Bayesian longitudinal low-rank regression models for imaging genetic data from longitudinal studies

Zhao-Hua Lu^a, Zakaria Khondker^b, Joseph G. Ibrahim^b, Yue Wang^b,
Hongtu Zhu^{c,*}, and for the Alzheimer's Disease Neuroimaging Initiative¹

^a*Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA*

^b*Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill,
NC, USA*

^c*Department of Biostatistics, University of Texas MD Anderson Cancer Center,
Houston, TX, USA*

Abstract

To perform a joint analysis of multivariate neuroimaging phenotypes and candidate genetic markers obtained from longitudinal studies, we develop a Bayesian longitudinal low-rank regression (L2R2) model. The L2R2 model integrates three key methodologies: a low-rank matrix for approximating the high-dimensional regression coefficient matrices corresponding to the genetic main effects and their interactions with time, penalized splines for characterizing the overall time effect, and a sparse factor analysis model coupled with random effects for capturing within-subject spatio-temporal correlations of longitudinal phenotypes. Posterior computation proceeds via an efficient Markov chain Monte Carlo algorithm. Simulations show that the L2R2 model outperforms several other competing methods. We apply the L2R2 model to investigate the effect of single nucleotide polymorphisms (SNPs) on the top 10 and top 40 previously reported Alzheimer disease-associated genes. We also identify associations between the interactions of these SNPs with patient age and the tissue volumes of 93 regions of interest from patients' brain images obtained from the Alzheimer's Disease Neuroimaging Initiative.

Preprint submitted to Neuroimage

December 27, 2016

*Corresponding author

Email address: h Zhu5@mdanderson.org (Hongtu Zhu)

¹Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http : //adni.loni.usc.edu/wp - content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

11 *Keywords:* Genetic variants; Longitudinal imaging phenotypes; Low-rank
12 regression; Markov chain Monte Carlo; Spatiotemporal correlation

13 **1. Introduction**

14 Many longitudinal neuroimaging studies concomitantly collect genetic
15 and recurrent imaging data to track individual changes in brain structure and
16 function over time. Several neurodegenerative disorders, including Alzheimer
17 disease (AD), are hypothesized to occur from abnormal development of the
18 brain, which may be caused by the additive and/or interactive effects of
19 various risk genes and environmental risk factors, each contributing small
20 individual effects. Thus, recurrent neuroimaging measures may lead to dis-
21 coveries of the genetic pathways and the causal genes associated with the spe-
22 cific brain changes underlying such neurodegenerative disorders (Scharinger
23 et al., 2010; Paus, 2010; Peper et al., 2007; Chiang et al., 2011a,b; Saykin
24 et al., 2015).

25 A standard statistical method used in longitudinal imaging and genetics
26 studies is the massive marginal association (MMA) framework (Li et al., 2013;
27 Zhang et al., 2014; Guillaume et al., 2014; Hibar and et al, 2011; Shen et al.,
28 2010; Bernal-Rusiel et al., 2013; Zhang et al., 2014). This approach repeat-
29 edly fits a linear mixed effects model (or generalized estimating equations) for
30 paired imaging phenotypes and genetic markers. Because the MMA frame-
31 work entails numerous comparisons, it can detect only phenotype-marker
32 pairs with extremely strong signals.

33 Several attempts have been made to more precisely investigate the effect
34 of multiple genotypes on longitudinal phenotypes. Chen and Wang (2011)
35 proposed functional mixed-effect models with penalized splines and vary-
36 ing coefficients, but they focused on small number of predictors and num-

37 ber of response variables in a low-dimensional setting. Wang et al. (2011)
 38 used a sparse multitask regression to examine the association between ge-
 39 netic markers and longitudinal neuroimaging phenotypes. However, their
 40 model focused on subjects with the same number of repeated measures and
 41 ignored the spatiotemporal correlations of imaging phenotypes. Therefore,
 42 the multitask regression model may lead to loss of statistical power to detect
 43 phenotype-marker pairs with moderate to weak signals. Vounou et al. (2011)
 44 and Silver et al. (2012) proposed that a sparse reduced-rank regression model
 45 using penalized regression can detect the main genetic effects on longitudinal
 46 phenotypes. They, however, did not account for the spatiotemporal associ-
 47 ation among the longitudinal phenotypes, which is important for estimation
 48 and prediction accuracy. Moreover, none of these studies explored SNP-age
 49 interactions, which can reveal dynamic genetic effects on phenotypes.

50 Several important statistical concerns are associated with the joint anal-
 51 ysis of neuroimaging phenotypes and a set of candidate genotypes obtained
 52 from longitudinal imaging and genetic studies. First, the number of regres-
 53 sion coefficients can be much larger than the sample size, denoted as N .
 54 Specifically, let d and p be the dimension of the responses and the number
 55 of covariates, respectively. Fitting a multivariate linear mixed effects model
 56 usually requires estimating a $d \times p$ matrix of regression coefficients, which
 57 can be much larger than N , even for moderately high d and p . Second, as il-
 58 lustrated in Fig. 1a, to improve prediction accuracy (Breiman and Friedman,
 59 1997), it is critically important to account for unstructured, within-subject
 60 spatial correlations among multivariate neuroimaging phenotypes. Third, as
 61 illustrated in Fig. 2a, to improve both estimation and prediction accuracy, it
 62 is also important to account for within-subject temporal correlation. Fourth,
 63 as shown in Fig. 2b,c, the temporal growth pattern varies across regions of

64 interest (ROIs) in the brain. Accounting for the overall longitudinal change
65 of ROIs is required to increase the detection power of the genetic effects.
66 Fifth, as shown in Fig. 2b, the genetic effects on ROI volumes can vary
67 across time.

68 Here, we have developed a Bayesian longitudinal low-rank regression
69 (L2R2) model for the joint analysis of high-dimensional longitudinal re-
70 sponses and covariates. We integrated multiple robust methods to explicitly
71 address the new challenges described previously. Our study has four major
72 methodological contributions that were previously undescribed:

- 73 (i) To the best of our knowledge, L2R2 is the first model of its kind for
74 jointly analyzing high-dimensional longitudinal responses and covari-
75 ates, although several approaches have been used for high-dimensional
76 responses and covariates in cross-sectional studies (Rothman et al.,
77 2010; Vounou et al., 2010; Zhu et al., 2014). The L2R2 model also
78 provides a set of standard inference tools (e.g. standard deviation) for
79 determining various unknown parameters. Zipunnikov et al. (2014)
80 proposed a functional principle components analysis for high dimen-
81 sional ($> 10,000$) longitudinal responses, where the intercepts and
82 slopes of time for all voxels were modeled by a few basis functions.
83 However, their methods cannot handle high-dimensional responses and
84 covariates simultaneously because the dimension of the covariance ma-
85 trix that requires time-consuming spectral decomposition equals the
86 product of the dimensions of the responses and covariates.
- 87 (ii) The temporal growth patterns of high dimensional responses were char-
88 acterized by a penalized spline method. Although the spline method
89 has been widely used in other studies (Eilers and Marx, 1996; Ruppert
90 et al., 2003; Wood, 2006; Wu and Zhang, 2006; Lang and Brezger, 2004;

91 Guo, 2002; Morris and Carroll, 2006; Greven et al., 2010), most of these
 92 studies focused on univariate responses. We included a new set of co-
 93 efficients for the high-dimensional gene-age interactions to model the
 94 genetic effects on longitudinal trajectories of responses, which were dif-
 95 ferent from previously reported coefficients corresponding to the main
 96 genetic effects that were time-invariant (Marttinen et al., 2014; Zhu
 97 et al., 2014). A low-rank regression model not only reduces the num-
 98 ber of unknown coefficients, but also captures the low-rank structure
 99 of the regression coefficient matrix.

100 (iii) To efficiently model the correlation structure and increase the detection
 101 power of association between response and covariates, we used a sparse
 102 factor model (Bhattacharya and Dunson, 2011) to capture the spatial
 103 correlation and the structured random effects to model the temporal
 104 correlation of the longitudinal response variables.

105 (iv) We have developed a downloadable L2R2 package, which is available
 106 at <http://odin.mdacc.tmc.edu/bigs2/> and <http://www.nitrc.org>.

107 2. Methods: L2R2 model formulation

Let $y_{ik}(t_{ij})$ be the j -th observation at time t_{ij} of the k th neuroimag-
 ing measurement (e.g. volume of ROI) for the i th subject for $i = 1, \dots, n$,
 $j = 1, \dots, m_i$ and $k = 1, \dots, d$, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ be a $p \times 1$ vector of
 genetic predictors (e.g. $p/2$ genetic markers and $p/2$ gene-age interactions),
 and $N = \sum_{i=1}^n m_i$ be the total number of observations. To address the chal-
 lenges of concomitantly analyzing high-dimensional longitudinal responses
 and covariates, we proposed the following L2R2 model: For each k , we have

$$y_{ik}(t) = \mathbf{x}_i^T \boldsymbol{\beta}_k + \mu_k(t) + \mathbf{w}_{2,i}(t)^T \boldsymbol{\gamma}_{2,k} + \mathbf{z}_i(t)^T \mathbf{b}_{ik} + \epsilon_{ik}(t), \quad (1)$$

where β_k is a $p \times 1$ vector of coefficients for the genetic predictors, $\mu_k(t)$ is a nonparametric function of t that characterizes the overall trajectories of the k th measurement across all subjects, $\mathbf{w}_{2,i}(t_{ij}) = (w_{2,i1}(t_{ij}), \dots, w_{2,iq_2}(t_{ij}))^T$ is a $q_2 \times 1$ vector of time-variant or time-invariant prognostic factors (e.g. age, gender), and $\gamma_{2,k}$ is a $q_2 \times 1$ vector of coefficients for the prognostic factors. Moreover, $\mathbf{b}_{ik}$ is a $p_b \times 1$ vector of latent random effects that characterize the within-subject correlation of the i th subject for the k th measurement, and $\mathbf{z}_i(t)$ is the related $p_b \times 1$ vector of covariates characterizing the correlation structure, and $\epsilon_{ik}(t)$ is an error term. Let $\mathbf{b}_i = (\mathbf{b}_{i1}, \dots, \mathbf{b}_{id})$ such that

$$\text{vec}(\mathbf{b}_i) \sim N(\mathbf{0}, \Sigma_b) \text{ and } \boldsymbol{\epsilon}_i(t) \sim N_d(\mathbf{0}, \Sigma_e = \Theta^{-1}),$$

108 where $\mathbf{0}$ is a vector of zero of appropriate dimension, and Σ_b and Σ_e are
 109 $p_b d \times p_b d$ and $d \times d$ covariance matrices, respectively. We assumed that $\mathbf{b}_i$
 110 and $\boldsymbol{\epsilon}_i(t) = (\epsilon_{i1}(t), \dots, \epsilon_{id}(t))^T$ are independent for all i and t , $\mathbf{b}_i$ and $\mathbf{b}_{i'}$
 111 are independent for $i \neq i'$, and $\boldsymbol{\epsilon}_i(t)$ and $\boldsymbol{\epsilon}_{i'}(t')$ are independent for $i \neq i'$ or
 112 $t \neq t'$.

113 To address the challenges described above, we first incorporated the gene-
 114 age interactions into $\mathbf{x}_i$ to model the dynamic genetic effects on individ-
 115 ual brain development trajectories. The individual effect of a risk gene on
 116 brain development may vary across time. We used data published by the
 117 Alzheimer's Disease Neuroimaging Initiative (ADNI) to show the estimated
 118 trajectories with local polynomial regression (LOESS, Cleveland and De-
 119 vlin, 1988) of subjects with and without the minor alleles for two selected
 120 phenotype-single nucleotide polymorphism (SNP) pairs (Fig. 2b,c). There
 121 was a strong interaction effect of the SNP rs769451 and age on the right
 122 hippocampus. Ignoring such interaction in model (1) may inflate errors, de-
 123 crease the detection power of important main genetic effects, and miss the

124 age-dependent genetic effects on imaging phenotypes. In contrast, the inter-
 125 action effect of SNP rs439401 and age on the left hippocampus was small. We
 126 need to efficiently distinguish the important interaction effects from others
 127 with negligible interaction effects.

