N, calculating $\mathbf{L}^{-T}$ and the eigenvectors of $V_{C,X}$ can be challenging, which makes the optimal weight matrix $\mathbf{W}$ very unstable.

2.1.1 Estimation Procedure for Optimal Weights

We develop an estimation procedure for estimating the optimal weights. This procedure consists of three major steps: (i) a pre-screening process for eliminating 'unrelated' measures; (ii) a shrinkage procedure for approximating $V_{C,X}$ and V_R; and (iii) a sparse principal component analysis (SPCA) procedure for calculating the eigenvalue-eigenvector pairs of $V_{C,X}$. Each step is implemented as follows.

The pre-screening procedure is to rank individual phenotypes according to marginal utility and eliminate 'unrelated' phenotypes when q is relatively large relative to N, say $q \geq N/3$. This procedure is to mimic various screening methods, such as sure independence screening (SIS), for discarding covariates in high-dimensional linear models [Fan and Lv, 2010]. In Step 1, we fit q marginal linear regression models to individual pheno-

types and the covariates of interest. In Step 2, we calculate the corresponding Wald-type test statistics under the same null hypothesis (2.6), and the respective p-values from a chi-square distribution with degrees of freedom r for each individual phenotype. In Step 3, after ordering the q p-values from the smallest to the largest, we only select the phenotypes with the first $q^* = [q/\log(q)] + 1$ if $q \leq N$, or the first $q^* = [N/\log(N)] + 1$ if $q > N$, where $[x]$ represents the largest integer smaller than x. Thus, we set the weights for those unselected individual phenotypes to be zero, or equivalently, we consider a reduced response vector, denoted as $\mathbf{y}_i^* = (\tilde{y}_{i1}, \ldots, \tilde{y}_{iq^*})^T$ or $\mathbf{Y}^*$.

The shrinkage procedure is to approximate $V_{C,X}$ and V_R as follows. In Step 1, we refit the multivariate linear regression in (2.1) with the selected individual phenotypes in $\mathbf{y}_i^*$ as responses conditional on $\mathbf{X}$. Let $\mathbf{B}^*$ be the regression parameter matrix for the selected individual phenotypes. We estimate $\mathbf{B}^*$ by its least square estimator, denoted by $\hat{\mathbf{B}}^*$, which equals $\hat{\mathbf{B}}^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}^*$. In Step 2, we estimate $\mathrm{Cov}(\mathbf{X})$ by using its empirical estimator, denoted by $\hat{\Sigma}_X$, and then approximate $V_B = \mathbf{B}^T\mathbf{C}^T(\mathbf{D}^{-T}\Sigma_X\mathbf{D}^{-1})_{(1,1)}\mathbf{C}\mathbf{B}$ by $\widehat{V_B} = \hat{\mathbf{B}}^{*T}\mathbf{C}^T(\mathbf{D}^{-T}\Sigma_X\mathbf{D}^{-1})_{(1,1)}\mathbf{C}\hat{\mathbf{B}}^*$. In Step 3, we calculate a shrinkage estimate of V_R by following Ledoit and Wolf [2004]. Let C_E be the sample covariance matrix of $\hat{\mathbf{E}}^* = (r_{jk}) = \mathbf{Y}^* - \mathbf{X}\hat{\mathbf{B}}^*$, $\mu_E = q^{-1}\mathrm{tr}(C_E)$ and $\rho = \min(1, N^{-2}\sum_{i=1}^{N}\mathrm{tr}[(\hat{\mathbf{e}}_i\hat{\mathbf{e}}_i^T - C_E)^2]/\mathrm{tr}[(C_E - \mu_E\mathbf{I}_q)^2])$, in which $\hat{\mathbf{e}}_i = \mathbf{y}_i^* - \hat{\mathbf{B}}^{*T}\mathbf{x}_i$. Finally, we approximate V_R and $V_{C,X}$ by using $\hat{V}_{R,S} = \rho\mu_E\mathbf{I}_q + (1-\rho)C_E$ and $\widehat{\mathbf{L}}^{-1}\widehat{V}_B(\widehat{\mathbf{L}}^{-T})$, respectively. We use $\hat{V}_{R,S}$ mainly due to its computational efficiency and relatively nice properties [Ledoit and Wolf, 2004].

The SPCA procedure is to estimate the sparse eigenvectors and eigenvalues of $\hat{V}_{R,S}$ by following Zou et al. [2006] as follows. The key idea of this SPCA process is to transform the eigenvalue-eigenvector problem into an elastic net problem [Zou et al., 2006], which can be solved neatly. We include the key steps here for completion. In Step 1, we choose a value of k so that the proportion of variance explained is greater than a certain threshold, such as 80% percent to truncate the eigenvalues. Then, we

calculate the loadings of the first k ordinary principal components of $\hat{V}_{R,S}$, denoted as α. In Step 2, given a fixed α, we solve the following naive elastic net problem: for $j = 1, \ldots, k$,

$$\hat{\gamma}_j = \underset{\gamma^*}{\text{argmin}} \gamma^{*T}(\hat{V}_{R,S} + \lambda_{2,j})\gamma^* - 2\alpha_j^T \hat{V}_{R,S}\gamma^* + \lambda_{1,j} \mid \gamma^* \mid_1, \tag{2.13}$$

where $| \cdot |_1$ denotes the L_1 norm. Moreover, $\lambda_{1,j}$ and $\lambda_{2,j}$ are tuning parameters and selected simultaneously by using a BIC-type selection criterion [Leng and Wang, 2009]. We calculate the BIC-type criterion given by

$$\text{BIC} = (\alpha_j - \hat{\gamma}_j)^T \hat{V}_{R,S}(\alpha_j - \hat{\gamma}_j) + df_{(\lambda_{1,j},\lambda_{2,j})} \times \frac{\log(q^*)}{q^*}, \tag{2.14}$$

where $df_{(\lambda_{1,j},\lambda_{2,j})}$ is the number of nonzero coefficients in $\hat{\gamma}_j$. In Step 3, for each fixed $\hat{\gamma}_j$, we calculate the singular value decomposition of $\hat{V}_{R,S}\hat{\gamma}_j = UDV^T$, and then we update $\alpha_j = UV^T$ for $j = 1, \ldots, k$. In Step 4, we repeat steps 2-3, until γ converges. In Step 5, we normalize γ, and then set $\hat{\mathbf{v}}_j = \gamma_j / \mid \gamma_j \mid$ for $j = 1, \ldots, k$. The optimal weight $\mathbf{w}_j$ is estimated by using $\hat{\mathbf{w}}_j = (\widehat{\mathbf{L}}^{-T})\hat{\mathbf{v}}_j$ for $j = 1, \ldots, k$ and $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_k]$.

Finally, to further reduce the dimension of the pre-screened $\mathbf{Y}^*$, we apply the SPCA procedure repeatedly to estimate $\mathbf{W}$ by selecting 'related' individual phenotypes suggested from the estimated weight matrix $\mathbf{W}$ obtained from the previous iteration. Specifically, we eliminate the responses corresponding to the zero rows in the sparse weight matrix $\mathbf{W}$ obtained from the SPCA procedure in order to reduce the screened response vector $\mathbf{Y}^*$ to an even smaller dimension. Subsequently, we rerun the shrinkage and SPCA procedures on the new $\mathbf{Y}^*$ to calculate the new weight matrix $\mathbf{W}$. This iteration process of weight estimation can be processed iteratively until $\mathbf{W}$ converges. Our simulation studies show that in most cases, the process converges in only two iterations.

2.1.2 Test Procedure for Testing Hypotheses

We develop several statistics of testing H_{0W} against H_{1W} for the PRM (2.4) as follows. Given the estimated weight matrix $\mathbf{W}$, we can calculate the ordinary least squares estimate of $\beta_{\mathbf{w}}$, given by $\hat{\beta}_{\mathbf{w}} = (\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^T)^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{y}_i^T \mathbf{W}$. Subsequently, to calculate a statistic for testing H_{0W} against H_{1W}, we calculate a $k \times k$ matrix, denoted by T_N, as follows:

$$T_N = (\mathbf{C}\hat{\beta}_{\mathbf{w}} - \mathbf{b}_0)^T \Sigma_{\tilde{\Omega}}^{-1} (\mathbf{C}\hat{\beta}_{\mathbf{w}} - \mathbf{b}_0), \tag{2.15}$$

where $\Sigma_{\tilde{\Omega}}$ is a consistent estimate of the covariance matrix of $\mathbf{C}\hat{\beta}_{\mathbf{w}} - \mathbf{b}_0$ given by

$$\Sigma_{\tilde{\Omega}} = \mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1} \sum_{i=1}^{N} a_i^2 \mathbf{x}_i \tilde{\epsilon}_i^T \tilde{\epsilon}_i \mathbf{x}_i^T (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{C}^T. \tag{2.16}$$

Moreover, $a_i = 1/\{1 - \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i\}$ and $\tilde{\epsilon}_i = \mathbf{W}^T\mathbf{y}_i - \tilde{\beta}_{\mathbf{w}}^T\mathbf{x}_i$, where $\tilde{\beta}_{\mathbf{w}}$ is the restricted least squares (RLS) estimate of β under H_0, and is given by

$$\tilde{\beta}_{\mathbf{w}} = \hat{\beta}_{\mathbf{w}} - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T]^{-1}(\mathbf{C}\hat{\beta}_{\mathbf{w}} - \mathbf{b}_0). \tag{2.17}$$

When $k = 1$, T_N is a Wald-type (or Hotelling's T^2) test statistic. When $k > 1$, we define three test statistics based on the functionals of T_N as follows:

$$W_N = \det(T_N), \quad \mathrm{Tr}_N = \mathrm{trace}(T_N), \quad \text{and} \quad \mathrm{Roy}_N = \max(\mathrm{eig}(T_N)), \tag{2.18}$$

where det, trace, and eig denote the determinant, trace and eigenvalues of a symmetric matrix, respectively. When $k = 1$, all these statistics reduce to T_N. For simplicity, we focus on Tr_N throughout the paper.

We present a wild bootstrap method to improve the finite sample performance of the test statistic Tr_N in (2.18) in testing the null hypothesis H_0. First, we fit model (2.1) under the null hypothesis (2.3) and calculate the estimated multivariate regression

coefficients under (2.3), denoted by $\widehat{\mathbf{B}}_*$, with corresponding residuals $\hat{\mathbf{e}}_i = \mathbf{y_i} - \widehat{\mathbf{B}}_*^{\mathbf{T}}\mathbf{x_i}$ for $i = 1, \ldots, N$. Then, we generate G bootstrap samples $\{(\mathbf{z}_i^{(g)}, \mathbf{x}_i) : i = 1, \ldots, N\}$ as follows:

$$\mathbf{z}_i^{(g)} = \widehat{\mathbf{B}}_*^T\mathbf{x_i} + \eta_i^{(g)}\hat{\mathbf{e}}_\mathbf{i} \quad \text{for} \quad i = 1, \ldots, N, \tag{2.19}$$

where $\eta_i^{(g)}$ are independently and identically distributed as a distribution d, in which d is chosen as

$$\eta_i^{(g)} = \begin{cases} 1, & \text{with probability } 0.5, \\ -1, & \text{with probability } 0.5. \end{cases} \tag{2.20}$$

For each generated wild-bootstrap sample, we repeat the estimation procedure for estimating the optimal weights and the calculation of the test statistic $\mathrm{Tr}_N^{(g)}$. Subsequently, the p-value of Tr_N is computed as $\sum_{g=1}^{G} \mathbf{1}(\mathrm{Tr}_N^{(g)} \geq \mathrm{Tr}_N)/G$, where $\mathbf{1}(\cdot)$ is an indicator function.

2.1.3 Summary

We summarize the key steps of the PRM as follows:

Step (i). Fit q marginal linear regression models with the univariate dependent variable as each single phenotype and the independent variables as the covariates of interest.

Step (ii). Calculate q Wald-type test statistics under the same null hypothesis (2.6) and their corresponding p-values.

Step (iii). Select the responses with the smallest $[\frac{q}{log(q)+1}] = q^*$ (or $[\frac{n}{log(n)+1}] = q^*$ if $n \leq q$) p-values and establish the shrunken response space $\mathbf{Y}^*$;

Step (iv). Apply SPCA to estimate the weight $\mathbf{W}$ based on $\mathbf{Y}^*$;

Step (v). Project $\mathbf{Y}$ to $\mathbf{W}^T\mathbf{Y}$ and regress $\mathbf{W}^T\mathbf{Y}$ by $\mathbf{X}$;

Step (vi). Calculate the Wald-type test statistic Tr_N;

Step (vii). Generate G bootstrap samples and repeat Steps (i) to (vi) for each bootstrap sample;

Step (viii). Approximate the p-value of Tr_N.

2.2 Results

2.2.1 Simulation Studies

We carried out two scenarios of simulation studies to examine the finite-sample performance of the PRM. The simulation studies were designed to establish the association between a relatively high-dimensional phenotype with a commonly used genetic marker (e.g., SNP), while adjusting for age and other environmental factors. The first scenario focuses on that q is relatively smaller than the sample size N. The second scenario focuses on that q is comparable to the sample size N.

We set q and then simulated the multivariate phenotype according to model (1). The random errors were simulated from a multivariate normal distribution with mean 0 and covariance matrix with diagonal elements equal to 1. For the off-diagonal elements in the covariance matrix, we categorized each component of the multivariate phenotype into three categories: high correlation (0.6), medium correlation (0.3), and very low correlation (0.1) with the corresponding number of components $(1, 1, q-2)$ in each category. Specifically, we set the correlation between the first and second random errors as 0.6, those between the first random error and all others to be 0.3, and others to be 0.1. In the covariate matrix, we included a SNP, a diagnostic status as a binary variable with probability 0.5, and 3 additional continuous covariates. We simulated the additive SNP effect under different minor allele frequencies (MAFs). We simulated the three additional continuous covariates from a multivariate normal distribution with mean 0, standard deviation 1, and equal correlation 0.3. Our hypothesis of interest is to test the SNP effect on the multivariate phenotype. We set the number of the repetitions to be

150 and the number of wild bootstrap samples to be 250.

Scenario I

In the first scenario, we set the sample size N to be 150 and the MAF to be 0.5. The q were chosen to be 5, 10, 20, 30, 80 and 100, respectively. The first five individual phenotypes were associated with the SNP, whose coefficients were independently generated from a normal distribution with mean 0.15 and variance 0.05, and the 5th phenotype was also associated with disease status with regression coefficient being 0.5. We applied both the PRM and Hotelling's T^2 test to each simulated dataset in order to examine the type I and II error rates under the 5% significance level. Inspecting Figure 1 reveals that the type I errors are well controlled for both methods. Moreover, as q increases, the power in detecting the SNP effect decreases faster for Hotelling's T^2 test compared with the PRM.

Scenario II

In the second scenario, we set q to be 50, 100, 150 and 200, respectively, and the sample size N to be 150, 200, 250 and 300, respectively. We generated the additive SNP effect under 6 different MAFs, which are 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5, respectively. We considered two scenarios of the SNP effect. In the first scenario, only the first individual phenotype is associated with the SNP effect with regression coefficient being 0.5 and the second individual phenotype is associated with the disease status effect with regression coefficient being 0.5. Other individual phenotypes are not associated with any covariate. The second scenario is that the first 10 individual phenotypes are associated with the SNP. We generated the corresponding regression coefficients independently from a normal distribution with mean 0.5 and standard deviation 0.15. Moreover, we

set the regression coefficient for the diagnosis status to be 0.5 for the 10th individual phenotype and all other regression coefficients to be zero.

We applied the PRM to the simulated data sets and compared it with two other methods including a component wise method (CWM) and a principal components regression (PCR) using a 5% significance level. The CWM method fits a single linear regression to each individual phenotype with the same set of covariates and uses the false discovery rate (FDR) to test the additive SNP effect. The PCR method extracts the first three principal components of the multivariate phenotype by using the PCA and then fits a multivariate linear model to the extracted principal components with the same set of covariates. The Hotelling's T^2 test is not considered here since it is invalid for $q > N$.

We observe that the type I error rates are well controlled and more stable in the PRM, compared to the CWM and PCR methods (Figures 2 and 3). When the SNP effect is sparse, the powers of the PRM are generally higher than the CWM method, particularly for SNPs with small MAF and it is uniformly better than the PCR method (Figures 4 and 5). As expected, increasing either the sample size N or the MAF enhances the statistical power in detecting the SNP effect, whereas increasing the number of responses q reduces the power in detecting the SNP effect. When more SNPs show impact on the phenotypes, PRM is still comparable to CWM and better then PCR when the MAF is small (Figures 6 and 7). With increasing MAF, all three methods perform equally well.

2.2.2 A neonatal study

The data set is from a neonatal study to assess the impact of common SNPs in putative psychiatric genes on early age brain development. The study recruited 237 pregnant women in their second trimester, who were free from abnormalities on fetal ultrasounds and major medical illness. Each subject had one time visit with a T1-weighted medical resonance image (MRI), demographic and genetic information assessment. The MRI

images were collected with a Siemens head-only 3T scanner using a 3D spoiled gradient (FLASH TR/TE/Flip Angle 15/7msec/25Âř) with spatial resolution 1 x 1 x 1 mm^3 voxel size. There are 47 regions of interest defined from the T1-weighted images by non-linear warping of a parcellation atlas template [Gilmore et al., 2007; Knickmeyer et al., 2008]. The demographic information includes gender, gestational age at birth in days, age after birth in days and intracranial volume (ICV) of the infants. There are 128 male and 109 female infants with average gestational age 264.0 (SD $\pm$18.91), age after birth in days of 30.2 (SD $\pm$17.80) and ICV 481799.9 (SD $\pm$61528.96). Moreover, 9 genetic variants expressed in SNPs from 6 genes were collected and genotyped by Genome Quebec using Sequenom iPLEX Gold Genotyping Technology.

We applied our PRM method to multivariate phenotype including the volumes of 47 regions of interest (ROIs) with covariates of interest including gender, gestational age, age after birth, ICV and the 9 SNPs with an additive effect. Each hypothesis tests a single SNP effect, while adjusting for other covariates including demographic information and other SNPs. We list the 9 SNPs with their corresponding genes and respective p-values in Table 1.

The results show that the SNPs rs6675281 and rs35753505 have a significant impact on early age brain development with p-values of 0.016 and 0.0136, respectively. This agrees with the existing literature. Specifically, DISC1 was known to be associated with mental illness, such as schizoprenia and bipolar disorder, and NRG1 was known to relate to brain tissue volume [Mata et al., 2009].

We also applied the PCR and CWM methods to the same data set with the same set of covariates for comparison. In the PCR application, the first three principal components of the 47 ROIs, which explain 74.4% of the variation, are regressed on the same group of covariates of interest and the same null hypotheses were tested for each SNP by Hotelling's T^2 test at the 0.05 significance level. None of the 9 SNPs were found to be significant for brain volume development. The details of the test results are given in the

supplementary document. When analyzing the same data set by CWM with multiple comparisons adjusted by FDR, none of the 9 SNPs are detected to be significant for the 47 ROIs at the same testing level.

2.2.3 Tables and Figures

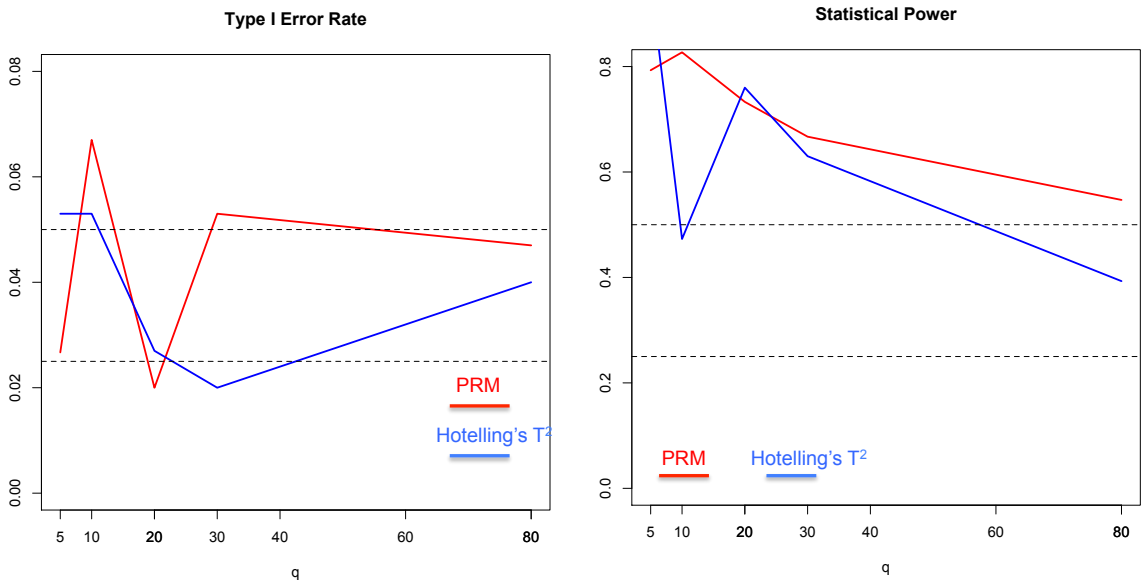

Figure 2.1: The comparison results of the PRM and Hotelling's T^2 test based on $N = 150$ and MAF=0.5: the *type I error* (the left panel) and *power* (the right panel). The upper and middle dashed lines in the left panel correspond to 0.05 and 0.025, respectively; and the upper and middle dashed lines in the right panel represent 0.5 and 0.25, respectively.

