

Groupwise Envelope Models for Imaging Genetic Analysis

Yeonhee Park^{1,*}, Zhihua Su^{2,**} and Hongtu Zhu^{1,***}

¹Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, U.S.A.

²Department of Statistics, University of Florida, Gainesville, FL 32611, U.S.A.

**email*: ypark3@mdanderson.org

***email*: zhihuasu@stat.ufl.edu

****email*: hzhu5@mdanderson.org

SUMMARY: Motivated by searching for associations between genetic variants and brain imaging phenotypes, the aim of this paper is to develop a groupwise envelope model for multivariate linear regression in order to establish the association between both multivariate responses and covariates. The groupwise envelope model allows for both distinct regression coefficients and distinct error structures for different groups. Statistically, the proposed envelope model can dramatically improve efficiency of tests and of estimation. Theoretical properties of the proposed model are established. Numerical experiments as well as the analysis of an imaging genetic data set obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study show the effectiveness of the model in efficient estimation. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

KEY WORDS: Dimension reduction, Envelope model, Grassmann manifold, Reducing subspace.

1. Introduction

To motivate the proposed methodology, we consider an imaging genetic data set from 745 subjects collected by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) study (<http://www.adni-info.org/>) in order to advance the discovery in detection, invention, prevention as well as treatments of the Alzheimer’s disease. Specifically, each subject has brain volumes of 93 regions of interest, single nucleotide polymorphisms (SNPs) on candidate genes of the Alzheimer’s Disease, and other covariates including gender, age, education level, marital status and handedness. Alzheimer’s Disease (AD) is characterized by death of nerve cells and accelerated cerebral atrophy, leading to the shrinkage of various brain volumes, such as hippocampus. Similar to a recent large-scale imaging study for schizophrenia in Franke et al. (2016), we are interested in characterizing the genetic influences of the top 40 AD candidate genes listed on the AlzGene database (www.alzgene.org) as of June 10, 2010 on structural brain phenotypes in ADNI.

A standard model in the imaging genetic literature (Vounou et al., 2010; Thompson et al., 2013; Sun et al., 2015) is the multivariate linear regression given by

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon}, \quad (1)$$

where \mathbf{Y} is a $r \times 1$ vector of multiple responses (e.g., brain volumes), \mathbf{X} is a $p \times 1$ vector of covariates, and the errors $\boldsymbol{\varepsilon}$ follows a distribution with mean $\mathbf{0}$ and positive definite covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{r \times r}$. Moreover, $\boldsymbol{\mu} \in \mathbb{R}^r$ and $\boldsymbol{\beta} \in \mathbb{R}^{r \times p}$ are unknown intercept and regression coefficients. It is common to calculate the ordinary least squares estimator of $\boldsymbol{\beta}$ by regressing each element in \mathbf{Y} on the predictors of interest. This method, however, ignores the relationship among different response components. A novel envelope modeling framework introduced in Cook et al. (2010) explicitly uses such relationship to identify a part of the responses that is immaterial to the

estimation of β , while bringing extraneous variation. This immaterial part is then accounted for in the subsequent estimation, making the estimation more efficient. After the original development, advances have been taken place to extend the scope of envelope model (Su and Cook, 2011, 2012, 2013; Cook et al., 2013; Cook and Zhang, 2015; Khare et al., 2016).

However, model (1) is not sufficient for addressing a more specific question of interest. Specifically, it is interesting to investigate how the associations between AD genetic variants and subcortical volume measures differ across male and female groups. Suppose that we observe imaging genetic data from subjects in L different groups. For each $l = 1, \dots, L$, the l -th group has $n_{(l)}$ observations and the total sample size is $n = \sum_{l=1}^L n_{(l)}$. By incorporating such group information, we can reformulate model (1) as

$$\mathbf{Y}_{(l)j} = \boldsymbol{\mu}_{(l)} + \boldsymbol{\beta}_{(l)}\mathbf{X}_{(l)j} + \boldsymbol{\varepsilon}_{(l)j} \text{ for } l = 1, \dots, L \text{ and } j = 1, \dots, n_l, \quad (2)$$

where $\mathbf{Y}_{(l)j} \in \mathbb{R}^r$ is the j th observed response vector in the l -th population, $\boldsymbol{\mu}_{(l)} \in \mathbb{R}^r$ is the mean of the l -th population, $\mathbf{X}_{(l)j} \in \mathbb{R}^p$ is the j th observed covariate vector in the l -th population, $\boldsymbol{\beta}_{(l)} \in \mathbb{R}^{r \times p}$ contains the regression coefficients for the l -th population, and $\boldsymbol{\varepsilon}_{(l)j}$ follows some distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{(l)}$. Throughout this paper, we use subscript (l) to denote the l -th population and subscripts without parenthesis to number the observations. Without loss of generality, we assume that $\mathbb{X}_{(l)} = (\mathbf{X}_{(l)1}^T, \dots, \mathbf{X}_{(l)n_{(l)}}^T)^T$, $l = 1, \dots, L$ are centered at $\mathbf{0}$ in the sample for each group. Model (2) is referred to as the standard model in later discussion.

The aim of this paper is to develop a new groupwise envelope modelling framework for model (2), which allows for distinct regression coefficients and the heteroscedastic error structure across groups. Compared with the existing literature (Su and Cook, 2013; Cook et al., 2010), we make at least three major contributions. First, we develop an efficient estimation method to estimate

distinct genetic-volume associations across groups under the heteroscedastic error structure. In contrast, the existing envelope models assume either a homogenous error structure or distinct means across groups without covariate \mathbf{X} . Second, we examine the asymptotic properties of the proposed estimates under some mild conditions. Third, our simulation studies and the ADNI data analysis confirm the efficiency gains obtained by using the groupwise envelope model. An alternative way to gain estimation efficiency in model (2) is to fit a separate envelope model to each group. However, as will be shown in Sections 3 and 5, efficiency is lost due to ignoring the common characteristics in response variables across groups.

The article is organized as follows. Section 2 reviews the envelope model and introduces the groupwise envelope model and its estimation procedure. Section 3 systematically investigates the asymptotic properties of all estimators. Simulation studies are conducted in Section 4. A real data analysis of the imaging genetic data set from ADNI is described in Section 5. Conclusion remarks are given in Section 6.

2. Methods

2.1 A Review of Envelope Models

We first introduce some notation. We use $\mathbf{P}_{\mathbf{S}}$ to denote the projection matrix onto $\text{span}(\mathbf{S})$ or \mathbf{S} if \mathbf{S} is a matrix or a subspace, and $\mathbf{Q}_{\mathbf{S}} = \mathbf{I} - \mathbf{P}_{\mathbf{S}}$. With a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\text{vec}(\mathbf{A}) \in \mathbb{R}^{mn}$ stacks the columns of \mathbf{A} into a vector. The Kronecker product is denoted by \otimes , $\mathbf{X} \sim \mathbf{Y}$ means \mathbf{X} and \mathbf{Y} has the same distribution, and $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ means that \mathbf{X} and \mathbf{Y} are independent.

The original envelope model (Cook et al., 2010) was developed for model (1). Under (1), we partition the response vector \mathbf{Y} into a material part and an immaterial part, where the distribution of the material part changes with the predictor \mathbf{X} and the distribution of the immaterial part does not.

More specifically, let \mathcal{S} be a subspace of \mathbb{R}^r , \mathbf{L} be an orthogonal basis of \mathcal{S} and \mathbf{L}_0 be an orthogonal basis of \mathcal{S}^\perp . The linear combinations of the responses $\mathbf{L}^T \mathbf{Y}$ and $\mathbf{L}_0^T \mathbf{Y}$ are called the material part and the immaterial part if the following two conditions are satisfy: (a) $\mathbf{L}_0^T \mathbf{Y} \mid \mathbf{X} \sim \mathbf{L}_0^T \mathbf{Y}$ and (b) $\text{cov}(\mathbf{L}^T \mathbf{Y}, \mathbf{L}_0^T \mathbf{Y} \mid \mathbf{X}) = \mathbf{0}$. Condition (a) indicates that the distribution of the immaterial part does not depend on \mathbf{X} , and condition (b) indicates that given \mathbf{X} , the material part and immaterial part are uncorrelated. Let $\mathcal{B} = \text{span}(\boldsymbol{\beta})$. Conditions (a) and (b) are also equivalent to: (I) $\mathcal{B} \subseteq \mathcal{S}$ and (II) $\boldsymbol{\Sigma} = \mathbf{P}_\mathcal{S} \boldsymbol{\Sigma} \mathbf{P}_\mathcal{S} + \mathbf{Q}_\mathcal{S} \boldsymbol{\Sigma} \mathbf{Q}_\mathcal{S}$ (Cook et al., 2010). Condition (I) indicates that the immaterial part does not contain information on $\boldsymbol{\beta}$ and condition (II) indicates that the variation $\boldsymbol{\Sigma}$ can be decomposed into the variation due to the material part and the variation due to the immaterial part. When $\boldsymbol{\Sigma}$ has the structure in condition (II), \mathcal{S} is a reducing subspace of $\boldsymbol{\Sigma}$ (Conway, 1990). Then, the $\boldsymbol{\Sigma}$ -envelope of \mathcal{B} , denoted by $\mathcal{E}_\boldsymbol{\Sigma}(\mathcal{B})$, is defined to be the smallest reducing subspace of $\boldsymbol{\Sigma}$ containing \mathcal{B} . Model (1) is called the envelope model when conditions (I) and (II) are imposed.