We considered the $p \times d$ coefficient matrix $\mathbf{B} = [\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_d]$ that characterizes the time-dependent and main genetic effects on the longitudinal phenotypes across all subjects and time points. Given the large d and p , the number of elements in $\mathbf{B}$ could be much larger than that of N . Moreover, as shown in Fig. 1a,b, the elements in $\mathbf{y}_i(t) = (y_{i1}(t), \dots, y_{id}(t))^T$ and $\mathbf{x}_i$ were highly correlated because of the spatial pattern of brain structure and the linkage disequilibrium structure of SNPs, respectively. Therefore, the estimate of $\mathbf{B}$ was unstable and often exhibited horizontal- and vertical-banded structures (Fig. 1c). These structures were more apparent after setting unimportant (i.e. insignificant) elements to zero (Fig. 1d). Applying the characteristics of the coefficient matrix, we introduced a low-rank model for $\mathbf{B}$ as follows:

$$\mathbf{B} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T = \sum_{l=1}^r \mathbf{B}_l = \sum_{l=1}^r \delta_l \mathbf{u}_l \mathbf{v}_l^T, \quad (2)$$

128 where r is the rank or number of layers of $\mathbf{B}$; $\mathbf{B}_l = \delta_l \mathbf{u}_l \mathbf{v}_l^T$ is the l th layer for
 129 $l = 1, \dots, r$; $\boldsymbol{\Delta} = \text{diag}(\delta_1, \dots, \delta_r)$; $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_r]$ is a $p \times r$ matrix; and
 130 $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_r]$ is a $d \times r$ matrix. Moreover, $\mathbf{u}_l$ and $\mathbf{v}_l$ may be viewed as the
 131 coefficient of the linear combination of $\mathbf{y}_i(t)$ and $\mathbf{x}_i$, respectively. Rather than
 132 investigating the correlation between all phenotype-SNP pairs, we examined
 133 the relation between important linear combinations of similar phenotypes
 134 and those of similar SNPs. Because only a few sets of genetic variants were
 135 expected to be associated with each longitudinal phenotype, a relatively small
 136 $r \ll \min(p, d)$ and a sparse $\mathbf{U}$ were considered sufficient to represent the

important structure of $\mathbf{B}$. Model (2) considerably reduced the number of unknown parameters of $\mathbf{B}$ from pd to $(d + p + 1)r$ and increased the power for detecting the important associations.

We used $\mu_k(t)$ to characterize the overall trajectory for the k th neuroimaging measure without consideration of any effects of genomic variables and prognostic factors. The solid curves in Fig. 2b,c depict the LOESS estimates of the development trajectories of the right and left hippocampal volumes, respectively. We used a penalized spline model (Eilers and Marx, 1996; Ruppert et al., 2003; Lang and Brezger, 2004) to model $\mu_k(t)$ with a polynomial of degree s given by

$$\mu_k(t) = \gamma_{k,0}t^0 + \cdots + \gamma_{k,s}t^s + \sum_{m=1}^{q_1-s-1} \gamma_{k,s+m}B_m(t; \mathbf{T}_0) = \mathbf{w}_1(t)^T \boldsymbol{\gamma}_{1,k}, \quad (3)$$

where $B_m(t; \mathbf{T}_0)$ is a basis function of t ; $\mathbf{T}_0$ is a vector of predetermined knots over the range of the t_{ij} 's (e.g. the sample percentiles); $\boldsymbol{\gamma}_{1,k} = (\gamma_{k,0}, \cdots, \gamma_{k,q_1-1})^T$; and $\mathbf{w}_1(t) = (1, t, \cdots, t^s, B_1(t; \mathbf{T}_0), \cdots, B_{q_1-s-1}(t; \mathbf{T}_0))^T$. Various spline basis functions may be considered for $B_m(t; \mathbf{T}_0)$, such as the B-spline basis (Eilers and Marx, 1996), truncated power basis (Ruppert et al., 2003), and wavelet basis (Morris and Carroll, 2006). Using $\mu_k(t)$ and the basis functions to flexibly capture the overall trajectories of the longitudinal ROIs measurements increases the detection power of the SNPs and SNP-age interactions and improves the prediction accuracy of responses at later time points.

Another issue in longitudinal studies that must be accounted for is the within-subject correlation among repeated measurements of each phenotype (Verbeke and Molenberghs, 2009; Hyun et al., 2016). Various correlation structures can be formulated through $\mathbf{z}_i(t)^T \mathbf{b}_{ik}$ and $\boldsymbol{\Sigma}_b$. We examined the trajectories of all phenotypes across all subjects for the longitudinal data set obtained from the ADNI. The within-subject association mainly occurred

155 from the subject-specific intercept. The trajectories of the left lateral ven-
 156 tricle volume across all subjects are depicted in Fig. 2a. To generate these
 157 trajectories, we used a univariate random effect for each ROI by setting
 158 $p_b = 1$, $\mathbf{z}_i(t) = 1$, and $\mathbf{\Sigma}_b = \tau_b \mathbf{I}_d$, where $\mathbf{I}_d$ is a $d \times d$ identity matrix. In this
 159 case, the random effects of different ROIs were assumed to be independent.

160 The correlation structure among the multivariate phenotypes was mod-
 161 eled with another component in (1). The covariance matrix of phenotypic
 162 measurements across all subjects and all time points is depicted in Fig. 1a.
 163 Phenotypic variables usually exhibit group correlation structures; therefore,
 164 we used latent factors to capture these correlation structures. Specifically,
 165 we considered an exploratory method of using the sparse factor model (Bhat-
 166 tacharya and Dunson, 2011) for $\boldsymbol{\epsilon}_i(t)$ as follows:

$$\boldsymbol{\epsilon}_i(t) = \mathbf{\Lambda} \boldsymbol{\eta}_i(t) + \boldsymbol{\xi}_i(t), \quad (4)$$

167 where $\mathbf{\Lambda}$ is a $d \times \infty$ factor loading matrix, $\boldsymbol{\eta}_i(t) \sim N_\infty(\mathbf{0}, \mathbf{I}_\infty)$, and $\boldsymbol{\xi}_i(t) \sim$
 168 $N(\mathbf{0}, \mathbf{\Sigma}_\xi)$ with $\mathbf{\Sigma}_\xi = \text{diag}(\sigma_{\xi,1}^2, \dots, \sigma_{\xi,d}^2)$. The structure of $\mathbf{\Lambda}$ and the number
 169 of factors were directly gleaned from the data.

The L2R2 in (1) with (2) - (4) incorporated can be rewritten in the matrix
 form

$$\mathbf{Y} = \mathbf{XU}\mathbf{\Delta}\mathbf{V}^T + \mathbf{W}\mathbf{\Gamma} + \mathbf{Z}\mathbf{b} + \mathbf{E}. \quad (5)$$

170 The notational definitions, prior distributions of the parameters and the tech-
 171 nical details of the posterior computation were provided in the appendix.

172 3. Simulation study

173 3.1. Simulation setup

174 We used simulations to examine the finite-sample performance of the
 175 L2R2 model. We simulated $\mathbf{Y}$ according to model (5). We considered four

176 cases with various dimensions and priors: (i) $p = 50$, $d = 50$; (ii) $p =$
 177 100 , $d = 100$; (iii) $p = 200$, $d = 100$; and (iv) $p = 400$, $d = 100$. The
 178 dimensions of these cases were comparable with the ADNI data set used in
 179 this study and with other related neuroimaging and genetic studies (Wang
 180 et al., 2011; Zhu et al., 2014; Marttinen et al., 2014). We chose the first
 181 $n = 100$ subjects from the ADNI data set and selected the top $p/2$ SNPs
 182 from the top 40 genes reported by AlzGene database (www.alzgene.org) as
 183 of June 10, 2010. Each subject was observed m_i times to replicate the ADNI
 184 data set, resulting in $N = 422$ records. We used the $p/2$ SNPs and their
 185 interactions with age to form the $\mathbf{X}$ in model (5). The $\mathbf{W}$ in (5) contains $q =$
 186 15 columns, including the intercept, the standardized intracranial volume,
 187 sex, education, handedness, and the basis function of age. The basis function
 188 consists of age, age², age³, and the B-spline basis given the j th percentile,
 189 $j = 0, 10, 20, \dots, 100$. The elements of $\mathbf{b}$ were independently generated from
 190 a $N(0, 1)$ distribution.

191 The following parameters in (5) were used in the simulation: The low-
 192 rank coefficient matrix $\mathbf{B}$ was generated with the true rank $r_0 = 3$ in a
 193 moderately sparse case with 65% zero elements and in an extremely sparse
 194 case with 95% zero elements. Specifically, we set $\mathbf{B} = \mathbf{U}\mathbf{\Delta}\mathbf{V}$, where $\mathbf{U} =$
 195 (u_{jl}) , $\mathbf{\Delta} = \text{diag}(\delta_u) = \text{diag}(100, 80, 60)$, and $\mathbf{V} = (v_{lk})$ as $p \times 3$, 3×3 ,
 196 and $3 \times d$ matrices, respectively. Moreover, we generated all elements u_{jl}
 197 and v_{lk} independently from a $N(0, 1) \times \text{binomial}(1, p_1)$ distribution and then
 198 normalized the columns of $\mathbf{U}$ and $\mathbf{V}$ to have zero mean and unit variance. The
 199 value of p_1 was tuned so that approximately 65% of the elements of B were
 200 zeros. Each element of $\mathbf{\Gamma}$ was independently generated as a $\gamma_{jk} \sim N(0, 1) \times$
 201 $\text{binomial}(1, 0.5)$. We simulated $\boldsymbol{\epsilon}_i(t) \sim N_d(0, \boldsymbol{\Sigma}_e)$, where the precision matrix
 202 $\boldsymbol{\Sigma}_e$ was first generated by a $d \times d$ matrix $\mathbf{A} = (a_{jj'})$ with $a_{jj} = 1$ and

203 $a_{jj'} = \text{uniform}(0, 1) \times \text{binomial}(1, p_2)$ for $j \neq j'$, setting $\Sigma_e = \mathbf{A}\mathbf{A}^T$, and
 204 standardizing Σ_e into a correlation matrix. The value of p_2 was tuned so
 205 that approximately 20% of the elements in Σ_e were zeros, yielding the mean
 206 of the absolute correlations of Σ_e at approximately 0.40.

207 For each case, 100 simulated data sets were generated. For each simulated
 208 data set, we ran the Gibbs sampler for 10,000 iterations after 5000 burn-in
 209 iterations. We chose noninformative priors for the hyperparameters and set
 210 $a_0 = b_0 = c_0 = d_0 = e_0 = f_0 = 10^{-6}$. For the covariance parameter
 211 Σ_e , we chose moderately informative priors to impose the positive-definite
 212 constraint.

213 Besides modeling the trajectories of the response variables to improve
 214 detection power of the genetic variables, the splines in $\mathbf{W}$ could also be used
 215 to predict the response variables at later time points. We calculated the
 216 mean of the response variables $\mathbf{X}_{pred}\mathbf{U}\Delta\mathbf{V}^T + \mathbf{W}_{pred}\mathbf{\Gamma} + \mathbf{Z}_{pred}\mathbf{b}$ for every
 217 the subjects at one year after their last observation, where $\mathbf{X}_{pred}$, $\mathbf{W}_{pred}$,
 218 and $\mathbf{Z}_{pred}$ contain the corresponding predictors for these subjects at the later
 219 time points. The prediction was based on $\mathbf{X}_{pred}\hat{\mathbf{U}}\hat{\Delta}\hat{\mathbf{V}}^T + \mathbf{W}_{pred}\hat{\mathbf{\Gamma}} + \mathbf{Z}_{pred}\hat{\mathbf{b}}$,
 220 where the hat represents the posterior means of the parameters and random
 221 effects. We compared the means and predicted means to check the prediction
 222 accuracy using the splines in $\mathbf{W}$.

223 3.2. Comparison with four other state-of-the-art methods

224 We compared the L2R2 model with four state-of-the-art methods that
 225 were developed to establish the association between high-dimensional re-
 226 sponses and predictors. First, we considered MMA analysis between the
 227 high-dimensional response variables and predictors. Each phenotype was re-
 228 gressed on the covariate matrix $\mathbf{W}$ and the residue was associated with each

229 SNP in $\mathbf{X}$ with regression. We used the lme4 R package (Bates et al., 2015)
 230 for MMA.

231 Second, we considered the group-sparse multitask regression and feature-
 232 selection (G-SMuRFS, Wang et al., 2011) method, which is a representation
 233 of associating high dimensional imaging and genetic variables with regular-
 234 ization methods. The longitudinal volumes of ROIs were used as the response
 235 variables, and the $\mathbf{X}$ and $\mathbf{W}$ in L2R2 were combined as the matrix of predic-
 236 tors in G-SMuRFS. The spline basis in $\mathbf{W}$ modeled the overall longitudinal
 237 trajectories of the ROI volumes, and the SNP-age interactions in $\mathbf{X}$ repre-
 238 sented the SNP effects on the longitudinal trajectories of ROIs. The large
 239 matrix of coefficients was estimated with penalty functions and optimiza-
 240 tion. Each covariate, SNP and SNP-age interaction formed a group in the
 241 penalty function. We used $\exp(-7)$, $\exp(-6)$, $\dots$, $\exp(2)$ as possible values
 242 of the tuning parameters in the G-SMuRFS method and performed a five-fold
 243 cross validation to choose the tuning parameters. G-SMuRFS did not make
 244 any assumption on the distribution and correlation structure of the response
 245 variables. In our longitudinal study, G-SMuRFS method did not explicitly
 246 account for the within-subject temporal correlations and spatial correlations.
 247 Also, G-SMuRFS did not use dimension reduction for the coefficient matrix
 248 like the low-rank representation of $\mathbf{B}$ in L2R2.

