Robust Bayesian variable selection for gene-environment interactions

Jie Ren[1,2], Fei Zhou[2], Xiaoxi Li[2], Shuangge Ma[3], Yu Jiang[4] and Cen Wu[*2]

[1] Department of Biostatistics, School of Medicine, Indiana University, Indianapolis, IN
[2] Department of Statistics, Kansas State University, Manhattan, KS
[3] Department of Biostatistics, Yale University, New Haven, CT
[4] Division of Epidemiology, Biostatistics and Environmental Health, School of Public Health, University of Memphis, Memphis, TN

∗ **Corresponding author**: Cen Wu, wucen@ksu.edu

Abstract

Gene-environment (G×E) interactions have important implications to elucidate the etiology of complex diseases beyond the main genetic and environmental effects. Outliers and data contamination in disease phenotypes of G×E studies have been commonly encountered, leading to the development of a broad spectrum of robust regularization methods. Nevertheless, within the Bayesian framework, the issue has not been taken care of in existing studies. We develop a fully Bayesian robust variable selection method for G×E interaction studies. The proposed Bayesian method can effectively accommodate heavy–tailed errors and outliers in the response variable while conducting variable selection by accounting for structural sparsity. In particular, for the robust sparse group selection, the spike–and–slab priors have been imposed on both individual and group levels to identify important main and interaction effects robustly. An efficient Gibbs sampler has been developed to facilitate fast computation. Extensive simulation studies and analysis of both the diabetes data with SNP measurements from the Nurses' Health Study and TCGA melanoma data with gene expression measurements demonstrate the superior performance of the proposed method over multiple competing alternatives.

Keywords: Bayesian variable selection, Gene-environment interactions, MCMC, Robust analysis, Sparse group selection

1 Introduction

Deciphering the genetic architecture of complex diseases is a challenging task, as it demands the elucidation of the coordinated function of multiple genetic factors, their interactions, as well as gene-environment interactions. How the genetic contributions to influence the variations in the disease phenotypes are mediated by the environmental factors reveals a unique perspective of the disease etiology beyond the main genetic effects and their interactions (or epistasis)[1;2]. Till now, G×E interaction analyses have been extensively conducted, especially within the framework of genetic association studies[3;4], to search for the important main and interaction effects that are associated with the disease trait[5].

This is the author's manuscript of the article published in final edited form as:

Ren, J., Zhou, F., Li, X., Ma, S., Jiang, Y., & Wu, C. (2023). Robust Bayesian variable selection for gene-environment interactions. Biometrics, 79(2), 684–694. https://doi.org/10.1111/biom.13670

With the availability of a large amount of genetic factors, such as SNPs or gene expressions, G×E interactions are of high dimensionality even though the preselected environmental factors are usually low dimensional. Therefore, variable selection has emerged as a powerful tool to identify G×E interactions associated with the phenotypic traits[6;7], and a surging amount of G×E studies have recently been conducted along this line, especially with regularization methods[8].

A prominent trend among these studies is to incorporate robustness in regularized identification of main and interaction effects in order to accommodate data contamination and heavy-tailed distributions in the disease phenotypes. Take the datasets analyzed in this article for example. The disease outcomes of interest are weight from the Nurses' Health Study (NHS) and (log-transformed) Breslow's depth from The Cancer Genome Atlas (TCGA) Skin Cutaneous Melanoma (SKCM) data. We plot the two in Figure 1, where the long tails can be clearly observed. In practice, such a heavy-tailed distribution is frequently encountered and arise due to multiple reasons. For instance, some phenotypes have skewness in nature. For the subjects' recruited for the NHS, their ages are in the range from 41 to 68 as the average age for the onset of type 2 diabetes is 45[9]. The subjects' weight among this age group does have a right-skewed tendency. In addition, in the study of complex diseases such as cancer, even patients of similar profiles may have different subtypes as rigorous accrual of patients is usually not affordable. The data from the major disease subtype can be viewed as being "contaminated" by other subtypes or outliers. As nonrobust approaches cannot efficiently accommodate data contamination and long tailed distributions, which inevitably leads to biased estimates and false identifications, the robust regularization methods have thus been extensively developed for G×E studies[7;8].

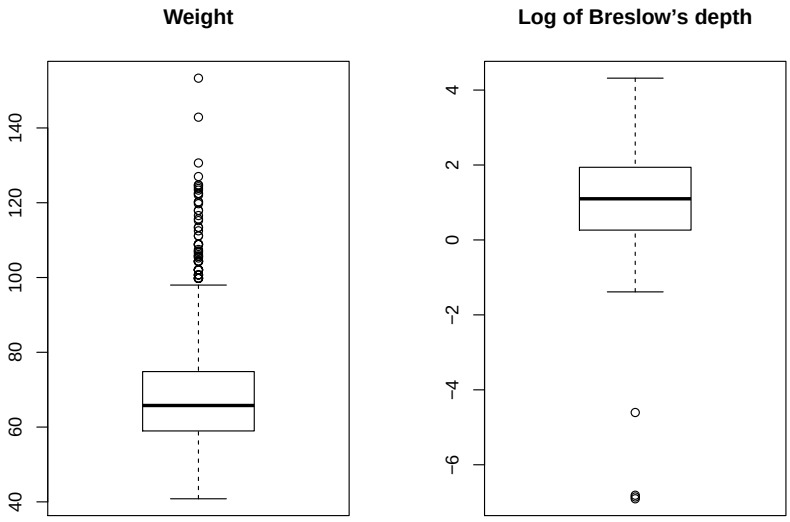

Figure 1: Distribution of the outcome variables for the NHS (left) and SKCM (right) data.