Let u denote the dimension of $\mathcal{E}_\boldsymbol{\Sigma}(\mathcal{B})$, $\boldsymbol{\Gamma} \in \mathbb{R}^{r \times u}$ be an orthogonal basis of $\mathcal{E}_\boldsymbol{\Sigma}(\mathcal{B})$, and $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times (r-u)}$ be an orthogonal basis of $\mathcal{E}_\boldsymbol{\Sigma}(\mathcal{B})^\perp$. The coordinate form of the envelope model is

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Gamma} \boldsymbol{\eta} \mathbf{X} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 = \boldsymbol{\Gamma} \boldsymbol{\Omega} \boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^T,$$

where $\boldsymbol{\beta} = \boldsymbol{\Gamma} \boldsymbol{\eta}$, $\boldsymbol{\eta} \in \mathbb{R}^{u \times p}$ carries the coordinates of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\Gamma}$, and $\boldsymbol{\Omega} = \boldsymbol{\Gamma}^T \boldsymbol{\Sigma} \boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}_0 = \boldsymbol{\Gamma}_0^T \boldsymbol{\Sigma} \boldsymbol{\Gamma}_0$ carry the coordinates of $\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$, respectively. When $u = r$, $\mathcal{E}_\boldsymbol{\Sigma}(\mathcal{B}) = \mathbb{R}^r$, the envelope model degenerates to the standard multivariate linear regression model. As shown in Cook et al. (2010), the envelope estimator of $\boldsymbol{\beta}$ is more efficient than or at least as efficient as the standard estimator. The efficiency gains can be substantial when $\|\boldsymbol{\Sigma}_1\| \ll \|\boldsymbol{\Sigma}_2\|$, where $\|\cdot\|$ denotes the spectral norm of a matrix or vector.

2.2 Formulation of Groupwise Envelop Model

Under model (2), let \mathcal{S} be a subspace of \mathbb{R}^r , \mathbf{L} be an orthogonal basis of \mathcal{S} and \mathbf{L}_0 be an orthogonal basis of \mathcal{S}^\perp . Then for each l , the material part $\mathbf{L}^T \mathbf{Y}_{(l)j}$ and the immaterial part $\mathbf{L}_0^T \mathbf{Y}_{(l)j}$ should satisfy (A) $\mathbf{L}_0^T \mathbf{Y}_{(l)j} \mid \mathbf{X}_{(l)} \sim \mathbf{L}_0^T \mathbf{Y}_{(l)j}$ and (B) $\mathbf{L}^T \mathbf{Y}_{(l)j} \perp \mathbf{L}_0^T \mathbf{Y}_{(l)j} \mid \mathbf{X}_{(l)}$. From conditions (A) and (B), \mathcal{S} is a reducing subspace of all $\Sigma_{(l)}$ and $\text{span}(\beta_{(l)}) \subseteq \mathcal{S}$ for $l = 1, \dots, L$. Therefore, we define the groupwise envelope to be the intersection of all such \mathcal{S} . More specifically, let $\mathcal{M} = \{\Sigma_{(1)}, \dots, \Sigma_{(L)}\}$ denote the collection of all covariance matrices, and $\mathcal{B} = \text{span}(\beta_{(1)}, \dots, \beta_{(L)})$. Then the \mathcal{M} -envelope of \mathcal{B} , denoted by $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$, is the smallest subspace that reduces each matrix in \mathcal{M} and contains \mathcal{B} . When $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$ appears in subscripts, it is shortened to \mathcal{E} . From the definition of $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$, we have

$$\text{span}(\beta_{(l)}) \subseteq \mathcal{E}_{\mathcal{M}}(\mathcal{B}) \quad \text{and} \quad \Sigma_{(l)} = \mathbf{P}_{\mathcal{E}} \Sigma_{(l)} \mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}} \Sigma_{(l)} \mathbf{Q}_{\mathcal{E}}, \quad \text{for } l = 1, \dots, L. \quad (3)$$

Model (2) is called the groupwise envelope model if conditions in (3) are imposed.

Let $\Gamma \in \mathbb{R}^{r \times u}$ be an orthogonal basis of $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$ and $\Gamma_0 \in \mathbb{R}^{r \times (r-u)}$ be its completion. The coordinate form of the groupwise envelope model is given by

$$\mathbf{Y}_{(l)j} = \boldsymbol{\mu}_{(l)} + \Gamma \boldsymbol{\eta}_{(l)} \mathbf{X}_{(l)j} + \boldsymbol{\varepsilon}_{(l)j} \quad \text{and} \quad \Sigma_{(l)} = \Gamma \boldsymbol{\Omega}_{(l)} \Gamma^T + \Gamma_0 \boldsymbol{\Omega}_0 \Gamma_0^T \quad (4)$$

for each $l = 1, \dots, L$, where $\beta_{(l)} = \Gamma \boldsymbol{\eta}_{(l)}$, $\boldsymbol{\eta}_{(l)} \in \mathbb{R}^{u \times p}$ carries the coordinate of $\beta_{(l)}$ with respect to Γ , and $\boldsymbol{\Omega}_{(l)} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(r-u) \times (r-u)}$ are symmetric matrices that carry the coordinates of $\Sigma_{(l)}$ with respect to Γ and Γ_0 , respectively. The groupwise envelope model degenerates to the envelope model in Cook et al. (2010) if $L = 1$.

For a fixed dimension u , the number of parameters in the groupwise envelope model (4) is $N(u) = Lr + Lup + Lu(u+1)/2 + (r-u)(r-u+1)/2 + u(r-u)$. This is because we need Lr parameters for all $\boldsymbol{\mu}_{(l)}$'s, Lup parameters for all $\boldsymbol{\eta}_{(l)}$'s, $Lu(u+1)/2$ parameters for all $\boldsymbol{\Omega}_{(l)}$'s, and

$(r - u)(r - u + 1)/2$ parameters for Ω_0 . The envelope subspace $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$ is on an $r \times u$ Grassmann manifold, which is the set of all u dimensional subspaces in an r dimensional space, so it has $u(r - u)$ free parameters.

2.3 Estimation Procedure

The groupwise envelope model does not require normality, but we will use the normal likelihood function as a pseudo likelihood function to calculate estimators. Technical details are included in Supplemental Section A. Let $\theta = (\mu, \eta, \Omega, \Omega_0)$ be a collection of parameters, where $\mu = (\mu_{(1)}, \dots, \mu_{(L)})$, $\eta = (\eta_{(1)}, \dots, \eta_{(L)})$, and $\Omega = (\Omega_{(1)}, \dots, \Omega_{(L)})$. For a fixed dimension u , $u = 0, \dots, r$, the normal log likelihood of the groupwise envelope model is given by

$$\begin{aligned} \ell(\theta) &= -\frac{nr}{2} \log(2\pi) - \frac{n}{2} \log |\Omega_0| - \frac{1}{2} \sum_{l=1}^L n_{(l)} \log |\Omega_{(l)}| \\ &\quad - \frac{1}{2} \sum_{l=1}^L \sum_{j=1}^{n_{(l)}} \{ \Gamma^T (\mathbf{Y}_{(l)j} - \mu_{(l)} - \Gamma \eta_{(l)} \mathbf{X}_{(l)j}) \}^T \Omega_{(l)}^{-1} \{ \Gamma^T (\mathbf{Y}_{(l)j} - \mu_{(l)} - \Gamma \eta_{(l)} \mathbf{X}_{(l)j}) \} \\ &\quad - \frac{1}{2} \sum_{l=1}^L \sum_{j=1}^{n_{(l)}} (\mathbf{Y}_{(l)j} - \mu_{(l)})^T \Gamma_0 \Omega_0^{-1} \Gamma_0^T (\mathbf{Y}_{(l)j} - \mu_{(l)}). \end{aligned} \quad (5)$$

When Γ is fixed, the estimators of $\mu_{(l)}$, $\eta_{(l)}$, $\Omega_{(l)}$, and Ω_0 , which maximize $\ell(\theta)$, can be written as explicit expressions of Γ . Let $\hat{\Sigma}_{\text{res},(l)} = (1/n_{(l)}) \mathbb{Y}_{(l)c}^T \mathbf{Q}_{\mathbb{X}_{(l)}} \mathbb{Y}_{(l)c}$, and $\hat{\Sigma}_{\mathbf{Y}} = (1/n) \sum_{l=1}^L \mathbb{Y}_{(l)c}^T \mathbb{Y}_{(l)c}$, where $\mathbb{X}_{(l)} \in \mathbb{R}^{n_{(l)} \times p}$ is the centered data matrix for \mathbf{X} and $\mathbb{Y}_{(l)c} \in \mathbb{R}^{n_{(l)} \times r}$ is the centered data matrix for \mathbf{Y} for group l . Substitute them back to $\ell(\theta)$, we get

$$\hat{\mathcal{E}}_{\mathcal{M}}(\mathcal{B}_{\Gamma}) = \underset{\text{span}(\Gamma) \in \mathcal{L}_{r \times u}}{\text{argmin}} \sum_{l=1}^L \frac{n_{(l)}}{n} \log |\Gamma^T \hat{\Sigma}_{\text{res},(l)} \Gamma| + \log |\Gamma^T \hat{\Sigma}_{\mathbf{Y}}^{-1} \Gamma|,$$

where $\mathcal{L}_{r \times u}$ denotes the $r \times u$ Grassmann manifold. To emphasize that Γ is an orthogonal basis of $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$, we put subscript Γ on \mathcal{B} . After we obtained $\hat{\mathcal{E}}_{\mathcal{M}}(\mathcal{B}_{\Gamma})$, $\hat{\Gamma}$ can be any orthogonal basis of $\hat{\mathcal{E}}_{\mathcal{M}}(\mathcal{B}_{\Gamma})$. For $l = 1, \dots, L$, the estimators for all other parameters are given as follows:

- $\hat{\boldsymbol{\mu}}_{(l)} = \bar{\mathbf{Y}}_{(l)}$, where $\bar{\mathbf{Y}}_{(l)} = (1/n_{(l)}) \sum_{j=1}^{n_{(l)}} \mathbf{Y}_{(l)j}$;
- $\hat{\boldsymbol{\eta}}_{(l)} = \hat{\boldsymbol{\Gamma}}^T (\mathbf{Y}_{(l)c}^T \mathbf{X}_{(l)}) (\mathbf{X}_{(l)}^T \mathbf{X}_{(l)})^{-1}$;
- $\hat{\boldsymbol{\Omega}}_{(l)} = \hat{\boldsymbol{\Gamma}}^T \hat{\boldsymbol{\Sigma}}_{\text{res},(l)} \hat{\boldsymbol{\Gamma}}$;
- $\hat{\boldsymbol{\Omega}}_0 = \hat{\boldsymbol{\Gamma}}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}} \hat{\boldsymbol{\Gamma}}_0$, where $\hat{\boldsymbol{\Gamma}}_0$ is the completion of $\hat{\boldsymbol{\Gamma}}$;
- $\hat{\boldsymbol{\beta}}_{(l)} = \hat{\boldsymbol{\Gamma}} \hat{\boldsymbol{\eta}}_{(l)} = \mathbf{P}_{\hat{\boldsymbol{\Gamma}}} \hat{\boldsymbol{\beta}}_{(l),\text{ols}}$, where $\hat{\boldsymbol{\beta}}_{(l),\text{ols}}$ is the ordinary least squares estimator of $\boldsymbol{\beta}_{(l)}$;
- $\hat{\boldsymbol{\Sigma}}_{(l)} = \hat{\boldsymbol{\Gamma}} \hat{\boldsymbol{\Omega}}_{(l)} \hat{\boldsymbol{\Gamma}}^T + \hat{\boldsymbol{\Gamma}}_0 \hat{\boldsymbol{\Omega}}_0 \hat{\boldsymbol{\Gamma}}_0^T$.