249 Third, we considered the Bayesian canonical correlation analysis (BCCA
 250 Klami et al., 2013) to obtain linear combinations of the response variables
 251 $\mathbf{Y}$ and the predictor $\mathbf{X}$. The coefficients of the linear combinations were
 252 conceptually similar to $\mathbf{U}$ and $\mathbf{V}$ in (5), and we denoted their estimates by
 253 $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$, respectively. The coefficient matrix $\mathbf{B}$ was estimated by $\tilde{\mathbf{U}}\tilde{\mathbf{V}}^T$.
 254 Gamma mixture multivariate normal distributions were assigned to $\mathbf{U}$ and
 255 $\mathbf{V}$, and the correlation structure of the response variables and the SNPs

256 were characterized by a factor model. We used the default setting of the
 257 companion R package and used eight components in the BCCA method.
 258 The R package cannot run a continuous predictor; therefore, we excluded
 259 the SNP-age interactions in $\mathbf{X}$. Also, the package does not process covariates
 260 like $\mathbf{W}$. Therefore, we regressed $\mathbf{Y}$ on $\mathbf{W}$ and applied the BCCA method
 261 to the residue and $\mathbf{X}$. Unlike the L2R2 model, the BCCA method does not
 262 account for within-subject structural correlation or the interactions between
 263 SNPs and age.

264 Fourth, we considered the Bayesian reduced-rank regression (BRRR)
 265 method developed by Marttinen et al. (2014). Because the BRRR software
 266 package does not allow continuous predictors, we excluded the SNP-age in-
 267 teraction terms in $\mathbf{X}$. We also set the rank to three and ran all of the other
 268 program parameters at their default settings. The BRRR method can process
 269 $\mathbf{W}$ directly, so we did not need to regress $\mathbf{Y}$ on $\mathbf{W}$ beforehand. In contrast
 270 with the L2R2 model, the BRRR method does not account for SNP-age
 271 interactions or within-subject structural correlation.

272 3.3. Evaluation

To evaluate the finite-sample classification performances of all five of the
 methods, we calculated the sensitivity and specificity scores of selecting the
 nonzero elements in the coefficient matrix $\mathbf{B}$ for each method. The following
 criteria were used for the selection of the coefficients in each method. As
 the G-SMuRFS method does not provide standard error estimate for each
 coefficient, the importance of the association between each pair of ROI and
 SNP/SNP-age interaction was determined by the absolute values of the coef-
 ficient estimates. For all other methods, the coefficients were selected if the
 absolute values of their standardized estimates (coefficient estimate over its

standard error commonly used to account for the estimation uncertainty) were larger than a threshold. The sensitivity and specificity scores for a given threshold T_0 were defined as

$$\text{Se}(T_0) = \frac{\text{TP}(T_0)}{\text{TP}(T_0) + \text{FN}(T_0)} \text{ and } \text{Sp}(T_0) = \frac{\text{TN}(T_0)}{\text{TN}(T_0) + \text{FP}(T_0)},$$

where $\text{TP}(T_0)$, $\text{FN}(T_0)$, $\text{TN}(T_0)$, and $\text{FP}(T_0)$ were the numbers of true positives, false negatives, true negatives, and false positives, respectively. Changing the threshold T_0 produced different sensitivity and specificity scores, generating the receiver operating characteristic (ROC) curves. In each ROC curve, the sensitivity was plotted against one minus specificity.

3.4. Results

The ROC curves of all four methods and the L2R2 model are depicted in Fig. 3. ROC curves that are closer to the upper-left corner are more likely to identify true positives and control for false positives.

In all simulation settings, the L2R2 model outperformed the four competing methods because the L2R2 model was designed to specifically address the challenges described above. Among the four previously developed methods, the G-SMuRFS method had the best performance. However, the G-SMuRFS method did not characterize the correlation structure among the response variables like the L2R2 model did. Therefore, the G-SMuRFS method potentially increased the number of false positives. Consequently, when the number of phenotypic variables increased, the difference between the L2R2 model and the G-SMuRFS method for $d = 100$ was larger in terms of main effects and interactions than that for $d = 50$.

The BRRR and BCCA methods cannot process SNP-age interactions because of the limitations of their companion R packages. Therefore, they

294 increase modeling error and reduce the power of detecting important main
 295 genetic effects in the presence of the SNP-age interactions. These limitations
 296 are worsened as the number of SNPs increases. Although the model setup of
 297 the BRRR method was similar to that of the L2R2 model, the prior setting
 298 for the BRRR method was more subtle and appeared to be more suitable
 299 for small numbers of responses and predictors. Thus, the BRRR method is
 300 very sensitive to the number of responses and predictors. The MMA method
 301 performed the worst, as it did not account for the correlation structure of
 302 responses and predictors or the spatio-temporal correlation structure of the
 303 longitudinal phenotypes.

304 Selected spline functions in the true coefficients ($\mathbf{\Gamma}$) and their correspond-
 305 ing estimates obtained from the L2R2 model and the G-SMuRFS method are
 306 shown in Fig. 4a-b. It is apparent from these splines that the bias of the
 307 L2R2 model was smaller than that of the G-SMuRFS method. The estimated
 308 trajectories with smaller bias enabled us to better predict the response vari-
 309 ables of later time points. A scatter plot of the true means at later time
 310 points versus the predicted means across all subjects and response variables
 311 in a randomly selected data set is shown in Fig. 4c. The true and predicted
 312 means were close to each other, indicating that model (5) can accurately
 313 predict longitudinal responses.

314 We examined the Markov chain Monte Carlo (MCMC) convergence through
 315 calculation of Gelman and Rubin’s shrink factors for all the parameter es-
 316 timates obtained from three MCMC chains with different start values of
 317 parameters (Fig. 4d). The MCMC chains appeared to converge after 2000
 318 iterations, as the shrink factors for all the parameters were less than 1.2.
 319 Generating 4000 MCMC samples after discarding 4000 burn-in samples for
 320 the L2R2 model to fit the four cases (i.e. $p=50, d=50$; $p=100, d=100$; $p=200,$

d=100; and p=400, d=100) took approximately 6, 14, 16, and 22 min, respectively, with a single core of an Intel Xeon E5520 CPU. In comparison, the G-SMuRFS method took 8, 18, 30, and 90 s to fit the four cases, respectively. However, the MCMC samples of the L2R2 model allowed us to make statistical inferences on all parameter estimates, whereas this could not be performed with the G-SMuRFS method. It is important to note that statistical inferences made under a high-dimensional setting require some resampling and splitting procedures (Meinshausen and Bühlmann, 2010). In this case, the G-SMuRFS method would require processing of thousands of subsamples, which may be computationally demanding.

4. Real data analysis

4.1. ADNI data description

The ADNI was launched in 2003 as a 5-year public-private partnership by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations. The primary goal of the ADNI is to test whether serial magnetic resonance imaging (MRI), positron emission tomography, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment and early AD. Determination of sensitive and specific markers of very early AD progression may aid researchers and clinicians in developing new treatments and monitoring their effectiveness, as well as reduce the time and cost of clinical trials.

344 4.2. Preprocessing

345 The MRI data were collected across a variety of MRI scanners with indi-
346 vidualized protocols for each scanner to obtain standard T1-weighted images,
347 volumetric 3-dimensional sagittal magnetization prepared gradient-echo se-
348 quences, or equivalent images with various resolutions. The typical imaging
349 protocol included inversion time=1000 ms, repetition time=2400 ms, flip
350 angle= 8° , and field of view=24 cm with a $256 \times 256 \times 170$ acquisition matrix
351 in the x -, y -, and z -dimensions, yielding a voxel size of $1.25 \times 1.26 \times 1.2$
352 mm³. Standard steps such as anterior commissure and posterior commissure
353 correction, skull stripping, cerebellum removing, intensity inhomogeneity cor-
354 rection, segmentation, and registration (Shen and Davatzikos, 2004) were
355 used to preprocess the MRI data. We then carried out automatic regional
356 labeling for the template and transferring the labels after the deformable reg-
357 istration of subject images. After labeling 93 ROIs, we computed the ROI
358 volumes for each subject.

359 The Human 610-Quad BeadChip (Illumina, Inc., San Diego, CA) was
360 used to genotype the subjects whose images in the ADNI database we ana-
361 lyzed, resulting in a set of 620,901 SNP and copy number variation markers.
362 Because the Apolipoprotein E (APOE) SNPs rs429358 and rs7412 are not in-
363 cluded in the Human 610-Quad Bead-Chip, they were genotyped separately.
364 These two SNPs define a three allele haplotype, namely the $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$
365 variants, and the presence of each variant was available in the ADNI database
366 for all subjects. We used EIGENSTRAT software (package 3.0) to calculate
367 the population stratification coefficients of all subjects. To reduce population
368 stratification effects, we used images from only white subjects who had at
369 least one imaging sample available (749 of 818 subjects in the ADNI data
370 set).

371 We also performed quality control on this initial set of genotypes (Wang
 372 et al., 2011). To input the missing genotypes into our analysis, we used
 373 MACH4 software (version 1.0.16) with default parameters to infer the haplo-
 374 type phase. We also included the APOE- ϵ_4 variant, coded as the number of
 375 observed ϵ_4 variants. We removed SNPs with more than 5% missing values
 376 and entered the mode for the missing SNPs. In the final quality-controlled
 377 genotype data, we removed SNPs with a minor allele frequency smaller than
 378 0.1 and a Hardy-Weinberg p-value $< 10^{-6}$.

379 4.3. Data analysis

380 The aim of our analysis was to apply the L2R2 model to establish an
 381 association between the SNPs in the top AD-risk genes reported by AlzGene
 382 (<http://www.alzgene.org/>) and the volumes of 93 brain ROIs collected by the
 383 ADNI. We used the L2R2 model to carry out formal statistical inferences,
 384 such as the identification of significant SNPs and SNP-age interactions. In
 385 this data analysis, we included data from $n = 749$ subjects and $N = 2817$
 386 MRI measurements, resulting in an unbalanced data set. Among the sub-
 387 jects, 41 had only one observation and another 67 had only two observations.
 388 A random effect was used to account for the subject-specific intercept. For
 389 the age effect, we used a quadratic penalized spline with 11 knots that were
 390 based on the percentiles of standardized age. We also included intracranial
 391 volume, sex, education, and handedness as covariates in $\mathbf{W}$ to adjust the
 392 ROI volumes.

393 We considered two sets of top AD-risk genes. First, we used the 114 SNPs
 394 in the top 10 genes reported by the AlzGene database. After we performed
 395 quality control to the data set, 87 SNPs, APOE- ϵ_4 , and their interaction
 396 with age were included in the model (1). Second, we selected the 1224

SNPs in the top 40 AD-risk genes by the AlzGene database. After applying quality control to the data, we included 1072 SNPs and their interactions with age in our analysis. A map of the linkage disequilibrium among the 1072 SNPs revealed a clear clustering pattern of SNPs (Fig. 1b). Specifically, the correlations among SNPs within each gene were large (> 0.7), whereas those among SNPs in different genes were relatively small.

We fitted the L2R2 model (1) to the ADNI data as follows: To determine the rank of $\mathbf{B}$, the L2R2 model was run up to $r = 10$ layers. By comparing the five different selection criteria in Zhu et al. (2014), we chose $r = 3$ layers as the optimal rank for the final data analysis. We ran the Gibbs sampler for 20,000 iterations after 20,000 burn-in iterations. For the G-SMuRFS method, we used the same $\mathbf{Y}$ matrix, combined the $\mathbf{W}$ and $\mathbf{X}$ matrices into a single predictor matrix, and then used the five-fold cross validation to choose the optimal penalty.

On the basis of the MCMC samples, we calculated the posterior median and median absolute deviations of $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{B}$, and used $1.4826 \times$ median absolute deviations to compute the robust standard errors for each element of $\mathbf{B}$. To determine the important coefficients, we used the standardized estimates $\mathbf{B}_{std}$, which were the absolute values of the estimates over their standard errors. To validate the low-rank structure of $\mathbf{B}$, we used the results of the top 10 genes and set the elements of $\mathbf{B}$ to zero if their absolute standardized estimates were larger than a threshold (Fig. 5). The sparsely distributed points along the horizontal and vertical directions indicated that the low-rank model was suitable for characterizing the association between the neuroimaging and genetic variables in the ADNI data set. We used a pragmatic approach to calculate the threshold for the Bayesian estimates, which resembled the Bonferroni correction in a frequentist approach. The

threshold was the $F(1 - 0.025/(pr/2 + dr + r))$, and F was the quantile function of standard normal distribution, where the denominator was the total number of parameters in the low-rank structure $\mathbf{U}_{int}$, $\mathbf{\Delta}$, and $\mathbf{V}$. $\mathbf{U}_{int}$ was the matrix of coefficients of interaction terms. The thresholds for the top 10 genes and top 40 genes were 3.912 and 4.339, respectively. We referred the SNPs that passed the thresholds as significant SNPs though the meaning is different from frequentist significance.

4.4. Results

We determined the longitudinal trajectories of the neuroimaging ROI volumes, the effects of the significant SNPs on such longitudinal trajectories, and the dynamic genetic effects on the trajectories. We estimated the trajectories of all standardized ROI volumes on the basis of $\hat{\mathbf{\Gamma}}$ and $\mathbf{W}$ (Fig. 6). ROIs with decreasing volumes occurred in many regions of the brain, including the left and middle temporal gyri, the left and superior temporal gyri, and the left and right amygdala. ROIs with increasing volumes occurred in the hollow areas of the brain and areas of white matter and included the left and right lateral ventricles, the left and right frontal lobe white matter, and the left and right temporal lobes. Moreover, the trajectories of most ROIs demonstrated structural symmetry.