Nevertheless, within the Bayesian framework, robust variable selection methods have not

been investigated for gene-environment interactions by far. In fact, our literature search indicates that only limited number of Bayesian variable selection methods have been developed for G×E studies, and none of them is robust[8]. Driven by the urgent need to conduct robust Bayesian analysis, we propose robust Bayesian variable selection methods tailored for interaction studies by adopting a Bayesian formulation of the least absolute deviation (LAD) regression to accommodate data contamination and long-tailed distributions in the phenotype. Such a formulation is a special case of the Bayesian quantile regression[10]. The LAD loss has been a very popular choice for developing robust regularization methods for data with structured sparsity, including networks[11;12] and sparse group structure[13]. Its computational convenience has been revealed within the Bayesian framework as efficient Gibbs sampler can be constructed when the loss is combined with LASSO, group LASSO and elastic net penalties[14]. Furthermore, following the strategy of bi–level selection from a non-robust Bayesian setting[15], we have developed the Bayesian LAD sparse group LASSO for robust G×E interaction studies. The spike-and-slab priors have been imposed on both the individual and group level to ensure the shrinkage of posterior estimates corresponding to unimportant main and interaction effects to zero exactly. Such a prior leads to the real sparsity and is superior over Laplacian types of shrinkage in terms of identification and prediction results[16–18].

In this study, our objective is to tackle the challenging task of developing a fully Bayesian robust variable selection method for G×E interactions, which has been well motivated from the success of regularization methods (especially those robust ones) in G×E studies and a lack of robust interaction analysis within the Bayesian framework. The significance of the proposed study lies in the following aspects. First, it advances from existing Bayesian G×E studies by incorporating robustness to accommodate data contamination and heavy-tailed distributions in the disease phenotype. Second, on a broader scope, although robust Bayesian quantile regression based variable selection has been proposed under LASSO, group LASSO and elastic net, the more complicated sparse group (or bi-level) structure, which is of particular importance in high dimensional data analysis in general[19], has not been fully understood yet. We are among the first to develop robust Bayesian sparse group LASSO for bi-level selection. Third, unlike existing Bayesian regularized quantile regression methods which build upon the priors under Laplacian type of shrinkage, we conduct efficient Bayesian regularization on both the individual and group levels by borrowing strength from the spike-and-slab priors, thus leading to better identification and prediction performance over the competing alternatives, as demonstrated in extensive simulation studies and case studies of NHS data with SNP measurements and TCGA melanoma data with gene expression measurements. To facilitate reproducible research and fast computation using our MCMC algorithms, we implement the proposed and alternative methods in C++, which are available from an open source R package roben[20] on CRAN.

2 Data and Model Settings

Use subscript i to denote the ith subject. Let (X_i, Y_i, E_i, W_i), $(i = 1, \ldots, n)$ be independent and identically distributed random vectors. Y_i is a continuous response variable representing the phenotypic trait. X_i is the p–dimensional vector of G factors. The environmental

factors and clinical covariates are denoted as the k- and q-dimensional vectors E_i and W_i, respectively. Considering the following model:

$$\begin{aligned} Y_i &= \sum_{t=1}^{q} \alpha_t W_{it} + \sum_{m=1}^{k} \theta_m E_{im} + \sum_{j=1}^{p} \gamma_j X_{ij} + \sum_{j=1}^{p}\sum_{m=1}^{k} \zeta_{jm} E_{im} X_{ij} + \epsilon_i \\ &= \sum_{t=1}^{q} \alpha_t W_{it} + \sum_{m=1}^{k} \theta_m E_{im} + \sum_{j=1}^{p} \left(\gamma_j X_{ij} + \sum_{m=1}^{k} \zeta_{jm} E_{im} X_{ij}\right) + \epsilon_i \\ &= \sum_{t=1}^{q} \alpha_t W_{it} + \sum_{m=1}^{k} \theta_m E_{im} + \sum_{j=1}^{p} \left(U_{ij}^{\top} \beta_j\right) + \epsilon_i, \end{aligned} \tag{1}$$

where α_t's, θ_m's, γ_j's and ζ_{jm}'s are the regression coefficients for the clinical covariates, environmental factors, genetic factors and G×E interactions, respectively. We define $\beta_j = (\gamma_j, \zeta_{j1}, \ldots, \zeta_{jk})^{\top} \equiv (\beta_{j1}, \ldots, \beta_{jL})^{\top}$ and $U_{ij} = (X_{ij}, X_{ij}E_{i1} \ldots, X_{ij}E_{ik})^{\top} \equiv (U_{ij1}, \ldots, U_{ijL})^{\top}$, where $L = k + 1$. The coefficient vector β_j represents all the main and interaction effects corresponding to the jth genetic measurement. The ϵ_i's are random errors. Without loss of generality, we assume that the data have been properly normalized so that the intercept can be omitted. Denote $U_i = (U_{i1}^{\top}, \ldots, U_{ip}^{\top})^{\top}$, $\alpha = (\alpha_1, \ldots, \alpha_q)^{\top}$, $\theta = (\theta_1, \ldots, \theta_k)^{\top}$ and $\beta = (\beta_1^{\top}, \ldots, \beta_p^{\top})^{\top}$. The vector β is of length $p \times L$. Then model (1) can be written in a more concise form as

$$Y_i = W_i^{\top}\alpha + E_i^{\top}\theta + U_i^{\top}\beta + \epsilon_i \tag{2}$$

2.1 Bayesian LAD Regression

The least absolute deviation (LAD) regression is well known for its robustness to long-tailed distributions in response. For a Bayesian formulation of LAD regression, we assume that ϵ_i's are i.i.d random variables from the Laplace distribution with density

$$f(\epsilon_i|\nu) = \frac{\nu}{2} \exp\{-\nu|\epsilon_i|\} \quad i = 1, \ldots, n, \tag{3}$$

where ν^{-1} is the scale parameter of the Laplace distribution. Let $Y = (Y_1, \ldots, Y_n)^{\top}$. With clinical covariates $W = (W_1, \ldots, W_n)^{\top}$, environment factors $E = (E_1, \ldots, E_n)^{\top}$, and genetic main effects and G×E interactions $U = (U_1, \ldots, U_n)^{\top}$, the likelihood function can be expressed as

$$f(Y|W, E, U, \alpha, \theta, \beta, \nu) = \prod_{i=1}^{n} \frac{\nu}{2} \exp\{-\nu\,|Y_i - \mu_i|\}, \tag{4}$$

where $\mu_i = W_i^{\top}\alpha + E_i^{\top}\theta + U_i^{\top}\beta$.