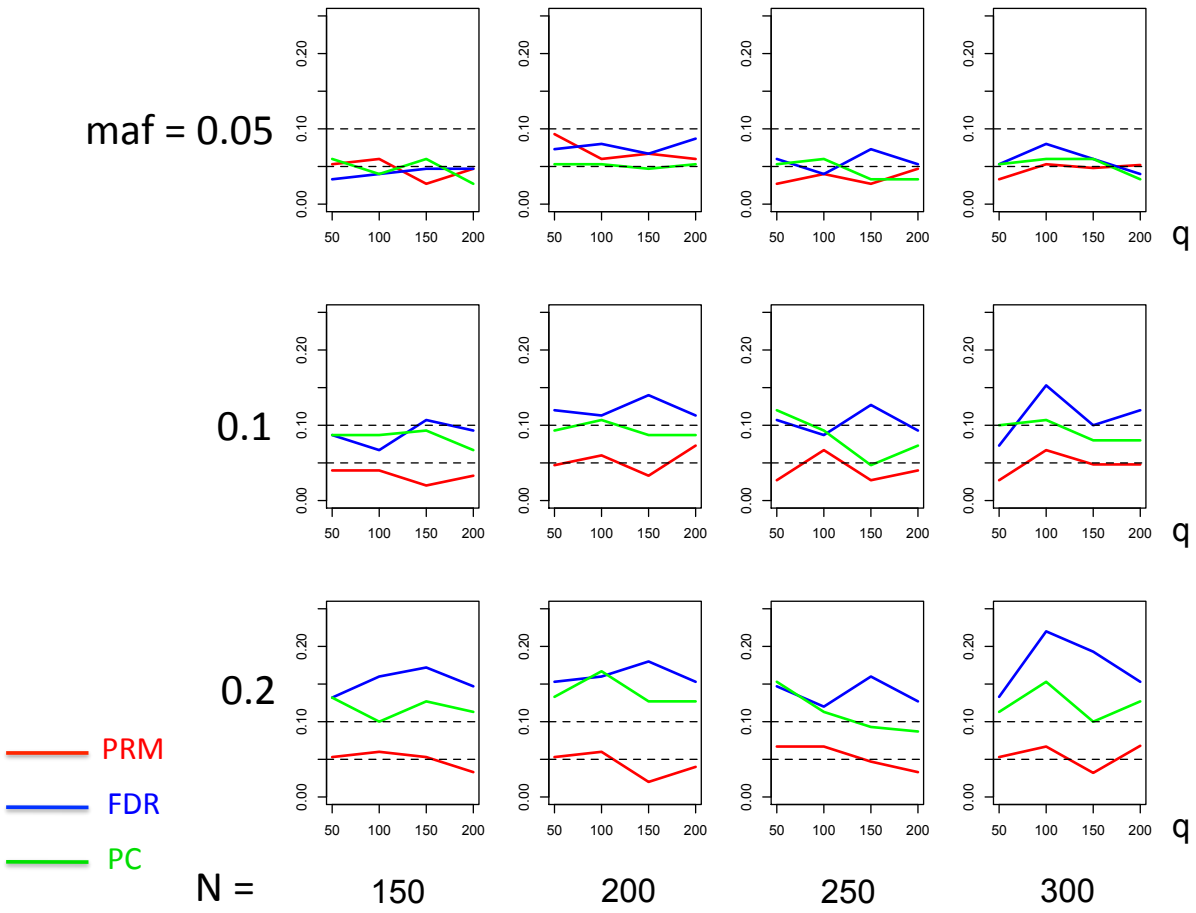

Figure 2.2: The *type I error* comparison results of the PRM, CWM, and PCR methods based on different sample sizes (150, 200, 250 and 300) and different *minor allele frequencies* (0.05, 0.1 and 0.2). The *horizontal axis* of each plot is the number of phenotypes q and the *vertical axis* is the type I error rate. The upper and middle dashed lines are 0.1 and 0.05, respectively.

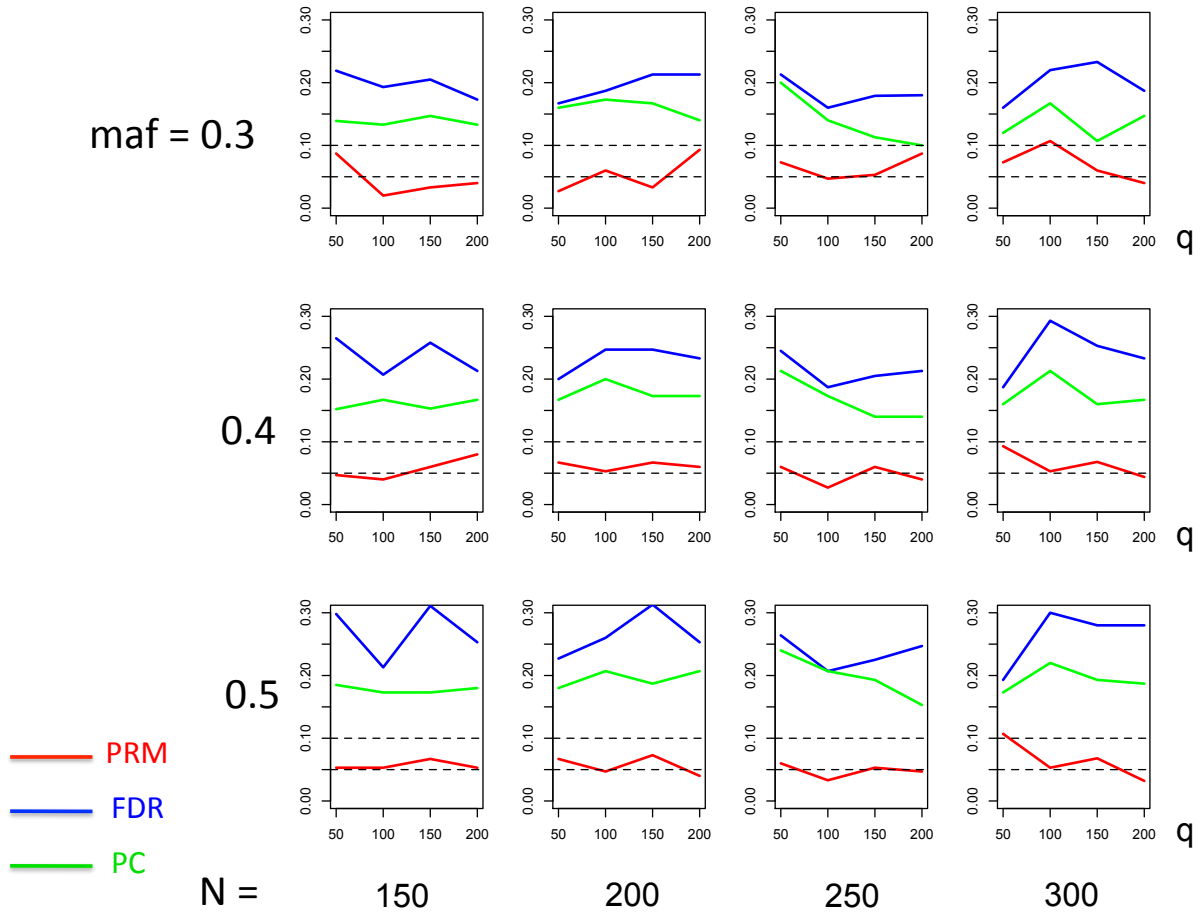

Figure 2.3: The *type I error* comparison results of the PRM, CWM, and PCR methods based on different sample sizes (150, 200, 250 and 300) and different *minor allele frequencies* (0.3, 0.4 and 0.5). The *horizontal axis* of each plot is the number of phenotypes q and the *vertical axis* is the type I error rate. The upper and middle dashed lines are 0.1 and 0.05, respectively.

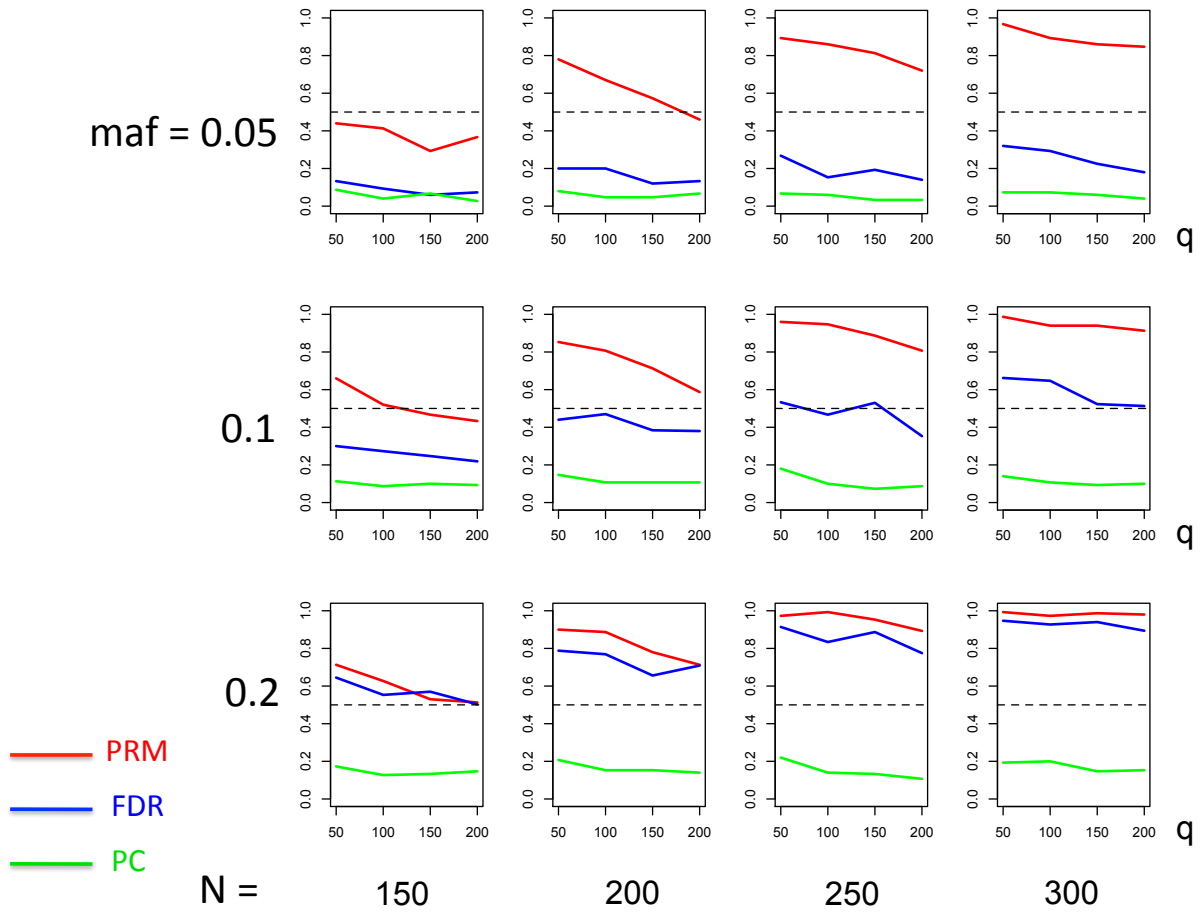

Figure 2.4: The *power* comparison results of the PRM, CWM, and PCR methods for the first scenario of sparse SNP effect based on different sample sizes (150, 200, 250 and 300) and different *minor allele frequencies* (0.05, 0.1 and 0.2). The *horizontal axis* of each plot is the number of phenotypes q and the *vertical axis* is the power. The dashed line represents a power of 50%.

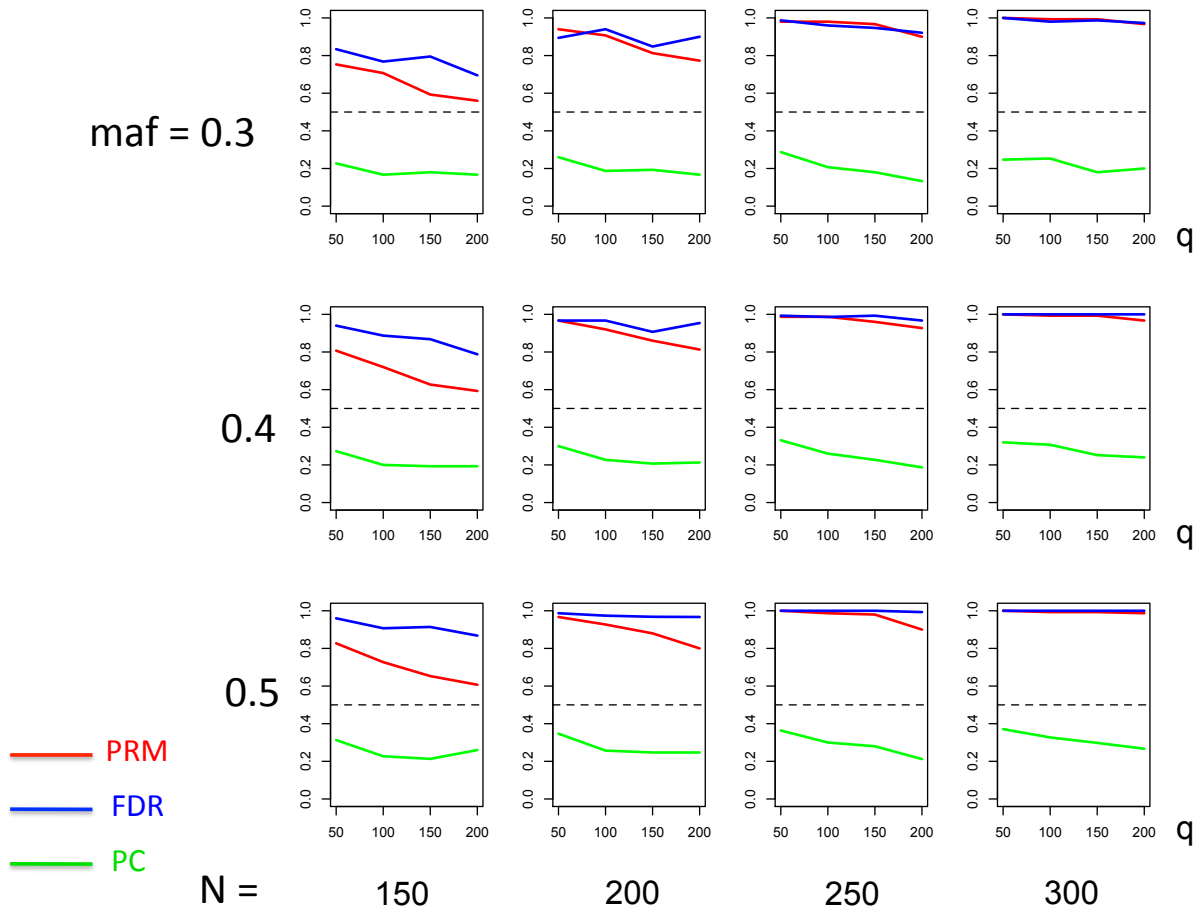

Figure 2.5: The *power* comparison results of the PRM, CWM, and PCR methods for the first scenario of sparse SNP effect based on different sample sizes (150, 200, 250 and 300) and different *minor allele frequencies* (0.3, 0.4 and 0.5). The *horizontal axis* of each plot is the number of phenotypes q and the *vertical axis* is the power. The dashed line represents a power of 50%.

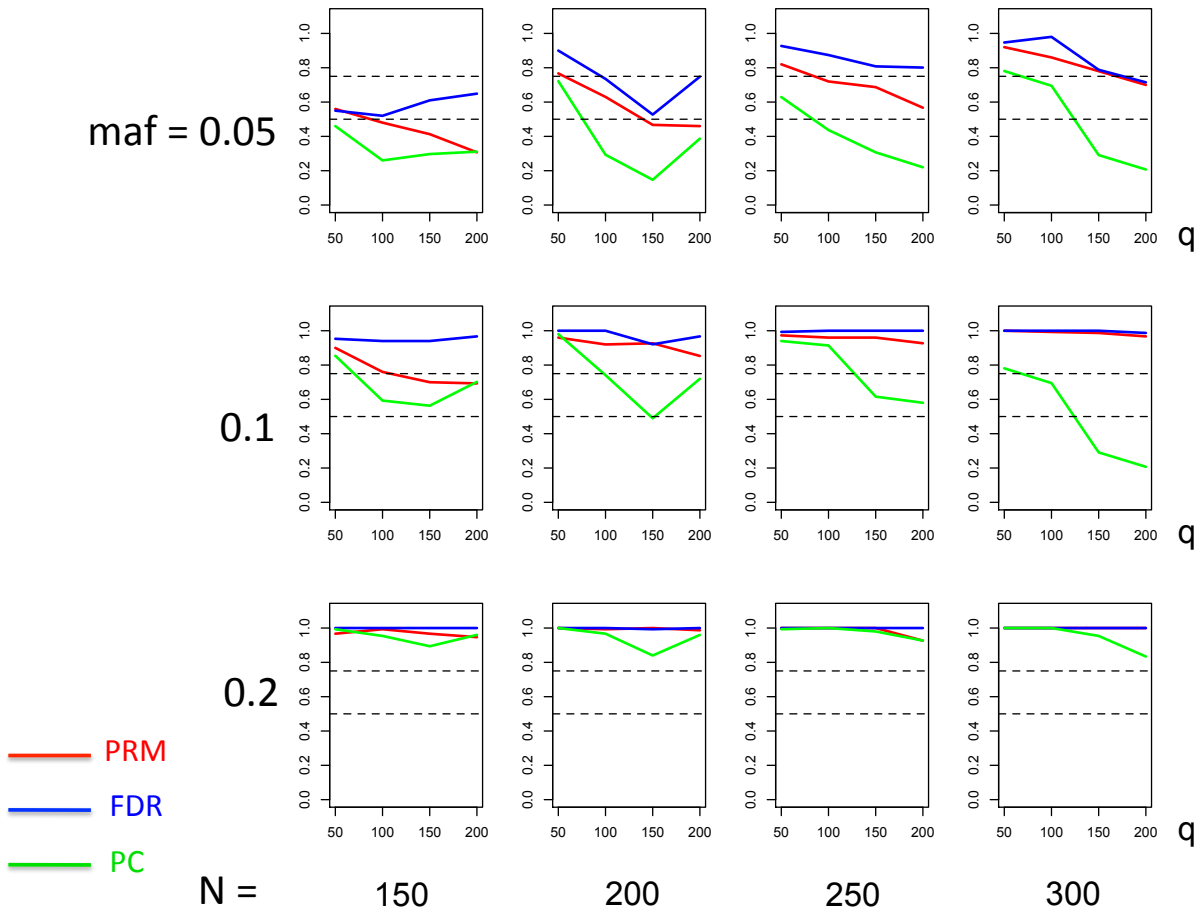

Figure 2.6: The *power* comparison results of the PRM, CWM, and PCR methods for multiple SNP effects based on different sample sizes (150, 200, 250 and 300) and different *minor allele frequencies* (0.05, 0.1 and 0.2). The *horizontal axis* of each plot is the number of phenotypes q and the *vertical axis* is the power. The upper and lower dashed lines represent the powers of 75% and 50%, respectively.

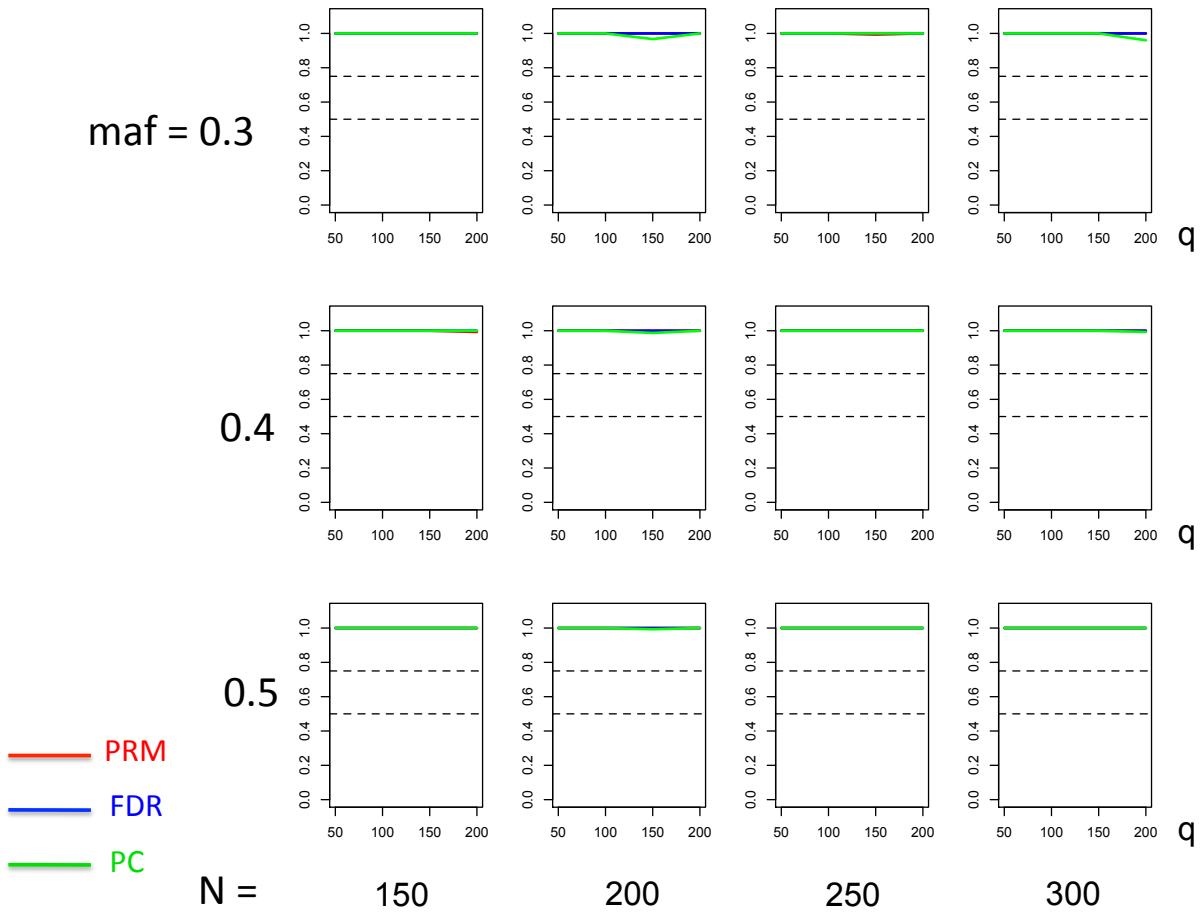

Figure 2.7: The *power* comparison results of the PRM, CWM, and PCR methods for the second scenario of multiple SNP effects based on different sample sizes (150, 200, 250 and 300) and different *minor allele frequencies* (0.3, 0.4 and 0.5). The *horizontal axis* of each plot is the number of phenotypes q and the *vertical axis* is the power. The upper and lower dashed lines represent the powers of 75% and 50%, respectively.