Because hippocampal atrophy and ventricular enlargement are consistent findings in patients with AD and mild cognitive impairment (Apostolova et al., 2012), we depicted the effects of some SNPs on the longitudinal trajectories of the left and right lateral ventricles and the left and right hippocampi (Fig. 7). Let A and a denote the major and minor alleles of a specific SNP, respectively. For each ROI, we depicted the mean trajectory associated with a selected SNP (black solid curves) and the mean trajectory corresponding to

450 the three allelic combinations (i.e. AA , Aa , and aa) of the SNP (Fig. 7). The
451 minor allele of SNP rs10501608 increased the rate of hippocampal atrophy,
452 whereas the SNP rs1354106 minor allele had the opposite effect. Moreover,
453 the minor allele of SNP rs10501608 increased the rate of ventricular enlarge-
454 ment, yet the SNP rs10501604 minor allele had the opposite effect.

455 We plotted the primary SNP-age interaction effects, whose absolute stan-
456 dardized estimates were larger than the predetermined threshold, on the lon-
457 gitudinal trajectories corresponding to the top 10 genes (upper row) and the
458 top 40 genes (lower row) (Fig. 8). The plots revealed a generally symmetric
459 pattern for the left and right hemispheres, which indicated the genetic effects
460 on the longitudinal change of ROI volumes were symmetric. Some asymmet-
461 ric longitudinal genetic effects also occurred. For example, in the analysis of
462 the SNPs from the top 40 genes, only the left nucleus accumbens and the
463 right subthalamic nucleus were associated with SNP rs2273684, whereas their
464 contralateral counterparts were not substantially correlated with any SNP.
465 With exception of the left and right lateral ventricles, the genetic effects on
466 most ROIs in the left and right hemispheres occurred in the same direction.
467 Because both the number of parameters and the threshold for significance
468 increased with the number of top genes, the number of significant SNP-age
469 interactions decreased with the number of top genes. However, the effects
470 of significant SNP-age interaction detected for the top 40 genes were more
471 substantial.

472 The significant SNP-age interaction effects of the ROIs depicted in Fig. 8
473 formed SNP-specific networks that we visualized with BrainNet Viewer (Xia
474 et al., 2013) in Fig. 9 and Fig. 10. Among all the networks, each SNP-
475 specific network demonstrated spatially adjacent patterns. Some SNPs were
476 associated with many ROIs, although other SNPs were associated only with

ROIs in a specific brain area. The networks could be considerably different for different SNPs in a gene. In each network the sizes of genetic effects on various ROIs was relatively homogeneous, whereas the those across different networks may be substantially different. The most significant SNPs were close to the ventricular system. Some ROIs in the frontal lobe were also associated with SNPs that exhibited small growth effects, but ROIs near the prefrontal cortex were not strongly associated with the SNPs in the top 10 and top 40 genes.

We used MMA based on linear mixed effects models where each SNP, its interaction with age, and the covariates in $\mathbf{W}$ were regressed on each ROI volume. The p-values of the significant coefficients are smaller than $0.05/(pr/2 + dr + r)$ (Bonferroni correction with the number of parameter in the L2R2), even though the true number of regression parameter is much larger (pd). The number of significant coefficients is smaller if the true number of parameters is used for correction. We compared the important estimated associations with L2R2. The two ROIs identified in the case with top 10 genes were the left and right lateral ventricles. Twelve ROIs were found in the case with top 40 genes, ten of which were found by L2R2. The number of significant associations identified by MMA were much less than that of L2R2. L2R2 provided more power in identifying important pairs of ROIs and SNP-age interactions and revealed more SNPs that may alter the longitudinal trajectories of ROIs.

The results of G-SMuRFS were not directly comparable to those of L2R2 and MMA because G-SMuRFS does not provide standard error estimates or a cut-off for all pairs of association as in L2R2 and MMA. The importance of the SNPs and SNP-age interactions was ranked by the sum of the absolute values of the estimated coefficients across all ROIs. Nonetheless, we found

the top ROIs and SNP-age interactions, and some of them overlap with those from L2R2 in the case with top 40 genes.

We built the networks of ROIs with two criteria to study the structure of the longitudinal volumes of ROIs. Let $\mathbf{B}_{int}$ be the part of $\mathbf{B}$ corresponding to the SNP-age interactions, and $\mathbf{B}_{int,std}$ be the standardized $\mathbf{B}_{int}$. Firstly, we selected the ROIs corresponding to the large diagonal elements of $\mathbf{B}_{int,std}^T \mathbf{B}_{int,std}$ to form a network. The change of each ROI in the network was heavily associated with the SNPs. The network was a collection of important ROIs which were affected by the time-dependent SNP effects. Secondly, we constructed networks based on the absolute values of every columns of the standardized $\mathbf{V}_{std}$, which is the standardized $\mathbf{V}$. Each network was an important combination of the longitudinal ROI measurements, and was strongly associated with the joint effect of a weighted combination of SNPs and age-SNP interactions quantified by the corresponding column of $\mathbf{U}$. The top 10 ROIs in the networks were listed in Table 1 and Table 2 for the top 10 and 40 genes, respectively. Such networks were mostly symmetric for the left and right hemispheres, which also illustrated the genetic effects on the longitudinal volumes of ROIs are largely symmetric.

The imaging phenotypes provided quantitative measurements of the morphometric changes of the ROIs, which may characterize the underlying degenerative process of AD and capture information mirroring patients' diagnostic status. We used a generalized linear-mixed model with logit link to determine the associations between longitudinal diagnostic status and the ROI volume. We found that 70 of 93 ROIs and 69 ROI and age interactions were significantly correlated with longitudinal diagnostic status after applying the false discovery rate correction with the significant level 0.05. Tissues with ROIs that correlated with longitudinal diagnostic status included the hip-

531 pocampus, cingulate cortex, amygdala, and lateral ventricles, in addition to
532 many other areas of the brain. Moreover, most of the ROIs identified by the
533 L2R2 model in Figure 8 were among the 69 ROIs that demonstrated age-
534 dependent interactions. The SNPs used in the L2R2 model were previously
535 reported in a genome-wide association study of the diagnostic status of pa-
536 tients with AD. Our findings show that the longitudinal trajectories of many
537 ROIs are highly associated with the longitudinal patterns of diagnostic sta-
538 tus; therefore, these ROIs may serve as consistent surrogates for measuring
539 AD progression.

540 The L2R2 model elucidated some ROI-SNP pairs that were reported
541 in previous genome-wide association studies on the basis of diagnostic sta-
542 tus, cognitive scores, and imaging measurements. For example, the SNP
543 rs1354106, which occurs in the *CD33* gene, was associated with a slower de-
544 clining rate of the AD assessment scale cognitive score (Sherva et al., 2014).
545 Biffi et al. (2010) reported that SNP rs1408077, which occurs in the *CR1*
546 gene, was associated with disease status. Moreover, the SNP rs1408077 was
547 related to the entorhinal cortex thickness (Biffi et al., 2010) and hippocam-
548 pal atrophy (Lazaris et al., 2015). The SNP rs6088662, which occurs in the
549 *PRNP* gene, was associated with reduced hippocampal volume and cognitive
550 performance in healthy individuals (Li et al., 2016).

551 Multivariate imaging measurements may provide additional information
552 compared with univariate measurements, such as diagnostic status. Some
553 ROIs that were not associated with diagnostic status in the generalized linear
554 mixed model were associated with certain SNPs in the L2R2 model, such as
555 the left and right thalamus, right medial frontal gyrus, and right lateral
556 front-orbital gyrus. More ROI measurements yielded additional variation
557 in the SNPs compared with that of diagnostic status. Thus, integrating

558 multivariate imaging measurements with genetic variables and diagnostic
559 status may provide a clearer picture of the biological processes occurring
560 in AD (Zhao and Castellanos, 2016).

561 *4.5. Random effects*

562 We examined the usefulness of incorporating random effects into our
563 model by investigating the distribution of the estimated random effects over
564 all subjects for each ROI. We generated histograms and quantile-quantile
565 plots of the random effects obtained from two randomly selected ROIs from
566 the L2R2 model that were fitted to the top 10 genes (Fig. 11). The variances
567 of the estimated random intercepts were substantial compared with the scale
568 of the response variables, which coincided with the subject-specific intercepts
569 of the neuroimaging measurements (Fig. 2a). Moreover, these distributions
570 did not severely deviate from the normal distribution.

571 To statistically test the presence of the subject-specific random effects
572 **Zb**, we refitted the data with a restricted L2R2 model without the presence
573 of **Zb** and then compared the log likelihood ratio of the full and restricted
574 L2R2 models. The inclusion of random effects decreased the log likelihood
575 function by 14,000 (or 5%), and the number of parameters increased by
576 93. Therefore, incorporating random effects into the L2R2 model was useful
577 for characterizing the correlation structure of the imaging phenotypes and
578 improved the detection of the association between imaging phenotypes and
579 genetic variants.

580 **5. Discussion**

581 We have developed a Bayesian L2R2 model to determine the association
582 between longitudinal imaging responses and covariates with applications in

583 imaging genetic data. We used this model to approximate the large associa-
584 tion matrix. We combined a sparse latent factor model and random effects
585 to flexibly capture the complex spatiotemporal correlation structure. We in-
586 corporated splines to capture the effect of aging and combined a traditional
587 coefficient estimation with a low-rank approach. The L2R2 model dramati-
588 cally reduced the number of parameters to be sampled and tested, leading to
589 a remarkably faster sampling scheme and efficient inference. The simulation
590 studies demonstrated that the L2R2 model has higher power in detecting
591 important time-dependent and main genetic effects. Age-dependent genetic
592 effects are useful for characterizing age-related degenerative disorders like
593 AD. When strong genetic effects modify changes in brain morphology over
594 time, it is critically important to account for SNP-age interaction effects and
595 to improve the detection of time-invariant main genetic effects. Although our
596 findings confirmed the important role of well-known genes, such as APOE- $\epsilon 4$,
597 in the pathology of AD, they also elucidated other potential candidate genes
598 that warrant further investigation.

599 The L2R2 model effectively established associations between phenotypic
600 markers present in neuroimaging data and the presence of high-risk AD al-
601 leles. However, many considerations still merit further research. First, it is
602 important to consider the joint effect of genetic markers and environmental
603 factors on high-dimensional imaging phenotypes (Thomas, 2010). Second, it
604 is of great interest to incorporate rare variant genetic markers (Bansal et al.,
605 2010) with the L2R2 model. Third, the key features of the L2R2 model may
606 be adapted to more complex data structures (e.g. twin and family sequenc-
607 ing studies) and other parametric and semiparametric models. Fourth, the
608 L2R2 model may be extended to combine different imaging phenotypes calcu-
609 lated from other imaging modalities (e.g. diffusion tensor imaging, functional

610 MRI, and electroencephalography) in concomitant imaging and genetic stud-
 611 ies. Fifth, group structures among SNPs (i.e. gene-based grouping) could be
 612 incorporated to improve the efficiency of modeling-correlated SNPs by choos-
 613 ing group priors for $\mathbf{U}$.

614 One important feature of L2R2 is the inclusion of SNP-age interactions
 615 to model the genetic effects on the longitudinal change of ROIs. The inter-
 616 actions can also be viewed as the age-dependent genetic effects. The total
 617 age-varying genetic effects of the j th SNP x_{ij} is $\beta_j + \beta_{2j}t_{ij}$, which assumes
 618 that the effect is a linear function of age. This assumption can be relaxed by
 619 modeling the effects as more flexible functions of age, e.g., polynomial func-
 620 tions and nonparametric functions, which is known as the varying-coefficient
 621 models (Hastie and Tibshirani, 1993; Fan and Zhang, 1999, 2008). Varying-
 622 coefficient models have been used to study the time-dependent genetic effects
 623 and gene-environment interactions (Gong and Zou, 2012; Wu and Cui, 2013;
 624 Li et al., 2015). Compared to L2R2, these studies only modeled univariate
 625 response variables. Extending L2R2 to incorporate more flexible varying ge-
 626 netic effects is helpful to better understand the genetic impact on the complex
 627 diseases at different age and under various states of environment factors.

628 It is worth noting that in Fig. 2b, the volume of right hippocampal
 629 formation appeared to increase in the three-year visit for subjects with minor
 630 allele of rs768451. This phenomenon is less expected in studies of Alzheimer’s
 631 disease but it may be explained by the following reasons. We included healthy
 632 subjects, MCI and AD patients from the ADNI data set in our analysis. The
 633 volumes for the healthy subjects may be relatively stable. Moreover, several
 634 studies in the literature (Erickson et al., 2011; Leavitt et al., 2014; Niemann
 635 et al., 2014) reported that exercises were associated with the increase of
 636 hippocampal volume in older adults. Also, the number of subjects in the

three-year visit was smaller than those in the previous visits, and the average of the hippocampal volumes for these subjects was larger than the average for all subjects at the initial visit. The dropout during the last visit may be informative. These factors warrant further investigation in the future studies.