To estimate the dimension of $\mathcal{E}_{\mathcal{M}}(\mathcal{B}_{\boldsymbol{\Gamma}})$, we apply the Bayesian information criterion (BIC). Let $l^*(u)$ be the maximized l for a fixed u , and $N(u)$ be the number of parameters discussed in Section 2.2. We choose a value u_{opt} that minimizes $\text{BIC}(u) = -2l^*(u) + \log(n)N(u)$.

3. Theoretical Properties

In this section, we examine the theoretical properties of the groupwise envelope estimators. We present the following theoretical properties, whose proofs are included in Supplemental Section B.

PROPOSITION 1: Under the groupwise envelope model (4), assume that the errors are independent and have finite fourth moments. Then $\hat{\boldsymbol{\beta}}_{(l)}$ is a \sqrt{n} consistent estimator of $\boldsymbol{\beta}_{(l)}$ and $\hat{\boldsymbol{\Sigma}}_{(l)}$ is a \sqrt{n} consistent estimator of $\boldsymbol{\Sigma}_{(l)}$ for $l = 1, \dots, L$.

Proposition 1 establishes the \sqrt{n} consistency of the groupwise envelope estimators. Notice that normality is not required even though the estimators are derived by maximizing the normal likelihood function.

PROPOSITION 2: Suppose that the conditions of Proposition 1 hold and $f_{(l)} = n_{(l)}/n$ does not change with n . Then, $\text{vec}(\hat{\boldsymbol{\beta}}_{(l)} - \boldsymbol{\beta}_{(l)})$ converges in distribution to a multivariate normal distribution

with mean $\mathbf{0}$ for each $l = 1, \dots, L$. Furthermore, under the normality, we have

$$\sqrt{n}[\text{vec}(\widehat{\boldsymbol{\beta}}_{(l)}) - \text{vec}(\boldsymbol{\beta}_{(l)})] \xrightarrow{d} N(0, \mathbf{V}_{(l)}),$$

where \xrightarrow{d} denotes convergence in distribution and $\mathbf{V}_{(l)} = f_{(l)}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}_{(l)}}^{-1} \otimes \boldsymbol{\Gamma} \boldsymbol{\Omega}_{(l)} \boldsymbol{\Gamma}^T + (\boldsymbol{\eta}_{(l)}^T \otimes \boldsymbol{\Gamma}_0) \mathbf{T}^{-1} (\boldsymbol{\eta}_{(l)} \otimes \boldsymbol{\Gamma}_0^T)$ for $l = 1, \dots, L$, in which \mathbf{T} is given by

$$\mathbf{T} = \sum_{l=1}^L f_{(l)} (\boldsymbol{\eta}_{(l)} \boldsymbol{\Sigma}_{\mathbf{X}_{(l)}} \boldsymbol{\eta}_{(l)}^T \otimes \boldsymbol{\Omega}_0^{-1} + \boldsymbol{\Omega}_{(l)} \otimes \boldsymbol{\Omega}_0^{-1} + \boldsymbol{\Omega}_{(l)}^{-1} \otimes \boldsymbol{\Omega}_0 - 2\mathbf{I}_u \otimes \mathbf{I}_{r-u}).$$

Proposition 2 provides the asymptotic distribution of the groupwise envelope estimator and derives explicit form of the asymptotic variance under the normality. The first term $f_{(l)}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}_{(l)}}^{-1} \otimes \boldsymbol{\Gamma} \boldsymbol{\Omega}_{(l)} \boldsymbol{\Gamma}^T$ is the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{(l)}$ for known $\boldsymbol{\Gamma}$, and the second term $(\boldsymbol{\eta}_{(l)}^T \otimes \boldsymbol{\Gamma}_0) \mathbf{T}^{-1} (\boldsymbol{\eta}_{(l)} \otimes \boldsymbol{\Gamma}_0^T)$ is the cost of estimating the envelope subspace $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$.

Suppose that we fit separate envelope model to the data in each group, and we denote their estimators as $\widehat{\boldsymbol{\beta}}_{(l),\text{seiv}}$ for $l = 1, \dots, L$. We have the following results.

PROPOSITION 3: Suppose that the conditions of Proposition 1 hold, then $\text{vec}(\widehat{\boldsymbol{\beta}}_{(l),\text{seiv}}) - \text{vec}(\boldsymbol{\beta}_{(l)})$ converges in distribution to a multivariate normal distribution with mean $\mathbf{0}$ for $l = 1, \dots, L$. Furthermore, under the normality, we have

$$\sqrt{n}[\text{vec}(\widehat{\boldsymbol{\beta}}_{(l),\text{seiv}}) - \text{vec}(\boldsymbol{\beta}_{(l)})] \xrightarrow{d} N(0, \mathbf{V}_{(l),\text{seiv}}),$$

where $\mathbf{V}_{(l),\text{seiv}} = f_{(l)}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}_{(l)}}^{-1} \otimes \boldsymbol{\Gamma} \boldsymbol{\Omega}_{(l)} \boldsymbol{\Gamma}^T + (\boldsymbol{\eta}_{(l)}^T \otimes \boldsymbol{\Gamma}_0) \mathbf{T}_2^{-1} (\boldsymbol{\eta}_{(l)} \otimes \boldsymbol{\Gamma}_0^T)$ for $l = 1, \dots, L$, in which $\mathbf{T}_2 = f_{(l)} (\boldsymbol{\eta}_{(l)} \boldsymbol{\Sigma}_{\mathbf{X}_{(l)}} \boldsymbol{\eta}_{(l)}^T \otimes \boldsymbol{\Omega}_0^{-1} + \boldsymbol{\Omega}_{(l)} \otimes \boldsymbol{\Omega}_0^{-1} + \boldsymbol{\Omega}_{(l)}^{-1} \otimes \boldsymbol{\Omega}_0 - 2\mathbf{I}_u \otimes \mathbf{I}_{r-u})$.

COROLLARY 4: $\mathbf{V}_{(l),\text{seiv}} \geq \mathbf{V}_{(l)}$ for $l = 1, \dots, L$.

Proposition 3 gives the asymptotic distribution of $\text{vec}(\widehat{\boldsymbol{\beta}}_{(l),\text{seiv}})$. Corollary 4 indicates that the groupwise envelope estimator is more efficient. A close examination reveals that $\mathbf{V}_{(l)}$ and $\mathbf{V}_{(l),\text{seiv}}$

differ only in terms of \mathbf{T} and \mathbf{T}_2 . This suggests that if $\mathbf{\Gamma}$ is known, the asymptotic variance for $\hat{\boldsymbol{\eta}}_{(l)}$ is the same. However, the cost of estimating $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$ is smaller for the groupwise envelope model. This is because that the groupwise envelope model uses all the data to estimate $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$, whereas the separate envelope model only uses the data from the l -th group to estimate $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$. We also notice that with finite sample, the envelope subspace $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$ calculated by fitting separate envelope model varies across groups.

4. Simulation Study

In this section, we use Monte Carlo simulations to evaluate the finite-sample performance of the groupwise envelope model (4). We generated the data from model (4) with two groups ($L = 2$), which have 40% and 60% of the observations. We set $r = 10$, $p = 3$, and $u = 1$. The matrix $(\mathbf{\Gamma}, \mathbf{\Gamma}_0)$ was obtained by normalizing an $r \times r$ matrix of independent normal variates, $\boldsymbol{\mu}_{(1)}$ was a vector of 3 and $\boldsymbol{\mu}_{(2)}$ was a vector of 10, $\boldsymbol{\eta}_{(1)}$ was a vector of independent χ_{10}^2 variates and $\boldsymbol{\eta}_{(2)}$ was a vector of independent χ_{20}^2 variates. Let $\mathbf{A} \in \mathbb{R}^{(r-u) \times (r-u)}$ be a matrix of independent normal $(5, 1^2)$ variates, $\Omega_{(1)}$ and $\Omega_{(2)}$ both be χ_1^2 variates, and $\mathbf{\Omega}_0 = \mathbf{A}\mathbf{A}^T$. The predictors were independent normal $(0, 5^2)$ variates for the first group and independent normal $(0, 10^2)$ variates for the second group. We varied the sample size from 100, 300, 1000 and 3000. For each sample size, 200 replications were generated. The standard model (2), the envelope model (Cook et al., 2010), separate envelope model and the groupwise envelope model (4) were fit to the data. Standard deviation of each element in $\boldsymbol{\beta}_{(1)}$ and $\boldsymbol{\beta}_{(2)}$ was calculated based on the 200 replications for each method at each sample size. We also computed the bootstrap standard deviations of each element in $\boldsymbol{\beta}_{(1)}$ and $\boldsymbol{\beta}_{(2)}$ based on 200 bootstrap samples. The results for a randomly chosen element in $\boldsymbol{\beta}_{(1)}$ and a randomly chosen element in $\boldsymbol{\beta}_{(2)}$ are summarized in Figure 1. For clarity, we did not

draw the line for the standard deviation of the standard model, but only displayed its asymptotic standard deviation.

[Figure 1 about here.]