Based on Kozumi and Kobayashi[21], the Laplace distribution is equivalent to the mixture of an exponential and a scaled normal distribution. Specifically, let z and $\tilde{u}$ be the standard normal and exponential random variables, respectively. If a random variable ϵ follows the Laplace distribution with parameter ν, then it can be represented as follows

$$\epsilon = \nu^{-1}\kappa\sqrt{\tilde{u}}z, \tag{5}$$

where $\kappa = \sqrt{8}$ is a constant. Therefore, the response Y_i can be rewritten as $Y_i = \mu_i + \nu^{-1}\kappa\sqrt{\tilde{u}_i}z_i$, where $z_i \sim N(0,1)$ and $\tilde{u}_i \sim \text{Exp}(1)$. Let $u = \nu^{-1}\tilde{u}$. Then u follows the exponential distribution $\text{Exp}(\nu^{-1})$. We thus have the following hierarchical representation of the Laplace likelihood:

$$\begin{aligned} Y_i &= \mu_i + \nu^{-\frac{1}{2}}\kappa\sqrt{u_i}z_i, \\ u_i|\nu &\overset{ind}{\sim} \nu \exp\left(-\nu u_i\right), \\ z_i &\overset{ind}{\sim} \text{N}(0,\ 1). \end{aligned}$$

This hierarchical representation allows us to express the likelihood function as a multivariate normal distribution, which is critical to construct a Gibbs sampler for efficient sampling of the regression coefficients corresponding to main and interaction effects robustly.

Remark: The Laplace distribution in Bayesian LAD regression can be treated as a special case of the asymmetric Laplace distribution (ALD) in Bayesian quantile regression[10;22]. In Bayesian quantile regression, we assume that ϵ_i follows the asymmetric Laplace distribution with density

$$f(\epsilon_i|\tau,\nu) = \tau(1-\tau)\nu \exp\left\{-\nu\rho_\tau(\epsilon_i)\right\} \quad i = 1,\ldots,n, \tag{6}$$

where the check loss function is $\rho_\tau(\epsilon_i) = \epsilon_i\left\{\tau - I(\epsilon_i < 0)\right\}$ for the τth quantile ($0 < \tau < 1$). Note that, when $\tau = 0.5$, the ALD in (6) reduces to a symmetric Laplace distribution defined in (3). Yu and Moyeed[10] have shown that maximizing a likelihood function under the asymmetric Laplace error distribution (6) is equivalent to minimizing the check loss function in quantile regression. Kozumi and Kobayashi[21] have proposed a Gibbs sampler for Bayesian quantile regression based on a location-scale mixture representation of the ALD. Specifically, with $\tilde{u}$ and z defined as above, the asymmetric Laplace error in (6) can be represented as

$$\epsilon = \nu^{-1}\psi z + \nu^{-1}\kappa\sqrt{\tilde{u}}z, \tag{7}$$

where

$$\psi = \frac{1-2\tau}{\tau(1-\tau)} \quad \text{and} \quad \kappa = \sqrt{\frac{2}{\tau(1-\tau)}}.$$

When $\tau = 0.5$, we have $\psi = 0$ and $\kappa = \sqrt{8}$, and equation (7) reduces to the Laplace error in (5).

2.2 Bayesian sparse group variable selection for G×E interactions

The proposed fully Bayesian sparse group variable selection is motivated by the following considerations. In model (1), the coefficient vector β_j corresponds to the main and interaction effects with respect to the jth genetic variant. Whether the genetic variant is associated with the phenotype or not can be determined by whether $\beta_j = 0$. A zero coefficient vector suggests that the variant does not have any effect on the disease outcome. If $\beta_j \neq 0$, then a further investigation on the presence of the main effect, the interaction or both is of interest, which can be facilitated by examining the nonzero component in β_j. Therefore, a tailored robust Bayesian variable selection method for G×E studies should accommodate the selection on both group (the entire vector of β_j) and individual (each component of β_j) levels at the same time.

In order to impose sparsity on both group and individual level to identify important main and interaction effects, we conduct the decomposition of β_j by following the reparameterization from[15]. Specifically, β_j is defined as

$$\beta_j = V_j^{\frac{1}{2}} b_j,$$

where $b_j = (b_{j1}, \ldots, b_{jL})^\top$ and $V_j^{\frac{1}{2}} = \text{diag}\{\omega_{j1}, \ldots, \omega_{jL}\}, \omega_{jl} \geq 0$ $(l = 1, ..., L)$. To determine whether the jth genetic variant has any effect at all, we conduct group-level selection on b_j by adopting the following multivariate spike–and–slab priors

$$\begin{aligned} b_j|\phi_j^b &\overset{ind}{\sim} \phi_j^b \, \text{N}_L(0, \mathbf{I}_L) + (1 - \phi_j^b)\delta_0(b_j), \\ \phi_j^b|\pi_0 &\overset{ind}{\sim} \text{Bernoulli}(\pi_0), \end{aligned} \tag{8}$$

where $\mathbf{I}_L$ is an identity matrix, $\delta_0(b_j)$ denotes a point mass at $0_{L\times 1}$ and $\pi_0 \in [0, 1]$. We introduce a latent binary indicator variable ϕ_j^b for each group $j(j = 1, \ldots, p)$ to tackle the group–level selection. In particular, when $\phi_j^b = 0$, the coefficient vector b_j has a point mass density at zero and all predictors representing the main and interaction effects in the jth group are excluded from the model, indicating that the jth genetic factor is not associated with the phenotype. On the other hand, when $\phi_j^b = 1$, the components in coefficient vector b_j have non-zero values.