6. Acknowledgments

This material was based upon work partially supported by the NSF grant DMS-1127914 to the Statistical and Applied Mathematical Science Institute. The research of Dr. Zhu was partially supported by NSF grants SES-1357666 and DMS-1407655, NIH grant MH086633, and a grant from Cancer Prevention Research Institute of Texas.

Apostolova, L.G., Green, A.E., Babakchanian, S., Hwang, K.S., Chou, Y.Y., Toga, A.W., Thompson, P.M., 2012. Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment (mci), and alzheimer disease. *Alzheimer Disease & Associated Disorders* 26, 17–27.

Bansal, V., Libiger, O., Torkamani, A., Schork, N.J., 2010. Statistical analysis strategies for association studies involving rare variants. *Nature Reviews Genetics* 11, 773–785.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48. doi:10.18637/jss.v067.i01.

Bernal-Rusiel, J., Greve, D., Reuter, M. and Fischl, B., Sabuncu, M.R., 2013. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *NeuroImage* 66, 249–260.

- 661 Bhattacharya, A., Dunson, D.B., 2011. Sparse bayesian infinite factor mod-
662 els. *Biometrika* 98, 291–306.
- 663 Biffi, A., Anderson, C.D., Desikan, R.S., Sabuncu, M., Cortellini, L., Schman-
664 sky, N., Salat, D., Rosand, J., 2010. Genetic variation and neuroimag-
665 ing measures in Alzheimer Disease. *Archives of neurology* 67, 677–685.
666 doi:10.1001/archneurol.2010.108.
- 667 Breiman, L., Friedman, J., 1997. Predicting multivariate responses in multi-
668 ple linear regression. *Journal of the Royal Statistical Society* 59, 3–54.
- 669 Chen, H., Wang, Y., 2011. A penalized spline approach to functional mixed
670 effects models analysis. *Biometrics* 67, 861–870.
- 671 Chiang, M.C., Barysheva, M., Toga, A.W., Medland, S.E., Hansell, N.K.,
672 James, M.R., McMahon, K.L., de Zubicaray, G.I., Martin, N.G., Wright,
673 M.J., Thompson, P.M., 2011a. Bdnf gene effects on brain circuitry repli-
674 cated in 455 twins. *NeuroImage* 55, 448–454.
- 675 Chiang, M.C., McMahon, K.L., de Zubicaray, G.I., Martin, N.G., Hickie,
676 I., Toga, A.W., Wright, M.J., Thompson, P.M., 2011b. Genetics of white
677 matter development: A dti study of 705 twins and their siblings aged 12
678 to 29. *NeuroImage* 54, 2308–2317.
- 679 Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: An ap-
680 proach to regression analysis by local fitting. *Journal of the American*
681 *Statistical Association* 83, 596–610. doi:10.1080/01621459.1988.10478639.
- 682 Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and
683 penalties. *Statistical Science* 11, 89–121. doi:10.1214/ss/1038425655.

684 Erickson, K.I., Voss, M.W., Prakash, R.S., Basak, C., Szabo, A., Chaddock,
 685 L., Kim, J.S., Heo, S., Alves, H., White, S.M., Wojcicki, T.R., Mailey,
 686 E., Vieira, V.J., Martin, S.A., Pence, B.D., Woods, J.A., McAuley, E.,
 687 Kramer, A.F., 2011. Exercise training increases size of hippocampus and
 688 improves memory. *Proceedings of the National Academy of Sciences* 108,
 689 3017–3022. doi:10.1073/pnas.1015950108.

690 Fan, J., Zhang, W., 1999. Statistical estimation in varying coefficient models.
 691 *Ann. Statist.* 27, 1491–1518. doi:10.1214/aos/1017939139.

692 Fan, J., Zhang, W., 2008. Statistical methods with varying coefficient models.
 693 *Stat. Interface* 1, 179–195.

694 Gong, Y., Zou, F., 2012. Varying coefficient models for mapping quantitative
 695 trait loci using recombinant inbred intercrosses. *Genetics* 190, 475–486.
 696 doi:10.1534/genetics.111.132522.

697 Greven, S., Crainiceanu, C., Caffo, B., Reich, D., 2010. Longitudinal func-
 698 tional principal component analysis. *Electron. J. Statist.* 4, 1022–1054.

699 Guillaume, B., Hua, X., Thompson, P.M., Waldorp, L., Nichols, T.E., 2014.
 700 Fast and accurate modelling of longitudinal and repeated measures neu-
 701 roimaging data. *NeuroImage* 94, 287–302.

702 Guo, W., 2002. Functional mixed effects models. *Biometrics* 58, 121–128.

703 Hastie, T.J., Tibshirani, R.J., 1993. Varying-coefficient models. *J. Roy.*
 704 *Statist. Soc. B.* 55, 757–796.

705 Hibar, D., et al, J.L.S., 2011. Voxelwise gene-wide association study (vge-
 706 newas): Multivariate gene-based association testing in 731 elderly subjects.
 707 *Neuroimage* 56, 1875–1891.

- Hyun, J.W., Li, Y., Huang, C., Styner, M., Lin, W., Zhu, H., Alzheimer's
Disease Neuroimaging Initiative and others, 2016. Stgp: Spatio-temporal
gaussian process models for longitudinal neuroimaging data. *NeuroImage*
134, 550–562.
- Khondker, Z., Zhu, H., Chu, H., Lin, W., Ibrahim, J., 2013. The bayesian
covariance lasso. *Statistics and its Interface*, 6, 243–259.
- Klami, A., Virtanen, S., Kaski, S., 2013. Bayesian canonical correlation
analysis. *The Journal of Machine Learning Research* 14, 965–1003.
- Lang, S., Brezger, A., 2004. Bayesian P-Splines. *Journal of Computational
and Graphical Statistics* 13, 183–212.
- Lazaris, A., Hwang, K.S., Goukasian, N., Ramirez, L.M., Eastman,
J., Blanken, A.E., Teng, E., Gylys, K., Cole, G., Saykin, A.J.,
Shaw, L.M., Trojanowski, J.Q., Jagust, W.J., Weiner, M.W., Apos-
tolova, L.G., 2015. Alzheimer risk genes modulate the relationship be-
tween plasma apoE and cortical PiB binding. *Neurology: Genetics* 1.
doi:10.1212/NXG.0000000000000022.
- Leavitt, V.M., Cirnigliaro, C., Cohen, A., Farag, A., Brooks, M., Wecht,
J.M., Wylie, G.R., Chiaravalloti, N.D., DeLuca, J., Sumowski, J.F.,
2014. Aerobic exercise increases hippocampal volume and improves mem-
ory in multiple sclerosis: Preliminary findings. *Neurocase* 20, 695–697.
doi:10.1080/13554794.2013.841951. pMID: 24090098.
- Li, J., Wang, Z., Li, R., Wu, R., 2015. Bayesian group lasso for
nonparametric varying-coefficient models with application to functional

731 genome-wide association studies. *Ann. Appl. Stat.* 9, 640–664. URL:
732 <http://dx.doi.org/10.1214/15-AOAS808>, doi:10.1214/15-AOAS808.

733 Li, M., Luo, X.j., Landén, M., Bergen, S.E., Hultman, C.M., Li, X.,
734 Zhang, W., Yao, Y.G., Zhang, C., Liu, J., Mattheisen, M., Cichon,
735 S., Mühleisen, T.W., Degenhardt, F.A., Nöthen, M.M., Schulze, T.G.,
736 Grigoriu-Serbanescu, M., Li, H., Fuller, C.K., Chen, C., Dong, Q., Chen,
737 C., Jamain, S., Leboyer, M., Bellivier, F., Etain, B., Kahn, J.P., Henry,
738 C., Preisig, M., Kutalik, Z., Castelao, E., Wright, A., Mitchell, P.B.,
739 Fullerton, J.M., Schofield, P.R., Montgomery, G.W., Medland, S.E., Gor-
740 don, S.D., Martin, N.G., Consortium, M., Group, T.S.B.S., Rietschel,
741 M., Liu, C., Kleinman, J.E., Hyde, T.M., Weinberger, D.R., Su, B.,
742 2016. Impact of a cis-associated gene expression SNP on chromosome
743 20q11.22 on bipolar disorder susceptibility, hippocampal structure and
744 cognitive performance. *The British Journal of Psychiatry* 208, 128–137.
745 doi:10.1192/bjp.bp.114.156976.

746 Li, Y., Gilmore, J.H., Shen, D., Styner, M., Lin, W., Zhu, H., 2013. Multi-
747 scale adaptive generalized estimating equations for longitudinal neuroimag-
748 ing data. *NeuroImage* 72, 91–105.

749 Marttinen, P., Pirinen, M., Sarin, A.P., Gillberg, J., Kettunen, J., Surakka,
750 I., Kangas, A.J., Soininen, P., O'Reilly, P., Kaakinen, M., Khnen, M.,
751 Lehtimäki, T., Ala-Korpela, M., Raitakari, O.T., Salomaa, V., Järvelin,
752 M.R., Ripatti, S., Kaski, S., 2014. Assessing multivariate gene-metabolome
753 associations with rare variants using Bayesian reduced rank regression.
754 *Bioinformatics* 30, 2026–2034. doi:10.1093/bioinformatics/btu140.

- 755 Meinshausen, N., Bühlmann, P., 2010. Stability selection. *Journal of the*
756 *Royal Statistical Society: Series B (Statistical Methodology)* 72, 417–473.
- 757 Morris, J.S., Carroll, R.J., 2006. Wavelet-based functional mixed models. *J.*
758 *R. Stat. Soc. Ser. B Stat. Methodol.* 68, 179–199.
- 759 Niemann, C., Godde, B., Voelcker-Rehage, C., 2014. Not only car-
760 diovascular, but also coordinative exercise increases hippocampal vol-
761 ume in older adults. *Frontiers in Aging Neuroscience* 6, 170.
762 doi:10.3389/fnagi.2014.00170.
- 763 Paus, T., 2010. Population neuroscience: Why and how. *Human Brain*
764 *Mapping* 31, 891–903.
- 765 Peper, J.S., Brouwer, R.M., Boomsma, D.I., Kahn, R.S., Pol, H.E.H., 2007.
766 Genetic influences on human brain structure: A review of brain imaging
767 studies in twins. *Human Brain Mapping* 28, 464–473.
- 768 Rothman, A.J., Levina, E., Zhu, J., 2010. Sparse multivariate regression with
769 covariance estimation. *Journal of Computational and Graphical Statistics*
770 19, 947–962.
- 771 Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric regression.
772 Cambridge University Press, Cambridge. doi:10.1017/CBO9780511755453.
- 773 Saykin, A.J., Shen, L., Yao, X., Kim, S., Nho, K., Risacher, S.L., Ramanan,
774 V.K., Foroud, T.M., Faber, K.M., Sarwar, N., Munsie, L.M., Hu, X.,
775 Soares, H.D., Potkin, S.G., M, T.P., Kauwe, J.S., Kaddurah-Daouk, R.,
776 Green, R.C., Toga, A.W., Weiner, M.W., Alzheimer’s Disease Neuroimag-
777 ing Initiative, 2015. Genetic studies of quantitative mci and ad phenotypes

778 in adni: Progress, opportunities, and plans. *Alzheimer's & Dementia* 11,
779 792–814.

780 Scharinger, C., Rabl, U., Sitte, H.H., Pezawas, L., 2010. Imaging genetics of
781 mood disorders. *NeuroImage* 53, 810–821.

782 Shen, D.G., Davatzikos, C., 2004. Measuring temporal morphological changes
783 robustly in brain mr images via 4-dimensional template warping. *NeuroIm-*
784 *age* 21, 1508–1517.

785 Shen, L., Kim, S., Risacher, S.L., Nho, K., Swaminathan, S., West, J.,
786 Foroud, T.M., Pankratz, N.D., Moore, J.H., Sloan, S.D., Huentelman,
787 M.J., Craig, D.W., DeChairo, B.M., Potkin, S.G., Jack, C.R., Weiner,
788 M.W., Saykin, A.J., ADNI., 2010. Whole genome association study of
789 brain-wide imaging phenotypes for identifying quantitative trait loci in
790 mci and ad: A study of the adni cohort. *Neuroimage* 53, 1051–1063.

791 Sherva, R., Tripodis, Y., Bennett, D.A., Chibnik, L.B., Crane, P.K.,
792 de Jager, P.L., Farrer, L.A., Saykin, A.J., Shulman, J.M., Naj, A.,
793 Green, R.C., 2014. Genome-wide association study of the rate of cog-
794 nitive decline in Alzheimer's disease. *Alzheimer's & Dementia* 10, 45–52.
795 doi:10.1016/j.jalz.2013.01.008.

796 Silver, M., Janousova, E., Hue, X., Thompson, P., Montana, G., 2012. Identi-
797 fication of gene pathways implicated in alzheimer's disease using longitudi-
798 nal imaging phenotypes with sparse regression. *Neuroimage* 63, 1681–1694.

799 Thomas, D., 2010. Gene–environment-wide association studies: emerging
800 approaches. *Nature Reviews Genetics* 11, 259–272.