Table 2.1: Selected SNPs with the corresponding genes and result for testing a single SNP effect while adjusting for demographic information and other SNPs

Gene	Abbreviation	SNP	P-value
Catechol-O-methyltransferase	COMT	rs4680	0.88
Disrupted-in-schizophrenia-1	DISC1	rs821616	0.75
		rs6675281	**0.016**
Neuregulin 1	NRG1	**rs35753505**	**0.0136**
		rs6994992	0.51
Estrogen Receptor Alpha	ESR1	rs9340799	0.44
		rs2234693	0.57
Brain-derived Neurotrophic Factor	BDNF	rs6265	0.60
Glutamate Decarboxylase 1	GAD1 (GAD67)	rs2270335	0.39

Chapter 3

Functional Mixed Effects Model - FMEM

3.1 Methods

Suppose that we observe imaging measures, clinical variables, and genetic markers from n unrelated subjects. Let $\mathcal{V}$ be the whole brain and v be a voxel in $\mathcal{V}$. For each individual i ($i = 1, \ldots, n$), a $N_V \times 1$ vector consisting of imaging measures is observed and denoted by $Y_i = \{y_i(v) : v \in \mathcal{V}\}$. For notational simplicity, we only consider univariate image measure and thus, N_V equals the number of voxels in $\mathcal{V}$. Moreover, a $K \times 1$ vector of clinical covariates $\mathbf{x}_i = (x_{i1}, \cdots, x_{iK})^T$ and an $G \times 1$ vector $\mathbf{g}_i = (g_{i1}, \cdots, g_{iG})^T$ for genetic data are also collected for each individual. For instance, imaging measures can be the shape representation of the surfaces of cortical or subcortical structures [Chung et al., 2008; Zhu et al., 2007], and genetic makers can be various polymorphism types, such as single nucleotide polymorphisms (SNPs), block substitutions, and copy number variants [Liu et al., 2007; Tzeng and Zhang, 2007; Wang and Chen, 2012].

Our FMEM consists of a mixed effects model (MEM) at each voxel and a jumping surface model (JSM) for varying coefficient functions across the brain. First, at each voxel v in $\mathcal{V}$, MEM is given by

$$y_i(v) = \mathbf{x}_i^T \beta(v) + h(\mathbf{g}_i; v) + e_i(v) = \mathbf{x}_i^T \beta(v) + \mathbf{z}_i^T \gamma(v) + e_i(v) \quad \text{for} \quad i = 1, \cdots, n, \quad (3.1)$$

where $\beta(v) = (\beta_1(v), \cdots, \beta_K(v))^T$ is a $K \times 1$ vector, the $h(\cdot\,; v)$s are unknown functions corresponding to the genetic effects, $\gamma(v) = (\gamma_1(v), \cdots, \gamma_L(v))^T$ is an $L \times 1$ vector for

genetic random effects, $\mathbf{z}_i$ is a pre-specified $L \times 1$ vector of functions of $\mathbf{g}_i$, and $e_i(v)$ is the measurement error. We assume that $e_i(v) \sim N(0, \sigma_e(v)^2)$, $\gamma(v) \sim N(0, \sigma_\gamma^2(v)\Gamma)$, and $\{e_i(v) : v \in \mathcal{V}\}$ are independent across i and independent of $\gamma(v)$ for all $v \in \mathcal{V}$, where Γ is an $L \times L$ identity matrix. Without loss of generality, we assume $\Gamma = \mathbf{I}_L$.

Model (3.1) can be regarded as an alternative representation of variance component models used in the literature [Liu et al., 2007; Tzeng and Zhang, 2007; Kang et al., 2010; Wang and Chen, 2012]. For instance, for the kernel machine framework [Liu et al., 2007], we can directly represent $h(\mathbf{g}_i; v)$ as the random weighted sum of a set of L orthonormal basis functions by using the Karhunen Loeve expansion. For a given voxel v, the covariance between two individuals i and j are $\sigma_\gamma^2(v)\mathbf{z}_i^T\mathbf{z}_j$. Moreover, for two different voxels v and v' in $\mathcal{V}$, we assume $\mathrm{Cov}(\gamma(v), \gamma(v')) = \sigma_\gamma(v)\sigma_\gamma(v')\rho_\gamma(v, v')\mathbf{I}_L$ and $\mathrm{Cov}(e_i(v), e_i(v')) = \sigma_e(v)\sigma_e(v')\rho_e(v, v')$, where $\rho_\gamma(v, v')$ and $\rho_e(v, v')$, respectively, characterize the spatial correlation between the genetic random effects and that between the measurement errors. Therefore, for any two voxels v and v', the covariance structure of $y_i(v)$ is given by

$$\Sigma_{y,i}(v, v') = \mathrm{Cov}(y_i(v), y_i(v')) = \sigma_\gamma(v)\sigma_\gamma(v')\rho_\gamma(v, v')\mathbf{z}_i^T\mathbf{z}_i + \sigma_e(v)\sigma_e(v')\rho_e(v, v'). \quad (3.2)$$

We propose JSM for the genetic varying coefficient function $\{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$ and the fixed effect varying coefficient functions $\{\beta_j(v) : v \in \mathcal{V}\}$ for $j = 1, \ldots, K$. It is common that imaging data $\{y_i(v) : v \in \mathcal{V}\}$ can be regarded as a noisy version of a piecewise-smooth function of $v \in \mathcal{V}$ with the possible existence of jumps or edges. In many neuroimaging datasets, those jumps or edges often reflect the functional and/or structural changes, such as white matter (WM) and grey matter (GM), across the brain. Therefore, the varying coefficient functions in model (3.1) may inherit the piecewise-smooth features from imaging data. Furthermore, it is more reasonable to assume that different varying coefficient functions have different jumps or edges, since different

covariates may play different roles in characterizing the piecewise-smooth pattern of the imaging data.

3.1.1 Two-stage Estimation Procedure

We propose a two-stage estimation procedure to estimate all varying coefficient functions and test their effects on imaging phenotypes. The key ideas of each stage are given as follows:

Stage (I). Spatially and adaptively estimate $\{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$ and test the null hypothesis $\sigma_\gamma^2(v) = 0$ across all voxels.

Stage (II). Directly apply the multiscale adaptive regression models (MARM) in [Li et al., 2011] to spatially and adaptively estimate $\beta = \{\beta(v) : v \in \mathcal{V}\}$ and then test associated hypotheses.

Since our primary interest lies in the genetic effect, we focus on Stage (I) and omit Stage (II) for the sake of space.

Stage I

The first stage consists of three major steps as follows:

Step (I.1). Calculate the restricted maximum likelihood (REML) estimator of $\eta(v) = (\sigma_\gamma^2(v), \sigma_e^2(v))$ across all voxels $v \in \mathcal{V}$.

Step (I.2). Spatially and adaptively re-estimate $\{\sigma_\gamma^2(v) : v \in V\}$ by incorporating information from neighboring voxels.

Step (I.3). Construct weighted likelihood ratio statistics and derive their approximate distributions to test the null hypothesis of $H_0(v) : \sigma_\gamma^2(v) = 0$ across all voxels.

In Step (I.1), we calculate the REML estimator of $\eta(v)$ across voxels. Let $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)$ be an $L \times n$ matrix, $\mathbf{Y}(v) = (y_1(v), \cdots, y_n(v))^T$ be an $n \times 1$ vector, and $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ be a $p \times n$ matrix. There exists an $(n-p) \times n$ matrix K_x such that $K_x \mathbf{X}^T = \mathbf{0}$ and $\text{rank}(K_x) = n - p$. A MEM for $\mathbf{Y}^*(v) = K_x \mathbf{Y}(v)$ is given by

$$\mathbf{Y}^*(v) = K_x \mathbf{Z}^T \gamma(v) + K_x E(v), \tag{3.3}$$

where $E(v) = (e_1(v), \cdots, e_n(v))^T$. Based on the distributional assumptions in (3.1), we have $\mathbf{Y}^*(v) \sim N(\mathbf{0}, \Sigma_{Y^*}(v))$, where $\Sigma_{Y^*}(v) = \sigma_\gamma^2(v) K_x \mathbf{Z}^T \mathbf{Z} K_x^T + \sigma_e^2(v) K_x K_x^T$. Thus, at each voxel v, the REML estimate of $\hat{\eta}(v)$, denoted by $\hat{\eta}(v)$, is to maximize the REML function given by

$$\ell_{REML}(\mathbf{Y}^*(v) | \mathbf{Z}, \eta(v)) = -0.5 \log |\Sigma_{Y^*}(v)| - 0.5 \mathbf{Y}^*(v)^T \Sigma_{Y^*}(v)^{-1} \mathbf{Y}^*(v). \tag{3.4}$$

Since our primary interest lies on $\sigma_\gamma^2(v)$, we fix $\sigma_e^2(v)$ as $\hat{\sigma}_e^2(v)$ from here on.

In Step (I.2), we construct a weighted REML function to estimate $\sigma_\gamma^2(v)$ by incorporating the spatial information in a neighborhood $B(v, h)$ for each voxel v with a specific radius h as follows:

$$L_{REML}(\sigma_\gamma^2(v) | \mathbf{Y}^*, B(v,h)) = \sum_{v' \in B(v,h)} \omega_\gamma(v, v'; h) \ell_{REML}(\mathbf{Y}^*(v') | \mathbf{Z}, \sigma_\gamma^2(v), \hat{\sigma}_e^2(v')), \tag{3.5}$$

where $\omega_\gamma(v, v'; h)$ is a weight function of voxels v, v', and the radius h. Then, we maximize $L_{REML}(\sigma_\gamma^2(v) | \mathbf{Y}^*, B(v,h))$ to calculate the weighted REML estimator of $\sigma_\gamma^2(v)$, denoted by $\hat{\sigma}_\gamma^2(v, h)$. The weight function $\omega_\gamma(v, v'; h)$ measures the data similarity between the two voxels v and v' such that $\sum_{v' \in B(v,h)} \omega_\gamma(v, v'; h) = 1$ and $\omega_\gamma(v, v'; h) \geq 0$. A large value of $\omega_\gamma(v, v'; h)$ means that the information contained in the voxels v and v' is very similar, whereas $\omega_\gamma(v, v'; h) \approx 0$ indicates that the data in voxel v' do not have too much information for $\sigma_\gamma(v)$. The adaptive weight $\omega_\gamma(v, v'; h)$ plays a critical role in

preventing over-smoothing estimation of $\sigma^2_\gamma(v)$ and preserving the edges of significant regions of $\{\sigma^2_\gamma(v) : v \in \mathcal{V}\}$.

In Step (I.3), to assess the synthetic genetic effect on imaging phenotype across all voxels, we formulate it as testing the following null and alternative hypotheses:

$$H_{0,\gamma}(v) : \sigma^2_\gamma(v) = 0 \quad \text{v.s.} \quad H_{1,\gamma}(v) : \sigma^2_\gamma(v) > 0. \tag{3.6}$$

We test (3.6) by using the weighted REML ratio statistic defined by

$$\text{RLRT}_{\sigma^2_\gamma}(v) = 2\{L_{REML}(\hat{\sigma}^2_\gamma(v) \mid \mathbf{Y}^*, B(v,h)) - L_{REML}(0 \mid \mathbf{Y}^*, B(v,h))\}. \tag{3.7}$$

Since all the subjects share the same random effect $\gamma(v)$, the standard asymptotic results in Stram and Lee [1994] are invalid and can perform very poorly even for the unweighted REML ratio statistics for testing random effects in model (3.1) [Crainiceanu and Ruppert, 2004]. However, we provide an exact null distribution for $\text{RLRT}_{\sigma^2_\gamma}(v)$ below.

Step (I.2): Adaptive Estimation of $\sigma^2_\gamma(v)$

Following the adaptive estimation (AET) procedure proposed in [Polzehl and Spokoiny, 2000; Li et al., 2011], we adaptively determine $\{\omega_\gamma(v, v'; h) : v, v' \in \mathcal{V}\}$ and then calculate $\hat{\sigma}_\gamma(v; h)$ as h increases from $h_0 = 0$ to a predetermined value $h_S = r_0$. A path

diagram of AET is given as follows:

$$
\begin{array}{ccccccc}
h_0 = 0 & < & h_1 & < & \cdots & < & h_S = r_0 \\
B(v, h_0) = \{v\} & \subset & B(v, h_1) & \subset & \cdots & \subset & B(v, h_S) \\
\Downarrow & & \Downarrow & \nearrow & \cdots & \nearrow & \Downarrow \\
\{\hat{\sigma^2_\gamma}(v) : v \in \mathcal{V}\} & \Rightarrow & \omega_\gamma(v, v'; h_1) & & \cdots & & \omega_\gamma(v, v'; h_S = r_0) \\
 & & \Downarrow & \nearrow & \cdots & \nearrow & \Downarrow \\
 & & \{\hat{\sigma^2_\gamma}(v; h_1) : v \in \mathcal{V}\} & & \cdots & & \{\hat{\sigma^2_\gamma}(v; h_S) : v \in \mathcal{V}\}.
\end{array}
$$

The key idea of AET is to build a sequence of nested spheres $B(v, h_s)$ for $h_0 = 0 < h_1 < \cdots < h_S = r_0$ at each voxel $v \in \mathcal{V}$ and then sequentially determine $\hat{\sigma}_\gamma(v, v'; h_s)$ for all $v' \in B(v, h_s)$ based on $\{\hat{\sigma}^2_\gamma(v', h_{s-1}) : v' \in B(v, h_s)\}$ for all $v \in \mathcal{V}$ and $s = 1, \ldots, S$. Since the tuning parameters of AET have been described in details in [Polzehl and Spokoiny, 2000; Li et al., 2011], we do not include them here for the sake of brevity.

The three key steps of AET, including weights adaptation, estimation, and termination checking, are presented as follows.

- In the weights *adaption* step (i), we select a series $\{h_s = c_h^s : s = 1, \cdots, S\}$ of radii with $c_h \in (1, 2)$, say $c_h = 1.125$. We then set $s = 1$ and $h_1 = c_h$. The adaptive weights are given by

$$\omega_\gamma(v, v'; h_s) = K_{loc}(||v - v'||_2/h_s) K_{st}(D_\gamma(v, v'; h_{s-1})/C_n), \tag{3.8}$$

  where $K_{loc}(u) = (1 - u)_+$ and $K_{st}(u) = \min(1, 2(1 - u^2))_+$ according to previous experience in the literature [Li et al., 2011], and $|| \cdot ||_2$ denotes the Euclidean norm of a vector (or a matrix). Moreover, $D_\gamma(v, v'; h_{s-1})$ is set as $\{\hat{\sigma^2_\gamma}(v; h_{s-1}) - \hat{\sigma^2_\gamma}(v'; h_{s-1})\}^2 / \widehat{\text{var}[\sigma^2_\gamma(v)]}$, where $\widehat{\text{var}[\sigma^2_\gamma(v)]}$ is estimated by using the inverse of the Fisher information matrix of $(\sigma^2_\gamma(v), \sigma^2_e(v))$ from the likelihood function (3.4) with $h = h_0$. Then these quantities are fixed for subsequent updates of h. Following

Li et al. [2011], we choose $C_n = n^{1/3}\chi^2(1)^{0.5}$ for $D_\gamma(v, v'; h_{s-1})$ defined in (4.13), where $\chi^2(1)^{0.5}$ is the 0.5-percentile of the $\chi^2(1)$ distribution. The adaptive weight $K_{st}(D_\gamma(v, v'; h_{s-1})/C_n)$ downweights the role of a voxel $v' \in B(v, h_s)$ in

$$L_{REML}(\sigma^2_\gamma(v)|\mathbf{Y}^*, B(v, h_s)) \tag{3.9}$$

if $D_\gamma(v, v'; h_{s-1})$ is large. The weight $K_{loc}(||v - v'||_2/h_s)$ gives less weight to the voxel $v' \in B(v, h_s)$, whose location is far from the voxel v.

- In the *estimation* step (ii), for each $v \in \mathcal{V}$ and for the radius h_s, we calculate $\hat{\sigma}_\gamma(v; h_s)$ by maximizing $L_{REML}(\sigma^2_\gamma(v)|\mathbf{Y}^*, B(v, h))$ defined in equation (3.5) given $\omega_\gamma(v, v'; h_s)$.

- In the *termination checking* step (iii), after the S_0-th iteration, we calculate a stopping criterion based on a distance between $\hat{\sigma^2_\gamma}(v; h_{S_0})$ and $\hat{\sigma^2_\gamma}(v; h_s)$ given by

$$D(\hat{\sigma^2_\gamma}(v; h_{S_0}), \hat{\sigma^2_\gamma}(v; h_s)) = \{\hat{\sigma^2_\gamma}(v; h_{S_0}) - \hat{\sigma^2_\gamma}(v; h_s)\}^2 \widehat{var[\sigma^2_\gamma(v)]}^{-1} \tag{3.10}$$

for $s > S_0$. Then, we compare $D(\hat{\sigma^2_\gamma}(v; h_{S_0}), \hat{\sigma^2_\gamma}(v; h_s))$ with a benchmark, denoted by $\tilde{C}(s)$, for $s > S_0$. If $D(\hat{\sigma^2_\gamma}(v; h_{S_0}), \hat{\sigma^2_\gamma}(v; h_s)) > \tilde{C}(s)$, then we set $\hat{\sigma^2_\gamma}(v) = \hat{\sigma^2_\gamma}(v, h_{s-1}) =$ and the estimation for this voxel v is terminated. If $s = S$ and $D(\hat{\sigma^2_\gamma}(v; h_{S_0}), \hat{\sigma^2_\gamma}(v; h_s)) \leq \tilde{C}(s)$, $\hat{\sigma^2_\gamma}(v)$ is set as $\hat{\sigma^2_\gamma}(v, h_S)$ and the estimation process terminates. The algorithm stops when the estimation is finished for all v in $\mathcal{V}$. If $s \leq S_0$ or $D(\hat{\sigma^2_\gamma}(v; h_{S_0}), \hat{\sigma^2_\gamma}(v; h_s)) \leq \tilde{C}(s)$ for $s < S_0 \leq S - 1$, then we go back to the weights adaptation step (i) with an increased radius $h = h_{s+1} = c_h^{s+1}$. Throughout the paper, we set $S_0 = 2$, $\tilde{C}(s) = \chi^2(p)^{0.7/(s-1)}$, and $S = 10$.

Note that $\tilde{C}(s)$ is a decreasing function in s which makes the stopping criteria more and more stringent when the radius increases in order to prevent from over-smoothing.

Step (I.3): Testing $H_{0,\gamma}(v) : \sigma^2_\gamma(v) = 0$

We perform the hypothesis testing in (3.6) by using the testing statistics $\mathrm{RLRT}_{\sigma^2_\gamma}(v)$ and the corresponding p-values. Let $\Omega = K_x \mathbf{Z}^T \mathbf{Z} K_x^T = U D_0 U^T$ be the spectral decomposition of Ω such that $D_0 = \mathrm{diag}(d_1, \cdots, d_{n-p})$ is the diagonal matrix of eigenvalues d_k and U is an $(n-p) \times (n-p)$ orthonormal matrix. Without loss of generality, we choose K_x such that $K_x K_x^T = \mathbf{I}_{n-p}$. We obtain the following theorem, whose proof is included in the appendix.

Theorem 1. *Under model (3.1), $RLRT_n(v)$ can be written as*

$$RLRT_n(v) = \sup_{\lambda(v) \geq 0} \{ \sum_{v' \in B(v,h)} \omega(v, v'; h) D(v'; \lambda(v)/\sigma^2_e(v')) \}, \tag{3.11}$$

where $\lambda(v) = \sigma^2_\gamma(v)$ and $D(v'; t)$ is given by

$$\sigma_e^{-2}(v') \mathbf{Y}^{*T}(v') U \, diag \left(\frac{t d_1}{1 + t d_1}, \cdots, \frac{t d_{n-p}}{1 + t d_{n-p}} \right) U^T \mathbf{Y}^*(v') - \sum_{l=1}^{n-p} \log(1 + t d_l). \tag{3.12}$$

Moreover, under the null hypothesis $H_{0,\gamma}(v)$, we have

$$D(v'; t) \stackrel{\mathcal{D}}{=} \sum_{l=1}^{n-p} \frac{\delta_l^2(v') t d_l}{1 + t d_l} - \sum_{l=1}^{n-p} \log(1 + t d_l), \tag{3.13}$$

where $\stackrel{\mathcal{D}}{=}$ means equality in distribution and the $\delta_l(v)$ are i.i.d $N(0, 1)$ random variables.

Although Theorem 1 provides an efficient way of approximating the null distribution of $\mathrm{RLRT}_n(v)$, a complex issue arises from the complex spatial correlations among the $\delta_l^2(v')$ across voxels $v' \in B(v, h)$. One approach for dealing with such an issue is to estimate the spatial correlation for any pair of voxels, which can be computationally intensive. To avoid calculating spatial correlations, we develop a wild bootstrap method to efficiently simulate the null finite sample distribution of $\mathrm{RLRT}_n(v)$. The detailed

steps of this bootstrap method are presented in Appendix. After the p-values for all voxels $v \in \mathcal{V}$ are computed, either a false discovery rate (FDR) method or random field theory (RFT) is applied to correct for multiple comparisons [Ge and et al, 2012].

3.2 Results

3.2.1 Simulation Studies

We simulated data at all $N_V = 5,808$ voxels on a $44 \times 44 \times 3$ phantom image. Each z-slice contains the same effect regions. At each voxel, we simulated the univariate imaging measure according to model (3.1) with $\beta(v) = (\beta_0(v), \beta_1(v), \beta_2(v), \beta_3(v))^T$ and $\mathbf{x}_i = (1, x_{i1}, x_{i2}, x_{i3})^T$. Moreover, the covariates x_{i1}, x_{i2}, and x_{i3} were generated from a Gaussian distribution with mean 40 and standard deviation 10, a Bernoulli distribution with success probability 0.5, and a Bernoulli distribution with success probability 0.3, respectively. These three covariates were designed to mimic the common clinical variables age, gender, and disease status. For a slice of a phantom image, the effect areas for $\beta_0(v)$ were divided into 16 regions with 4 different values ranging from 0.02 to 0.08, increasing by 0.02 (Figure 8(a)); for $\beta_1(v)$, the effect regions were divided into 25 regions ranging from $10^{-2.5}$ to $10^{-12.5}$, decreasing by a rate of $10^{-2.5}$ (Figure 8(b)); for $\beta_2(v)$, the whole space was separated into 3 regions with values 0, 0.05, and 0.1 (Figure 8(c)); the effect area of $\beta_3(v)$ on a slice of phantom image was divided into 9 regions with values ranging from 0 to 0.1, increasing by differences of 0.025 (Figure 8(d)).