To further determine whether there is an important main genetic effect, G×E interaction or both, we impose sparsity within the group j by assigning the following spike–and–slab priors on each ω_{jl} ($j = 1, \ldots, p$ and $l = 1, \ldots, L$)

$$\begin{aligned} \omega_{jl}|\phi_{jl}^w &\overset{ind}{\sim} \phi_{jl}^w \, \text{N}^+\left(0, s^2\right) + (1 - \phi_{jl}^w)\delta_0(\omega_{jl}), \\ \phi_{jl}^w|\pi_1 &\overset{ind}{\sim} \text{Bernoulli}(\pi_1), \end{aligned} \tag{9}$$

where $\text{N}^+(0, s^2)$ denotes a normal distribution, $\text{N}(0, s^2)$, truncated below at 0. When the binary indicator variable $\phi_{jl}^w = 0$, ω_{jl} is set to zero by the point mass function $\delta_0(\omega_{jl})$. Within the jth group, when the component $\omega_{jl} = 0$, we have $\beta_{jl} = 0$ and the corresponding U_{jl} is excluded from the model, even when $b_j \neq 0$. This implies that the jth genetic variant does not have the main effect (if l=1) or the interaction effect with the $(l-1)$th environment factor (if $l > 1$). The β_{jl} is non-zero if and only if the vector $b_j \neq 0$ and the individual element $\omega_{jl} \neq 0$.

In (8) and (9), π_0 and π_1 control the sparsity on the group and individual level, respectively. Their values should be carefully tuned. Fixing their values at 0.5 makes the prior essentially non-informative since equal prior probabilities are given to all the submodels. Instead of fixing π_0 and π_1, we assign conjugate beta priors $\pi_0 \sim \text{Beta}(a_0, b_0)$ and $\pi_1 \sim \text{Beta}(a_1, b_1)$, which can automatically account for the uncertainty in choosing π_0 and π_1. We fixed parameters $a_0 = b_0 = a_1 = b_1 = 1$, so that the priors are essentially non-informative. For computation convenience, we assign a conjugate Inverse–Gamma hyperprior on s^2

$$s^2 \sim \text{Inv-Gamma}(1, \eta)$$

η is estimated with the Monte Carlo EM algorithm[15;23]. For the gth EM update,

$$\eta^{(g)} = \frac{1}{E_{\eta^{(g-1)}}\left[\frac{1}{s^2}|Y\right]},$$

where the posterior expectation of $\frac{1}{s^2}$ is estimated from the MCMC samples based on $t^{(g-1)}$. To maintain conjugacy, we place a Gamma prior on ν,

$$\nu \sim \text{Gamma}(c, d).$$

where c and d are set to small values.

2.3 Gibbs sampler

The joint posterior distribution of all the unknown parameters conditional on data can be expressed as

$$\begin{aligned}
&\pi(\alpha, \theta, b_j, \omega_{jl}, \nu, u_i, \pi_0, \pi_1, s^2|Y) \\
&\qquad \propto \prod_{i=1}^{n} (2\pi\kappa^2\nu^{-1}u_i)^{-\frac{1}{2}} \exp\left\{ -\frac{\left(Y_i - W_i^\top\alpha - E_i^\top\theta - \sum_{j=1}^{p}\left(U_{ij}^\top\beta_j\right)\right)^2}{2\kappa^2\nu^{-1}u_i} \right\} \\
&\qquad \times \prod_{i=1}^{n} \nu \exp(-\nu u_i)\, \nu^{c-1} \exp\{-d\nu\} \\
&\qquad \times \exp\Big(-\frac{1}{2}\theta^\top\Sigma_{\theta 0}^{-1}\theta\Big) \exp\Big(-\frac{1}{2}\alpha^\top\Sigma_{\alpha 0}^{-1}\alpha\Big) \\
&\qquad \times \prod_{j=1}^{p} \left(\pi_0(2\pi)^{-\frac{L}{2}} \exp\left\{ -\frac{1}{2} b_j^\top b_j \right\} \mathbf{I}_{\{b_j \neq 0\}} + (1-\pi_0)\delta_0(b_j) \right) \\
&\qquad \times \prod_{j=1}^{p}\prod_{l=1}^{L} \left(\pi_1 2(2\pi s^2)^{-\frac{1}{2}} \exp\left\{ -\frac{\omega_{jl}^2}{2s^2} \right\} \mathbf{I}_{\{\omega_{jl} > 0\}} + (1-\pi_1)\delta_0(\omega_{jl}) \right) \\
&\qquad \times \pi_0^{a_0 - 1}(1-\pi_0)^{b_0 - 1} \\
&\qquad \times \pi_1^{a_1 - 1}(1-\pi_1)^{b_1 - 1} \\
&\qquad \times (s^2)^{-2} \exp(-\frac{\eta}{s^2}).
\end{aligned}$$

Define the coefficient vector without the jth group as $\beta_{(j)} = (\beta_1^\top, \ldots, \beta_{j-1}^\top, \beta_{j+1}^\top, \ldots, \beta_p^\top)$ and the corresponding part of the design matrix as $U_{(j)}$. Likewise, define the coefficient vector without the lth element in the jth group as $\beta_{(jl)}$ and the corresponding design matrix as $U_{(jl)}$. Let $l_j^b = p(b_j \neq 0|\text{rest})$, then the conditional posterior distribution of b_j is a multivariate spike–and–slab distribution:

$$b_j|\text{rest} \sim l_j^b\, \text{N}_L(\mu_{b_j},\, \Sigma_{b_j}) + (1 - l_j^b)\, \delta_0(b_j), \tag{10}$$