- 801 Verbeke, G., Molenberghs, G., 2009. Linear Mixed Models for Longitudinal
802 Data. Springer Series in Statistics, Springer New York.
- 803 Vounou, M., Janousova, E., Wolz, R., Stein, J.L., Thompson, P.M., Rueckert,
804 D., Montana, G., ADNI, 2011. Sparse reduced-rank regression detects
805 genetic associations with voxel-wise longitudinal phenotypes in alzheimer’s
806 disease. *NeuroImage* 60, 700–716.
- 807 Vounou, M., Nichols, T.E., Montana, G., the Alzheimer’s Disease Neuroimag-
808 ing Initiative, 2010. Discovering genetic associations with high-dimensional
809 neuroimaging phenotypes: A sparse reduced-rank regression approach.
810 *Neuroimage* 53, 1147–1159.
- 811 Wang, H., Nie, F., Huang, H., Kim, S., Nho, K., Risacher, S.L., Saykin,
812 A.J., Shen, L., 2011. Identifying quantitative trait loci via group-sparse
813 multitask regression and feature selection: an imaging genetics study of
814 the adni cohort. *Bioinformatics* 28, 229–237.
- 815 Wood, S.N., 2006. Generalized additive models: An introduction with R.
816 Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton,
817 FL.
- 818 Wu, C., Cui, Y., 2013. A novel method for identifying non-
819 linear gene–environment interactions in case–control associ-
820 ation studies. *Human Genetics* 132, 1413–1425. URL:
821 [http://dx.doi.org/10.1007/s00439-](http://dx.doi.org/10.1007/s00439-013-1350-z)
822 [013-1350-z](http://dx.doi.org/10.1007/s00439-013-1350-z), doi:10.1007/s00439-
013-1350-z.
- 823 Wu, H.L., Zhang, J.T., 2006. Nonparametric Regression Methods for Longi-
824 tudinal Data Analysis. John Wiley & Sons, Inc., Hoboken, New Jersey.

- 825 Xia, M., Wang, J., He, Y., 2013. Brainnet viewer: A network vi-
826 sualization tool for human brain connectomics. PLoS ONE 8, 1–15.
827 doi:10.1371/journal.pone.0068910.
- 828 Zhang, Y., Xu, Z., Shen, X., Pan, W., 2014. Testing for associ-
829 ation with multiple traits in generalized estimation equations, with
830 application to neuroimaging data. NeuroImage 96, 309 – 325.
831 doi:10.1016/j.neuroimage.2014.03.061.
- 832 Zhao, Y., Castellanos, F.X., 2016. Annual research review: Discovery science
833 strategies in studies of the pathophysiology of child and adolescent psy-
834 chiatric disorders - promises and limitations. Journal of Child Psychology
835 and Psychiatry 57, 421–439. doi:10.1111/jcpp.12503.
- 836 Zhu, H., Khondker, Z., Lu, Z., Ibrahim, J.G., 2014. Bayesian Generalized
837 Low Rank Regression Models for Neuroimaging Phenotypes and Genetic
838 Markers. Journal of the American Statistical Association 109, 977–990.
839 doi:10.1080/01621459.2014.923775.
- 840 Zipunnikov, V., Greven, S., Shou, H., Caffo, B.S., Reich, D.S., Crainiceanu,
841 C.M., 2014. Longitudinal high-dimensional principal components analysis
842 with application to diffusion tensor imaging of multiple sclerosis. Ann.
843 Appl. Stat. 8, 2175–2202. doi:10.1214/14-AOAS748.

844 7. Appendix: technical details

845 7.1. Matrix formulation of L2R2

The L2R2 can be rewritten in the matrix form (5)

$$\mathbf{Y} = \mathbf{X}\mathbf{U}\mathbf{\Delta}\mathbf{V}^T + \mathbf{W}\mathbf{\Gamma} + \mathbf{Z}\mathbf{b} + \mathbf{E},$$

846 where $\mathbf{y}_i(t) = (y_{i1}(t), \dots, y_{id}(t))^T$, $\mathbf{y}_i = (\mathbf{y}_i(t_{i1}), \dots, \mathbf{y}_i(t_{im_i}))^T$, $\mathbf{Y} = (\mathbf{y}_1^T, \dots,$
847 $\mathbf{y}_n^T)^T$, $\mathbf{1}_{m_i}$ be a $m_i \times 1$ vector of ones, and $\mathbf{t}_{m_i} = (t_{i1}, \dots, t_{i,m_i})^T$. Let $\mathbf{x}_i^*$ be
848 the SNPs for the i th subject, $\mathbf{X}_{main} = (\mathbf{x}_1^* \otimes \mathbf{1}_{m_1}^T, \dots, \mathbf{x}_n^* \otimes \mathbf{1}_{m_n}^T)^T$, $\mathbf{X}_{int} =$
849 $(\mathbf{x}_1^* \otimes \mathbf{t}_{m_1}^T, \dots, \mathbf{x}_n^* \otimes \mathbf{t}_{m_n}^T)^T$, and $\mathbf{X} = (\mathbf{X}_{main}, \mathbf{X}_{int})$. Moreover, let $\mathbf{Z}_i =$
850 $(\mathbf{z}_i(t_{i1}), \dots, \mathbf{z}_i(t_{im_i}))^T$, $\mathbf{Z}$ be an $N \times np_b$ block diagonal matrix with diagonal
851 elements $\mathbf{Z}_1, \dots, \mathbf{Z}_n$, and $\mathbf{b} = (\mathbf{b}_{ik})$ be an $np_b \times d$ block matrix with elements
852 $\mathbf{b}_{ik}$. We defined $\mathbf{w}_i(t) = (\mathbf{w}_1(t)^T, \mathbf{w}_{2,i}(t)^T)^T$, $\mathbf{W}_i = (\mathbf{w}_i(1), \dots, \mathbf{w}_i(t_{im_i}))^T$, and
853 $\mathbf{W} = (\mathbf{W}_1^T, \dots, \mathbf{W}_n^T)^T$. Let $\mathbf{\epsilon}_i(t) = (\epsilon_{i1}(t), \dots, \epsilon_{id}(t))^T$, $\mathbf{E}_i = (\mathbf{\epsilon}_i(t_{i1}), \dots,$
854 $\mathbf{\epsilon}_i(t_{im_i}))^T$, and $\mathbf{E} = (\mathbf{E}_1^T, \dots, \mathbf{E}_n^T)^T$. Consequently, $\mathbf{Y}$ is an $N \times d$ matrix
855 of imaging measurements; $\mathbf{X}$ is an $N \times p$ matrix of genetic predictors; $\mathbf{W}$
856 is an $N \times q$ matrix of covariates with fixed effects; $\mathbf{Z}$ is an $N \times np_b$ ma-
857 trix of covariates with random effects; and $\mathbf{E}$ is an $N \times d$ matrix of error
858 terms, where $q = q_1 + q_2$. $\mathbf{\Gamma}$ is a $q \times d$ matrix and its first q_1 row consists of
859 $\mathbf{\Gamma}_1 = (\gamma_{1,1}, \dots, \gamma_{1,d})$, and its last q_2 columns consist of $\mathbf{\Gamma}_2 = (\gamma_{2,1}, \dots, \gamma_{2,d})$.
860 In model (5), we primarily made statistical inferences from $\mathbf{B}$ (equivalently
861 $\mathbf{U}, \mathbf{\Delta}, \mathbf{V}$) and $\mathbf{\Gamma}$.

862 7.2. Priors

We consider priors on the elements of $\mathbf{B}$. Let $\text{Ga}(a, b)$ be a gamma dis-
tribution with scale a and shape b . We choose L_2 priors on the parameters
at each layer $\mathbf{B}_l$ as follows:

$$\delta_l \sim N(0, \tau_\delta^{-1}), \quad \tau_\delta \sim \text{Ga}(a_0, b_0), \quad \mathbf{u}_l \sim N_p(0, p^{-1}I_p), \text{ and } \mathbf{v}_l \sim N_d(0, d^{-1}I_d),$$

863 where a_0 and b_0 are prespecified hyperparameters. Although, we have used
 864 the same precision parameter for each element of $\mathbf{u}_l$, group information can be
 865 incorporated by choosing separate precision parameters for each group. The
 866 number of predictors p (or d) is included in the hyperprior of $\mathbf{u}_l$ (or $\mathbf{v}_l$) to have
 867 a positive definite covariance matrix of high dimensional $\mathbf{u}_l$ (or $\mathbf{v}_l$). Moreover,
 868 this data driven approach for the priors of $\mathbf{u}_l$ and $\mathbf{v}_l$ requires no additional
 869 hyperparameters to choose. because we standardize all predictors to have
 870 zero mean and unit variance, a single prior suffices for all elements of $\mathbf{u}_l$.
 871 Moreover, because we rescale all of the responses, we use the same dispersion
 872 for all components of $\mathbf{v}_l$. Since we focus on exploiting the potential two-way
 873 correlations among the estimated coefficients, we choose the L_2 priors, which
 874 tend to borrow strength from correlated neighbors and force the coefficients
 875 towards each other to produce two highly correlated coefficients. Moreover,
 876 the posterior computations are simpler and faster under the L_2 priors.

We consider priors on the elements of $\mathbf{\Gamma} = (\gamma_{jk})$. We also choose the L_2
 prior on the γ_{jk} s as follows:

$$\gamma_{jk} \sim N(0, \tau_\gamma^{-1}) \text{ and } \tau_\gamma \sim \text{Ga}(c_0, d_0),$$

where c_0 and d_0 are hyperparameters. For the subject-specific random coefficients, we also choose independent and identically distributed normal priors as

$$\tau_b \sim \text{Ga}(e_0, f_0),$$

877 where e_0 and f_0 are hyperparameters.

878 For the sparse factor model (4), we assign the multiplicative gamma process shrinkage prior (Bhattacharya and Dunson, 2011) on $\mathbf{\Lambda}$ to automatically
 879 determine the dimension of the factors needed to characterize the error covariance structure. The shrinkage prior increasingly shrinks the factor loadings
 881

882 towards zero with the column index and avoids identifiability of the order of
 883 the factor in $\boldsymbol{\eta}_i(t)$. Specifically, these priors are summarized as follows:

$$\begin{aligned}\mathbf{\Lambda} &= \{\lambda_{kh}\}, k = 1, \dots, d, h = 1, \dots, \infty, \\ \lambda_{kh} | \phi_{kh}, \tau_{\lambda h} &\sim N(0, \phi_{kh}^{-1} \tau_{\lambda h}^{-1}), \phi_{kh} \sim \text{Ga}(v/2, v/2), \tau_{\lambda h} = \prod_{l=1}^h \psi_l, \\ \psi_1 &\sim \text{Ga}(a_1, 1), \psi_l \sim \text{Ga}(a_2, 1), l \geq 2, \sigma_{\xi, k}^{-2} \sim \text{Ga}(a_{\sigma k}, b_{\sigma k}),\end{aligned}$$

884 where v , a_1 , a_2 , $a_{\sigma k}$, and $b_{\sigma k}$ are prefixed hyperparameters, $\tau_{\lambda h}$ is a global
 885 shrinkage parameter for the h -th column, and the ϕ_{kh} s are local shrinkage
 886 parameters for the elements in the h -th column. When $a_2 > 1$, the $\tau_{\lambda h}$ s
 887 increase stochastically with the column index h , which shrinks the elements
 888 of the loading matrix to zero as the column index progresses and determines
 889 the effective dimension of $\mathbf{\Lambda}$.

890 7.3. Posterior computation

The joint posterior for the L2R2 model with the above priors can be written as

$$\begin{aligned}p(\mathbf{U}, \mathbf{\Delta}, \mathbf{V}, \mathbf{\Gamma}, \mathbf{b}, \mathbf{\Theta} | \text{data}) &\propto \tau_{\delta}^{a_0 + \frac{1}{2}r - 1} \tau_{\gamma}^{c_0 + \frac{1}{2}q - 1} \tau_b^{e_0 + \frac{1}{2}np_b d - 1} e^{-b_0 \tau_{\delta} - d_0 \tau_{\gamma} - f_0 \tau_b} \quad (6) \\ &\exp \left\{ -\frac{1}{2} \sum_{l=1}^r (\tau_{\delta} \delta_l^2 + d \mathbf{v}_l^T \mathbf{v}_l + p \mathbf{u}_l^T \mathbf{u}_l) \right\} + \exp \left\{ -\frac{1}{2} \text{Tr}(\tau_{\gamma} \mathbf{\Gamma} \mathbf{\Gamma}^T + \tau_b \mathbf{b} \mathbf{b}^T) \right\} \\ &\exp \left\{ -\frac{1}{2} \text{Tr}(\mathbf{Y} - \mathbf{XU} \mathbf{\Delta} \mathbf{V} - \mathbf{W} \mathbf{\Gamma} - \mathbf{Zb}) \mathbf{\Theta} (\mathbf{Y} - \mathbf{XU} \mathbf{\Delta} \mathbf{V} - \mathbf{W} \mathbf{\Gamma} - \mathbf{Zb})^T \right\},\end{aligned}$$

891 where $\text{Tr}(\cdot)$ is the trace of the matrix.