The genetic information was simulated according to the SNP data obtained from the public accessible data of the Alzheimer's Disease Neuroimage Initiative (ADNI). It is an ongoing longitudinal study with primary purpose of exploring the genetic and neuroimaing information associated with late-onset Alzheimer's disease (LOAD). The study recruited elderly subjects older than 65 years of age consisting about 400 subjects with mild cognitive impairment (MCI), about 200 subjects with Alzheimer (AD), and around 200 healthy controls. Each subject was followed for at least 3 years. During

the study period, the subjects were assessed with magnetic resonance imaging (MRI) measures and psychiatric evaluation to determine the diagnosis status at each time point. The genetic information was also collected from each subject at baseline and it is genotyped by the Illumina 610 Quad array with more than 620,000 single nucleotide polymorphysm (SNPs). More information of ADNI is provided in the real data analysis result Section 3.2. We simulated the genetic information based on the two following scenarios.

- Scenario I. To preserve the linkage disequilibrium among SNPs, we utilize all of the SNPs on chromosome 1 from 197 Caucasian controls to generate the genetic effect. After eliminating the SNPs with minor allele frequency (MAF) less than 5%, there were 31554 out of 45627 SNPs left. Then we randomly chose 20 SNPs and 100 subjects among the 197 healthy controls as the simulated genetic data $\mathbf{z}_i$ in (3.1). In this case, $n = 100$. If any of these 20 SNPs have MAF less than 5%, the genetic data was resampled until all of the 20 SNPs have MAF $\geq 5\%$.

- Scenario II. To evaluate the performance of FMEM in the case of high LD, we selected the SNPs from the same gene in the second scenario. Searching the SNPs on the gene PICALM, which is found to be relevant to Alzheimer's disease in many studies [Harold et al., 2009] using the gene list "glist-hg18" provided by PLINK, there were 23 SNPs on PICALM with MAF larger than 5%. After eliminating the missing values, there are 176 healthy controls with complete genotype data at these 23 SNPs. We randomly selected 7 SNPs from 75 healthy controls to be $\mathbf{z}_i$ in (3.1). Although there is strong LD among these 7 SNPs, no SNP has perfect correlation (1 or -1) with any other SNP in these 75 subjects. In this case, n equals 75.

In both scenarios, the SNP effects were assumed to be additive. The $\gamma(v)$ was generated from a multivariate Gaussian distribution with mean zero and covariance

matrix $\sigma_\gamma^2(v)\mathbf{I}_L$. Different $\sigma_\gamma^2(v)$ values, which represent different signal-to-noise ratios, were chosen to examine the performance of our method at different signal-to-noise ratios and also to test whether FMEM can perform well for different shapes. See Figure 9 (b) and Figure 9 (e) for Scenarios I and II. Moreover, we overlay some of the effect areas of $\beta_3(v)$ and $\sigma_\gamma^2(v)$ in order to account for the fact that the brain phenotype is an intermediate expression of disease progression. The $\{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$ of the effect regions in Scenario I were ranging from 0.005 to 0.025, increasing by 0.0025, whereas the $\{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$ of effect regions in Scenario II were ranging from 0.005 to 0.045, increasing by 0.005. The random error $e_i(v)$ was independently distributed as a univariate Gaussian distribution with mean 0 and standard deviation 3 for all voxels. We set the number of bootstrap samples M and the number of repetitions to be 200.

Tables 2 and 3 summarize the estimation results of $\sigma_\gamma^2(v)$ obtained from FMEM and traditional voxel-wise method for both scenarios. It includes the average absolute value of the bias, the root mean square (RMS), standard deviation (SD), and the ratio of RMS over SD. The difference between estimation of RMS and SD is that RMS is estimated using the empirical mean and SD is calculated using theoretical mean. As shown in both tables, FMEM outperforms voxel-wise method with respect to smaller estimation bias which leads to more accurate hypothesis testing conclusion. Note that compared with voxel-wise method, the RMSs and SDs are also smaller for FMEM. This indicates much more stable estimation.

We tested the hypotheses $H_0 : \sigma_\gamma^2(v) = 0$ and $H_1 : \sigma_\gamma^2(v) \neq 0$ for all voxels in $\mathcal{V}$ based on both FEFM and voxel-wise method and evaluated their performance in cluster-based thresholding [M. Silver and ADNI, 2011]. Specifically, we first thresholded the p-values for all voxels in $\mathcal{V}$ by using an initial p-value 0.01 suggested by M. Silver and ADNI [2011] to identify clusters of contiguous supra-threshold voxels. Then, the thresholded clusters were matched with the 9 separated activated areas in Figure 9 (b) or 9 (e). If a specific thresholded cluster overlaps with at least one voxel in any of the 9 effect regions, we call

such cluster as a "true positive". In contrast, if a specific thresholded cluster does not overlap with any voxels of the 9 effect regions, we call the cluster a "false positive". We summarized the hypothesis testing results by the average dice overlap ratio (DOR), the average number of false positive clusters, and the average size in the number of voxels of false positive clusters. DOR is the ratio between the number of true positive clusters over the true number of effect areas, which is 9 in this simulation setting. Thus, the higher DOR means the higher the probability of detecting true effect regions. As shown in Tables 4 and 5, if we set the cluster size threshold at 1 voxel, FMEM has smaller DOR and smaller number of false positive clusters compared with voxel-wise method. When the cluster size threshold increases to 10 voxels, FMEM has a similar DOR value as that of the no threshold case, whereas the DOR of the voxel-wise approach reduces by about 20%. Table 6 summarizes the number of significant voxels identified by the two methods in each effect region of Scenarios I and II In Table 6, FMEM identifies less voxels in the non-effect regions, while detecting more voxels in effect regions in both scenarios. Finally, we conclude that FMEM outperforms voxel-wise method in both detecting true effect regions and controlling the false positive error rate.

3.2.2 ADNI Data Analysis

The aim of this ADNI data analysis is to use FMEM to identify brain regions affected by candidate genes, thereby hoping to shed light on the pathological interactions between these causal genes and brain function. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5- year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the pro-

gression of mild cognitive impairment (MCI) and early AlzheimerâĂŹs disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California âĂŞ San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see $www.adni-info.org$.

The data we employed to evaluate the performance of FMEM was from the first phase (ADNI), which was conducted mainly to search for the causal SNPs associated with the progression of Alzheimer's disease and to establish an alternative diagnosis standard using MRI brain images. About 800 subjects with age older than 65 were recruited and followed at least 3 years. The 800 subjects included 200 healthy controls, 400 subjects with different levels of mild cognitive impairment (MCI), and 200 subjects with Alzheimer's disease (AD). Besides the SNPs and the T1 weighted MRI imaging measurements, the subjects were assessed with demographic information and psychiatric examination scores to determine the diagnosis status at each scheduled visit.

The T1-weighted MRI images were preprocessed by standard image processing steps including AC (anterior commissure) and -PC (posterior commissure) correction, bias field correction, skull-stripping, intensity inhomogeneity correction, cerebellum removal,

segmentation, and nonlinear registration. After segmentation, the brain was segmented into four different tissues: grey matter (GM), white matter (WM), ventricle (VN), and cerebrospinal fluid (CSF). We quantified the local volumetric group differences by generating RAVENS maps [Davatzikos et al., 2001] for the whole brain and each of the segmented tissue type (GM, WM, VN, and CSF) respectively, using the deformation field we obtained during registration. RAVENS methodology is based on a volume-preserving spatial transformation, which ensures that no volumetric information is lost during the process of spatial normalization, since this process changes an individual′s brain morphology to conform it to the morphology of the Jacob template.

We are interested in detecting meaningful brain regions of interest that are associated with several candidate genes. We included only the subjects whose diagnosis status were healthy control and Alzheimer's disease at the baseline and had no status change during the study period. After screening, the total number of subjects we included was 372 (195 HCs and 177 ADs). The clinical covariates of interest included in our analysis were gender, baseline age, square of baseline age, handedness, education, baseline intracranial volume, and the risk of APOE. Specifically, the handedness was treated as a binary variable, the education information was the self-reported years of education by the subjects, and the risk of APOE is assumed to be additive. In detail, the risk of APOE for a subject was 3 if he/she carries $\varepsilon 4$ at both alleles; it was 2 if he/she carries $\varepsilon 3$ and $\varepsilon 4$ in two alleles, the risk would be considered 0 if the two APOE alleles were the combination of $\varepsilon 2$ and $\varepsilon 3$, and other combination of APOE alleles are assumed to have risk 1.

Many genes have been reported to be causal in the progression of Alzheimer's disease. We selected three candidate causal genes including CR1 on chromosome 1, CD2AP on chromosome 6, and PICALM on chromosome 11 due to their strong association with the progression of Alzheimer's disease [Harold et al., 2009; Naj et al., 2011; Lambert et al., 2009]. Specifically, PICALM encodes the protein phosphatidylinositol-binding clathrin

assembly and is highly correlated with the emergence of late-on-set AD, which is possibly due to the perturbation at synapse triggering its function change [Harold et al., 2009]. The gene CD2AP encodes the CD2-asscociated protein and involves in the process of cell membrane, including endocytosis, that plays critical roles in neurodegeneration and Aβ clearance from the brain [Naj et al., 2011]. The gene CR1 encodes complement component (3b/4b) receptor 1 and the pathways involving CR1 are involved in the AD process, specifically in clearance of Aβ peptides which is the primary composition of amyloid plaques [Lambert et al., 2009].

We first matched the SNPs in ADNI with the gene list "glist-hg18" provided by PLINK [Purcell and et al, 2007] and were able to locate 16, 15, and 23 SNPs on the selected CR1, CD2AP, and PICALM genes, respectively. All these SNPs pass the quality control procedure with MAF $> 5\%$ and the Hardy Weinberg Equilibrium (HWE) test p-value> 0.01. The MAFs of SNPs of the selected genes vary from 0.1 to 0.5 and are presented in the supplementary document. After deleting missing values, there are 335, 299 and 328 subjects corresponding to the CR1, CD2AP, and PICALM genes, respectively, and their associated demographic information is presented in the supplementary document.

For each selected gene, we fitted FMEM (3.1) with $\mathbf{z}$ coded as the numbers of minor alleles to detect its associated significant brain regions of interest (ROIs). For comparison, we fitted the same model by using the classical voxel-wise method to analyze the same data sets. To formally detect significant ROIs, by following Ge and et al [2012], we used a cluster-form of threshold of 0.1% with a minimum voxel clustering value of 50 voxels. FMEM is able to to detect 45, 45, and 27 significant clusters for CR1, CD2AP, and PICALM, respectively, whereas the standard voxel-wise method can only identify 6, 14, and 2 significant clusters for CR1, CD2AP, and PICALM, respectively. Then, we overlapped these significant clusters with the 96 predefined ROIs in the Jacob template and were able to detect several predefined ROIs for each cluster. As shown in the

supplementary document, we were able to detect several major ROIs, such as the hippocampus, the putamen, and the fusiform. The hippocampus is known to be associated with memory and cognition. The fusiform is associated with color recognition, word and body recognition and the putamen is associated with motor skills. Figure 11 shows the $-\log_{10}(p)$ map of several selected slices with significant clusters for testing the genetic effect of CD2AP on RAVEN images identified by FMEM.

3.2.3 Tables and Figures

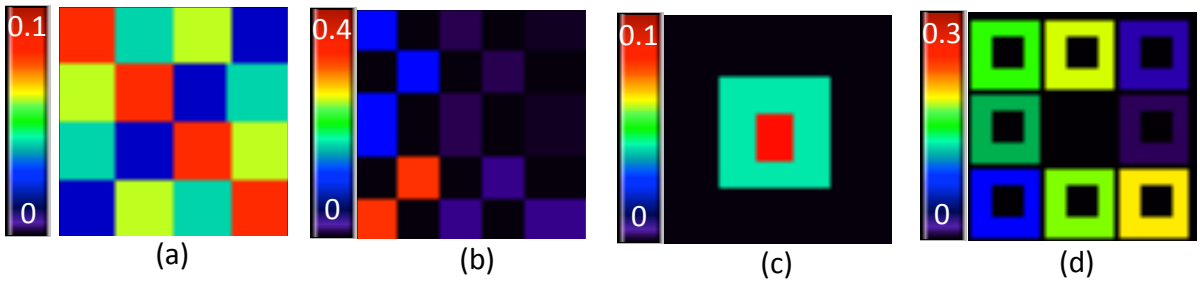

Figure 3.1: Simulation setting: (a) True image of β_0; (b) true image of β_1, in which the colors represent the values of $\beta_1(v) \times 10^4$; (c) true image of β_2; and (d) true image of β_3.

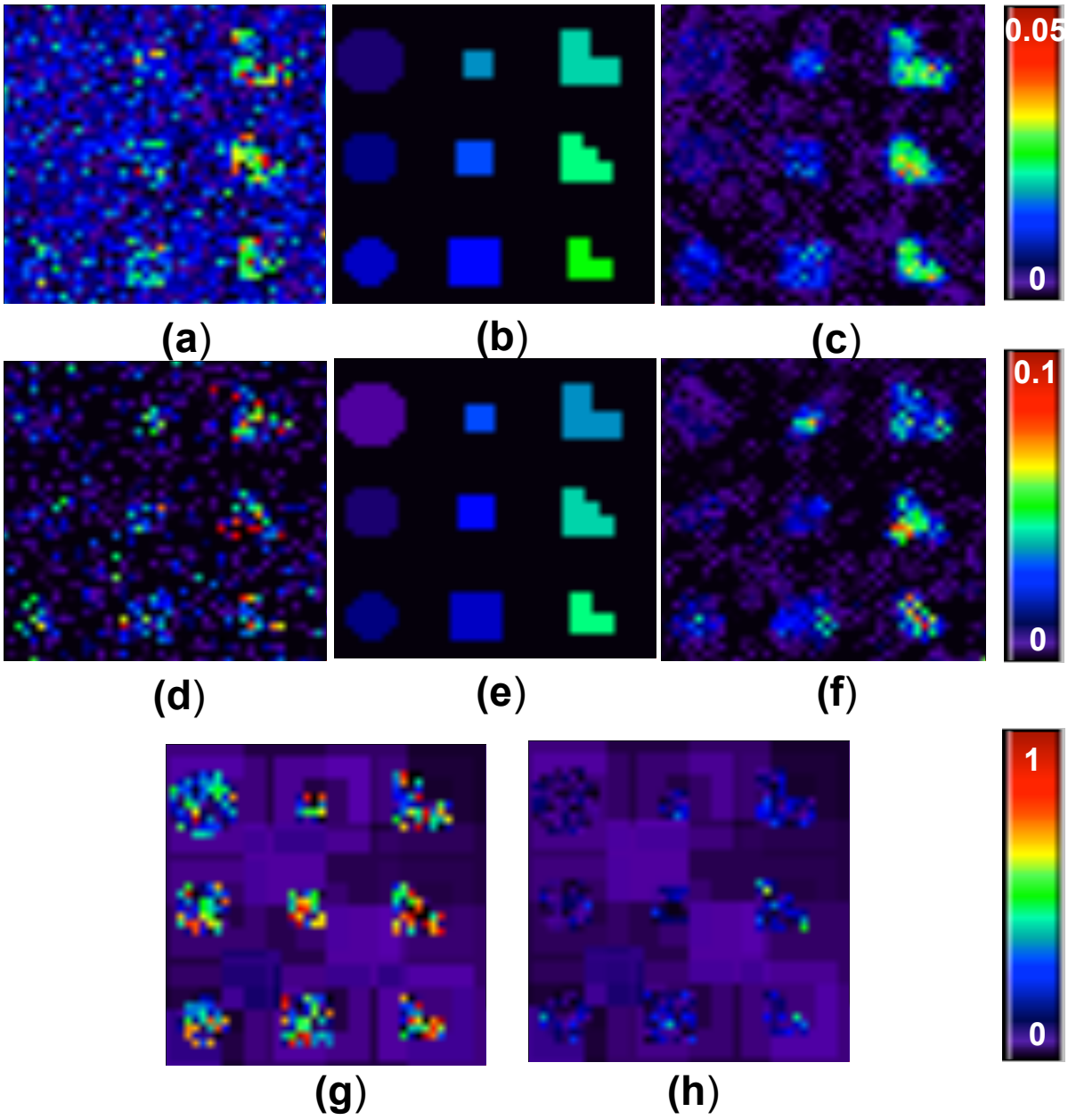

Figure 3.2: Simulation results for estimation accuracy: Scenario I : (a) estimated $\sigma^2_\gamma(v)$ by using *voxel-wise* approach; (b) true $\sigma^2_\gamma(v)$ image; and (c) estimated $\sigma^2_\gamma(v)$ by using *FMEM*. Scenario II : (d) estimated $\sigma^2_\gamma(v)$ by using *voxel-wise* approach; (e) true $\sigma^2_\gamma(v)$ image; (f) estimated $\sigma^2_\gamma(v)$ by using *FMEM*. Signal to noise ratio (SNR) images for (g) Scenario I and (h) Scenario II

.

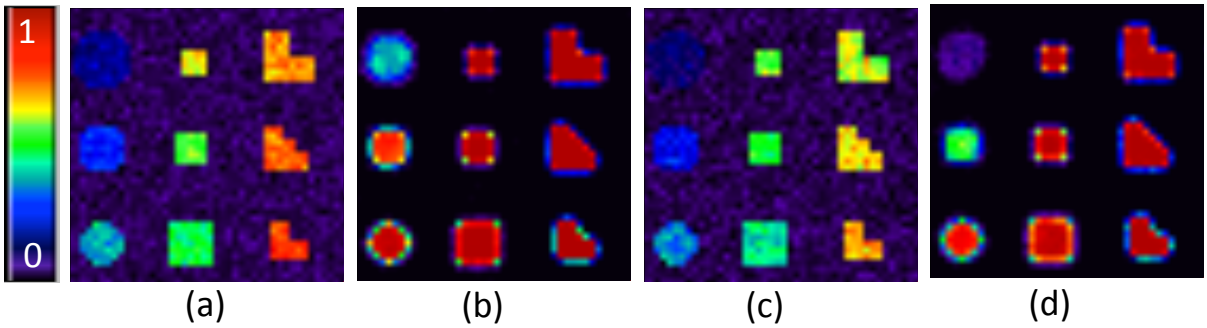

Figure 3.3: Simulation results for testing the genetic effect: Scenario I: the rejection rate image at a selected slice by using (a) *voxel-wise* approach and (b) *FMEM*; Scenario II: the rejection rate image by using (c) *voxel-wise* approach and (d) *FMEM*.

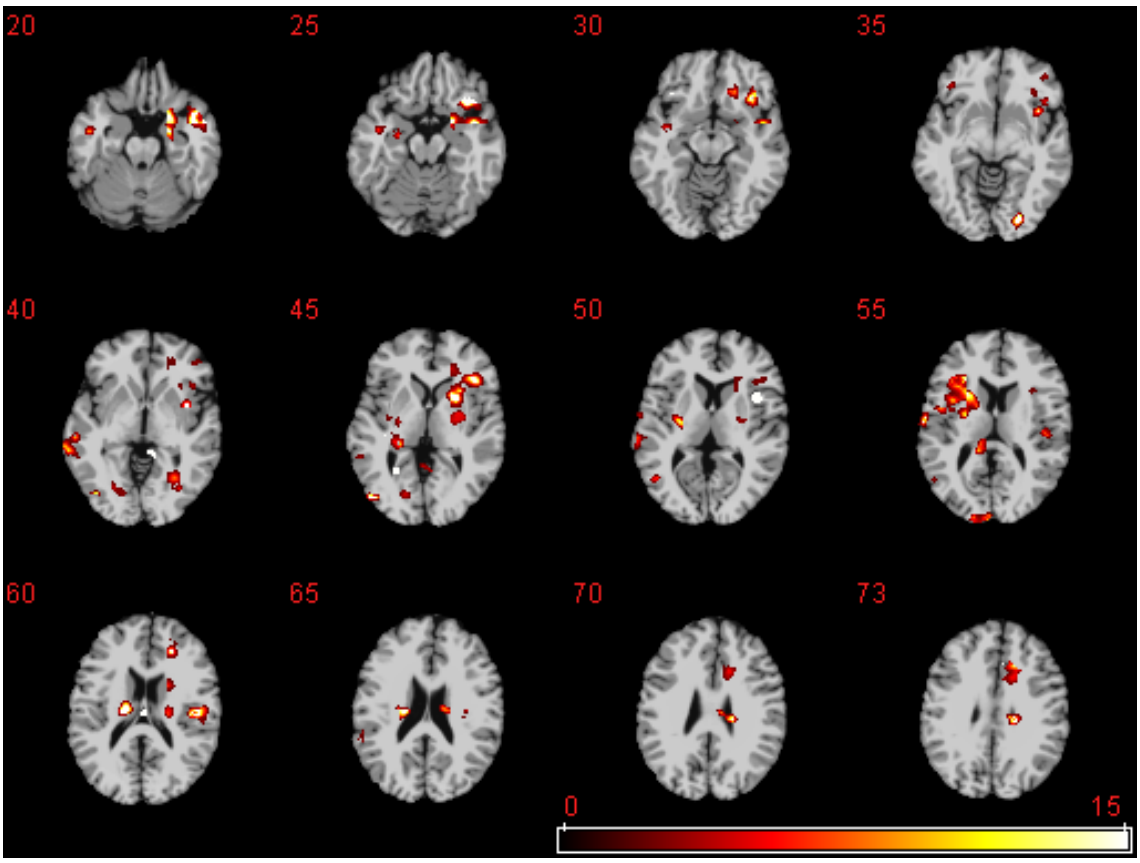

Figure 3.4: ADNI data analysis: the $-\log_{10}(p)$ map of testing the genetic effect of **CD2AP** on RAVEN images by using FMEM from 12 selected slices.

Table 3.1: The estimation results of $\sigma^2_\gamma(v)$ in *Scenario I* using FMEM and voxel-wise method in terms of average absolute value of bias (BIAS), root mean squares (RMS), standard deviation (SD), and the ratio between RMS and SD (RE).

	FEFM				Voxel-based			
$\sigma^2_\gamma(v)$	|BIAS|	RMS	SD	RE	|BIAS|	RMS	SD	RE
0	0.001	0.002	0.002	1	0.007	0.005	0.005	1
0.005	2.36e-06	0.003	0.003	1	0.005	0.005	0.005	1
0.0075	0.0005	0.003	0.003	1	0.005	0.006	0.006	1
0.01	0.001	0.004	0.004	1	0.006	0.008	0.008	1
0.0125	0.001	0.004	0.004	1	0.006	0.008	0.008	1
0.015	0.002	0.005	0.005	1	0.008	0.010	0.010	1
0.0175	0.003	0.005	0.005	1	0.008	0.010	0.010	1
0.020	0.002	0.006	0.006	1	0.010	0.012	0.012	1
0.0225	0.003	0.006	0.006	1	0.010	0.013	0.013	1
0.025	0.004	0.006	0.006	1	0.010	0.014	0.014	1

Table 3.2: The estimation results of $\sigma^2_\gamma(v)$ in *Scenario II* by using FMEM and voxel-wise method in terms of average absolute value of bias (BIAS), root mean squares (RMS), standard deviation (SD), and the ratio between RMS and SD (RE).