where $\Sigma_{b_j} = \left(\nu\kappa^{-2}\sum_{i=1}^n u_i^{-1}V_j^{\frac{1}{2}}U_{ij}U_{ij}^\top V_j^{\frac{1}{2}} + \mathbf{I}_L\right)^{-1}$, $\mu_{b_j} = \Sigma_{b_j}\nu\kappa^{-2}\sum_{i=1}^n u_i^{-1}V_j^{\frac{1}{2}}U_{ij}\tilde{y}_{ij}$ and $\tilde{y}_{ij} = y_i - W_i^\top\alpha - E_i^\top\theta - U_{(j)}^\top\beta_{(j)}$. The l_j^b can be derived as

$$l_j^b = \frac{\pi_0}{\pi_0 + (1-\pi_0)|\Sigma_{b_j}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\|\Sigma_{b_j}^{\frac{1}{2}}\nu\kappa^{-2}\sum_{i=1}^n u_i^{-1}V_j^{\frac{1}{2}}U_{ij}\tilde{y}_{ij}\|_2^2\right\}}.$$

The posterior distribution (10) is a mixture of a multivariate normal and a point mass at 0. Specifically, at the gth iteration of MCMC, $b_j^{(g)}$ is drawn from $N(\mu_{b_j}, \Sigma_{b_j})$ with probability l_j^b and is set to 0 with probability $1-l_j^b$. If $b_j^{(g)}$ is set to 0, we have $\phi_j^{b(g)} = 0$, which suggests that the jth genetic variant is not associated with the phenotype at the gth iteration. Otherwise, $\phi_j^{b(g)} = 1$.

In addition to the multivariate spike–and–slab distribution on the group level, on the individual level, the conditional posterior distribution of ω_{jl} is also spike-and-slab. Let $l_{jl}^w = p(\omega_{jl} \neq 0|\text{rest})$, we have

$$\omega_{jl}|\text{rest} \sim l_{jl}^w\, \mathrm{N}^+(\mu_{\omega_{jl}}, \sigma_{\omega_{jl}}^2) + (1-l_{jl}^w)\delta_0(\omega_{jl}),$$

where $\sigma_{\omega_{jl}}^2 = \left(\frac{1}{s^2} + \nu\kappa^{-2}\sum_{i=1}^n u_i^{-1}U_{ijl}^2 b_{jl}^2\right)^{-1}$, $\mu_{\omega_{jl}} = \sigma_{\omega_{jl}}^2\nu\kappa^{-2}\sum_{i=1}^n u_i^{-1}b_{jl}U_{ijl}\tilde{y}_{ijl}$ and $\tilde{y}_{ijl} = y_i - W_i^\top\alpha - E_i^\top\theta - U_{(jl)}^\top\beta_{(jl)}$. It can be shown that

$$l_{jl}^w = \frac{\pi_1}{\pi_1 + (1-\pi_1)\frac{1}{2}s(\sigma_{\omega_{jl}}^2)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\sigma_{\omega_{jl}}^2\left(\nu\kappa^{-2}\sum_{i=1}^n u_i^{-1}b_{jl}U_{ijl}\tilde{y}_{ijl}\right)^2\right\}\left[\Phi\left(\frac{\mu_{\omega_{jl}}}{\sigma_{\omega_{jl}}}\right)\right]^{-1}},$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable. At the gth iteration, the value of $\phi_{jl}^{w(g)}$ can be determined by whether the $\omega_{jl}^{(g)}$ is set to 0 or not. Recall that $\phi_{jl}^{w(g)} = 0$ implies that the jth genetic variant does not have the main effect (if l=1) or the interaction effect with the $(l-1)$th E factor (if $l > 1$).

The full conditional distribution for u_i is Inverse-Gaussian:

$$u_i|\text{rest} \sim \text{Inverse-Gaussian}(\mu_{u_i}, \lambda_{u_i}),$$

where the shape parameter $\lambda_{u_i} = 2\nu$, mean parameter $\mu_{u_i} = \sqrt{\frac{2\kappa^2}{(Y_i-\tilde{y}_i)^2}}$ and $\tilde{y}_i = Y_i - W_i^\top\alpha - E_i^\top\theta - U_i^\top\beta$.

With the conjugate Inverse–Gamma prior, the posteriors of s^2 is still an Inverse–Gamma distribution

$$s^2|\text{rest} \sim \text{Inv-Gamma}\left(1 + \frac{1}{2}\sum_{j,l}\mathbf{I}_{\{\omega_{jl}\neq 0\}},\ \eta + \frac{1}{2}\sum_{j,l}\omega_{jl}^2\right).$$

With conjugate Beta priors, π_0 and π_1 have beta posterior distributions

$$\pi_0|\text{rest} \sim \text{Beta}\left(a_0 + \sum_{j=1}^p \mathbf{I}_{\{b_j\neq 0\}},\ b_0 + \sum_{j=1}^p \mathbf{I}_{\{b_j=0\}}\right),$$

$$\pi_1|\text{rest} \sim \text{Beta}\left(a_1 + \sum_{j,l}\mathbf{I}_{\{\omega_{jl}\neq 0\}},\ b_1 + \sum_{j,l}\mathbf{I}_{\{\omega_{jl}=0\}}\right).$$

Last, the full conditional distribution for ν is Gamma distribution

$$\nu|\text{rest} \sim \text{Gamma}\left(s_\nu,\, r_\nu\right),$$

where the shape parameter $s_\nu = c + \frac{3n}{2}$ and the rate parameter $r_\nu = d + \sum_{i=1}^{n} u_i + (2\kappa^2)^{-1} \sum_{i=1}^{n} u_i^{-1} \tilde{y}_i^2$. Under our prior setting, conditional posterior distributions of all unknown parameters have closed forms by conjugacy. Therefore, efficient Gibbs sampler can be constructed for the posterior distribution.