We have the following observations from Figure 1. The groupwise envelope estimator is consistent and its standard deviation approaches to the asymptotic standard deviation as sample size increases, which agrees with Proposition 1. It is also observed that the groupwise envelope model achieves substantial efficiency gains over the standard model. Take the element in the right panel of Figure 1 for example, the standard deviation of the groupwise envelope estimator is already smaller than the asymptotic standard deviation of the standard estimator with $n = 100$. This means by using the groupwise envelope model, with 100 samples we have achieved the efficiency of taking infinity number of samples under the standard model. We also notice that the bootstrap standard deviation is a good estimation of the sample standard deviation. The envelope model with constant covariance structure (Cook et al., 2010) has standard deviation about five times as large as the groupwise envelope model, indicating that accommodating the groupwise error structure brings extra efficiency gains. The separate envelope model also has larger standard deviations than the groupwise envelope model, as asserted in Corollary 4. The difference is more pronounced in one group than the other.

We investigated the numerical properties of the groupwise envelope model under non-normal errors in Web Appendix A and Web Appendix B. In Web Appendix A, we considered the estimation standard deviation of the groupwise envelope model under different non-normal error distributions including t distribution with degrees of freedom 6, uniform distribution defined on the unit interval, and chi-squared distribution with degrees of freedom 4. We examined the selection performance of BIC under non-normal errors in Web Appendix B.

5. The Alzheimer's Disease Neuroimaging Initiative

We applied the groupwise envelope model to the imaging genetic data set obtained from ADNI study as described in Section 1. We used the image processing pipeline and quality control methods described in Zhu et al. (2014) to process the structural Magnetic Resonance Imaging (MRI) data and genetic data downloaded from the ADNI publicly available database (<http://adni.loni.usc.edu/>). Our problem of interest is to investigate the genetic effects of the SNPs on the top 40 AD candidate genes on the brain volumes of 93 regions of interest (ROI), whose names and abbreviation are given in the supplementary material in Zhu et al. (2014) across male and female groups. The top 40 AD candidate genes are listed on the AlzGene database (www.alzgene.org) as of June 10, 2010 on the brain volumes of 93 ROIs across male and female groups. The selection of the genes is described in Section 4.1 in Zhu et al. (2014). To correct for normal variation in head size, we used the proportion method to calculate ROI tissue-to-intracranial volume ratios and then took logarithm of these ratios. We selected the 1071 SNPs on the 37 top genes and Apolipoprotein E (APOE) $\epsilon 4$. We first performed a principal component analysis on the 1071 SNP predictors. Figure 2 presents the 1071×1071 correlation matrix of all 1071 SNPs. We selected 205 principal components (PCs) whose loadings have corresponding eigenvalues greater than 1, and these 205 PCs explain 89.76% of the total variation of all SNP predictors.

[Figure 2 about here.]

We fit a linear regression of the logarithm of the ratios for the brain volumes on five PCs obtained from the entire SNP data accounting for the population stratification. The residuals are used for multivariate responses in the groupwise envelop model (4). We used gender (male versus female) as the group variable and the SNP PCs and all covariates except gender as predictors. Note that the responses are unitless. The covariates include APOE $\epsilon 4$, age (years, from 55 to 91), education

level (years, from 6 to 20), marriage status (married, widowed, divorced, other) and handedness (right-handed, other). All variables were standardized. BIC suggests $u = 1$. For each element in $\beta_{(l)}$, we computed the ratio of the bootstrap standard deviation under the standard model versus the bootstrap standard deviation under the groupwise envelope model. As demonstrated in Figure 1, this ratio is a good approximation to the ratio of actual standard deviations. The ratios range from 4.56 to 136.80, with an average of 29.71 for the male group, and from 6.06 to 351.82 with an average of 56.99 for the female group. This indicates that the groupwise envelope model obtains substantial efficiency gains over the standard model (2). A boxplot of the ratios is given in Figure 3. The efficiency gains can be explained by the covariance structure: $\hat{\Omega}_{(1)} = 1.27 \times 10^{-4}$, $\hat{\Omega}_{(2)} = 0.1579$ and $\|\hat{\Omega}_0\| = 19.29$. This indicates that the variation of the immaterial part is much larger than that of the material part, and by identifying and accounting for the immaterial variation, the groupwise envelope model achieves substantial efficient estimation in this case. The efficiency gains in estimation lead to better prediction performance. The prediction error is estimated by the average of 50 five-fold cross validations with random splits, and the identity inner product is used to bind the responses. The standard model has a prediction error of 21.27, and the groupwise envelope model has a prediction error of 9.65, which is more than a 50% reduction.

[Figure 3 about here.]

We also fitted the separate envelope model to the data. BIC suggested $u = 1$ for both male and female groups. The ratios of the bootstrap standard deviation under the standard model to that under the separate envelope model range from 1.05 to 16.63 with an average of 3.16 for the male group, and range from 1.01 to 13.93 with average of 2.68 for the female group. The ratios are also displayed in Figure 3. The boxplot reveals that the groupwise envelope model and the separate envelope model are both more efficient than the standard model. But by using the information from

all the groups in the estimation of $\mathcal{E}_{\mathcal{M}}(\mathcal{B})$, the groupwise envelope model achieves substantially more efficiency gains than the separate envelope model, which confirms the results in Corollary 4.

To investigate the genetic effects of the SNPs on the brain volumes of 93 ROIs through the groupwise envelope model, we looked at a submatrix of $\widehat{\beta}$, say $\widehat{\beta}_{\text{sub}}$, which consisted of columns corresponding to the 205 PCs and APOE $\epsilon 4$. And we calculated $c_{j,i} = |\widehat{\beta}_{\text{sub},j,i}| / \|\widehat{\beta}_{\text{sub},j}\|$, where $\widehat{\beta}_{\text{sub},j,i}$ denotes the (j, i) th element of $\widehat{\beta}_{\text{sub}}$ and $\widehat{\beta}_{\text{sub},j}$ denotes the j th row of $\widehat{\beta}_{\text{sub}}$, $j = 1, \dots, 93$ and $i = 1, \dots, 206$. In particular, $j = 30$ and $j = 69$ correspond to the right and left part of hippocampal formation region, which are important ROIs. Figure 4 shows $c_{30,i}$ and $c_{69,i}$ for the hippocampal formation region. In particular, the 206th element in the horizontal axis corresponds to APOE $\epsilon 4$. We have $c_{30,206} = c_{69,206} = 0.077$ for the male group and $c_{30,206} = c_{69,206} = 0.110$ for the female group.

[Figure 4 about here.]

To consider the relationship between SNPs and ROI volumes in the coefficient matrix, we focused on the PCs with large coefficients based on the 98th quantile of the absolute values of the estimated regression coefficients. After thresholding, 42 PCs were selected for the male group and 49 PCs were selected for the female group. The heatmaps of regression coefficients for each group under the groupwise envelope model and the standard model are displayed in Figure 5. The horizontal axes in Figure 5 include the selected PCs as well as APOE $\epsilon 4$. Under the groupwise envelope model, it is easier to identify the regions with diminishing regression coefficients.

[Figure 5 about here.]

We also tested if a coefficient equals to 0 for all elements in β . An ROI is chosen if at least one of the p -values in the corresponding coefficients is smaller than 0.05. Based on this crite-

tion, all ROIs are selected under the standard model. Under the groupwise envelope model, 41 ROIs are selected for the male group. And eight additional ROIs *lat.f-o.gy.R*, *sup.f.gy.R*, *hiopp.R*, *caud.neuc.R*, *mid.t.gy.L*, *prec.gy.L*, *par.lb.WM.R*, and *ant.caps.R* are selected for female group, i.e. 49 ROIs are selected for female group. This indicates that the unselected ROIs are significant under the homogeneous model, but not under the model that considers heteroscedasticity. The selected ROIs are listed in Table 1. Among the selected ROIs, (i) *me.f-o.gy.L/R*, *lat.f-o.gy.R*, *inf.f.gy.L/R*, *f.lob.WM.L*, *sup.f.gy.L/R* and *me.f.gy.R* are related to the function of management (making a decision and carrying out tasks), attention (interest and concentration) and working memory; (ii) *tmp.pl.R*, *sup.gy.R*, *sup.t.gy.L/R*, *mid.t.gy.L/R*, *inf.t.gy.L* are related to memory and language; (iii) *sup.p.lb.L/R* and *par.lb.WM.R* are related to the sense of the space and size; and (iv) *sup.o.gy.L/R*, *mid.o.gy.R*, *me.o.gy.L/R* are located in the back of the brain playing an important role in vision.

[Table 1 about here.]

We then use the polygenic score to study the genetic relationship between the SNPs and the brain volumes on 93 ROIs. The genetic effects in our setting are contained in the 205 PCs obtained from the SNPs as well as APOE $\epsilon 4$. Let l be the gender indicator, $l = 1$ for male and $l = 2$ for female. Then for group l , $\mathbf{Y}_l = (Y_{l,1}, \dots, Y_{l,93})^T$ is the response vector and \mathbf{Z}_l denotes the vector of 205 PCs and all the covariates. For each response $Y_{l,j}$, $j = 1, \dots, 93$, we fit the linear regression model of $Y_{l,j}$ on \mathbf{Z}_l , and obtained the estimated coefficients $\hat{\beta}_{l,j}$ and the p-values for each element in $\hat{\beta}_{l,j}$. We set the significance level at 0.05. Suppose $\mathbf{Z}_{l,j}^*$ are the significant genetic effects and $\hat{\beta}_{l,j}^*$ are their coefficients, we can construct the polygenic score as $\hat{\mathbf{S}}_{l,j} = \hat{\beta}_{l,j}^* \mathbf{Z}_{l,j}^*$. We then fit the regression of $Y_{l,j}$ on $\hat{\mathbf{S}}_{l,j}$, and calculated the total sum of squares $\text{SST}_{l,j}$ and regression sum of squares $\text{SSR}_{l,j}$. We then computed $R_1 = \sum_{j=1}^{93} \text{SSR}_{1,j} / \sum_{j=1}^{93} \text{SST}_{1,j}$ and $R_2 = \sum_{j=1}^{93} \text{SSR}_{2,j} / \sum_{j=1}^{93} \text{SST}_{2,j}$. Based on the value of R_1 and R_2 , the genetic effects of SNPs explain 9.55% and 9.60% of the variation

in the male and female groups, respectively. The association of the polygenic score with the phenotypical variation is tested based on the chi-squared distribution with non-centrality parameter λ_l and degrees of freedom 1, where $\lambda_l = (n_{(l)}R_l^2)/(1 - R_l^2)$, $n_{(1)} = 441$ and $n_{(2)} = 304$. Then the power of the two-tailed chi-squared test with significance level α is obtained from

$$1 - \Phi[\Phi^{-1}(1 - \alpha/2) - \sqrt{\lambda_l}] + \Phi[\Phi^{-1}(\alpha/2) - \sqrt{\lambda_l}], \quad l = 1, 2,$$

where Φ is the cumulative distribution function for the standard normal distribution. We found both groups have power 1 with $\alpha = 0.05$ in the testing.