	FEFM				Voxel-based			
$\sigma^2_\gamma(v)$	|BIAS|	RMS	SD	RE	|BIAS|	RMS	SD	RE
0	0.002	0.003	0.003	1	0.003	0.008	0.008	1
0.005	0.0001	0.004	0.004	1	0.007	0.126	0.126	1
0.010	0.001	0.006	0.006	1	0.011	0.016	0.016	1
0.015	0.002	0.008	0.008	1	0.014	0.020	0.02	1
0.020	0.002	0.010	0.010	1	0.024	0.017	0.017	1
0.025	0.003	0.020	0.020	1	0.020	0.029	0.029	1
0.030	0.005	0.013	0.013	1	0.023	0.032	0.032	1
0.035	0.004	0.014	0.014	1	0.026	0.035	0.035	1
0.040	0.015	0.006	0.006	1	0.028	0.040	0.040	1
0.045	0.007	0.016	0.016	1	0.031	0.040	0.040	1

Table 3.3: The dice overlap ratio (DOR), average number of false positive cluster, and average size of false positive cluster for *Scenario I* with different cluster size thresholds.

		FEFM		**Voxel-based**	
Threshold		Mean	SD	Mean	SD
Voxel Size = 1	DOR	0.94	0.05	0.99	0.02
	False Positive Cluster Number	1.88	6.12	21.30	12.29
	False Positive Cluster Size	1.03	0.04	1.06	0.06
Voxel Size = 10	DOR	0.91	0.04	0.83	0.10
	False Positive Cluster Number	0	0	0	0
	False Positive Cluster Size	NA	NA	NA	NA

Table 3.4: The dice overlap ratio (DOR), average number of false positive cluster, and average size of false positive cluster for *Scenario II* with different cluster size thresholds.

		FEFM		**Voxel-based**	
Threshold		Mean	SD	Mean	SD
Voxel Size = 1	DOR	0.86	0.06	0.996	0.02
	False Positive Cluster Number	1.35	4.85	15.45	12.92
	False Positive Cluster Size	1.07	0.08	1.05	0.07
Voxel Size = 10	DOR	0.85	0.07	0.78	0.11
	False Positive Cluster Number	0	0	0	0
	False Positive Cluster Size	NA	NA	NA	NA

Table 3.5: The global power calculation of number of significant voxels detected by voxel-wise approach and FMEM in both scenarios.

Number of Voxels	4749	207	135	111	147	75	48	144	111	81
Scenario I										
$\sigma^2_\gamma(v)$	0	0.005	0.0075	0.010	0.0125	0.015	0.0175	0.020	0.0225	0.025
Voxel-based	123.91	36.3	38.25	43.08	72.245	42.98	31.535	104.32	86.1	65.97
FMEM	68.43	50.52	77.32	83.09	133.7	69.51	45.57	142.26	110.07	80.54
Scenario II										
$\sigma^2_\gamma(v)$	0	0.005	0.010	0.015	0.020	0.025	0.030	0.035	0.040	0.045
Voxel-based	110.93	27.08	33.105	40.05	65.92	39.06	27.51	90.69	73.42	56.59
FMEM	76.57	8.11	46.05	71.82	124.42	66.15	43.182	139.51	108.35	79.49

Chapter 4

Score Test for Functional Mixed Effects Model

4.1 Method

4.1.1 Weighted Score Test Statistic

Suppose we observe imaging measures, clinical variables, and genetic markers from n unrelated subjects. Let $\mathcal{V}$ be a selected brain region and v be a voxel in $\mathcal{V}$. For each individual i $(i = 1, 2...., n)$, we observe an $N_V \times 1$ vector of imaging measures, denoted by $Y_i = \{y_i(v) : v \in \mathcal{V}\}$, a $K \times 1$ vector of clinical covariates $\mathbf{x}_i = (x_{i1}, \cdots, x_{iK})^T$, and an $L \times 1$ vector $\mathbf{z}_i = (z_{i1}, \cdots, z_{iL})^T$ for genetic data. For notational simplicity, only univariate image measure is considered here, and thus N_V equals to the number of points in $\mathcal{V}$.

At each voxel v in $\mathcal{V}$, $y_i(v)$ can be modeled as a linear mixed effects model given by

$$y_i(v) = \mathbf{x}_i^T \beta(v) + \mathbf{z}_i^T \gamma(v) + e_i(v) \quad \text{for} \quad i = 1, \cdots, n, \tag{4.1}$$

where $\beta(v) = (\beta_1(v), \cdots, \beta_K(v))^T$ is a $K \times 1$ vector, $\gamma(v) = (\gamma_1(v), \cdots, \gamma_L(v))^T$ is an $L \times 1$ vector of genetic random effects, and $e_i(v)$ is the measurement error. We assume that $e_i(v) \sim N(0, \sigma_e(v)^2)$, $\gamma(v) \sim N(0, \sigma_\gamma^2(v)\mathbf{I}_L)$, and $\{e_i(v) : v \in \mathcal{V}\}$ are independent across i and independent of $\gamma(v)$ for all $v \in \mathcal{V}$, where $\mathbf{I}_L$ is an $L \times L$ identity matrix. Let $\Sigma_Y(v) = \sigma_\gamma^2(v)\mathbf{Z}^T\mathbf{Z} + \sigma_e^2(v)\mathbf{I}_n$ be the marginal variance of $\mathbf{Y}(v) = (y_1(v), \cdots, y_n(v))^T$, where $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)$ is an $L \times n$ matrix. Moreover, we define the projection matrix

of model (4.1) to be

$$P_X(v) = \Sigma_Y(v)^{-1} - \Sigma_Y(v)^{-1}\mathbf{X}(\mathbf{X}^T\Sigma_Y(v)^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma_Y(v)^{-1},$$

where $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ is a $p \times n$ matrix. When $\sigma_\gamma(v) = 0$, $\sigma_e^2(v)P_X(v)$ reduces to $L_X = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. The REML log-likelihood function of $\eta(v) = (\sigma_\gamma^2(v), \sigma_e^2(v))$, denoted by $\ell_{REML}(\mathbf{Y}^*(v)|\mathbf{Z}, \eta(v))$, is given by

$$\ell_{REML}(\eta(v); \mathbf{Y}(v)) = -0.5\log|\Sigma_Y(v)| - 0.5\log|\mathbf{X}^T\Sigma_Y(v)^{-1}\mathbf{X}| - 0.5\mathbf{Y}(v)^T P_X(v)\mathbf{Y}(v). \tag{4.2}$$

To investigate the genetic effect on imaging phenotype, we test

$$H_{0,\gamma} : \sigma_\gamma^2(v) = 0 \quad \text{v.s.} \quad H_{1,\gamma} : \sigma_\gamma^2(v) > 0. \tag{4.3}$$

The regular score test for (4.3) is given by

$$\mathrm{U}_{\sigma_\gamma^2}(v|\hat{\sigma}_e^2(v)) = \left.\frac{\partial \ell_{REML}(\sigma_\gamma^2(v)|\mathbf{Y})}{\partial \sigma_\gamma^2(v)}\right|_{\sigma_\gamma^2(v)=0}, \tag{4.4}$$

where $\hat{\sigma}_e^2(v)$ is the estimate of $\sigma_e^2(v)$ under the null hypothesis. Let $\Omega_X = L_X\mathbf{Z}^T\mathbf{Z}L_X^T$, then we have

$$\mathrm{U}_{\sigma_\gamma^2}(v|\hat{\sigma}_e^2) = 0.5\{\hat{\sigma}_e^{-4}(v)\mathbf{Y}(v)^T\Omega_X\mathbf{Y}(v) - tr(\Omega_X)\hat{\sigma}_e^{-2}(v)\}. \tag{4.5}$$

Since the major variation of $U_{\sigma_\gamma^2}$ comes from the first term on the right hand side of (4.5), by following Tzeng and Zhang [2007], we obtain a regular score test as

$$\mathrm{SC}_{\sigma_\gamma^2}(v|\hat{\sigma}_e^2) = 0.5\{\hat{\sigma}_e^{-4}(v)\mathbf{Y}(v)^T\Omega_X\mathbf{Y}(v)\}. \tag{4.6}$$

To motivate our new test statistic, we give a scratch of our idea. Let a_j be a constant

and A_j be a random variable for $j = 1, \ldots, J$ such that $a_j \geq 0$ and $\sum_{j=1}^{J} a_j = 1$, where J is a positive integer. It is also assumed that under both $H_{0,\gamma}$ and $H_{1,\gamma}$, A_js have the same first and second order moments. Therefore, for any j, the variance of A_j is not smaller than the variance of the weighted sum $\sum_{j=1}^{J} a_j A_j$, since we have

$$\begin{aligned} Var(\sum_{j=1}^{J} a_j A_j) &= \sum_{j=1}^{J} a_j^2 Var(A_j) + \sum_{j\neq k}^{j=J,k=J} a_j a_k Cov(A_j, A_k) \\ &\leq \sum_{j=1}^{J} a_j^2 Var(A_j) + \sum_{j\neq k}^{j=J,k=J} a_j a_k Var(A_j) = Var(A_j). \end{aligned} \tag{4.7}$$

Under certain regularity conditions, it follows from the central limit theorem that

$$\frac{A_j - E_0(A_j)}{\sqrt{Var(A_j)}} + \frac{E_0(A_j) - E(A_j)}{\sqrt{Var(A_j)}} \xrightarrow{d} N(0,1), \tag{4.8}$$

$$\begin{aligned} &\frac{\sum_{j=1}^{J} a_j A_j - E_0(\sum_{j=1}^{J} a_j A_j)}{\sqrt{Var(\sum_{j=1}^{J} a_j A_j)}} + \frac{E_0(\sum_{j=1}^{J} a_j A_j) - E(\sum_{j=1}^{J} a_j A_j)}{\sqrt{Var(\sum_{j=1}^{J} a_j A_j)}} \\ &= \frac{\sum_{j=1}^{J} a_j A_j - E_0(A_j)}{\sqrt{Var(\sum_{j=1}^{J} a_j A_j)}} + \frac{E_0(A_j) - E(A_j)}{\sqrt{Var(\sum_{j=1}^{J} a_j A_j)}} \xrightarrow{d} N(0,1), \end{aligned} \tag{4.9}$$

where $\xrightarrow{d}$ denotes convergence in distribution and $E_0(\cdot)$ denotes the expectation taken with respect to the null hypothesis $H_{0,\gamma}$. Thus, it follows from (4.7) that

$$\frac{|E_0(A_j) - E(A_j)|}{\sqrt{\mathrm{Var}(A_j)}} \leq \frac{|E_0(A_j) - E(A_j)|}{\sqrt{\mathrm{Var}(\sum_{j=1}^{J} a_j A_j)}} \quad \text{for any} \quad j, \tag{4.10}$$

which yields that the weighted test statistic $\sum_{j=1}^{J} a_j A_j$ is statistically more powerful than A_j.

Given the functional feature of imaging data, imaging data often consists of several homogeneous regions and imaging data within each homogeneous region have similar distributional properties, such as first and second moments. Based on this observation, we propose a weighted score test statistic with smaller variance as a modification for

(4.6) as follows:

$$\mathrm{WSC}_{\sigma^2_\gamma}(v \mid \hat{\sigma}^2_e, h) = 0.5 \sum_{v' \in B(v,h_v)} \omega(v, v'; h_v)\{\hat{\sigma}_e^{-4}(v')\mathbf{Y}(v')^T \Omega_X \mathbf{Y}(v')\}, \tag{4.11}$$

where $\omega(v, v'; h_v)$ is an nonnegative weight function of v, v', and h_v for characterizing the similarity between voxels v and v', $B(v, h_v)$ is a ball with center at voxel v and radius h_v, and $\sum_{v' \in B(v,h_v)} \omega(v, v'; h_v) = 1$.

4.1.2 Adaptive Estimation for Weight and Neighborhood

Given a voxel v, the adaptive weight $\omega(v, v'; h_v)$ and radius h_v should reflect the similarity of genetic effect between voxel v and its neighboring voxel v'. Therefore, to properly determine the weight for each neighboring voxel and the radius h_v in (4.11) under the joint parameter space of null hypothesis and alternative hypothesis $\{\sigma^2_\gamma(v) : \sigma_\gamma(v)^2 \geq 0\}$, we follow the adaptive estimation procedure developed in Chapter 3. The key steps of FMEM are given as follows:

> Step (E.1). Calculate the restricted maximum likelihood (REML) estimator of $\eta(v) = (\sigma^2_\gamma(v), \sigma^2_e(v))$ across all voxels $v \in \mathcal{V}$.
>
> Step (E.2). Spatially and adaptively re-estimate $\{\sigma^2_\gamma(v) : v \in V\}$ by incorporating information from neighboring voxels.

Step (E.1): Regular Estimation of $\sigma^2_\gamma(v)$ and $\sigma^2_e(v)$

In Step (E.1), we calculate the REML estimator of $\eta(v)$ across voxels via the REML function in (4.2). At each voxel v, the REML estimate of $\hat{\eta}(v)$, denoted by $\hat{\eta}(v)$, is to maximize $\ell_{REML}(\eta(v); \mathbf{Y}(v))$. Since our primary interest focuses on $\sigma^2_\gamma(v)$, $\sigma^2_e(v)$ is fixed as $\hat{\sigma}^2_e(v)$ for all voxels from now on.

Step (E.2): Adaptive Estimation of $\sigma_\gamma^2(v)$

In Step (E.2), a weighted REML function for estimating $\sigma_\gamma^2(v)$ is constructed as follows:

$$L_{REML}(\sigma_\gamma^2(v)|\mathbf{Y}^*, B(v,h)) = \sum_{v' \in B(v,h)} \omega_\gamma(v, v'; h) \ell_{REML}(\mathbf{Y}^*(v')|\mathbf{Z}, \sigma_\gamma^2(v), \hat{\sigma}_e^2(v')). \quad (4.12)$$

Following the adaptive estimation (AET) procedure proposed in [Polzehl and Spokoiny, 2000; Li et al., 2011], we adaptively determine $\{\omega_\gamma(v, v'; h) : v, v' \in \mathcal{V}\}$ and then calculate $\hat{\sigma}_\gamma(v; h)$ as h increases from $h_0 = 0$ to a predetermined value $h_S = r_0$. The key idea of AET is to build a sequence of nested spheres $B(v, h_s)$ for $h_0 = 0 < h_1 < \cdots < h_S = r_0$ at each voxel $v \in \mathcal{V}$ and then sequentially estimate $\hat{\sigma}_\gamma(v, v'; h_s)$ for all $v' \in B(v, h_s)$ based on $\{\hat{\sigma}_\gamma^2(v', h_{s-1}) : v' \in B(v, h_s)\}$ for all $v \in \mathcal{V}$ and $s = 1, \ldots, S$. Since the parameters of AET have been described in details in [Polzehl and Spokoiny, 2000; Li et al., 2011], we do not include them here for the sake of simplicity.

The three key steps of AET, including weights adaptation, estimation, and termination checking are presented as follows.

- In the weights *adaption* step (i), we prefix a series $\{h_s = c_h^s : s = 1, \cdots, S\}$ of radii with $c_h \in (1, 2)$, say $c_h = 1.15$. We then set $s = 1$ and $h_1 = c_h$. In the weight adaptation step, the adaptive weights are given by

$$\omega_\gamma(v, v'; h_s) = K_{loc}(||v - v'||_2/h_s) K_{st}(D_\gamma(v, v'; h_{s-1})/C_n), \quad (4.13)$$

  where $||\cdot||_2$ denotes the Euclidean norm of a vector (or a matrix), $K_{loc}(u) = (1-u)_+^{0.5}$ and $K_{st}(u) = exp(\frac{-u}{2})\mathbf{1}(0 < u < 5)$, in which $\mathbf{1}(\cdot)$ is an indicator function. The selection of kernel function has been discussed in the litera-

ture [Tabelow et al., 2006; Li et al., 2011]. Moreover, $D_\gamma(v, v'; h_{s-1})$ is set as $\{\hat{\sigma^2_\gamma}(v; h_{s-1}) - \hat{\sigma^2_\gamma}(v'; h_{s-1})\}^2 / \widehat{var[\sigma^2_\gamma(v)]}$, where $\widehat{var[\sigma^2_\gamma(v)]}$ is estimated by using the inverse of Fisher information matrix of $(\sigma^2_\gamma(v), \sigma^2_e(v))$ from the likelihood function (4.2) as $h = h_0$. Then they are fixed for subsequent update of h. Following Li et al. [2011], we choose $C_n = n^{1/3}\chi^2(1)^{0.5}$ for $D_\gamma(v, v'; h_{s-1})$ defined in (4.13), where $\chi^2(1)^{0.5}$ is the 0.5-percentile of the $\chi^2(1)$ distribution. The adaptive weight $K_{st}(D_\gamma(v, v'; h_{s-1})/C_n)$ downweights the role of a voxel $v' \in B(v, h_s)$ in $L_{REML}(\sigma^2_\gamma(v)|\mathbf{Y}^*, B(v, h_s))$ if $D_\gamma(v, v'; h_{s-1})$ is large. The weight $K_{loc}(||v-v'||_2/h_s)$ gives less weight to the voxel $v' \in B(v, h_s)$, whose location is far from the voxel v.

- In the *estimation* step (ii), for each $v \in \mathcal{V}$ and the radius h_s, we calculate the $\hat{\sigma}_\gamma(v; h_s)$ by maximizing $L_{REML}(\sigma^2_\gamma(v)|\mathbf{Y}^*, B(v, h_s))$ defined in equation (4.12) given $\omega_\gamma(v, v'; h_s)$s'.

- In the *termination checking* step (iii), after the S_0-th iteration, we start to calculate a stopping criterion based on a distance between $\hat{\sigma^2_\gamma}(v; h_{S_0})$ and $\hat{\sigma^2_\gamma}(v; h_s)$ given by

$$D(\hat{\sigma^2_\gamma}(v; h_{S_0}), \hat{\sigma^2_\gamma}(v; h_s)) = \{\hat{\sigma^2_\gamma}(v; h_{S_0}) - \hat{\sigma^2_\gamma}(v; h_s)\}^2 \widehat{var[\sigma^2_\gamma(v)]}^{-1} \tag{4.14}$$

for $s > S_0$. Then, we compare $D(\hat{\sigma^2_\gamma}(v; h_{S_0}), \hat{\sigma^2_\gamma}(v; h_s))$ with a benchmark, denoted by $\tilde{C}(s)$, for $s > S_0$. If $D(\hat{\sigma^2_\gamma}(v; h_{S_0}), \hat{\sigma^2_\gamma}(v; h_s)) > \tilde{C}(s)$, then we set $\hat{\sigma^2_\gamma}(v, h_{s-1}) = \hat{\sigma^2_\gamma}(v)$, $w(v, v'; h_v) = w(v, v'; h_{s-1})$ and $h_v = h_{s-1}$; and the estimation for this voxel v is terminated. If $s = S$ and $D(\hat{\sigma^2_\gamma}(v; h_{S_0}), \hat{\sigma^2_\gamma}(v; h_s)) \leq \tilde{C}(s)$, $\hat{\sigma^2_\gamma}(v)$, $w(v, v'; h_v)$ and h_v are set as $\hat{\sigma^2_\gamma}(v, h_S)$, $w(v, v'; h_S)$ and h_S, respectively, and the estimation process terminates. Once the estimations for all v in $\mathcal{V}$ are finished, we then stop. If $s \leq S_0$ or $D(\hat{\sigma^2_\gamma}(v; h_{S_0}), \hat{\sigma^2_\gamma}(v; h_s)) \leq \tilde{C}(s)$ for $s < S_0 \leq S-1$, then we go back to the weights adaptation step (i) with increased radius $h = h_{s+1} = c_h^{s+1}$. Throughout the paper, we set $S_0 = 2$, $\tilde{C}(s) = \chi^2(p)^{0.7/(s-1)}$, and $S = 10$.

4.1.3 p-value Calculation based on Gamma Approximation

After determining all weights and the radius at each voxel $v \in \mathcal{V}$, we plug them into the weighted score test (4.11). For $\sigma_e^2(v)$ in (4.11), we use its REML estimate $\hat{\sigma}_e^2(v) = \mathbf{Y}(v)^T L_X \mathbf{Y}(v)/(n-p)$ under the null hypothesis. Following Tzeng and Zhang [2007], we approximate the distribution of the test statistic (4.11) by a Gamma(a_v, b_v) distribution. To estimate the parameters a_v and b_v, we use the method of moment (MoM) to match (a_v, b_v) with the first two moments of (4.11) under the null hypothesis. The first moment of $E_{H_{0,\gamma}}(\mathrm{WSC}_{\sigma_\gamma^2}(v \mid \hat{\sigma}_e^2, h))$ is given by

$$\hat{E}_{H_{0,\gamma}}(\mathrm{WSC}_{\sigma_\gamma^2}(v \mid \hat{\sigma}_e^2, h)) = \ 0.5\ tr(\Omega_X) \sum_{v' \in B(v,h_v)} \omega(v, v'; h_v)\sigma_e^2(v') \equiv \hat{\mathcal{E}}_v. \tag{4.15}$$

For the purpose of simplicity and computational efficiency, we assume the correlation between the voxels v' and v'' to be $\rho_{v',v''}$ and estimate it by using $\hat{\rho}_{v',v''} = \mathbf{Y}^T(v')L_X\mathbf{Y}^T(v'')/(\hat{\sigma}_e(v')\hat{\sigma}_e(v'')(n$ Therefore, under the null hypothesis, we have

$$Cov(\mathbf{Y}_{v',v''}) = \begin{bmatrix} \sigma_e^2(v')I_n & \rho_{v',v''}\sigma_e(v')\sigma_e(v'')I_n \\ \rho_{v',v''}\sigma_e(v')\sigma_e(v'')I_n & \sigma_e^2(v'')I_n \end{bmatrix}. \tag{4.16}$$

Furthermore, it follows from [Bao and Ullah, 2010] that the second moment $E_{H_{0,\gamma}}(\mathrm{WSC}_{\sigma_\gamma^2}(v \mid \sigma_e^2, h))^2$ is given by

$$\begin{aligned}
&\hat{E}_{H_{0,\gamma}}\{\mathrm{WSC}_{\sigma_\gamma^2}(v \mid \sigma_e^2, h)^2\} = \ 0.25\ tr(\Omega_X)^2\{\sum_{v' \in B(v,h_v)} \omega^2(v, v'; h_v)\sigma_e^{-4}(v')\} \\
&\quad + 0.5\ tr(\Omega_X^2)\{\sum_{v' \in B(v,h_v)} \omega^2(v, v'; h_v)\sigma_e^{-4}(v') + \sum_{v' \neq v"} \omega(v, v'; h_v)\omega(v, v"; h_v)\sigma_e^{-2}(v')\sigma_e^{-2}(v'')\rho_{v',v''}^2\} \\
&= Var_{H_{0,\gamma}}(\mathrm{WSC}_{\sigma_\gamma^2}(v \mid \hat{\sigma}_e^2, h)) + \{\hat{E}_{H_{0,\gamma}}(\mathrm{WSC}_{\sigma_\gamma^2}(v \mid \hat{\sigma}_e^2, h))\}^2 \equiv \hat{\mathcal{S}}_v^2 + \hat{\mathcal{E}}_v^2.
\end{aligned} \tag{4.17}$$

The derivation of (4.17) is provided in Appendix. Thus, it follows from (4.15) and (4.17) that

$$\hat{a_v} = \frac{\hat{\mathcal{E}_v}^2}{\hat{\mathcal{S}}_v^2} \quad \text{and} \quad \hat{b_v} = \frac{\hat{\mathcal{S}_v}^2}{\hat{\mathcal{E}}_v}. \tag{4.18}$$

Based on Gamma$(\hat{a_v}, \hat{b_v})$, we can approximate the p-value of $\text{WSC}_{\sigma_\gamma^2}(v \mid \sigma_e^2, h)$ at each voxel in $\mathcal{V}$. Subsequently, one may perform False Discovery Rate (FDR) method or Random Field Theory (RFT) to account for multiple comparisons.