We term the proposed robust Bayesian sparse group variable selection with spike and slab priors as RBSG–SS, with direct competitors RBG–SS, RBL–SS and ones without spike and slab priors: RBSG, RBG and RBL. With the non-robust counterpart, there are 12 methods under comparison, which have all been implemented in the C++ based R package roben[20] available from CRAN. It is worth mentioning that besides RBSG–SS, RBG–SS, RBL–SS and RBSG have also been proposed for the first time. A summary of all the methods is provided below.

2.4 A summary of proposed and alternative methods

All the methods under comparison can be grouped according to three criteria: with or without robustness, with or without spike-and-slab priors, and the types of structured sparsity (individual-, group- and bi-level) accommodated through variable selection. We first describe the robust Bayesian methods with spike-and-slab priors: RBSG–SS, RBG–SS and RBL–SS, which have all been proposed for the first time. Among them, RBSG–SS is the "golden" method developed for conducting robust sparse group variable selection for G×E interactions with spike-and-slab priors on both the group and individual levels. Besides, RBG–SS and RBL–SS are robust Bayesian group level and individual level selection with spike–and–slab priors, respectively. The spike–and–slab prior has only been imposed on the group level in RBG–SS. Compared to RBSG–SS, it does not induce within group sparsity. On the other hand, RBL–SS conducts individual-level selection without accounting for group structure. An immediate family of robust methods related to the three are RBSG, RBG and RBL, which do not adopt spike–and–slab priors and cannot shrink coefficients corresponding to the main and interaction effects to zero exactly. While RBG and RBL can be directly derived based on Li et al.[14], RBSG, robust Bayesian sparse group selection, has not been investigated in existing studies so far.

We have also included six non-robust methods for comparison. Among them, BSG–SS, BG–SS and BL–SS are the non–robust counterparts of RBSG–SS, RBG–SS and RBL–SS, respectively. In particular, the BSG–SS conducts (non–robust) Bayesian sparse group selection with spike-and-slab priors on group and individual level simultaneously, while variable selection has only been conducted on group (individual) level through RBG–SS (RBL–SS) under the spike–and–slab priors. In addition, BSG, BG and BL can be viewed as the benchmarks without incorporating spike–and–slab priors corresponding to BSG–SS, BG–SS and BL–SS. They can also be considered as the non–robust counterpart corresponding to RBSG, RBG and RBL. All the six non–robust alternatives can be readily derived based on existing studies.

For clarification, we list all the methods under comparison in Table 4 in the Appendix. Our contribution includes developing the 4 robust Bayesian variable selection approaches,

RBSG–SS, RBG–SS, RBL–SS and RBSG among the first time. For all the rest of the approaches, a modification to the methods from the references provided in Table 4 by including clinical covariates is necessary. Otherwise, these methods cannot be adopted for a direct comparison with the four newly developed ones.

3 Simulation

We comprehensively evaluate the proposed and alternative methods through simulation studies. Under all the settings, the responses are generated from model (1) with $n = 500, q = 3, p = 100$ and $k = 5$, which leads to a total dimension of 608 with 105 main effects, 500 interactions and 3 additional clinical covariates. The genetic main effects and G×E interactions form 100 groups with group size $L = 6$. We consider six error distributions for ϵ_i's: N(0, 1)(Error 1), Laplace(μ,b) with the mean $\mu = 0$ and the scale parameter $b = 2$ (Error 2), 10%Laplace(0,1) + 90%Laplace(0,$\sqrt{5}$) (Error 3), 90%N(0,1) + 10%Cauchy(0, 1) (Error 4), t-distribution with 2 degrees of freedom ($t(2)$) (Error 4), LogNormal(0,1) (Error 5). All of them are heavy-tailed distributions except the first one.

We assess the performance in terms of identification and prediction accuracy. For methods incorporating spike–and–slab priors, we consider the median probability model (MPM)[15;24] to identify important effects. In particular, for the proposed RBSG–SS, we define $\phi_{jl} = \phi_j^b \phi_{jl}^w$ for the lth predictor in the jth group. At the gth MCMC iterations, this predictor is included in the model if the indicator $\phi_{jl}^{(g)}$ is 1. Suppose we have collected G posterior samples from the MCMC after burn-ins, then the posterior probability of including the lth predictor from the jth group in the final model is

$$p_{jl} = \hat{\pi}(\phi_{jl} = 1|y) = \frac{1}{G}\sum_{g=1}^{G} \phi_{jl}^{(g)}, \quad j = 1, \ldots, p \text{ and } l = 1, \ldots, L. \tag{11}$$

A higher posterior inclusion probability p_{jl} can be interpreted as a stronger empirical evidence that the corresponding predictor has a non-zero coefficient and is associated with the phenotype. The MPM model is defined as the model consisting of predictors with at least $\frac{1}{2}$ posterior inclusion probability. When the goal is to select a single model, Barbieri and Berger[24] recommend using MPM because of its optimal prediction performance. Meanwhile, the 95% credible interval (95%CI)[25] is adopted for methods without spike–and–slab priors.

Prediction performance is evaluated using the mean prediction errors on an independently generated testing dataset under the same data generating model over 100 replicates. For all robust approaches, the prediction error is defined as mean absolute deviations (MAD). MAD can be computed as $\frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$. The prediction error for non–robust ones is defined as the mean squared error (MSE), i.e., $\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.

The G factors are simulated in the following 4 examples(settings). In the first example, a gene expression matrix with $n = 500$ and $p = 100$ has been generated from a multivariate normal distribution with marginal mean 0, marginal variance 1 and an auto-regression correlation structure ($\rho = 0.3$). In the second example, the single-nucleotide polymorphism (SNP) data are obtained by dichotomizing the gene expression values (from the first setting) at the 1st and 3rd quartiles, with the 3–level (0,1,2) for genotypes (aa,Aa,AA) respectively.