892 We propose a straightforward Gibbs sampler for posterior computation,
 893 which converges rapidly. Starting from the initiation step, the Gibbs sampler
 894 at each iteration proceeds as follows:

1. For $l = 1, \dots, r$, update $\mathbf{u}_l$ from the full conditional distributions

$$p(\mathbf{u}_l | -) \sim N_p(\delta_l \Sigma_{\mathbf{u}_l} \mathbf{X}^T \mathbf{Y}_{B,l} \Theta \mathbf{v}_l, \Sigma_{\mathbf{u}_l}),$$

895 where $\mathbf{Y}_{B,l} = \mathbf{Y} - \mathbf{X} \sum_{l' \neq l}^r \mathbf{B}_{l'} - \mathbf{W} \Gamma - \mathbf{Z} \mathbf{b}$ and $\Sigma_{\mathbf{u}_l} = \{p \mathbf{I}_p + \delta_l^2 (\mathbf{v}_l^T \Theta \mathbf{v}_l) \mathbf{X}^T \mathbf{X}\}^{-1}$.

2. Update $\mathbf{v}_l$ from

$$p(\mathbf{v}_l | -) \sim N_d(\delta_l \Sigma_{\mathbf{v}_l} \Theta \mathbf{Y}_{B,l}^T \mathbf{X} \mathbf{u}_l, \Sigma_{\mathbf{v}_l}),$$

896 where $\Sigma_{\mathbf{v}_l} = \{d \mathbf{I}_d + \delta_l^2 (\mathbf{u}_l^T \mathbf{X}^T \mathbf{X} \mathbf{u}_l) \Theta\}^{-1}$.

3. Update δ_l from

$$p(\delta_l | -) \sim N(\sigma_{\delta_l}^2 \mathbf{u}_l^T \mathbf{X}^T \mathbf{Y}_{B,l} \Theta \mathbf{v}_l, \sigma_{\delta_l}^2),$$

897 where $\sigma_{\delta_l}^2 = \{\tau_\delta + (\mathbf{v}_l^T \Theta \mathbf{v}_l) (\mathbf{u}_l^T \mathbf{X}^T \mathbf{X} \mathbf{u}_l)\}^{-1}$.

4. Update τ_δ from

$$p(\tau_\delta | -) \sim \text{Ga}(a_0 + 0.5r, b_0 + 0.5 \sum_{l=1}^r \delta_l^2).$$

5. Update γ_k (i.e. the k th column of Γ) from

$$p(\gamma_k | -) \sim N_d(\Sigma_{\gamma_k} \mathbf{W}^T \{\mathbf{y}_{\Gamma,k} \theta_{kk} + (\mathbf{Y}_{\Gamma,-k} - \mathbf{W} \Gamma_{-k}) \boldsymbol{\theta}_k, \Sigma_{\gamma_k}),$$

898 where $\Sigma_{\gamma_k} = \{\theta_{kk} \mathbf{W}^T \mathbf{W} + \tau_\gamma \mathbf{I}_q\}^{-1}$. $\mathbf{y}_{\Gamma,k}$ is the k th column of $\mathbf{Y}_\Gamma =$
 899 $\mathbf{Y} - \mathbf{X} \mathbf{B} - \mathbf{Z} \mathbf{b}$; $\mathbf{Y}_{\Gamma,-k}$ is the matrix after dropping the k th column of
 900 $\mathbf{Y}_\Gamma$; θ_{kk} is the element for the k th row and k th column of Θ ; $\boldsymbol{\theta}_k$ is
 901 the k th column of Θ after dropping θ_{kk} ; and Θ_{-kk} is the matrix after
 902 dropping k th row and k th column of Θ .

6. Update τ_γ from

$$p(\tau_\gamma | -) \sim \text{Ga}(c_0 + 0.5qd, d_0 + 0.5 \sum_{j=1}^q \sum_{k=1}^d \gamma_{jk}^2).$$

7. Let $\tilde{\mathbf{b}}_k$ be the k th column of $\mathbf{b}$. Update $\tilde{\mathbf{b}}_k$ from

$$p(\tilde{\mathbf{b}}_k|-) \sim N(\Sigma_{b_k} \mathbf{Z}^T \{\mathbf{y}_{b,k} - (\mathbf{Y}_{b,-k} - \mathbf{Z}\mathbf{b}_{-k})\boldsymbol{\theta}_k\}, \Sigma_{b_k}^2),$$

903 where $\Sigma_{b_k} = (\theta_{kk} \mathbf{Z}^T \mathbf{Z} + \tau_b I_{np_b})^{-1}$. $\mathbf{y}_{b,k}$ is the k th column of $\mathbf{Y}_b =$
 904 $\mathbf{Y} - \mathbf{XB} - \mathbf{W}\boldsymbol{\Gamma}$, and $\mathbf{Y}_{b,-k}$ is the matrix after dropping the k th column
 905 of $\mathbf{Y}_b$.

8. Update τ_b from

$$p(\tau_b|-) \sim \text{Ga}(e_o + 0.5nq^*d, d_0 + 0.5 \sum_{i=1}^n \sum_{j^*=1}^{q^*} \sum_{k=1}^d b_{ij^*k}^2).$$

9. Let k_* be a integer large enough to approximate the number of factors
 in the factor model (4), and $\boldsymbol{\epsilon}_{it} = \boldsymbol{\Lambda}^* \boldsymbol{\eta}_{it} + \boldsymbol{\xi}_{it}$ be the corresponding
 approximation. Update $\boldsymbol{\Lambda}_k^*$, which is the k th row of $\boldsymbol{\Lambda}^*$, from its full
 conditional distribution

$$p(\boldsymbol{\Lambda}_k^*|-) \sim \mathbf{N}((\sigma_{\xi,k}^{-2} \boldsymbol{\Omega}^T \boldsymbol{\Omega} + \mathbf{D}_k^{-1})^{-1} \boldsymbol{\Omega}^T \sigma_k^{-2} \mathbf{E}_k, (\sigma_{\xi,k}^{-2} \boldsymbol{\Omega}^T \boldsymbol{\Omega} + \mathbf{D}_k^{-1})^{-1}),$$

906 where $\boldsymbol{\Omega} = (\boldsymbol{\eta}_{11}, \dots, \boldsymbol{\eta}_{1m_1}, \dots, \boldsymbol{\eta}_{nm_n})^T$, $\mathbf{E}_k$ is the k th column of $\mathbf{E} = \mathbf{Y} -$
 907 $\mathbf{XB} - \mathbf{W}\boldsymbol{\Gamma} - \mathbf{Z}\mathbf{b}$, and $D_k = \text{diag}(\phi_{k1}^{-1} \tau_{\lambda 1}^{-1}, \dots, \phi_{kk_*}^{-1} \tau_{\lambda k_*}^{-1})$ for $k = 1, \dots, d$.

10. Update ϕ_{kh} from its full conditional distribution

$$p(\phi_{kh}|-) \sim \text{Ga}(\frac{v+1}{2}, \frac{v + \lambda_{kh}^2 \tau_{\lambda h}}{2}).$$

11. Update $\sigma_{\xi,k}^{-2}$, $k = 1, \dots, d$, from its full conditional distribution

$$p(\sigma_{\xi,k}^{-2}|-) \sim \text{Ga}(a_{\sigma k} + \frac{n}{2}, b_{\sigma k} + \frac{1}{2} \sum_{i=1}^n \sum_{t=1}^{m_i} (y_{ik} - \boldsymbol{\Lambda}_k^{*\top} \boldsymbol{\eta}_{it})^2).$$

12. Update ψ_1 from its full conditional distribution

$$p(\psi_1|-) \sim \text{Ga}(a_1 + \frac{1}{2} dk_*, 1 + \frac{1}{2} \sum_{g=1}^{k_*} \tau_{\lambda g}^{(h)} \sum_{k=1}^d \phi_{kg} \lambda_{kg}^2).$$

13. Update ψ_h , $h \geq 2$ from its full conditional distribution

$$p(\psi_h|-) \sim \text{Ga}(a_2 + \frac{1}{2}d(k_* - h + 1), 1 + \frac{1}{2} \sum_{g=h}^{k_*} \tau_{\lambda g}^{(h)} \sum_{k=1}^d \phi_{kg} \lambda_{kg}^2),$$

908 where $\tau_{\lambda g}^{(h)} = \prod_{t=1, t \neq h}^g \psi_t$ for $h = 1, \dots, k_*$.

14. Update η_i for $i = 1, \dots, n$ from the conditionally independent posteriors

$$p(\boldsymbol{\eta}_{it}|-) \sim N((I_{k_*} + \boldsymbol{\Lambda}^T \boldsymbol{\Sigma}_{\xi}^{-1} \boldsymbol{\Lambda})^{-1} \boldsymbol{\Lambda}_{k_*}^T \boldsymbol{\Sigma}_{\xi}^{-1} \boldsymbol{\epsilon}_i, (I_{k_*} + \boldsymbol{\Lambda}^T \boldsymbol{\Sigma}_{\xi}^{-1} \boldsymbol{\Lambda})^{-1}),$$

909 where $\boldsymbol{\epsilon}_i$ is the i th row of $\mathbf{E}$.

910 *7.4. Full conditional distributions of L2R2*

911 *7.4.1. Full conditional distribution for Δ*

From equation (6), we can write

$$\begin{aligned} p(\delta_l|-) &\propto \exp\left(-\frac{1}{2}\tau_{\delta}\delta_l^2\right) \text{etr}\left\{(\mathbf{Y}_{B,l} - \delta_l \mathbf{X}\mathbf{u}_l \mathbf{v}_l) \boldsymbol{\Theta} (\mathbf{Y}_{B,l} - \delta_l \mathbf{X}\mathbf{u}_l \mathbf{v}_l)^T\right\} \\ &\propto \exp\left(\delta_l \mathbf{u}_l^T \mathbf{X}^T \mathbf{Y}_{B,l} \boldsymbol{\Theta} \mathbf{v}_l - \frac{1}{2}\delta_l^2(\tau_{\delta} + \mathbf{v}_l \boldsymbol{\Theta}^T \mathbf{v}_l)(\mathbf{u}_l^T \mathbf{X}^T \mathbf{X} \mathbf{u}_l)\right); \\ p(\tau_{\delta}|-) &\propto \tau_{\delta}^{a_0 + \frac{1}{2}r-1} \exp\left(-\frac{1}{2}(b_0 + \sum_{l=1}^r \tau_{\delta} \delta_l^2)\right). \end{aligned}$$

This implies

$$p(\delta_l|-) \sim N(\sigma_{\delta_l}^2 \mathbf{u}_l^T \mathbf{X}^T \mathbf{Y}_{B,l} \boldsymbol{\Theta} \mathbf{v}_l, \sigma_{\delta_l}^2) \text{ and } p(\tau_{\delta}|-) \sim \text{Ga}(a_0 + \frac{1}{2}r, b_0 + \frac{1}{2} \sum_{l=1}^r \delta_l^2),$$

912 where $\sigma_{\delta_l}^2 = \{\tau_{\delta} + (\mathbf{v}_l \boldsymbol{\Theta}^T \mathbf{v}_l)(\mathbf{u}_l^T \mathbf{X}^T \mathbf{X} \mathbf{u}_l)\}^{-1}$ and $\mathbf{Y}_{B,l} = \mathbf{Y} - \mathbf{W}\boldsymbol{\Gamma} - \mathbf{Z}\mathbf{b} -$
 913 $\sum_{l' \neq l} \delta_{l'} \mathbf{X}\mathbf{u}_{l'} \mathbf{v}_{l'}^T$.

914 7.4.2. Full conditional distribution for $\mathbf{U}$

From equation (6), we can write

$$\begin{aligned} p(\mathbf{u}_l|-) &\propto \exp\left(-\frac{1}{2}p\mathbf{u}_l^T\mathbf{u}_l\right) \text{etr}\left\{(\mathbf{Y}_{B,l} - \delta_l\mathbf{X}\mathbf{u}_l\mathbf{v}_l)\boldsymbol{\Theta}(\mathbf{Y}_{B,l} - \delta_l\mathbf{X}\mathbf{u}_l\mathbf{v}_l)^T\right\} \\ &\propto \text{etr}\left(\mathbf{u}_l^T\delta_l\mathbf{X}^T\mathbf{Y}_{B,l} - \frac{1}{2}\mathbf{u}_l^T(p\mathbf{I}_p + \delta_l^2\mathbf{v}_l^T\boldsymbol{\Theta}\mathbf{v}_l\mathbf{X}^T\mathbf{X})\mathbf{u}_l\right). \end{aligned}$$

This implies

$$p(\mathbf{u}_l|-) \sim N_p(\delta_l\boldsymbol{\Sigma}_{\mathbf{u}_l}\mathbf{X}^T\mathbf{Y}_{B,l}\boldsymbol{\Theta}\mathbf{v}_l, \boldsymbol{\Sigma}_{\mathbf{u}_l}),$$

915 where $\boldsymbol{\Sigma}_{\mathbf{u}_l} = \{p\mathbf{I}_p + \delta_l^2(\mathbf{v}_l^T\boldsymbol{\Theta}\mathbf{v}_l)\mathbf{X}^T\mathbf{X}\}^{-1}$.