6. Conclusion

We have proposed a groupwise envelope model, which is an efficient model for estimating regression coefficients for heterogeneous groups. Since the interest of research in heterogeneity soars, such as the great attention in precision medicine, the development of models dealing with heterogeneity is desired in multivariate response analysis. The groupwise envelope model allows for distinct regression coefficients and heteroscedastic error structure for different groups. Our simulation studies and ADNI data analysis demonstrates the efficiency gains obtained by the groupwise envelope model, compared to both standard model and separate envelope model. The groupwise envelope model leads to a better understanding of the genetic effects of the top 40 AD candidate genes for male and female groups on brain volumes of 93 ROIs in the ADNI dataset.

For future research, a sparse groupwise envelope model that pinpoints the immaterial responses is desired for real applications, as it is more interpretable. If we consider the spatial structure of the brains, we would have a multi-dimensional array (tensor) response. Li and Zhang (2016) developed a tensor response envelope model that achieves efficient estimation in tensor regression. We can expand our methodology to handle heteroscedastic error structure in the tensor envelope model. In

addition, a Bayesian version of this model that incorporates prior information from earlier studies is also worth exploration. As longitudinal data and missing data appear in Alzheimer's study, a groupwise envelope model that can handle these data structures would also be of practical use.

SUPPLEMENTARY MATERIALS

Web Appendix A and Web Appendix B referenced in Section 4 are available with this paper at the Biometrics website on Wiley Online Library. An R package implementing the groupwise envelope model with simulating data is available at <https://github.com/BIG-S2/GENV>.

ACKNOWLEDGMENT

We are grateful to the editor and two referees for their insightful suggestion and comments that helped us improve the paper. We thank Professors Dennis Cook and Hani Doss for helpful discussions. This material was based upon work partially supported by the NSF grant DMS-1127914 to the Statistical and Applied Mathematical Science Institute. This work was partially supported by NIH grants MH086633 and MH092335, NSF grants SES-1357666, DMS-1407460, and DMS-1407655, and a grant from Cancer Prevention Research Institute of Texas. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

REFERENCES

Conway, J. (1990). *A course in functional analysis*. New York: Springer.

- Cook, R. D., Helland, I. S., and Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society: Series B* **75**, 851–877.
- Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression (with discussion). *Statistica Sinica* **20**, 927–1010.
- Cook, R. D. and Zhang, X. (2015). Foundations for envelope models and methods. *Journal of the American Statistical Association* **110**, 599–611.
- Franke, B., Stein, J. L., Ripke, S., Anttila, V., Hibar, D. P., Van Hulzen, K. J., Arias-Vasquez, A., Smoller, J. W., Nichols, T. E., Neale, M. C., et al. (2016). Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. *Nature neuroscience* .
- Khare, K., Pal, S., and Su, Z. (2016). A bayesian approach for envelope models. *Annals of Statistics* (to appear) .
- Li, L. and Zhang, X. (2016). Parsimonious tensor response regression. *Journal of the American Statistical Association* (to appear) .
- Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of the American Statistical Association* **81**, 142–149.
- Su, Z. and Cook, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika* **98**, 133–146.
- Su, Z. and Cook, R. D. (2012). Inner envelopes: efficient estimation in multivariate linear regression. *Biometrika* **99**, 687–702.
- Su, Z. and Cook, R. D. (2013). Estimation of multivariate means with heteroscedastic errors using envelope models. *Statistica Sinica* **23**, 213–230.
- Sun, Q., Zhu, H., Liu, Y., and Ibrahim, J. G. (2015). Sprem: sparse projection regression model for high-dimensional linear regression. *Journal of the American Statistical Association* **110**,

289–302.

Thompson, P. M., Ge, T., Glahn, D. C., Jahanshad, N., and Nichols, T. E. (2013). Genetics of the connectome. *NeuroImage* **80**, 475–488.

Vounou, M., Nichols, T. E., Montana, G., and ADNI (2010). Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *NeuroImage* **53**, 1147–1159.

Zhu, H., Khondker, Z., Lu, Z., and Ibrahim, J. G. (2014). Bayesian generalized low rank regression models for neuroimaging phenotypes and genetic markers. *Journal of the American Statistical Association* **109**, 977–990.

Received xxx xxx. Revised xxx xxx. Accepted xxx xxx.

APPENDIX

A: Estimation of the groupwise envelope model

we first hold Γ fixed and derive the estimators of parameters $\boldsymbol{\mu}_{(l)}$, $\boldsymbol{\eta}_{(l)}$, $\boldsymbol{\Omega}_{(l)}$, $l = 1, \dots, L$, and $\boldsymbol{\Omega}_0$ as a function of Γ . The derivative of the log likelihood $l = l(\boldsymbol{\theta})$ in (5) with respect to $\boldsymbol{\mu}_{(l)}$ is

$$\frac{\partial l}{\partial \boldsymbol{\mu}_{(l)}} = n_{(l)} \boldsymbol{\Gamma} \boldsymbol{\Omega}_{(l)}^{-1} (\boldsymbol{\Gamma}^T \bar{\mathbf{Y}}_{(l)} - \boldsymbol{\Gamma}^T \boldsymbol{\mu}_{(l)} - \boldsymbol{\eta}_{(l)} \bar{\mathbf{X}}_{(l)}) + n_{(l)} \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0^{-1} \boldsymbol{\Gamma}_0^T (\bar{\mathbf{Y}}_{(l)} - \boldsymbol{\mu}_{(l)}). \quad (\text{A.1})$$

Setting the derivative in (A.1) to $\mathbf{0}$ and using the fact that $\bar{\mathbf{X}}_{(l)}$ is centered at $\mathbf{0}$, we have $\hat{\boldsymbol{\mu}}_{(l)} = \bar{\mathbf{Y}}_{(l)}$.

Substitute $\hat{\boldsymbol{\mu}}_{(l)}$ into the likelihood, we now consider the derivative of l in (5) with respect to $\boldsymbol{\eta}_{(l)}$:

$$\frac{\partial l}{\partial \boldsymbol{\eta}_{(l)}} = \sum_{j=1}^{n_{(l)}} \boldsymbol{\Omega}_{(l)}^{-1} (\boldsymbol{\Gamma}^T \mathbf{Y}_{(l)j} - \boldsymbol{\Gamma}^T \bar{\mathbf{Y}}_{(l)} - \boldsymbol{\eta}_{(l)} \mathbf{X}_{(l)j}) \mathbf{X}_{(l)j}^T. \quad (\text{A.2})$$

We set the the derivative in (A.2) to be $\mathbf{0}$ and obtain

$$\hat{\boldsymbol{\eta}}_{(l)} = \left(\sum_{j=1}^{n_{(l)}} \boldsymbol{\Gamma}^T (\mathbf{Y}_{(l)j} - \bar{\mathbf{Y}}_{(l)}) \mathbf{X}_{(l)j}^T \right) \left(\sum_{j=1}^{n_{(l)}} \mathbf{X}_{(l)j} \mathbf{X}_{(l)j}^T \right)^{-1} = \boldsymbol{\Gamma}^T (\mathbf{Y}_{(l)c}^T \bar{\mathbf{X}}_{(l)}) (\bar{\mathbf{X}}_{(l)}^T \bar{\mathbf{X}}_{(l)})^{-1},$$

where $\mathbb{X}_{(l)}$ is an $n_{(l)} \times p$ matrix with its i th row being $\mathbf{X}_{(l),i}^T$, $\mathbb{Y}_{(l)c}$ is an $n_{(l)} \times r$ matrix with its i th row being $\mathbf{Y}_{(l),i}^T - \bar{\mathbf{Y}}_{(l)}^T$. Substitute $\hat{\boldsymbol{\mu}}_{(l)}$ and $\hat{\boldsymbol{\eta}}_{(l)}$ to the log likelihood function, we have

$$l = -\frac{nr}{2} \log(2\pi) - \frac{n}{2} \log |\boldsymbol{\Omega}_0| - \frac{1}{2} \sum_{l=1}^L n_{(l)} \log |\boldsymbol{\Omega}_{(l)}| \quad (\text{A.3})$$

$$- \frac{1}{2} \sum_{l=1}^L \text{tr}(\mathbf{Q}_{\mathbb{X}_{(l)}} \mathbb{Y}_{(l)c} \boldsymbol{\Gamma} \boldsymbol{\Omega}_{(l)}^{-1} \boldsymbol{\Gamma}^T \mathbb{Y}_{(l)c}^T \mathbf{Q}_{\mathbb{X}_{(l)}}) - \frac{1}{2} \sum_{l=1}^L \text{tr}(\mathbb{Y}_{(l)c} \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0^{-1} \boldsymbol{\Gamma}_0^T \mathbb{Y}_{(l)c}^T).$$

We can easily get the estimators of $\boldsymbol{\Omega}_{(l)}$ and $\boldsymbol{\Omega}_0$ by taking the derivatives, and the estimators are

$$\hat{\boldsymbol{\Omega}}_{(l)} = \frac{1}{n_{(l)}} \boldsymbol{\Gamma}^T \mathbb{Y}_{(l)c}^T \mathbf{Q}_{\mathbb{X}_{(l)}} \mathbb{Y}_{(l)c} \boldsymbol{\Gamma} = \boldsymbol{\Gamma}^T \hat{\boldsymbol{\Sigma}}_{\text{res},(l)} \boldsymbol{\Gamma} \quad \text{and} \quad \hat{\boldsymbol{\Omega}}_0 = \frac{1}{n} \sum_{l=1}^L \boldsymbol{\Gamma}_0^T \mathbb{Y}_{(l)c}^T \mathbb{Y}_{(l)c} \boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}} \boldsymbol{\Gamma}_0. \quad (\text{A.4})$$