In the third setting, the SNP data are simulated under a pairwise linkage disequilibrium (LD) structure. Let the minor allele frequencies (MAFs) of two neighboring SNPs with risk alleles A and B be r_1 and r_2, respectively. The frequencies of four haplotypes are as $p_{AB} = r_1 r_2 + \delta$, $p_{ab} = (1-r_1)(1-r_2) + \delta$, $p_{Ab} = r_1(1-r_2) - \delta$, and $p_{aB} = (1-r_1)r_2 - \delta$, where δ denotes the LD. Assuming Hardy-Weinberg equilibrium and given the allele frequency for A at locus 1, we can generate the SNP genotype (AA, Aa, aa) from a multinomial distribution with frequencies $(r_1^2, 2r_1(1-r_1), (1-r_1)^2)$. The genotypes at locus 2 can be simulated according to the conditional genotype probability matrix in Cui et al.[26]. We have $\delta = r_p\sqrt{r_1(1-r_1)r_2(1-r_2)}$ with MAFs 0.3 and pairwise correlation $r_p = 0.6$. In the last example, we consider a more practical correlation structure by extracting the first 100 SNPs from the NHS data analyzed in the case study, so the correlation is based on the real data. For each simulation replicate, we randomly sample 500 subjects from the dataset.

For E factors, five continuous variables are generated from a multivariate normal distribution with marginal mean 0, marginal variance 1 and AR correlation structure with $\rho = 0.5$. We then dichotomize one of them at 0 to create a binary E factor. Besides, we simulate three clinical covariates from a multivariate normal distribution and AR correlation structure with $\rho = 0.5$, and dichotomize one of them at 0 to create a binary variable.

For the clinical covariates and environmental main effects, the coefficients α_t's and θ_m's are generated from Uniform$[0.8, 1.5]$. For genetic main effect and G×E interactions, we randomly selected 25 β_{jl}'s in 9 groups to have non-zero values that are generated from Uniform$[0.3, 0.9]$. All other β_{jl}'s are set to zeros.

We have collected the posterior samples from the Gibbs sampler running 15,000 iterations while discarding the first 7,500 samples as burn-ins. The Bayesian estimates are calculated using the posterior medians. Simulation results for the gene expression data in Example 1 are tabulated in Table (1) and (2). We can observe that the performance of methods that adopt spike–and–slab priors in Table (1) is consistently better than methods without spike–and–slab priors in Table (2). Although, methods without spike–and–slab priors have slightly lower FPs than their counterparts with spike–and–slab priors under some error distributions, they tend to have much lower TPs and higher prediction errors under all the error distributions. For example, under Error2, RBSG identifies 14.48(SD 2.04) out of the 25 true positives, much lower than the true positives of 21.66(SD 1.72) from RBSG–SS. Meanwhile, its false positives 0.64(SD 0.85) is only slightly lower than the FP of RBSG–SS (1.32(SD 1.33)). The prediction error of RBSG, 2.57 with a SD of 0.11, is also inferior than that of the RBSG–SS (2.15(SD 0.10)). Such an advantage can also be observed by comparing other methods in Table (1) with their counterparts (without spike–and–slab priors) from Table (2).

Among all the methods with spike–and–slab priors, as shown in Table (1), the proposed RBSG–SS has the best performance in both identification and prediction in the presence of data contamination and heavy–tailed errors. Under the mixture Laplace error (Error 3), RBSG–SS identifies 21.28(SD 2.24) true positives, with a small number of false positives, 1.48(SD 1.34). RBG–SS has a true positive of 24.80(SD 0.73), however, the number of false positives, 30.64(SD 4.23), is much higher. This is due to the fact that RBG–SS only conducts group level selection and does not impose the within-group sparsity. Compared to RBSG–SS, RBL–SS ignores the group structure, leading to fewer true positives of 18.14(SD2.68). In terms of prediction, RBSG–SS has the smallest L1 error, 2.29(0.12), among all the 3 robust methods with spike–and–slab priors. Although the difference in prediction error

Table 1: Simulation results in Example 1. $(n, q, k, p) = (500, 2, 5, 100)$. mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.

		RBSG–SS	RBG–SS	RBL–SS	BSG–SS	BG–SS	BL–SS
Error 1	TP	24.97(0.18)	25.00(0.00)	24.93(0.25)	24.97(0.18)	25.00(0.00)	24.93(0.25)
	FP	1.30(1.24)	29.60(2.42)	1.30(1.44)	0.47(0.68)	29.00(0.00)	0.43(0.73)
	Pred	0.83(0.03)	0.86(0.03)	0.84(0.04)	1.07(0.07)	1.13(0.07)	1.08(0.08)
Error 2	TP	21.66(1.72)	24.84(0.55)	18.58(2.14)	19.98(1.95)	24.58(0.86)	15.54(2.04)
	FP	1.32(1.33)	30.96(4.27)	1.62(1.64)	1.82(1.53)	30.98(4.83)	0.92(0.94)
	Pred	2.15(0.10)	2.17(0.09)	2.24(0.12)	9.32(0.97)	8.98(0.79)	10.09(1.08)
Error 3	TP	21.28(2.24)	24.80(0.73)	18.14(2.68)	19.00(2.61)	24.40(1.09)	14.24(2.39)
	FP	1.48(1.34)	30.64(4.23)	1.42(1.63)	2.04(1.73)	30.20(4.73)	1.18(1.16)
	Pred	2.29(0.12)	2.32(0.11)	2.41(0.12)	11.11(1.12)	10.59(0.95)	12.02(1.12)
Error 4	TP	23.80(1.30)	24.93(0.37)	21.80(1.94)	16.20(6.45)	21.83(5.24)	12.53(5.79)
	FP	0.53(0.86)	29.47(2.56)	0.20(0.41)	3.73(4.61)	35.77(23.92)	1.93(2.49)
	Pred	1.50(0.14)	1.52(0.13)	1.53(0.14)	12.48(6.56)	12.34(7.27)	13.35(6.72)
Error 5	TP	24.33(0.76)	25.00(0.00)	22.93(1.20)	22.93(1.26)	25.00(0.00)	18.00(2.17)
	FP	0.26(0.45)	29.00(0.00)	0.13(0.35)	4.30(3.40)	34.80(8.11)	1.23(1.55)
	Pred	1.16(0.10)	1.18(0.10)	1.18(0.10)	4.75(1.24)	4.78(1.23)	5.18(1.34)

between RBSG–SS and RBG–SS is not distinct, considering the much smaller number of false positive main and interaction effects, we can fully observe the advantage of RBSG–SS over RBG–SS in prediction.