916

917 7.4.3. Full conditional distribution for $\mathbf{V}$

From equation (6), we can write

$$\begin{aligned} p(\mathbf{v}_l|-) &\propto \exp\left(-\frac{1}{2}\mathbf{v}_l^T(d\mathbf{I}_d)\mathbf{v}_l\right) \text{etr}((\mathbf{Y}_{B,l} - \delta_l\mathbf{X}\mathbf{u}_l\mathbf{v}_l)\boldsymbol{\Theta}(\mathbf{Y}_{B,l} - \delta_l\mathbf{X}\mathbf{u}_l\mathbf{v}_l)^T) \\ &\propto \exp\left(\mathbf{v}_l^T\delta_l\boldsymbol{\Theta}\mathbf{Y}_{B,l}^T\mathbf{X}\mathbf{u}_l - \frac{1}{2}\mathbf{v}_l^T(p\mathbf{I}_p + \delta_l^2\mathbf{u}_l^T\mathbf{X}^T\mathbf{X}\mathbf{u}_l\mathbf{v}_l\mathbf{X}^T\mathbf{X})\mathbf{v}_l\right). \end{aligned}$$

This gives

$$p(\mathbf{v}_l|-) \sim N_d(\delta_l\boldsymbol{\Sigma}_{\mathbf{v}_l}\boldsymbol{\Theta}\mathbf{Y}_{B,l}^T\mathbf{X}\mathbf{u}_l, \boldsymbol{\Sigma}_{\mathbf{v}_l}),$$

918 where $\boldsymbol{\Sigma}_{\mathbf{v}_l} = \{d\mathbf{I}_d + \delta_l^2(\mathbf{u}_l^T\mathbf{X}^T\mathbf{X}\mathbf{u}_l)\boldsymbol{\Theta}\}^{-1}$.

919 7.4.4. Sampling $\boldsymbol{\Gamma}$ by columns

Sampling the coefficient matrix one element at a time would be time consuming and less attractive in high-dimensional settings. Another approach is to convert the whole matrix into a vector and sample the vector at once, but this requires a covariance matrix with dimension $qd \times qd$, which can be quite large and requires considerable memory making it infeasible. We propose a columnwise sampling scheme, which allows for computationally

efficient sampling with a feasible dimension of the conditional covariance matrix (Khondker et al., 2013). Let $\mathbf{Y}_\Gamma = \mathbf{Y} - \mathbf{XU}\Delta\mathbf{V} - \mathbf{Zb}$, we can write $\mathbf{Y}_\Gamma = \mathbf{W}\Gamma + \mathbf{E}$. For $k = 1, \dots, d$, we can reorder and partition

$$\mathbf{Y}_\Gamma = (\mathbf{y}_{\Gamma,k} \ \mathbf{Y}_{\Gamma,-k}), \ \Gamma = (\gamma_k \ \Gamma_{-k}), \text{ and } \Theta = \begin{pmatrix} \theta_{kk} & \boldsymbol{\theta}_k \\ \boldsymbol{\theta}_k^\top & \Theta_{-kk} \end{pmatrix}.$$

In the above partition, $\mathbf{y}_{\Gamma,k}$ is the k th column of $\mathbf{Y}_\Gamma$, and $\mathbf{Y}_{\Gamma,-k}$ and Γ_{-k} are the matrices after dropping the k th column of $\mathbf{Y}_\Gamma$ and Γ , respectively. θ_{kk} is the element for the k th row and k th column of Θ ; $\boldsymbol{\theta}_k$ is the k th column of Θ after dropping θ_{kk} ; and Θ_{-kk} is the matrix after dropping the k th row and k th column of Θ . We can write

$$\begin{aligned} p(\gamma_k | -) &\propto \text{etr} \left\{ -\frac{1}{2}(\mathbf{Y}_\Gamma - \mathbf{W}\Gamma)^\top (\mathbf{Y}_\Gamma - \mathbf{W}\Gamma) \Theta \right\} \exp \left(-\frac{1}{2} \tau_\gamma \gamma_k^\top \mathbf{I}_q \gamma_k \right) \\ &\propto \text{etr} \left(-\frac{1}{2}(\mathbf{y}_{\Gamma,k} - \mathbf{W}\gamma_k)^\top (\mathbf{y}_{\Gamma,k} - \mathbf{W}\gamma_k) \theta_{kk} \right) \\ &\quad \times \text{etr} \left((\mathbf{y}_{\Gamma,k} - \mathbf{W}\gamma_k)^\top (\mathbf{Y}_{\Gamma,-k} - \mathbf{W}\Gamma_{-k}) \boldsymbol{\theta}_k \right) \exp \left(-\frac{1}{2} \gamma_k^\top (\tau_\gamma \mathbf{I}_q) \gamma_k \right) \\ &\propto \exp \left(\gamma_k^\top \mathbf{W}^\top \{ \mathbf{y}_{\Gamma,k} \theta_{kk} + (\mathbf{Y}_{\Gamma,-k} - \mathbf{W}\Gamma_{-k}) \boldsymbol{\theta}_k \} - \frac{1}{2} \gamma_k^\top (\theta_{kk} \mathbf{W}^\top \mathbf{W} + (\tau_\gamma \mathbf{I}_q)) \gamma_k \right). \end{aligned}$$

This gives us

$$p(\gamma_k | -) \sim N_d(\Sigma_{\gamma_k} \mathbf{W}^\top \{ \mathbf{y}_{\Gamma,k} \theta_{kk} + (\mathbf{Y}_{\Gamma,-k} - \mathbf{W}\Gamma_{-k}) \boldsymbol{\theta}_k \}, \Sigma_{\gamma_k}),$$

920 where $\Sigma_{\gamma_k} = \{ \theta_{kk} \mathbf{W}^\top \mathbf{W} + (\tau_\gamma \mathbf{I}_q) \}^{-1}$.

921 7.4.5. Sampling $\mathbf{b}$

Given that $\text{vec}(\mathbf{b}_i) \sim N(\mathbf{0}, \Sigma_b)$ with $\Sigma_b = \tau_b \mathbf{I}_d$, the full conditionals for $\mathbf{b}$ can be derived in a similar way to Γ . Let $\tilde{\mathbf{b}}_k$ be the k th column of $\mathbf{b}$. Update $\tilde{\mathbf{b}}_k$ from

$$p(\tilde{\mathbf{b}}_k | -) \sim N(\Sigma_{b_k} \mathbf{Z}^\top \{ \mathbf{y}_{b,k} - (\mathbf{Y}_{b,-k} - \mathbf{Zb}_{-k}) \boldsymbol{\theta}_k \}, \Sigma_{b_k}^2),$$

922 where $\Sigma_{b_k} = (\theta_{kk} \mathbf{Z}^T \mathbf{Z} + \tau_b I_{np_b})^{-1}$. $\mathbf{y}_{b,k}$ is the k th column of $\mathbf{Y}_b = \mathbf{Y} - \mathbf{X}\mathbf{B} -$
923 $\mathbf{W}\mathbf{\Gamma}$; and $\mathbf{Y}_{b,-k}$ and $\mathbf{b}_k$ is the matrix after dropping the k th column of $\mathbf{Y}_b$
924 and $\mathbf{b}$, respectively.

925 Full conditionals for all other parameters are straightforward.

Table 1: Top-ranked ROIs based on the diagonal of $B_{bin}^T B_{bin}$ and columns of V based on the top 10 genes.

$B_{int,std}^T B_{int,std}$	V_1	V_2	V_3
sup.t.gy.R	mid.t.gy.R	caud.neuc.L	prec.L
sup.t.gy.L	mid.t.gy.L	thal.L	per.cort.L
mid.t.gy.L	sup.t.gy.R	caud.neuc.R	tmp.pl.L
mid.t.gy.R	hiopp.R	glob.pal.L	per.cort.R
inf.t.gy.L	amyg.R	post.limb.R	ent.cort.R
parah.gy.L	unc.L	putamen.L	mid.f.gy.R
inf.f.gy.R	unc.R	nuc.acc.R	ling.gy.R
lat.f-o.gy.L	inf.t.gy.R	post.limb.L	sup.f.gy.L
inf.f.gy.L	sup.t.gy.L	insula.R	pec.R
unc.R	inf.t.gy.L	subtha.nuc.L	sup.p.lb.L

Table 2: Top-ranked ROIs based on the diagonal of $B_{bin}^T B_{bin}$ and columns of V based on the top 40 genes.

$B_{int,std}^T B_{int,std}$	V_1	V_2	V_3
insula.R	hiopp.R	per.cort.R	glob.pal.R
insula.L	hiopp.L	sup.p.lb.L	post.limb.R
amyg.R	inf.t.gy.R	per.cort.L	thal.L
hiopp.L	sup.t.gy.R	tmp.pl.R	caud.neuc.R
sup.t.gy.R	unc.L	sup.p.lb.R	caud.neuc.L
unc.L	mid.t.gy.R	me.f.gy.L	nuc.acc.R
unc.R	insula.R	lat.ve.R	ant.caps.R
mid.t.gy.L	amyg.R	pstc.gy.L	glob.pal.L
hiopp.R	mid.t.gy.L	prec.L	subtha.nuc.L
mid.t.gy.R	inf.o.gy.L	tmp.pl.L	subtha.nuc.R

Fig. 1: The characteristics of the ROIs and SNPs from the ADNI data set and their association.

(a) Lateral ventricle left (b) rs769451, hippocam- (c) rs439401, hippocam-
pal formation right pal formation left

Fig. 2: The longitudinal characteristics of the ROI volumes in the ADNI data set. (a) The trajectories of the volumes of left lateral ventricle for all subjects. (b) and (c) The LOESS estimates of two volume trajectories based on the subjects with different SNP alleles.

Fig. 3: Simulation results depicting the estimated ROC curves of the true positive rates and false positive rates for **B**. The four columns correspond to the settings (i) $p = 50$, $d = 50$; (ii) $p = 100$, $d = 100$; (iii) $p = 200$, $d = 100$; (iv) $p = 400$, $d = 100$. The first row shows the coefficients of the SNPs and the second row shows the coefficients of the interactions between the SNPs and subject age. The lines represent the L2R2 (black solid), BRRR (red dotted), BCCA (purple dashed), G-SMuRFS (blue dot-dash), and MMA (green dashed).

Fig. 4: Simulation results from one randomly selected replication with $p = 50$, $d = 50$, and moderately sparse $\mathbf{B}$. (a) and (b) The true and estimated splines for a randomly selected standardized ROI volume. (c) The scatterplot showing the true mean versus the predicted mean at one year after the subject's last observation using model (5). (d) The Gelman and Rubin's shrink factors at different iterations for all the parameters in $\mathbf{B}$.

(a) Thresholded main genetic effect (b) Thresholded SNP-age effect

Fig. 5: The thresholded estimates of the main genetic effects and the SNP-age interaction effects on ROIs based on the ADNI data set with the top 10 genes.

Fig. 6: The estimated spline functions for (a) and (b) all the ROIs, (c) and (d) selected ROIs with declining volumes, and (e) and (f) selected ROIs with increasing volumes based on the SNPs from the ADNI data set with the top 10 genes and top 40 genes, respectively.

(a) Hippocampal right, rs10501608 (b) Hippocampal right, rs1354106

(c) Lateral ventricle left, rs10501608 (d) Lateral ventricle left, rs10501604

Fig. 7: The effects of SNP alleles on the trajectories of two ROIs in the ADNI analysis. Black solid line represents the overall effects characterized by the basis functions and coefficients in $\mathbf{\Gamma}$. The blue dashed, magenta dot-dash, and red dotted lines are the trajectories of the ROIs altered by the SNP with AA, Aa, and aa, respectively. ‘A’ represents the major allele and ‘a’ represents the minor allele of a SNP.

Fig. 9: The SNP-specific networks of ROIs based on the ADNI data set with the top 10 genes. The first row depicts the locations of ROIs that are correlated with the SNPs rs10501604, rs475639, rs677909, and rs10501608 in gene *PICALM*; the second row shows those of rs3826656, rs3865444, rs988337, rs1354106 in gene *CD33*; the third row shows those of rs3752237 and rs3752240 in gene *ABCA7* and rs12734030 and rs650877 in gene *CR1*; the last row shows those of APOE34 (*APOE*), rs880436 (*BIN1*) and rs6458573 (*CD2AP*). The sizes of the dots represent the absolute magnitudes of the regression coefficients. Warmer and cooler colors represent ROIs located in superior and inferior anatomical sites of the brain, respectively.

Fig. 10: The SNP-specific networks of ROIs based on the ADNI data set with the top 40 genes, which show the locations of ROIs that are correlated with SNPs rs4513489 (*CCR2*), rs2273684 (*PRNP*) and rs6088662 (*PRNP*). The sizes of the dots represent the absolute magnitudes of the regression coefficients. Warmer and cooler colors represent ROIs located in superior and inferior anatomical sites of the brain, respectively.

Fig. 11: Histograms and QQ plots of the posterior means of the random effects for two random selected ROIs over all subjects in the analysis of ADNI data set with the top 10 genes.