Substitution of (A.4) into (A.3) gives

$$l = \text{constant} - \frac{1}{2} \sum_{l=1}^L n_{(l)} \log |\boldsymbol{\Gamma}^T \hat{\boldsymbol{\Sigma}}_{\text{res},(l)} \boldsymbol{\Gamma}| - \frac{n}{2} \log |\boldsymbol{\Gamma}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}} \boldsymbol{\Gamma}_0|. \quad (\text{A.5})$$

By Lemma 6.2 of Cook et al. (2010), we rewrite the function in (A.5) as

$$l(\boldsymbol{\Gamma}) = \text{constant} - \frac{n}{2} \log |\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}| - \frac{1}{2} \sum_{l=1}^L n_{(l)} \log |\boldsymbol{\Gamma}^T \hat{\boldsymbol{\Sigma}}_{\text{res},(l)} \boldsymbol{\Gamma}| - \frac{n}{2} \log |\boldsymbol{\Gamma}^T \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}^{-1} \boldsymbol{\Gamma}|.$$

Then the objective function for $\boldsymbol{\Gamma}$ is

$$\sum_{l=1}^L \frac{n_{(l)}}{n} \log |\boldsymbol{\Gamma}^T \hat{\boldsymbol{\Sigma}}_{\text{res},(l)} \boldsymbol{\Gamma}| + \log |\boldsymbol{\Gamma}^T \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}^{-1} \boldsymbol{\Gamma}|,$$

and $\hat{\boldsymbol{\Gamma}}$ can be obtained by minimizing the preceding objective function.

B: Proofs of theoretical results in Section 3

Proof of Proposition 1 and Proposition 2 For preparation, if $\mathbf{A} \in \mathbb{R}^{m \times m}$ is a symmetric matrix, $\text{vech}(\mathbf{A}) \in \mathbb{R}^{m(m+1)/2}$ denotes the vector that stacks the lower triangle of \mathbf{A} into a vector. The notations $\mathbf{E}_m \in \mathbb{R}^{m^2 \times m(m+1)/2}$ and $\mathbf{C}_m \in \mathbb{R}^{m(m+1)/2 \times m^2}$ are expansion and contraction operators that connect vec and vech : $\text{vec}(\mathbf{A}) = \mathbf{E}_m \text{vech}(\mathbf{A})$ and $\text{vech}(\mathbf{A}) = \mathbf{C}_m \text{vec}(\mathbf{A})$.

We use Proposition 4.1 in Shapiro (1986) to prove Proposition 1. Let \mathbf{h} denote the parameters under the standard model, and let ϕ denote the parameters under the groupwise envelope model. Then $\mathbf{h} = \text{vec}[\{\boldsymbol{\mu}_{(1)}^T, \dots, \boldsymbol{\mu}_{(L)}^T, \text{vec}^T(\boldsymbol{\beta}_{(1)}), \dots, \text{vec}^T(\boldsymbol{\beta}_{(L)}), \text{vech}^T(\boldsymbol{\Sigma}_{(1)}), \dots, \text{vech}^T(\boldsymbol{\Sigma}_{(L)})\}^T]$, and $\phi = \text{vec}\{\boldsymbol{\mu}_{(1)}^T, \dots, \boldsymbol{\mu}_{(L)}^T, \text{vec}^T(\boldsymbol{\eta}_{(1)}), \dots, \text{vec}^T(\boldsymbol{\eta}_{(L)}), \text{vech}^T(\boldsymbol{\Omega}_{(1)}), \dots, \text{vech}^T(\boldsymbol{\Omega}_{(L)}), \text{vec}^T(\boldsymbol{\Gamma}), \text{vech}^T(\boldsymbol{\Omega}_0)\}^T$. We use \mathbf{J} to denote the Fisher information matrix under the standard model, and \mathbf{G} to denote the gradient matrix. Then

$$\mathbf{G} \equiv \frac{\partial \mathbf{h}}{\partial \phi^T} = \begin{pmatrix} \mathbf{g}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{g}_2 & \mathbf{0} & \mathbf{g}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{g}_4 & \mathbf{g}_5 & \mathbf{g}_6 \end{pmatrix},$$

where $\mathbf{g}_1 = \mathbf{I}_{Lr}$, \mathbf{g}_2 is a $Lpr \times Lpu$ block diagonal matrix whose all diagonal blocks are $\mathbf{I}_p \otimes \boldsymbol{\Gamma}$, $\mathbf{g}_3 = (\boldsymbol{\eta}_{(1)} \otimes \mathbf{I}_r, \dots, \boldsymbol{\eta}_{(L)} \otimes \mathbf{I}_r)^T$, \mathbf{g}_4 is a $Lr(r+1)/2 \times ur$ block diagonal matrix whose all diagonal blocks are $\mathbf{C}_r(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\mathbf{E}_u$, $\mathbf{g}_5 = \{2(\boldsymbol{\Omega}_{(1)}\boldsymbol{\Gamma}^T \otimes \mathbf{I}_r - \boldsymbol{\Gamma}^T \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)\mathbf{C}_r^T, \dots, 2(\boldsymbol{\Omega}_{(L)}\boldsymbol{\Gamma}^T \otimes \mathbf{I}_r - \boldsymbol{\Gamma}^T \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)\mathbf{C}_r^T\}^T$, and $\mathbf{g}_6 = \{\mathbf{E}_{r-u}^T(\boldsymbol{\Gamma}_0^T \otimes \boldsymbol{\Gamma}_0^T)\mathbf{C}_r^T, \dots, \mathbf{E}_{r-u}^T(\boldsymbol{\Gamma}_0^T \otimes \boldsymbol{\Gamma}_0^T)\mathbf{C}_r^T\}^T$. Now we match Shapiro's notations with our notations. Shapiro's $\boldsymbol{\theta}$ is our ϕ ; Shapiro's $\boldsymbol{\xi}$ is our \mathbf{h} ; Shapiro's $\hat{\boldsymbol{x}}$ is the standard estimator of \mathbf{h} ; Shapiro's $\boldsymbol{\Delta}$ is our gradient matrix \mathbf{G} ; Shapiro's \mathbf{V} is our \mathbf{J} ; and Shapiro's discrepancy function \mathbf{F} is $l_{\max} - l$, where l is the log likelihood function and l_{\max} is the maximum value of l attained when \mathbf{h} is the standard estimator of \mathbf{h} . Specifically,

$$\begin{aligned} l_{\max} - l &= l_{\max} + \frac{nr}{2} \log(2\pi) + \frac{1}{2} \sum_{l=1}^L [n_{(l)} \log |\boldsymbol{\Sigma}_{(l)}| \\ &\quad + \text{tr}\{(\mathbb{Y}_{(l)} - \mathbf{1}_{n_{(l)}}\boldsymbol{\mu}_{(l)}^T - \mathbb{X}_{(l)}\boldsymbol{\beta}_{(l)}^T)\boldsymbol{\Sigma}_{(l)}^{-1}(\mathbb{Y}_{(l)} - \mathbf{1}_{n_{(l)}}\boldsymbol{\mu}_{(l)}^T - \mathbb{X}_{(l)}\boldsymbol{\beta}_{(l)}^T)^T\}] \\ &= l_{\max} + \frac{nr}{2} \log(2\pi) + \frac{1}{2} \sum_{l=1}^L \left(n_{(l)} \log |\boldsymbol{\Sigma}_{(l)}| + \text{tr}[\boldsymbol{\Sigma}_{(l)}^{-1}\{n\hat{\boldsymbol{\Sigma}}_{\text{res}(l)} \right. \\ &\quad \left. + (\boldsymbol{\beta}_{(l)} - \hat{\boldsymbol{\beta}}_{(l),\text{ols}})\mathbb{X}_{(l)}^T\mathbb{X}_{(l)}(\hat{\boldsymbol{\beta}}_{(l)} - \hat{\boldsymbol{\beta}}_{(l),\text{ols}})^T + n_{(l)}(\boldsymbol{\mu}_{(l)} - \bar{\mathbf{Y}}_{(l)})(\boldsymbol{\mu}_{(l)} - \bar{\mathbf{Y}}_{(l)})^T\} \right]. \end{aligned}$$

It is easy to see that $l_{\max} - l$ satisfies the conditions 1 – 4 in Section 3 of Shapiro (1986).

Since \mathbf{J} is full rank, we have $\text{rank}(\mathbf{G}^T \mathbf{J} \mathbf{G}) = \text{rank}(\mathbf{J})$. As the standard estimator of \mathbf{h} is \sqrt{n} consistent and converges in distribution to a normal distribution with mean $\mathbf{0}$ and covariance \mathbf{J}^{-1} , all the conditions in Proposition 4.1 of Shapiro (1986) are satisfied. Thus, the groupwise envelope estimator $\hat{\mathbf{h}}$ is a \sqrt{n} consistent estimator of \mathbf{h} , and $\sqrt{n}(\hat{\mathbf{h}} - \mathbf{h})$ has asymptotically normal distribution with mean $\mathbf{0}$.