4.2 Results

4.2.1 Simulation Studies

We simulated data at all $N_V = 5,808$ voxels on a $44 \times 44 \times 3$ phantom image. Each z-slice contains the same effect regions. At each voxel, we simulated the univariate imaging measure according to model (4.1) with $\beta(v) = (\beta_0(v), \beta_1(v), \beta_2(v), \beta_3(v))^T$ and $\mathbf{x}_i = (1, x_{i1}, x_{i2}, x_{i3})^T$. Moreover, the covariates x_{i1}, x_{i2}, and x_{i3} were generated from a Gaussian distribution with mean 40 and standard deviation 10, a Bernoulli distribution with success probability 0.5, and a Bernoulli distribution with success probability 0.3, respectively. These three covariates were designed to mimic the common clinical variables age, gender, and disease status. For a slice of a phantom image, the effect areas for $\beta_0(v)$ were divided into 16 regions with 4 different values ranging from 0.02 to 0.08, increasing by 0.02 (Figure 8(a) as in previous section); for $\beta_1(v)$, the effect regions were divided into 25 regions ranging from $10^{-2.5}$ to $10^{-12.5}$, decreasing by a rate of $10^{-2.5}$ (Figure 8(b)); for $\beta_2(v)$, the whole space was separated into 3 regions with values 0, 0.05, and 0.1 (Figure 8(c)); the effect area of $\beta_3(v)$ on a slice of phantom image was divided into 9 regions with values ranging from 0 to 0.1, increasing by differences of 0.025 (Figure 8(d)).

The genetic information was simulated according to the SNP data obtained from the public accessible data of the Alzheimer's Disease Neuroimage Initiative (ADNI).

It is an ongoing longitudinal study with primary purpose of exploring the genetic and neuroimaing information associated with late-onset Alzheimer's disease (LOAD). The study recruited elderly subjects older than 65 years of age consisting about 400 subjects with mild cognitive impairment (MCI), about 200 subjects with Alzheimer (AD), and around 200 healthy controls. Each subject was followed for at least 3 years. During the study period, the subjects were assessed with magnetic resonance imaging (MRI) measures and psychiatric evaluation to determine the diagnosis status at each time point. The genetic information was also collected from each subject at baseline and it is genotyped by the Illumina 610 Quad array with more than 620,000 single nucleotide polymorphysm (SNPs). More information of ADNI is provided in the real data analysis result Section 3.2. We simulated the genetic information based on the two following scenarios.

- Scenario I. To preserve the linkage disequilibrium among SNPs, we utilize all of the SNPs on chromosome 1 from 197 Caucasian controls to generate the genetic effect. After eliminating the SNPs with minor allele frequency (MAF) less than 5%, there were 31554 out of 45627 SNPs left. Then we randomly chose 20 SNPs and 100 subjects among the 197 healthy controls as the simulated genetic data $\mathbf{z}_i$ in (4.1). In this case, $n = 100$. If any of these 20 SNPs have MAF less than 5%, the genetic data was resampled until all of the 20 SNPs have MAF $\geq 5\%$.

- Scenario II. To evaluate the performance of weighted score test in the case of high LD, we selected the SNPs from the same gene in the second scenario. Searching the SNPs on the gene PICALM, which is found to be relevant to Alzheimer's disease in many studies [Harold et al., 2009] using the gene list "glist-hg18" provided by PLINK, there were 23 SNPs on PICALM with MAF larger than 5%. After eliminating the missing values, there are 176 healthy controls with complete genotype data at these 23 SNPs. We randomly selected 7 SNPs from 75 healthy controls to be $\mathbf{z}_i$ in (4.1). Although there is strong LD among these 7 SNPs, no SNP has

perfect correlation (1 or -1) with any other SNP in these 75 subjects. In this case, n equals 75.

In both scenarios, the SNP effects were assumed to be additive. The $\gamma(v)$ was generated from a multivariate Gaussian distribution with mean zero and covariance matrix $\sigma_\gamma^2(v)\mathbf{I}_L$. Different $\sigma_\gamma^2(v)$ values, which represent different signal-to-noise ratios, were chosen to examine the performance of our method at different signal-to-noise ratios and also to test whether weighted score test we proposed can perform well for different shapes. See Figure 12 (a) and Figure 12 (b) for Scenarios I and II. Moreover, we overlay some of the effect areas of $\beta_3(v)$ and $\sigma_\gamma^2(v)$ in order to account for the fact that the brain phenotype is an intermediate expression of disease progression. The $\{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$ of the effect regions in Scenario I were ranging from 0.005 to 0.025, increasing by 0.0025, whereas the $\{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$ of effect regions in Scenario II were ranging from 0.005 to 0.045, increasing by 0.005. The random error $e_i(v)$ was independently distributed as a univariate standard Gaussian distribution for all voxels. The sample slice from a sample dataset of signal-to-noise ratio for both scenarios are shown in Figure 12(c) and 12(d).

We tested the hypotheses $H_0 : \sigma_\gamma^2(v) = 0$ and $H_1 : \sigma_\gamma^2(v) \neq 0$ for all voxels in $\mathcal{V}$ based on both weighted score test in (4.11) and regular score test (4.6). The computational time of a sample dataset in *Scenario II* using a 2.4 GHz single CPU machine with 48 GB memory is 1380.2 seconds while implementing the complete weight estimation and hypothesis testing procedure for all 5808 voxels; whereas it is 4591.3 seconds under the same computing conditions for 15% of voxels using likelihood ratio test described in Chapter 3. The finite sample performance is also evaluated using cluster-based thresholding M. Silver and ADNI [2011]. Specifically, we first thresholded the p-values for all voxels in $\mathcal{V}$ by using an initial p-value 0.01 suggested by M. Silver and ADNI [2011] to identify clusters of contiguous supra-threshold voxels. Then, the thresholded clusters were matched with the 9 separated activated areas in Figure 12(a) or 12(b). If a specific thresholded cluster overlaps with at least one voxel in any of the 9 effect regions, we call

such cluster as a "true positive". In contrast, if a specific thresholded cluster does not overlap with any voxels of the 9 effect regions, we call the cluster a "false positive". We summarized the hypothesis testing results by the average dice overlap ratio (DOR), the average number of false positive clusters, and the average size in the number of voxels of false positive clusters. DOR is the ratio between the number of true positive clusters over the true number of effect areas, which is 9 in this simulation setting. Thus, the higher DOR means the higher the probability of detecting true effect regions. As shown in Tables 7 and 8, if we set the cluster size threshold at 1 voxel, Weighted score test has smaller DOR and smaller number of false positive clusters compared with voxel-wise method. When the cluster size threshold increases to 10 voxels, weighted score test has a similar DOR value as that of the no threshold case, whereas the DOR of the voxel-wise approach reduces by about 20%. Table 9 summarizes the number of significant voxels identified by the two methods in each effect region of Scenarios I and II In Table 9, weighted score test detects more voxels in effect regions in both scenarios while maintaining reasonable rate of false positive detection of null voxels. Finally, we conclude that weighted score test outperforms voxel-wise method in both detecting true effect regions and controlling the false positive error rate.

4.2.2 ADNI Data Analysis

The aim of this ADNI data analysis is to use FMEM to identify brain regions affected by candidate genes, thereby hoping to shed light on the pathological interactions between these causal genes and brain function. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5- year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers,

and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AlzheimerâĂŹs disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California âĂŞ San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see $www.adni-info.org$.

The data we employed to evaluate the performance of FMEM was from the first phase (ADNI), which was conducted mainly to search for the causal SNPs associated with the progression of Alzheimer's disease and to establish an alternative diagnosis standard using MRI brain images. About 800 subjects with age older than 65 were recruited and followed at least 3 years. The 800 subjects included 200 healthy controls, 400 subjects with different levels of mild cognitive impairment (MCI), and 200 subjects with Alzheimer's disease (AD). Besides the SNPs and the T1 weighted MRI imaging measurements, the subjects were assessed with demographic information and psychiatric examination scores to determine the diagnosis status at each scheduled visit.

The T1-weighted MRI images were preprocessed by standard image processing steps including AC (anterior commissure) and -PC (posterior commissure) correction, bias

field correction, skull-stripping, intensity inhomogeneity correction, cerebellum removal, segmentation, and nonlinear registration. After segmentation, the brain was segmented into four different tissues: grey matter (GM), white matter (WM), ventricle (VN), and cerebrospinal fluid (CSF). We quantified the local volumetric group differences by generating RAVENS maps [Davatzikos et al., 2001] for the whole brain and each of the segmented tissue type (GM, WM, VN, and CSF) respectively, using the deformation field we obtained during registration. RAVENS methodology is based on a volume-preserving spatial transformation, which ensures that no volumetric information is lost during the process of spatial normalization, since this process changes an individual′s brain morphology to conform it to the morphology of the Jacob template.

We are interested in detecting meaningful brain regions of interest that are associated with several candidate genes. We included only the subjects whose diagnosis status were healthy control and Alzheimer's disease at the baseline and had no status change during the study period. After screening, the total number of subjects we included was 206 (107 HCs and 99 ADs). The clinical covariates of interest included in our analysis were gender, baseline age, square of baseline age, handedness, education, baseline intracranial volume, and the risk of APOE. Specifically, the handedness was treated as a binary variable, the education information was the self-reported years of education by the subjects, and the risk of APOE is assumed to be additive. In detail, the risk of APOE for a subject was 3 if he/she carries $\varepsilon 4$ at both alleles; it was 2 if he/she carries $\varepsilon 3$ and $\varepsilon 4$ in two alleles, the risk would be considered 0 if the two APOE alleles were the combination of $\varepsilon 2$ and $\varepsilon 3$, and other combination of APOE alleles are assumed to have risk 1.

Many genes have been reported to be causal in the progression of Alzheimer's disease, especially to the area of hippocampus which plays an important role of consolidation of information from short-term memory and long-term memory. We selected the candidate causal gene TOMM40 on chromosome 19 [Potkin et al., 2009] which encodes the protein "Translocase of Outer Mitochondrial Membrane 40 homolog (yeast)" and is shown to be

associated with the increasing risk of late-onset Alzheimer's disease. We first matched the SNPs in ADNI with the gene list "glist-hg18" provided by PLINK [Purcell and et al, 2007] and were able to locate 3 SNPs on TOMM40, including rs157580, rs2075650 and rs8106922. All these SNPs pass the quality control procedure with MAF $> 5\%$ and the Hardy Weinberg Equilibrium (HWE) test p-value> 0.01. The MAFs of rs157580, rs2075650 and rs8106922 are around 0.35, 0.24 and 0.33. The associated demographic information is presented in the Table 10.

For the gene, we computed WSC in (4.11) and its approximated p-values for each voxel with $\mathbf{z}$ coded as the numbers of minor alleles to detect its associated significant brain regions of interest (ROIs). For comparison, we also computed regular score test in (4.6) to analyze the same dataset. To formally detect significant ROIs, following Ge and et al [2012], we used a cluster-form of threshold of 0.1% with a minimum voxel clustering value of 50 voxels. Weighted score test is able to to detect the brain region of *right insula* with the cluster size 95 voxels for TOMM40, after overlapping the significant cluster with the 96 predefined ROIs in the Jacob template; whereas the standard voxel-wise method cannot identify any brain regions. The brain regions of insulae are believed to be involved in consciousness and play a role in diverse functions usually linked to emotion or the regulation of the body's homeostasis. These functions include perception, motor control, self-awareness, cognitive functioning, and interpersonal experience. Figure 14 shows the $-\log_{10}(p)$ map of several selected slices with significant clusters for testing the genetic effect of TOMM40 on RAVEN images identified by weighted score test. Additionally, the computational time to complete the real data analysis using weighted likelihood ratio test proposed in previous section *Functional Mixed Effect Model* under the same computational condition is around 103.3 hours while it only takes 67 hours using weighted score test proposed in this section.

4.2.3 Tables and Figures

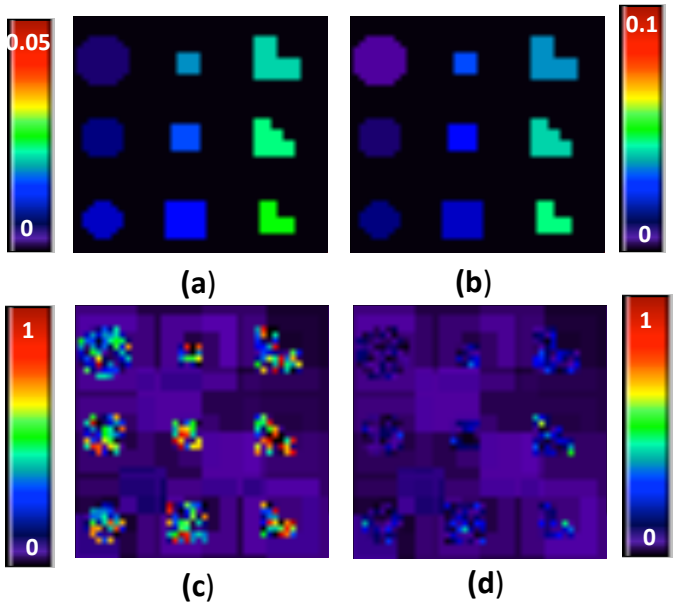

Figure 4.1: Simulation setting. True $\sigma^2_\gamma(v)$ image (a) Scenario I and (b) Scenario II; signal to noise ratio (SNR) images for (c) Scenario I and (d) Scenario II.

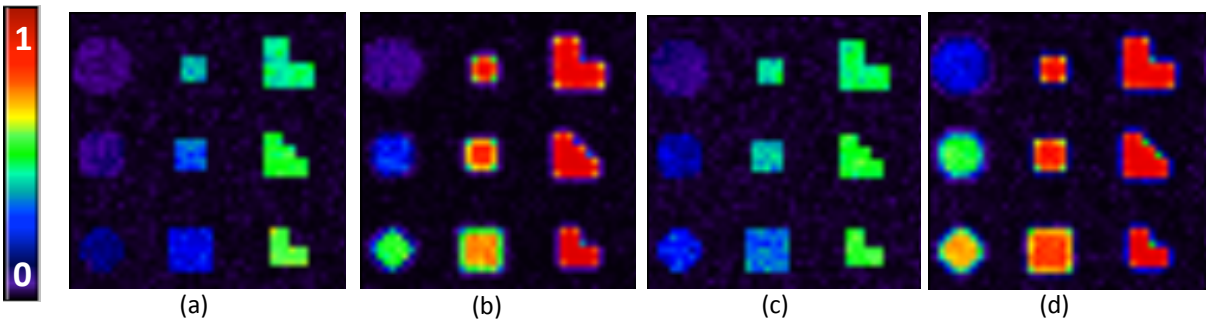

Figure 4.2: Simulation results for testing the genetic effect: Scenario I: the rejection rate image at a selected slice by using (a) *regular* score test and (b) *weighted* score test; Scenario II: the rejection rate image by using (c) *regular* score test and (d) *weighted* score test.

Table 4.1: The dice overlap ratio (DOR), average number of false positive cluster, and average size of false positive cluster for *Scenario I* with different cluster size thresholds.

		FEFM		**Voxel-based**	
Threshold		Mean	SD	Mean	SD
Voxel Size = 1	DOR	0.94	0.06	0.80	0.08
	False Positive Cluster Number	0.135	0.38	0.05	0.22
	False Positive Cluster Size	2.08	0.28	2	0
Voxel Size = 10	DOR	0.75	0.07	0.31	0.06
	False Positive Cluster Number	0	0	0	0
	False Positive Cluster Size	NA	NA	NA	NA

Table 4.2: The dice overlap ratio (DOR), average number of false positive cluster, and average size of false positive cluster for *Scenario II* with different cluster size thresholds.

		FEFM		**Voxel-based**	
Threshold		Mean	SD	Mean	SD
Voxel Size = 1	DOR	0.998	0.02	0.97	0.05
	False Positive Cluster Number	0.62	0.89	0.11	0.31
	False Positive Cluster Size	2.49	0.76	2	0
Voxel Size = 10	DOR	0.91	0.06	0.49	0.11
	False Positive Cluster Number	0	0	0	0
	False Positive Cluster Size	NA	NA	NA	NA

Table 4.3: The global power calculation of number of significant voxels detected by voxel-wise approach and FMEM in both scenarios.

Number of Voxels	4749	207	135	111	147	75	48	144	111	81
Scenario I										
$\sigma^2_\gamma(v)$	0	0.005	0.0075	0.010	0.0125	0.015	0.0175	0.020	0.0225	0.025
Voxel-based	46.69	8.605	13.66	19.515	38.765	26.72	21.03	70.09	64.15	52.805
FMEM	52.95	15.12	35.38	52.03	102.41	59.19	40.225	133.97	105.47	77.94
Scenario II										
$\sigma^2_\gamma(v)$	0	0.005	0.010	0.015	0.020	0.025	0.030	0.035	0.040	0.045
Voxel-based	54.77	20.495	27.54	31.99	52.905	31.87	22.935	74.785	61.805	47.82
FMEM	132.54	47.305	69.045	73.47	118.97	63.98	42.37	134.39	104.77	77.355

Table 4.4: Demographic information of the selected 206 subjects. For the *categorical* variables: gender, handedness and risk of APOE, the numbers are the frequencies for the corresponding groups. For the *continuous* variables: baseline age, baseline ICV and years of education, the numbers are the mean (standard deviation) for the corresponding groups.

		Healthy Control	Alzheimer's Disease
Total		107	99
Male (Female)		57 (50)	55(44)
Right (Left) Hand Users		97 (10)	92(7)
Risk of APOE	3	0	21
	2	29	37
	1	69	38
	0	9	3
Baseline Age		75.9 (5.1)	76.1 (7.1)
Baseline ICV		1.269×10^6 (1.258×10^5)	1.284×10^6 (1.524×10^5)
Years of Education		16.1 (2.9)	14.6 (3.11)

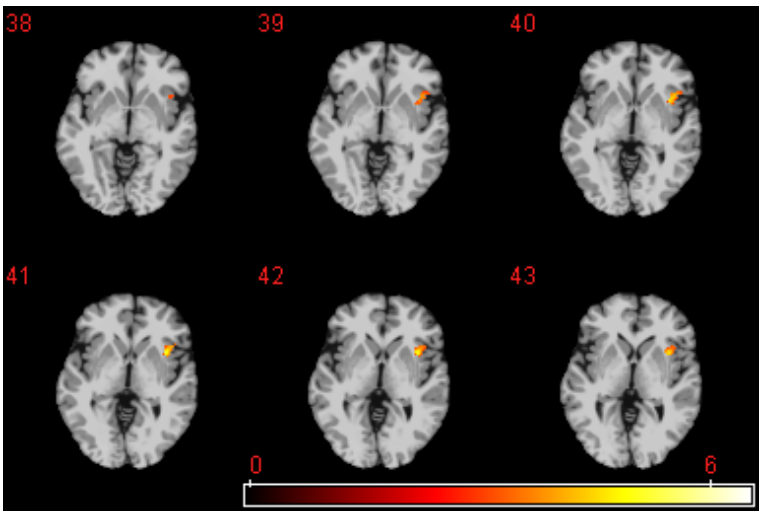

Figure 4.3: ADNI data analysis: the $-\log_{10}(p)$ map of testing the genetic effect of **TOMM40** on RAVEN images by using FMEM from 6 selected slices.

Chapter 5

Conclusion

For general correlated multivariate phenotype, we have developed the PRM which provides a more effective analysis for the association delineation between multivariate phenotypes and covariates of interest. The proposed methodology is demonstrated in a study investigating the impact of candidate SNPs on early age brain development. Analysis results obtained from the PRM successfully identified two previously reported SNPs while none of them were detected by either CWM or PCR. This phenomenon is consistent with the results in the simulation studies showing that compared to the two other methods, the PRM tends to have higher power for detecting the association between high dimensional phonetypes and the covariates of interest with better type I error control. Hence we expect that this novel statistical tool will assist scientists in exploring new findings with more effective and reliable statistical results in the high dimensional data settings. Future work includes establishing the asymptotic properties of the PRM under mild conditions, considering ultra-high dimensional phenotypes and genomic data, as well as extending the PRM to longitudinal and familial studies.

For structure MRI responses, we have developed FMEM with two test statistics - weighted likelihood ratio test and weighted score test - to carry out an association analysis between neuroimaging phenotypes and a group of genetic markers, while adjusting for the clinical variables of interest. We have proposed a multiscale adaptive procedure with three features: being spatial, being hierarchical, and being adaptive. Moreover, weighted score test provides a new philosophy with efficient computation of smoothing

the test statistics instead smoothing estimators in most traditional approaches. Our simulation results have shown substantial gain in parameter estimation precision and statistical power in detecting true effect ROIs compared with the voxel-wise method. We expect that FMEM will provide a more efficient tool of exploiting meaningful associations between brain ROIs and genetic makers. Given different possible etiology of the selected genes, we believe such information may provide important guidance for personalized definition of disease status and hence inform the development of personalized treatment.

User-friendly software to implement the PRM will be available to public for non-profit purposes on our group website: http://www.bios.unc.edu/research/bias/software.html.

Appendix I

Additional Results for PRM

Table A1.1: Type I Errors - Part 1

q	50	100	150	200	50	100	150	200
p	N=150				N=200			
0.05	0.053	0.06	0.027	0.047	0.093	0.06	0.067	0.06
0.1	0.04	0.04	0.02	0.033	0.047	0.06	0.033	0.073
0.2	0.053	0.06	0.053	0.033	0.053	0.06	0.02	0.04
Principle Component p								
0.05	0.06	0.04	0.06	0.027	0.053	0.053	0.047	0.053
0.1	0.087	0.087	0.093	0.067	0.093	0.107	0.087	0.087
0.2	0.132	0.1	0.127	0.113	0.133	0.167	0.127	0.127
Multiple Comparison p								
0.05	0.033	0.04	0.047	0.047	0.073	0.08	0.067	0.087
0.1	0.087	0.067	0.107	0.093	0.12	0.113	0.14	0.113
0.2	0.132	0.16	0.172	0.147	0.153	0.16	0.18	0.153

This table gives the results of the *type I error* based on the sample sizes $N = 150$ and *200* for PRM multiple comparison and principle component methods, with *minor allele frequencies* 0.05, 0.1 and 0.2 to simulate the distribution of the genetic groups and number of responses q.

Table A1.2: Type I Errors - Part 2

q	50	100	150	200	50	100	150	200
p	N=150				N=200			
0.3	0.087	0.02	0.033	0.04	0.027	0.06	0.033	0.093
0.4	0.047	0.04	0.06	0.08	0.067	0.053	0.067	0.06
0.5	0.053	0.053	0.067	0.053	0.067	0.047	0.073	0.04
Principle Component p								
0.3	0.139	0.133	0.147	0.133	0.16	0.173	0.167	0.14
0.4	0.152	0.167	0.153	0.167	0.167	0.2	0.173	0.173
0.5	0.185	0.173	0.173	0.18	0.18	0.207	0.187	0.207
Multiple Comparison p								
0.3	0.219	0.193	0.205	0.173	0.265	0.207	0.258	0.213
0.4	0.265	0.207	0.258	0.213	0.2	0.247	0.247	0.233
0.5	0.298	0.213	0.311	0.253	0.227	0.26	0.313	0.253

The table gives the results of the *type I error* based on the sample sizes $N = 150$ and *200* for PRM multiple comparison and principle component methods, with *minor allele frequencies* 0.3, 0.4 and 0.5 to simulate the distribution of the genetic group and number of responses q.

Table A1.3: Type I Errors - Part 3

q	50	100	150	200	50	100	150	200
p	N=250				N=300			
0.05	0.027	0.04	0.027	0.047	0.033	0.053	0.048	0.052
0.1	0.027	0.067	0.027	0.04	0.027	0.067	0.048	0.048
0.2	0.067	0.067	0.047	0.033	0.053	0.067	0.032	0.068
Principle Component p								
0.05	0.053	0.06	0.033	0.03	0.053	0.06	0.06	0.033
0.1	0.12	0.093	0.047	0.073	0.1	0.107	0.08	0.08
0.2	0.153	0.113	0.093	0.087	0.113	0.153	0.1	0.127
Multiple Comparison p								
0.05	0.06	0.04	0.073	0.053	0.053	0.08	0.06	0.04
0.1	0.107	0.087	0.127	0.093	0.073	0.153	0.1	0.12
0.2	0.153	0.16	0.18	0.153	0.147	0.12	0.16	0.127

The table gives the results of the *type I error* based on sample sizes $N = 250$ and *300* for PRM, multiple comparison and principle component methods, with *minor allele frequencies* 0.05, 0.1 and 0.2 to simulate the distribution of the genetic group and number of responses q.

Table A1.4: Type I Errors - Part 4

q	50	100	150	200	50	100	150	200
p	N=250				N=300			
0.3	0.073	0.047	0.053	0.087	0.073	0.107	0.06	0.04
0.4	0.06	0.027	0.06	0.04	0.093	0.053	0.068	0.044
0.5	0.06	0.033	0.053	0.047	0.107	0.053	0.068	0.032
Principle Component p								
0.3	0.2	0.14	0.113	0.1	0.12	0.167	0.107	0.147
0.4	0.213	0.173	0.14	0.14	0.16	0.213	0.16	0.167
0.5	0.24	0.207	0.193	0.153	0.173	0.22	0.193	0.187
Multiple Comparison p								
0.3	0.213	0.16	0.179	0.18	0.16	0.22	0.233	0.187
0.4	0.245	0.187	0.205	0.213	0.187	0.293	0.253	0.233
0.5	0.264	0.207	0.225	0.247	0.193	0.3	0.28	0.28

The table gives the results of the *type I error* based on sample sizes $N = 250$ and *300* for PRM, multiple comparison and principle component methods, with *minor allele frequencies* 0.3, 0.4 and 0.5 to simulate the distribution of the genetic group and number of responses q.

Table A1.5: Power - Part 1

q	50	100	150	200	50	100	150	200
p	N=150				N=200			
0.05	0.44	0.413	0.293	0.367	0.78	0.67	0.573	0.46
0.1	0.66	0.52	0.467	0.433	0.853	0.807	0.713	0.587
0.2	0.713	0.627	0.53	0.513	0.9	0.887	0.78	0.713
Principle Component p								
0.05	0.087	0.04	0.067	0.027	0.08	0.047	0.043	0.067
0.1	0.113	0.087	0.1	0.093	0.147	0.107	0.107	0.107
0.2	0.173	0.127	0.133	0.147	0.207	0.153	0.153	0.14
Multiple Comparison p								
0.05	0.133	0.093	0.06	0.073	0.2	0.2	0.12	0.133
0.1	0.3	0.273	0.247	0.219	0.44	0.47	0.384	0.38
0.2	0.645	0.553	0.57	0.503	0.788	0.769	0.656	0.709

The table gives the results of the *power* based on sample sizes $N = 150$ and *200* for PRM, multiple comparison and principle component methods, with *minor allele frequencies* 0.05, 0.1 and 0.2 to simulate the distribution of the genetic group and number of responses q.

Table A1.6: Power - Part 2

q	50	100	150	200	50	100	150	200
p	N=150				N=200			
0.3	0.753	0.707	0.593	0.56	0.94	0.907	0.813	0.773
0.4	0.807	0.72	0.627	0.593	0.967	0.92	0.86	0.813
0.5	0.827	0.727	0.653	0.607	0.967	0.927	0.88	0.8
Principle Component p								
0.3	0.227	0.167	0.18	0.167	0.26	0.187	0.193	0.167
0.4	0.273	0.2	0.193	0.193	0.3	0.227	0.207	0.213
0.5	0.313	0.227	0.213	0.26	0.347	0.257	0.247	0.247
Multiple Comparison p								
0.3	0.834	0.768	0.795	0.695	0.894	0.94	0.848	0.9
0.4	0.94	0.887	0.868	0.788	0.967	0.967	0.907	0.954
0.5	0.96	0.907	0.914	0.868	0.987	0.974	0.968	0.967

The table gives the results of the *power* of sample sizes $N = 150$ and *200* for PRM, multiple comparison and principle component methods, with *minor allele frequencies* 0.3, 0.4 and 0.5 to simulate the distribution of the genetic group and number of responses q.

Table A1.7: Power - Part 3

q	50	100	150	200	50	100	150	200
p	N=250				N=300			
0.05	0.893	0.86	0.813	0.72	0.967	0.893	0.86	0.847
0.1	0.96	0.947	0.887	0.807	0.987	0.94	0.94	0.913
0.2	0.973	0.993	0.953	0.893	0.993	0.973	0.987	0.98
Principle Component p								
0.05	0.067	0.06	0.033	0.033	0.073	0.073	0.06	0.04
0.1	0.18	0.1	0.073	0.087	0.14	0.107	0.093	0.1
0.2	0.22	0.14	0.133	0.107	0.193	0.2	0.147	0.153
Multiple Comparison p								
0.05	0.268	0.153	0.193	0.14	0.32	0.293	0.225	0.18
0.1	0.533	0.467	0.53	0.353	0.662	0.647	0.523	0.513
0.2	0.914	0.834	0.887	0.775	0.947	0.927	0.94	0.894

The table gives the results of the *power* of sample sizes $N = 250$ and *300* for PRM, multiple comparison and principle component methods, with *minor allele frequencies* 0.05, 0.1 and 0.2 to simulate the distribution of the genetic group and number of responses q.

Table A1.8: Power - Part 4

q	50	100	150	200	50	100	150	200
p	N=250				N=300			
0.3	0.98	0.98	0.967	0.9	1	0.993	0.993	0.967
0.4	0.987	0.987	0.96	0.927	1	0.993	0.993	0.967
0.5	1	0.987	0.98	0.9	1	0.993	0.993	0.987
Principle Component p								
0.3	0.287	0.207	0.18	0.133	0.247	0.253	0.18	0.2
0.4	0.331	0.26	0.227	0.187	0.32	0.307	0.252	0.24
0.5	0.364	0.3	0.28	0.212	0.371	0.327	0.298	0.267
Multiple Comparison p								
0.3	0.987	0.96	0.947	0.921	1	0.98	0.987	0.973
0.4	0.993	0.986	0.993	0.967	1	1	1	1
0.5	1	1	1	0.993	1	1	1	1

The table gives the results of the *power* of sample sizes $N = 250$ and *300* for PRM, multiple comparison and principle component methods, with *minor allele frequencies* 0.3, 0.4 and 0.5 to simulate the distribution of the genetic group and number of responses q.

Table A1.9: P-values from PCA method to analyze neonatal data

SNP	P-value	SNP	P-value	SNP	P-value
rs4680	0.99	rs821616	0.79	rs6675281	0.95
rs35753505	0.99	rs6994992	0.99	rs9340799	0.79
rs2234693	0.92	rs6265	0.99	rs2270335	0.98

Appendix II

Supplementary Tables and Additional Results of FMEM

Table A2.1: Descriptive statistics of **SNRs** for 10 regions of $\sigma_\gamma^2(v)$ from a simulated data of *scenario I*, in which the SNPs are extracted from the chromosome 1 in ADNI.

$\sigma_\gamma^2(v)$	Number of voxels	Mean	Standard Deviation	Min	Max
0	4749	0.124	0.062	0.024	0.412
0.0050	207	0.819	0.572	0.003	2.745
0.0075	135	1.010	0.774	0.006	3.340
0.0100	111	1.123	0.815	0.021	3.714
0.0125	147	1.270	1.014	0.043	4.303
0.0150	75	1.262	1.026	0.025	4.074
0.0175	48	1.334	1.150	0.034	5.930
0.0200	144	1.436	1.060	0.019	3.940
0.0225	111	1.687	1.191	0.036	4.886
0.0250	81	1.877	1.366	0.020	6.674

Table A2.2: Descriptive statistics of **SNRs** for 10 regions of $\sigma^2_\gamma(v)$ from a simulated data set of *scenario II*, in which the SNPs are extracted from the gene PICALM in ADNI.

$\sigma^2_\gamma(v)$	Number of voxels	Mean	Standard Deviation	Min	Max
0	4749	0.111	0.055	0.022	0.357
0.005	207	0.189	0.136	0.001	0.699
0.010	135	0.234	0.188	0.001	0.854
0.015	111	0.301	0.246	0.003	1.092
0.020	147	0.349	0.274	0.003	1.307
0.025	75	0.337	0.282	0.004	1.111
0.030	48	0.365	0.257	0.0005	0.996
0.035	144	0.424	0.283	0.003	1.324
0.040	111	0.491	0.443	0.010	1.008
0.045	81	0.500	0.403	0.034	1.909

Table A2.3: The minor allele frequency (MAF) in % of selected SNPs on TOMM40, PICALM, CR1 and CD2AP

PICALM							
rs618679	22.1	rs2077815	22.6	rs666692	48.9	rs527162	22.1
rs10898427	21.2	rs11234495	24.8	rs664629	37.0	rs680119	40.7
rs510566	22.9	rs10501602	8.8	rs1774523	19.4	rs10501608	17.2
rs10501604	17.4	rs713346	25.2	rs1941375	27.3	rs10792821	21.0
rs475639	48.0	rs669336	18.8	rs677909	31.6	rs642949	37.5
rs7938033	42.7	rs10792820	25.3	rs11234532	10.2		

CR1				**CD2AP**			
rs1571344	13.7	rs2025935	31.5	rs9296559	26.9	rs9473121	32.8
rs4310446	18.1	rs11117959	19.3	rs12523687	10.0	rs9395267	10.4
rs10127904	29.4	rs650877	19.9	rs1385741	39.0	rs9296562	40.1
rs3737002	27.5	rs11118131	19.0	rs6458573	32.9	rs13191654	22.4
rs677066	21.9	rs6691117	21.8	rs3818866	26.3	rs9395285	26.6
rs3818361	19.7	rs6701713	19.7	rs6936622	11.0	rs1485780	26.9
rs12734030	17.9	rs12034383	41.0	rs1485781	32.9	rs9349417	26.9
rs1408077	19.1	rs10779339	49.7	rs10948367	26.9		

Table A2.4: Demographic information of the 328 subjects in the dataset investigating the effects of **PICALM**. For the *categorical* variables: gender, handedness and risk of APOE, the numbers are the frequencies for the corresponding groups. For the *continuous* variables: baseline age, baseline ICV and years of education, the numbers are the mean (standard deviation) for the corresponding groups.

		Healthy Contol	Alzheimer's Disease
Total		176	152
Male (Female)		99 (77)	84(68)
Right (Left) Hand Users		163 (13)	142(10)
Risk of APOE	3	2	31
	2	41	68
	1	111	49
	0	22	4
Baseline Age		76.1 (5.1)	75.5 (7.6)
Baseline ICV		1.28×10^6 (1.24×10^5)	1.27×10^6 (1.45×10^5)
Years of Education		16.3 (2.7)	14.9 (3.2)

Table A2.5: Demographic information of the 299 subjects in the dataset investigating the effects of **CD2AP**. For the *categorical* variables: gender, handedness and risk of APOE, the numbers are the frequencies for the corresponding groups. For the *continuous* variables: baseline age, baseline ICV and years of education, the numbers are the mean (standard deviation) for the corresponding groups.

		Healthy Contol	Alzheimer's Disease
Total		155	144
Male (Female)		84 (71)	77(67)
Right (Left) Hand Users		143 (12)	135(9)
Risk of APOE	3	3	38
	2	37	64
	1	97	49
	0	18	3
Baseline Age		76.1 (5.2)	75.1 (7.5)
Baseline ICV		1.27×10^6 (1.25×10^5)	1.27×10^6 (1.46×10^5)
Years of Education		16.1 (2.8)	14.7 (3.1)

Table A2.6: The detailed significant brain regions affected by **CD2AP** using FMEM. The regions with the $*$ means the regions are detected by FMEM and voxel-based method; the regions with the $**$ means only detected by voxel-based method; $\bullet$ means not applicable.

Brain Region	Number of Voxels (right)	Number of Voxels (left)
right left Superior temporal gryus	167	$\bullet$
Inferior temporal gryus	*369	378
Precentral gryus	*170	439
Middle frontal gryus	127	*295
Postcentral gryus	60	183
Insula	53	74
Putamen	*200	123
Fusiform	117	306
Inferior parietal but supramarginal and angular gyri	87	$\bullet$
Angular	433	$\bullet$
Inferior frontal gyrus - triangular part	*180	*79
Inferior occipital gyrus	269	$\bullet$
Superior frontal gyrus	71	$\bullet$
Supplementary Motor Area	117	$\bullet$
Postcentral gryus	*104	*183
Superior frontal gyrus and medial	188	159
Anterior cingulate and paracingulate gyri	428	88
Median cingulate and paracingulate gyri	149	*116
Calcarine fissure and surrounding cortex	53	266
Cuneus	151	345
Superior occipital gyrus	$\bullet$	317
Middle occipital gyrus	$\bullet$	*144
Precuneus	61	$\bullet$
Paracentral Lobule	86	$\bullet$
Pallidum	$\bullet$	58
Caudate	$\bullet$	197
Lingual	$\bullet$	*114
Inferior frontal gyrus - opercular part	$\bullet$	*78
Inferior occipital gyrus	$\bullet$	133
Middle temporal gyrus	$\bullet$	394
Inferior frontal gyrus - orbital part	$\bullet$	61
Temporal pole: superior temporal gryus	**74	$\bullet$
Temporal pole: middle temporal gryus	**61	$\bullet$

Table A2.7: The detailed significant brain regions affected by **CR1** using FMEM. The regions with the $*$ means the regions are detected by FMEM and voxel-based method; the regions with the $**$ means only detected by voxel-based method; $\bullet$ means not applicable.

Brain Region	Number of Voxels (right)	Number of Voxels (left)
Superior temporal gryus	167	$\bullet$
Inferior temporal gryus	150	378
Precentral gyrus	170	*439
Middle frontal gyrus	127	295
Postcentral gryus	60	**65
Insula	53	74
Putamen	200	*123
Fusiform	117	306
Inferior temporal gryus	219	$\bullet$
Inferior parietal but supramarginal and angular gyri	87	$\bullet$
Angular	433	$\bullet$
Inferior frontal gyrus, triangular part	180	*79
Inferior occipital gyrus	269	133
Superior frontal gryus	71	$\bullet$
Supplementary Motor Area	117	$\bullet$
Postcentral gryus	104	183
Superior frontal gyrus, medial	188	159
Anterior cingulate and paracingulate gyri	428	88
Median cingulate and paracingulate gyri	209	116
Calcarine fissure and surrounding cortex	53	266
Cuneus	151	345
Superior occipital gyrus	$\bullet$	317
Middle occipital gyrus	144	$\bullet$
Precuneus	61	$\bullet$
Paracentral Lobule	86	$\bullet$
Pallidum	58	$\bullet$
Caudate	$\bullet$	197
Lingual	$\bullet$	114
Inferior frontal gyrus, opercular part	$\bullet$	78
Middle temporal gryus	$\bullet$	*394
Inferior frontal gyrus, orbital part	$\bullet$	61

Table A2.8: The detailed significant brain regions affected by **PICALM** using FMEM. The regions with the $*$ means the regions are detected by FMEM and voxel-based method; the regions with the $**$ means only detected by voxel-based method; $\bullet$ means not applicable.

Brain Region	Number of Voxels (right)	Number of Voxels (left)
Inferior frontal gyrus - triangular part	206	$\bullet$
Inferior frontal gyrus - orbital part	213	$\bullet$
Insula	363	$\bullet$
Hippocampus	72	$\bullet$
ParaHippocampal gyrus	72	$**$60
Amygdala	68	$\bullet$
Fusiform	205	$\bullet$
Putamen	144	$\bullet$
Superior temporal gyrus	790	$\bullet$
Temporal pole: superior temporal gryus	472	$\bullet$
Middle temporal gyrus	441	$\bullet$
Temporal pole: middle temporal gryus	127	$\bullet$
Inferior temporal gyrus	402	$\bullet$
Precentral gryus	519	$\bullet$
Middle frontal gryus	588	$\bullet$
Postcentral gryus	404	$\bullet$
SupraMarginal gryus	350	$\bullet$
Superior occipital gyrus	72	$\bullet$
Middle occipital gyrus	834	$\bullet$
Inferior parietal but supramarginal and angular gyri	239	$\bullet$
Angular	491	$\bullet$
Rolandic Operculum	76	$\bullet$
Inferior frontal gyrus - opercular part	112	$\bullet$
Superior frontal gyrus - orbital part	75	$\bullet$
Middle frontal gyrus	375	$\bullet$
Middle frontal gyrus - orbital part	338	$\bullet$
Superior frontal gyrus	66	$\bullet$
Median cingulate and paracingulate gyri	$\bullet$	$**$51

Appendix III

Proof of (3.11)

It is shown that the REML estimation of covariance is invariant to the choice of K_x (Searle et al, 1992, Ch.6). Without lost of generality, we may assume $K_x K_x^T = \mathbf{I}_{n-p}$, where $\mathbf{I}_{n-p}$ is an $(n-p)\times(n-p)$ identity matrix. Also, since the matrix $K_x \mathbf{Z}^T \mathbf{I}_L \mathbf{Z} K_x^T = \Omega$ is semi-positive, there exists an $(n-p)\times(n-p)$ orthogonal matrix U and a diagonal matrix $D_0 = \mathrm{diag}\{d_1, \ldots, d_{n-p}\}$ such that $\Omega = U D_0 U^T$, where the columns of U are the eigenvectors of Ω and $d_i, i = 1, \cdots, n-p$ are the corresponding eigenvalues. Under the null hypothesis $\sigma_\gamma^2(v) = 0$, $\mathbf{Y}^*(v')$ is distributed as $N(0, \sigma_e^2(v')\mathbf{I}_{n-p})$. Therefore, we have the following results:

- $\sigma_e(v')^{-1} U^T \mathbf{Y}^*(v') \overset{d}{\sim} N(\mathbf{0}_{n-p}, \mathbf{I}_{n-p})$ and $\frac{1}{\sigma_e(v')\sqrt{1+\lambda(v')d_i}} u_i^T \mathbf{Y}^*(v') \overset{d}{\sim} N(0, \frac{1}{1+\lambda(v')d_i})$, where u_i is the ith eigenvector of Ω. Then, we have

$$\frac{1}{\sigma_e(v')}\{\mathbf{I}_{n-p} + \lambda(v')D_0\}^{-\frac{1}{2}} U^T \mathbf{Y}^*(v') \overset{d}{\sim} N(\mathbf{0}_{n-p}, [\mathbf{I}_{n-p} + \lambda(v')D_0]^{-1}). \tag{1}$$

- The determinant of $\sigma_e^2(v')\{\mathbf{I}_{n-p} + \lambda(v')\Omega\}$ equals $\sigma_e(v')^{2(n-p)} \prod_{i=1}^{n-p}\{1 + \lambda(v')d_i\}$.

- Since the actual dimension of Ω is at most L, we have $d_i = 0$ for $i = L+1, \ldots, n-p$.