When the errors are normally distributed, then \mathbf{J} has a closed form:

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_3 \end{pmatrix},$$

where $f_{(l)} = n_{(l)}/n$ is the proportion of the l -th population, \mathbf{J}_1 is a $Lr \times Lr$ block diagonal matrix whose diagonal blocks are $f_{(1)}\boldsymbol{\Sigma}_{(1)}^{-1}, \dots, f_{(L)}\boldsymbol{\Sigma}_{(L)}^{-1}$, \mathbf{J}_2 is a $Lpr \times Lpr$ block diagonal matrix whose diagonal blocks are $f_{(1)}\boldsymbol{\Sigma}_{\mathbf{X}_{(1)}} \otimes \boldsymbol{\Sigma}_{(1)}^{-1}, \dots, f_{(L)}\boldsymbol{\Sigma}_{\mathbf{X}_{(L)}} \otimes \boldsymbol{\Sigma}_{(L)}^{-1}$, and \mathbf{J}_3 is a $\{Lr(r+1)/2\} \times \{Lr(r+1)/2\}$ block diagonal matrix whose diagonal blocks are $(1/2)f_{(1)}\mathbf{E}_r^T(\boldsymbol{\Sigma}_{(1)}^{-1} \otimes \boldsymbol{\Sigma}_{(1)}^{-1})\mathbf{E}_r, \dots, (1/2)f_{(L)}\mathbf{E}_r^T(\boldsymbol{\Sigma}_{(L)}^{-1} \otimes \boldsymbol{\Sigma}_{(L)}^{-1})\mathbf{E}_r$. The asymptotic variance of $\hat{\mathbf{h}}$ under the groupwise envelope model then has a closed form $\mathbf{V}_{\text{env}} = \mathbf{G}(\mathbf{G}^T \mathbf{J} \mathbf{G})^\dagger \mathbf{G}^T$, where \mathbf{A}^\dagger denotes the Moore-Penrose generalized inverse of a matrix \mathbf{A} . After some straightforward calculations, the asymptotic variance of $\text{vec}(\hat{\boldsymbol{\beta}}_{(l)})$ is $(1/f_{(l)})\boldsymbol{\Sigma}_{\mathbf{X}_{(l)}}^{-1} \otimes \boldsymbol{\Gamma} \boldsymbol{\Omega}_{(l)} \boldsymbol{\Gamma}^T + (\boldsymbol{\eta}_{(l)}^T \otimes \boldsymbol{\Gamma}_0) \left\{ \sum_{l=1}^L f_{(l)} (\boldsymbol{\eta}_{(l)} \boldsymbol{\Sigma}_{\mathbf{X}_{(l)}} \boldsymbol{\eta}_{(l)}^T \otimes \boldsymbol{\Omega}_0^{-1} + \boldsymbol{\Omega}_{(l)} \otimes \boldsymbol{\Omega}_0^{-1} + \boldsymbol{\Omega}_{(l)}^{-1} \otimes \boldsymbol{\Omega}_0 - 2\mathbf{I}_u \otimes \mathbf{I}_{r-u}) \right\}^{-1} (\boldsymbol{\eta}_{(l)} \otimes \boldsymbol{\Gamma}_0^T)$, and the asymptotic variance of $\hat{\boldsymbol{\mu}}_{(l)}$ is $(1/f_{(l)})\boldsymbol{\Sigma}_{(l)}$, for $l = 1, \dots, L$.

Proof of Proposition 3 Proof of Proposition 3 follows directly from (5.7) in Cook et al. (2010).

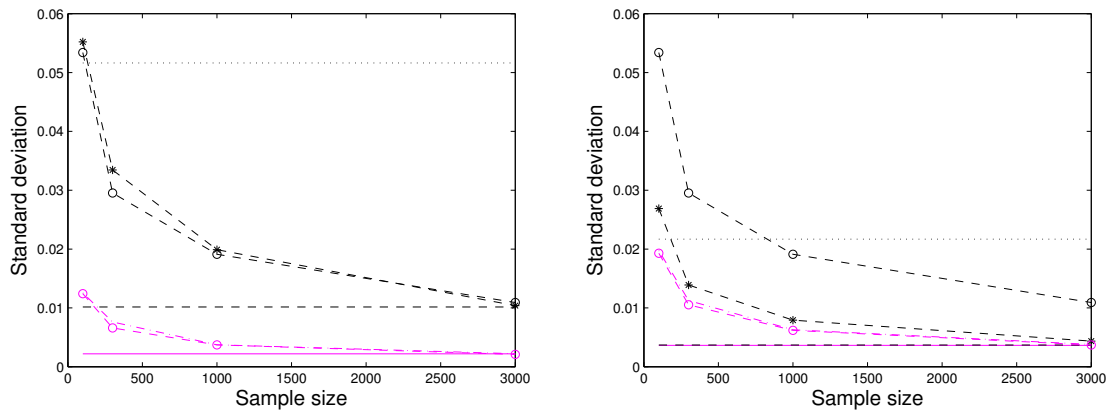


Figure 1: Left panel: Standard deviations for a random picked element in $\beta_{(1)}$. Right panel: Standard deviations for a random picked element in $\beta_{(2)}$. The magenta $-o-$, $-\cdot-$ and $—$ lines mark the actual, bootstrap and asymptotic standard deviations of the groupwise envelope model. The black $-*-$ and $- -$ lines mark the actual and asymptotic standard deviation of the separate envelope model. The black $-o-$ line marks the standard deviations of the envelope estimator in Cook et al. (2010). The black $\cdot\cdot\cdot$ line marks the asymptotic standard deviations of the standard model.

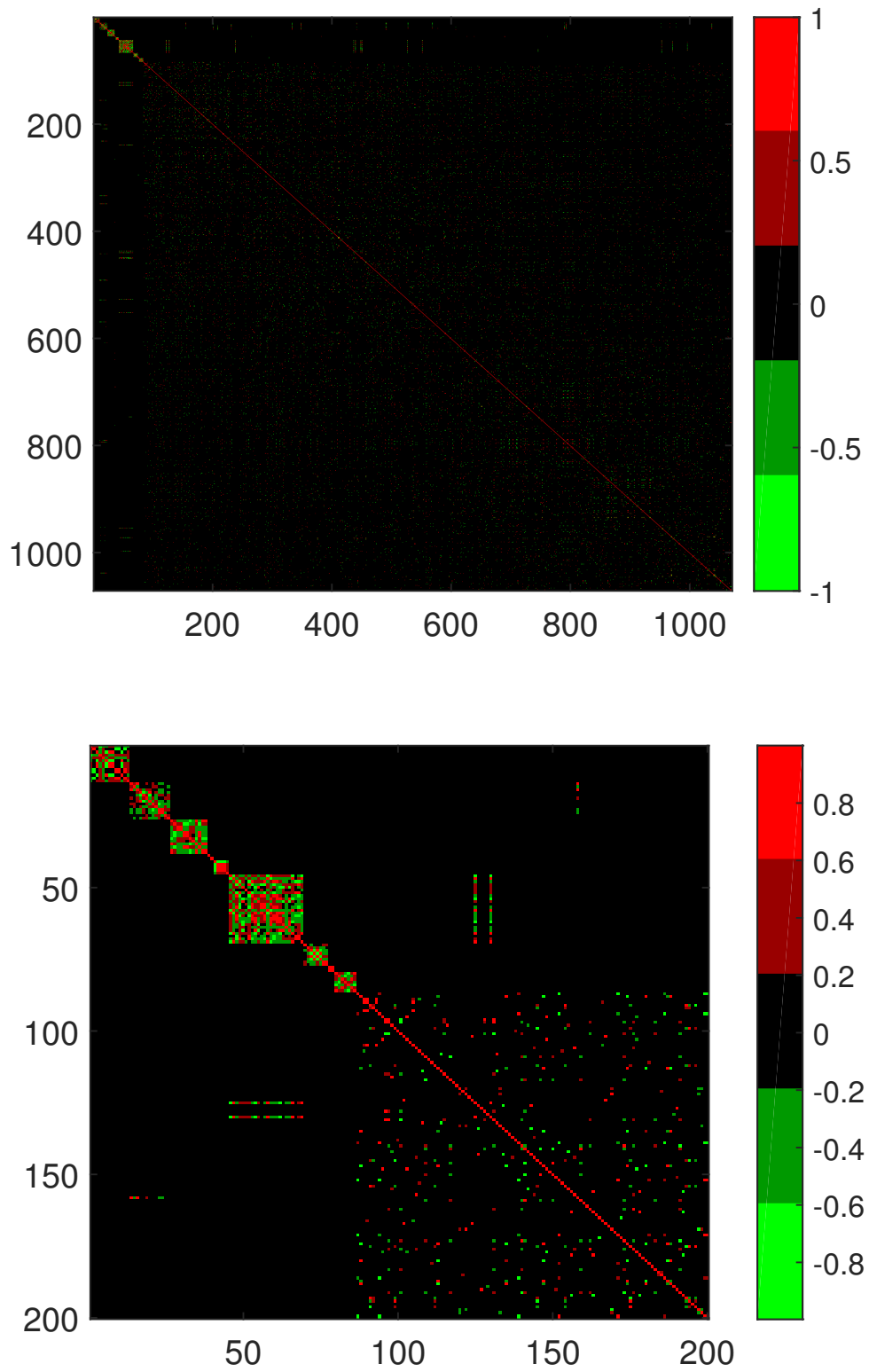


Figure 2: Plot of the correlation matrix. The top panel shows the correlation of 1071 SNPs and the bottom panel is the enlarged version for the first 200 SNPs.

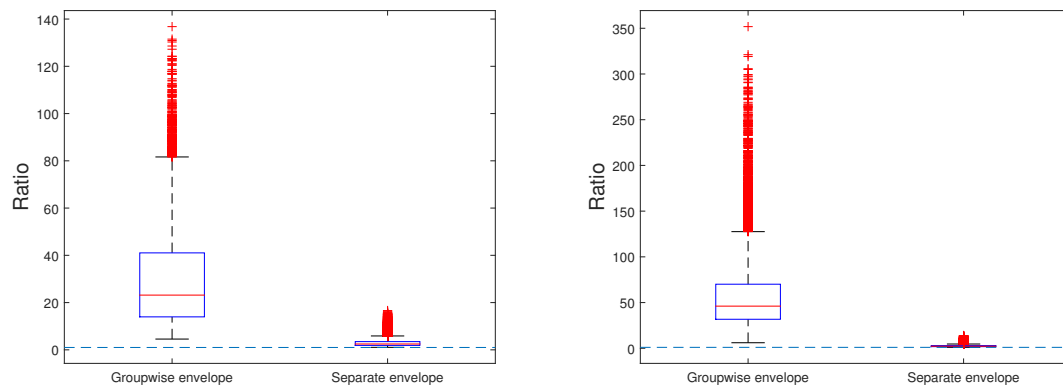


Figure 3: Boxplot of the ratio of the bootstrap standard deviation under the standard model to the bootstrap standard deviation under the groupwise envelope model or separate envelope model. The left panel is boxplot of the ratios for the male group, and the right panel is boxplot of the ratios for the female group. The horizontal blue line is at 1.

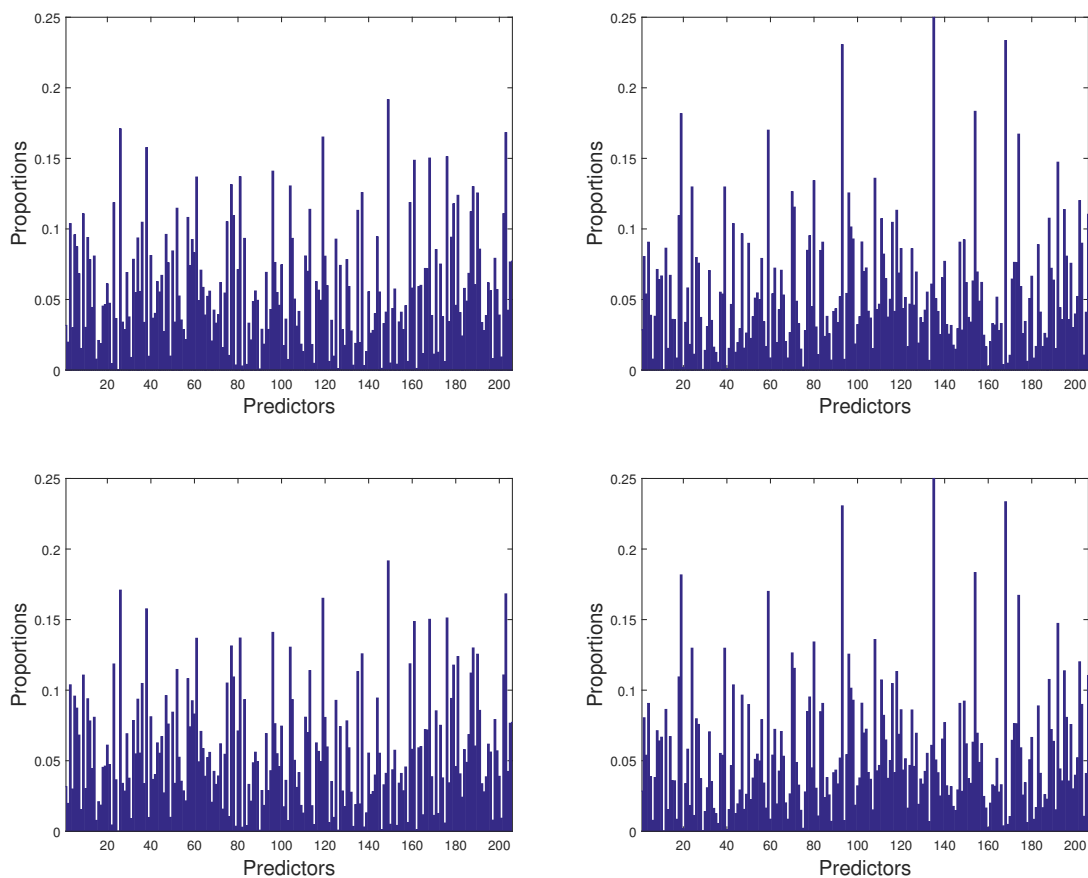


Figure 4: Plot of the proportions of the absolute value of coefficient for SNPs in the Euclidean norm of the coefficient vector corresponding to the region: the last on the horizontal axis denotes APOE $\epsilon 4$ and the remaining is for 205 PCs. The first row is for left hippocampal formation region, and the second row is for the right hippocampal formation region. The left panel is for the male group, and the right panel is for the female group.

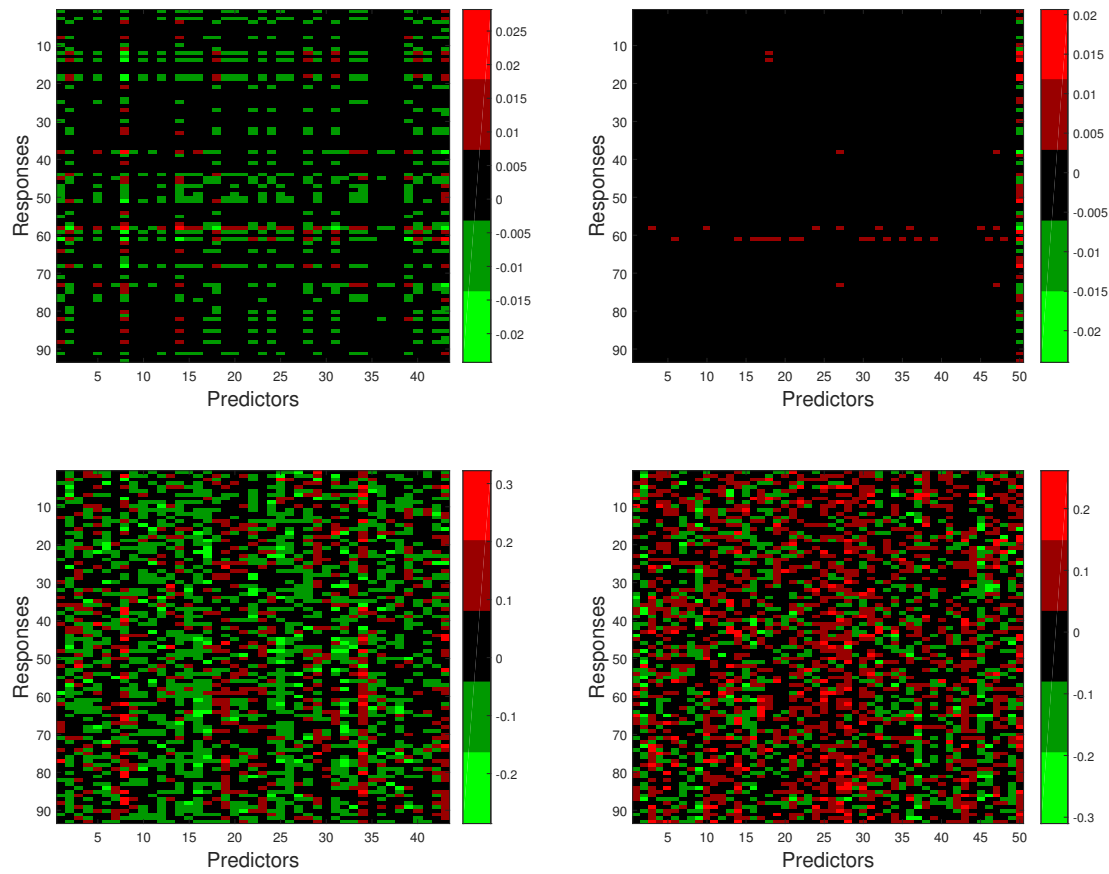


Figure 5: Heatmaps of the regression coefficients: the last on the horizontal axis denotes APOE $\epsilon 4$ and the remaining is for the selected PCs. The first row is for groupwise envelope model, and the second row is for the standard model (2). The left panel is for the male group, and the right panel is for the female group.

Table 1: The selected ROIs based on the p -value of β_{red} . The italic ROIs in groupwise envelope model are regions additionally selected from female group except 41 common selected ROIs.

Groupwise Envelope model		Standard model		
me.f-o.gy.R	pstc.gy.R	me.f-o.gy.R	sup.o.gy.R	tmp.pl.L
lat.ve.L	ling.gy.R	mid.f.gy.R	caud.neuc.L	ent.cort.L
insula.R	me.f.gy.R	lat.ve.L	sup.gy.L	inf.o.gy.R
lat.ve.R	amyg.L	insula.R	ant.caps.L	sup.o.gy.L
glob.pal.R	me.o.gy.L	prec.gy.R	oc.lb.WM.R	lat.o.t.gy.R
glob.pal.L	mid.t.gy.R	lat.f-o.gy.R	mid.f.gy.L	ent.cort.R
inf.f.gy.L	corp.col	cing.R	sup.p.lb.L	hiopp.L
ang.gyr.R	sup.t.gy.R	lat.ve.R	caud.neuc.R	thal.L
tmp.pl.R	me.o.gy.R	me.f.gy.L	cun.L	par.lb.WM.R
nuc.acc.R	thal.R	sup.f.gy.R	prec.L	insula.L
f.lob.WM.L	<i>lat.f-o.gy.R</i>	glob.pal.R	par.lb.WM.L	pstc.gy.R
subtha.nuc.L	<i>sup.f.gy.R</i>	glob.pal.L	tmp.lb.WM.R	ling.gy.R
sup.o.gy.R	<i>hiopp.R</i>	putamen.L	sup.gy.R	me.f.gy.R
caud.neuc.L	<i>caud.neuc.R</i>	inf.f.gy.L	sup.t.gy.L	amyg.L
sup.p.lb.L	<i>mid.t.gy.L</i>	putamen.R	unc.L	me.o.gy.L
prec.L	<i>prec.gy.L</i>	f.lob.WM.R	mid.o.gy.R	parah.gy.R
sup.gy.R	<i>par.lb.WM.R</i>	parah.gy.L	mid.t.gy.L	ant.caps.R
sup.t.gy.L	<i>ant.caps.R</i>	ang.gyr.R	ling.gy.L	mid.t.gy.R
unc.L		tmp.pl.R	sup.f.gy.L	occ.pol.R
mid.o.gy.R		subtha.nuc.R	nuc.acc.L	corp.col
ling.gy.L		nuc.acc.R	oc.lb.WM.L	amyg.R
sup.f.gy.L		unc.R	pstc.gy.L	inf.t.gy.R
nuc.acc.L		cing.L	inf.f.gy.R	sup.t.gy.R
inf.f.gy.R		fornix.L	prec.gy.L	mid.o.gy.L
me.f-o.gy.L		f.lob.WM.L	tmp.lb.WM.L	ang.gyr.L
per.cort.R		pec.R	me.f-o.gy.L	me.o.gy.R
sup.p.lb.R		subtha.nuc.L	per.cort.R	cun.R
per.cort.L		post.limb.L	sup.p.lb.R	lat.o.t.gy.L
inf.t.gy.L		post.limb.R	lat.f-o.gy.L	thal.R
ent.cort.L		hiopp.R	per.cort.L	occ.pol.L
ent.cort.R		inf.o.gy.L	inf.t.gy.L	fornix.R