Then, $\ell_{REML}(\mathbf{Y}^*(v') \mid \mathbf{Z}, \sigma_e^2(v'), \sigma_\gamma^2(v) > 0) - \ell_{REML}(\mathbf{Y}^*(v') \mid \mathbf{Z}, \sigma_e^2(v'), \sigma_\gamma^2(v) = 0)$ can be approximated by

$$\begin{aligned}
&\frac{1}{2\sigma_e^2(v')}(\mathbf{Y}^*(v')^T\mathbf{Y}^*(v') - \mathbf{Y}^*(v')^T\{U[\mathbf{I}_{n-p} + \lambda(v')\Omega]U^T\}^{-1}\mathbf{Y}^*(v')) - \frac{1}{2}\sum_{i=1}^{L}\log(1+\lambda(v')d_i) \\
&\overset{\mathcal{D}}{=} \frac{1}{2}\{\sum_{i=1}^{n-p} z_i^2(v') - \sum_{i=1}^{L}\log(1+\lambda(v')d_i) - \sum_{i=1}^{n-p}\frac{z_i^2(v')}{1+\lambda(v')d_i}\} \\
&\overset{\mathcal{D}}{=} \frac{1}{2}\{\sum_{i=1}^{L}\frac{z_i^2(v')\lambda(v')d_i}{1+\lambda(v')d_i} - \sum_{i=1}^{L}\log(1+\lambda(v')d_i)\} = \frac{1}{2}\, D(v'),
\end{aligned} \tag{2}$$

where $z_i^2(v')$'s are mutually independent for all $i = 1, \cdots n-p$ and normally distributed with mean 0 and variance 1. Finally, we have

$$\begin{aligned}
\mathrm{RLRT}_n(v) &= 2 \sup_{\sigma_\gamma^2(v)\geq 0} \{L_{REML}(\sigma_\gamma^2(v) \mid \mathbf{Y}^*, B(v,h)) - L_{REML}(0 \mid \mathbf{Y}^*, B(v,h))\} \\
&= 2 \sup_{\sigma_\gamma^2(v)\geq 0} \sum_{v'\in B(v,h)} \omega(v,v',h)\{\ell_{REML}(\mathbf{Y}^*(v') \mid \mathbf{Z}, \hat{\sigma}_e^2(v'), \sigma_\gamma^2(v) > 0) \\
&\qquad - \ell_{REML}(\mathbf{Y}^*(v') \mid \mathbf{Z}, \hat{\sigma}_e^2(v'), \sigma_\gamma^2(v) = 0)\} \\
&\stackrel{\mathcal{D}}{=} \sup_{\sigma_\gamma^2(v)\geq 0} \{ \sum_{v'\in B(v,h)} \omega(v,v',h) D(v')\}.
\end{aligned} \tag{3}$$

Wild Bootstrap Procedure

Let $\Sigma_{Y^*}(v) = \sigma_\gamma^2(v)\Omega + \sigma_e^2(v)\mathbf{I}_{n-p}$, we transform $\mathbf{Y}^*(v)$ into $\widetilde{\mathbf{Y}}^*(v) = U^T\Sigma_{Y^*}(v)^{-1/2}\mathbf{Y}^*(v)$ such that $\widetilde{\mathbf{Y}}^*(v) \sim N(\mathbf{0}, \mathbf{I}_{n-p})$. It follows from $\Omega = UD_0U^T$ that

$$\widetilde{\mathbf{Y}}^*(v) = (\tilde{y}_1(v), \cdots, \tilde{y}_{n-p}(v))^T = \{\sigma_e^2(v)\mathbf{I}_{n-p} + \sigma_\gamma^2(v)D\}^{-1/2}U^T\mathbf{Y}^*(v). \tag{4}$$

We develop a wild bootstrap procedure as follows.

- Step 1. Define a grid $0 = \lambda_1(v) < \lambda_2(v) < \cdots < \lambda_M(v)$ of M possible values for $\lambda(v)$.
- Step 2. Generate $\xi_l^{(s)}$ independently from $N(0,1)$ for $l = 1, \ldots, n-p$.
- Step 3. For every grid point $\lambda_m(v)$, calculate an approximation of $D(v'; \lambda_m(v)/\hat{\sigma}_e^2(v))$ as follows:

$$\begin{aligned}
\widehat{D}^{(s)}(v'; \lambda_m(v)/\hat{\sigma}_e^2(v)) &= \sum_{l=1}^{n-p} \frac{\xi_l^{(s)}\{\tilde{y}_l^2(v') - 1\} d_l\lambda_m(v)/\hat{\sigma}_e^2(v') + d_l\lambda_m(v)/\hat{\sigma}_e^2(v')}{1 + d_l\lambda_m(v)/\hat{\sigma}_e^2(v')} \\
&- \sum_{l=1}^{n-p} \log(1 + d_l\lambda_m(v)/\hat{\sigma}_e^2(v')).
\end{aligned} \tag{5}$$

Remark that (5) is slightly different from (3.12) in terms of random samples gen-

eration. In (3.12), the samples are generated independently among voxels which does not take the spatial correlation in imaging measurements into consideration. To address this limitation, in (5), the samples are generated from original imaging data instead.

- Step 4. Calculate

$$\mathrm{RLRT}_n^{(s)}(v) = \sup_{1\le m\le M}\{\sum_{v'\in B(v,h)} \omega(v,v';h)\widehat{D}^{(s)}(v';\lambda_m(v)/\hat{\sigma}_e^2(v'))\}. \tag{6}$$

- Step 5. Repeat Steps 1-5 for S iterations.

- Step 6. Approximate the p-value of $\mathrm{RLRT}_n(v)$ based on an approximation given by

$$\mathrm{RLRT}_n(v) \overset{D}{\approx} a_0 u_0 d_0, \tag{7}$$

where $u_0 \sim Bernoulli(1-p_0)$ and $d_0 \sim \chi_1^2$ are independent random variables, $p_0 = P(u_0 = 0)$ and a_0 are unknown constants, and $\overset{D}{\approx}$ denotes an approximate equality in distribution. Considering computational efficiency, instead of generating many bootstrap samples, we estimate the parameters a_0 and p_0 in the approximation (7) as follows:

$$\hat{a}_0 = \sum_{v'\in B(v,h^S)} \omega(v,v';h^S)\tilde{a}_0 \quad , \quad \hat{p}_0 = \sum_{v'\in B(v,h^S)} \omega(v,v';h^S)\tilde{p}_0, \tag{8}$$

where

$$\tilde{p}_0 = 1 - 3\frac{\{\sum_{s=1}^S \mathrm{RLRT}_n^{(s)}(v)\}^2}{S\sum_{s=1}^S \mathrm{RLRT}_n^{(s)}(v)^2} \quad \text{and} \quad \tilde{a}_0 = \frac{\sum_{s=1}^S RLRT_n^{(s)}(v)}{S(1-\hat{p})}. \tag{9}$$

Computationally, the above algorithm is quite efficient. First, we only need to compute $\tilde{\mathbf{Y}}^*(v)$ once for all voxels. Second, Ω is fixed and thus it is straightforward to

compute its eigenvalues and eigenfunctions. Third, all $\hat{\sigma}_e^2(v)$ are calculated once at Step (I.1). Fourth, to choose grid points of $\lambda(v)$, we set $M = 50$ and $\lambda_M(v) = 3 \max_{v' \in \mathcal{V}} \hat{\sigma}_\gamma^2(v')$.

Appendix IV

Derivation of Weighted Score Test of FMEM

Let $\mathbf{b}$ be a $p \times 1$ multivariate normal random vector with mean μ and covariance matrix I_p, where I_p is a $p \times p$ identity matrix. Then, for two semi-positive matrices B_1 and B_2, we have

$$E(b^T B_1 b \; b^T B_2 b) = E(b^T B_1 b) E(b^T B_2 b) + 4\mu^T B_1 B_2 \mu + 2tr(B_1 B_2). \tag{10}$$

Let v' and v'' are two different voxels in $B(v, h_v)$. Stack $\mathbf{Y}(v')$ and $\mathbf{Y}(v'')$ together into $\mathbf{Y}^T_{v'.v''} = (\mathbf{Y}(v'), \mathbf{Y}(v"))^T$ and we denote $Cov(\mathbf{Y}_{v'.v''}) = \Sigma_{v',v''}$. Then under the null hypothesis, we have

$$\mathbf{Y}_{v',v''} \sim N_{2n} \left(\begin{bmatrix} \mathbf{X} & 0_{n\times p} \\ 0_{n\times p} & \mathbf{X} \end{bmatrix} \begin{bmatrix} \beta(v') \\ \beta(v'') \end{bmatrix}, \; \Sigma_{v',v''} \right). \tag{11}$$

By transforming $\mathbf{Y}_{v',v''}$ to $\tilde{\mathbf{Y}}_{v',v''}$, we have

$$\tilde{\mathbf{Y}}_{v',v''} = \begin{bmatrix} \tilde{\mathbf{Y}}(v') \\ \tilde{\mathbf{Y}}(v'') \end{bmatrix} = \Sigma^{-1/2} \begin{bmatrix} \mathbf{Y}(v') \\ \mathbf{Y}(v'') \end{bmatrix} \sim N_{2n} \left(\Sigma^{-1/2}_{v',v''} \begin{bmatrix} \mathbf{X} & 0_{n\times p} \\ 0_{n\times p} & \mathbf{X} \end{bmatrix} \begin{bmatrix} \beta(v') \\ \beta(v'') \end{bmatrix}, \; I_n \right). \tag{12}$$

We choose B_1 and B_2 as follows:

$$B_1 = \Sigma^{1/2}_{v',v''} \begin{bmatrix} I_n & 0_n \\ 0_n & 0_n \end{bmatrix} \begin{bmatrix} \Omega_X & 0_n \\ 0_n & 0_n \end{bmatrix} \begin{bmatrix} I_n & 0_n \\ 0_n & 0_n \end{bmatrix} \Sigma^{1/2}_{v',v''}, \tag{13}$$

$$B_2 = \Sigma^{1/2}_{v',v''} \begin{bmatrix} 0_n & 0_n \\ 0_n & I_n \end{bmatrix} \begin{bmatrix} 0_n & 0_n \\ 0_n & \Omega_X \end{bmatrix} \begin{bmatrix} 0_n & 0_n \\ 0_n & I_n \end{bmatrix} \Sigma^{1/2}_{v',v''}, \tag{14}$$

Therefore, we can obtain $\mathbf{Y}^T(v')\Omega_X \mathbf{Y}(v') = \tilde{\mathbf{Y}}^T_{v',v''} B_1 \tilde{\mathbf{Y}}^T_{v',v''}$ and $\mathbf{Y}^T(v'')\Omega_X \mathbf{Y}(v'') =$

$\tilde{\mathbf{Y}}^T_{v',v''} B_2 \tilde{\mathbf{Y}}^T_{v',v''}$. Hence, applying (10) with the fact that $\Omega_X \mathbf{X}\beta(v) = 0$, we have

$$E_{H_{0,\gamma}}\{(\mathbf{Y}^T(v')\Omega_X \mathbf{Y}(v'))^2\} = tr(\Omega_X)^2 + 2\ tr(\Omega_X^2) \tag{15}$$

$$E_{H_{0,\gamma}}(\mathbf{Y}^T(v')\Omega_X \mathbf{Y}(v')\mathbf{Y}^T(v'')\Omega_X \mathbf{Y}(v'')) = tr(B_1)tr(B_2) + 2\ tr(B_1 B_2) = \Psi_{v',v''} \tag{16}$$

According to the assumption under the null hypothesis that

$$Cov(\mathbf{Y}_{v',v''}) = \begin{bmatrix} \sigma_e^2(v') I_n & \rho_{v',v''}\sigma_e(v')\sigma_e(v'') I_n \\ \rho_{v',v''}\sigma_e(v')\sigma_e(v'') I_n & \sigma_e^2(v'') I_n \end{bmatrix} = \Sigma_{v',v''}, \tag{17}$$

(16) can be simplified as

$$\Psi_{v',v''} = \hat{\sigma}_e^2(v')\hat{\sigma}_e^2(v'')tr(\Omega_X)^2 + 2\ \hat{\rho}^2_{v',v''}\hat{\sigma}_e^2(v')\hat{\sigma}_e^2(v'')\ tr(\Omega_X^2). \tag{18}$$

Finally, we can get (4.17) based on the above fact.

Bibliography

Bao, Y. and Ullah, A. (2010). Expectation of quadratic forms in normal and nonnormal variables with econometric applications. *Journal of Statistical Planning and Inference*, 140:1193–1205.

Braskie, M. and et al, N. J. (2011). Common alzheimer's disease risk variant within the clu gene affects white matter microstructure in young adults. *Journal of Neuroscience*, 31:6764–6770.

Chun, H. and Keles, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J. Roy. Statist. Soc. Ser. B*, 72:3–25.

Chung, M., Dalton, K., and Davidson, R. (2008). Tensor-based cortical surface morphometry via weighed spherical harmonic representation. *IEEE Transactions on Medical Imaging*, 27:1143–1151.

Chung, M. K., Worsley, K. J., Nacewicz, B. M., Dalton, K. M., and Davidson, R. J. (2010). General multivariate linear modeling of surface shapes using surfstat. *NeuroImage*, 53:491–505.

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

Davatzikos, C., Genc, A., Xu, D., and Resnick, S. (2001). Voxel-based morphometry using the ravens maps: Methods and validation using simulated longitudinal atrophy. *NeuroImage*, 14:1361–1369.

Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimen- sional feature space. *Statistica Sinica*, 20:101–148.

Formisano, E., Martino, F. D., and Valente, G. (2008). Multivariate analysis of fmri time series: classification and regression of brain responses using machine learning. *Magnetic Resonance Imaging*, 26:921 – 934.

Friston, K. J. (2007). *Statistical Parametric Mapping: the Analysis of Functional Brain Images*. Academic Press, London.

Ge, T. and et al, J. F. (2012). Increasing power for voxel-wise genome-wide association studies: The random field theory, least square kernel machines and fast permutation procedures. *Neuroimage*, 63:858–873.

Gilmore, J. H., Lin, W., Prastawa, M., Looney, C. B., Vetsa, Y. S. K., Knickmeyer, R. C., Evans, D. D., Smith, J. K., Hamer, R. M., Lieberman, J., and Gerig, G. (2007). Regional gray matter growth, sexual dimorphism, and cerebral asymmetry in the neonatal brain. *Journal of Neuroscience*, 27:1255–1260.

Harold, D., Abraham, R., Hollingworth, P., and et al, R. S. (2009). Genome-wide association study identifies variants at clu and picalm associated with alzheimer's disease. *Nat Genet*, 41:1088–1093.

Heller, R., Golland, Y., Malach, R., and Benjaminia, Y. (2007). Conjunction group analysis: an alternative to mixed/random effect analysis. *Neuroimage*, 37:1178–1185.

Hibar, D. and et al, J. L. S. (2011). Voxelwise gene-wide association study (vgenewas): Multivariate gene-based association testing in 731 elderly subjects. *Neuroimage*, 56:1875âĂŞ1891.

Huettel, S. A., Song, A. W., and McCarthy, G. (2004). *Functional Magnetic Resonance Imaging*. Sinauer Associates, Inc, London.

Kang, H. M., J. H, S., Service, S. K., Zaitlen, N. A., Kong, S., Freimer, N. B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42:348–354.

Kherif, F., Poline, J. B., Flandin, G., Benali, H., Simon, O., Dehaene, S., and Worsley, K. (2002). Multivariate model specification for fmri data. *Neuroimage*, 16:1068–1083.

Klei, L., Luca, D., Devlin, B., and Roeder, K. (2008). Pleiotropy and principle components of heritability combine to increase power for association. *Genetic Epidemiology*, 32:9–19.

Knickmeyer, R. C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J., Hamer, R., Lin, W., Gerig, G., and Gilmore, J. (2008). A structural mri study of human brain development from birth to 2 years. *J Neurosci.*, 28:12176–12182.

Krishnan, A., Williams, L. J., McIntosh, A. R., and Abdi, H. (2011). Partial least squares (pls) methods for neuroimaging: a tutorial and review. *NeuroImage*, 56:455–475.

Lambert, J., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M. J., Tavernier, B., Letenneur, L., Bettens, K., Berr, C., Pasquier, F., and et al (2009). Genome-wide association study identifies variants at clu and cr1 associated with alzheimer's disease. *Nature Genetics*, 41:1094–1099.

Lange, C., van Steen, K., Andrew, T., Lyon, H., DeMeo, D. L., Raby, B., Murphy, A., Silverman, E. K., MacGregor, A., Weiss, S. T., and Laird, N. M. (2004). A family-based association test for repeatedly measured quantitative traits adjusting for unknown environmental and/or polygenic effects. *Stat Appl Genet Mol Biol*, 3:1–17.

Lazar, N., Luna, B., Sweeney, J., and Eddy, W. (2002). Combiningbrains: a survey of methods for statistical pooling of information. *NeuroImage*, 16:538–550.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365–411.

Leng, C. and Wang, H. (2009). On general adaptive sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 18:201–215.

Lenroot, R. and Giedd, J. (2006). Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. *Neurosci Biobehav Rev.*, 30:718–729.

Li, Y., Zhu, H., Shen, D., Lin, W., Gilmore, J. H., and Ibrahim, J. G. (2011). Multiscale adaptive regression models for neuroimaging data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73:559–578.

Liu, D., Lin, X., and Ghosh, D. (2007). Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics*, 63:1079–1088.

M. Silver, G. Montana, T. E. N. and ADNI (2011). False positives in neuroimaging genetics using voxel-based morphometry data. *NeuroImage*, 54:992–1000.

Mata, I., Perez-Iglesias, R., Roiz-Santianez, R., Tordesillas-Gutierrez, D., Gonzalez-Mandly, A., Vazquez-Barquero, J. L., and Crespo-Facorro, B. (2009). A neuregulin 1 variant is associated with increased lateral ventricle volume in patients with first-episode schizophrenia. *Biological Psychiatry*, 65:535–540.

Mukhopadhyay, I., Feingold, E., Weeks, D. E., and Thalamuthu, A. (2010). Association tests using kernel-based measures of multi-locus genotype similarity between individuals. *Genetic Epidemiology*, 34:213–221.

Naj, A. C., Jun, G., Beecham, G. W., Wang, L., Vardarajan, B. N., Buros, J., Gallins, P. J., Buxbaum, J. D., Jarvik, G. P., Crane, P. K., Larson, E. B., and et al., T. D. B. (2011). Common variants at ms4a4/ms4a6e, cd2ap, cd33 and epha1 are associated with late-onset alzheimer's disease. *Nature Genetics*, 43:436–441.

Ott, J. and Rabinowitz, D. (1999). A principle-components approach based on heritability for combining phenotype information. *Hum Heredity*, 49:106–111.

Polzehl, J. and Spokoiny, V. G. (2000). Adaptive weights smoothing with applications to image restoration. *J. R. Statist. Soc. B*, 62:335–354.

Potkin, S. G., Guffanti, G., Lakatos, A., Turner, J. A., and et al, F. K. (2009). Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for alzheimer's disease. *PLoS ONE*, 40:e6501.

Purcell, S. and et al, B. N. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81:559–575.

Rowe, D. and Hoffmann, R. (2006). Multivariate statistical analysis in fmri. *IEEE Eng Med Biol Med*, 25:60–64.

Shen, L., Kim, S., Risacher, S. L., Nho, K., Swaminathan, S., West, J., Foroud, T. M., Pankratz, N. D., Moore, J. H., Sloan, S. D., Huentelman, M. J., Craig, D. W., DeChairo, B. M., Potkin, S. G., Jack, C. R., Weiner, M. W., Saykin, A. J., and ADNI. (2010). Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in mci and ad: A study of the adni cohort. *Neuroimage*, 53:1051âĂŞ1063.

Stein, J., Hua, X., Lee, S., Ho, A. J., Leow, A. D., Toga, A. W., Saykin, A. J., Shen, L., Foroud, T., Pankratz, N., Huentelman, M. J., Craig, D. W., Gerber, J. D., Allen, A. N., Corneveaux, J. J., Dechairo, B. M., Potkin, S. G., Weiner, M. W., Thompson, P. M., and ADNI. (2010). Voxelwise genome-wide association study (vgwas). *Neuroimage*, 53:1160âĂŞ1174.

Stram, D. and Lee, J. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50:1171–1177.

Styner, M., Gerig, G., Lieberman, J., Jones, D., and Weinberger, D. (2003). Statistical shape analysis of neuroanatomical structures based on medial models. *Medical Image Analysis*, 3:207–220.

Styner, M., Lieberman, J., Pantazis, D., and Gerig, G. (2004). Boundary and medial shape analysis of the hippocampus in schizophrenia. *Medical Image Analysis*, 4:197–203.

Tabelow, K., Polzehl, J., Voss, H. U., and Spokoiny, V. (2006). Analyzing fmri experiments with structural adaptive smoothing procedures. *NeuroImage*, 33:55–62.

Taylor, J. and Worsley, K. (2008). Random fields of multivariate test statistics, with applications to shape analysis. *Annals of Statistics*, 36:1–27.

Teipel, S. J., Born, C., Ewers, M., Bokde, A. L. W., Reiser, M. F., Ḿoller, H. J., and Hampel, H. (2007). Multivariate deformation-based analysis of brain atrophy to predict alzheimer's disease in mild cognitive impairment. *NeuroImage*, 38:13–24.

Tzeng, J. and Zhang, D. (2007). Haplotype-based association analysis via variance component score test. *The American Journal of Human Genetics*, 81:927–938.

Vounou, M., Nichols, T. E., Montana, G., and the Alzheimer's Disease Neuroimaging Initiative (2010). Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *Neuroimage*, 53:1147âĂŞ1159.

Wang, Y. and Chen, H. (2012). On testing a variance component in a linear mixed effects model with multiple variance components. *Biometrics*.

Worsley, K. J., Taylor, J. E., Tomaiuolo, F., and Lerch, J. (2004). Unified univariate and multivariate random field theory. *NeuroImage*, 23:189–195.

Yang, Q., Wu, H., Guo, C., and Fox, C. S. (2010). Analyze multivariate phenotypes in genetic association studies by combining univeriate association tests. *Genetic Epidemiology*, 34:444–454.

Zhu, H. T., Ibrahim, J. G., Tang, N., Rowe, D., Hao, X., Bansal, R., and Peterson, B. S. (2007). A statistical analysis of brain morphology using wild bootstrapping. *IEEE Trans Med Imaging*, 26:954–966.

Zou, H., Hastie, T., and Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15:262–286.