Moreover, a cross–comparison between the robust and non–robust methods further demonstrates the necessity of developing robust Bayesian methods. For instance, under the error of t distribution with 2 degrees of freedom (Error 4), RBSG–SS has identified 23.80(SD 1.30) true main and interaction effects with only 0.53(SD 0.86) false positives. Its direct non–robust competitor, BSG–SS, leads to a true positive of 16.20(SD 6.45) with 3.73(SD 4.61) false effects. The superior performance of RBSG–SS over the other two non–robust methods, BG–SS and BL–SS, is also clear. Although a comparison between the prediction errors of robust and non–robust methods is not feasible as the two are computed under the L1 and least square errors, the identification results convincingly suggest the advantage of robust methods over non-robust ones,

Similar patterns have been observed in Table 6, 7, 8, 9, 10 and 11 for Examples 2, 3 and 4, respectively, in the Appendix. Overall, based on the investigations over all the methods through comprehensive simulation studies, we can establish the advantage of conducting robust Bayesian bi–level selection incorporating spike–and–slab priors.

We demonstrate the sensitivity of RBSG–SS for variable selection to the choice of the

Table 2: Simulation results in Example 1. $(n, q, k, p) = (500, 2, 5, 100)$. mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.

		RBSG	RBG	RBL	BSG	BG	BL
Error 1	TP	21.87(1.38)	24.67(0.76)	21.97(1.40)	22.93(1.34)	24.93(0.37)	23.07(1.23)
	FP	2.63(1.94)	55.33(15.76)	3.07(2.35)	2.43(1.77)	83.47(20.07)	11.20(4.34)
	Pred	1.15(0.05)	1.37(0.06)	1.15(0.05)	1.73(0.12)	2.29(0.19)	2.21(0.17)
Error 2	TP	14.48(2.04)	23.06(1.96)	14.42(2.12)	15.18(2.06)	24.02(1.48)	15.48(2.30)
	FP	0.64(0.85)	32.26(7.41)	0.74(0.88)	2.20(1.55)	85.78(20.06)	14.06(4.41)
	Pred	2.57(0.11)	2.85(0.13)	2.57(0.11)	12.43(1.15)	15.92(1.68)	16.55(1.69)
Error 3	TP	13.74(2.65)	22.52(2.38)	13.80(2.66)	14.30(2.70)	23.92(1.37)	14.62(2.69)
	FP	0.68(0.68)	34.24(8.93)	0.80(0.83)	2.74(1.48)	97.40(19.78)	15.96(4.30)
	Pred	2.71(0.12)	3.00(0.14)	2.71(0.12)	14.36(1.35)	18.52(1.70)	19.25(1.84)
Error 4	TP	16.90(3.12)	21.83(3.04)	16.90(3.36)	11.70(5.86)	20.70(5.74)	12.07(5.44)
	FP	0.33(0.48)	27.97(8.48)	0.27(0.45)	3.10(2.64)	88.50(28.58)	14.83(5.52)
	Pred	1.85(0.15)	2.10(0.17)	1.85(0.15)	16.25(9.88)	22.78(17.05)	24.20(18.91)
Error 5	TP	16.26(2.28)	23.42(2.01)	16.42(2.16)	13.80(3.37)	23.24(2.25)	14.24(3.05)
	FP	0.32(0.62)	29.38(7.54)	0.32(0.65)	3.00(2.14)	94.72(27.12)	16.26(4.84)
	Pred	2.20(0.14)	2.49(0.17)	2.21(0.14)	15.94(4.43)	20.73(5.12)	21.66(5.48)

hyper–parameters for π_0, and π_1 in the Appendix. The results are tabulated in Table 5, showing that the MPM model is insensitive to different specification of the hyper-parameters. Following Li et al.[25], we assess the convergence of the MCMC chains by the potential scale reduction factor (PSRF)[27;28]. PSRF values close to 1 indicate that chains converge to the stationary distribution. Gelman et al.[29] recommend using PSRF$\leq$ 1.1 as the cutoff for convergence, which has been adopted in our study. We compute the PSRF for each parameter and find the convergence of all chains after the burn-ins. For the purpose of demonstration, Figure 2 shows the pattern of PSRF the first five groups of coefficients in Example 1 under Error 2. The figure clearly shows the convergence of the proposed Gibbs sampler.

4 Real Data Analysis

4.1 Nurses' Health Study (NHS) data

Nurses' Health Study (NHS) is one of the largest investigations into the risk factors for major chronic diseases in women. As part of the Gene Environment Association Studies initiative (GENEVA), the NHS provides SNP genotypes data as well as detailed information on dietary and lifestyle variables. Obesity level is one of the most important risk factors for Type 2

diabetes mellitus (T2D), a chronic disease due to both genetic and environmental factors. In this study, we analyze the NHS type 2 diabetes data to identify main and interaction effects associated with obesity. We use weight as the response and focus on SNPs on chromosome 10. We consider five environment factors, including the total physical activity (act), glycemic load (gl), cereal fiber intake (ceraf), alcohol intake (alcohol) and a binary indicator of whether an individual has a history of high cholesterol (chol). All these environmental exposures have been suggested to be associated with obesity and diabetes[30]. In addition, we include three clinical covariates: height, age and a binary indicator of whether an individual has a history of hypertension (hbp). In NHS study, about half of the subjects are diagnosed of type 2 diabetes and the other half are controls without the disease. We only use health subjects in this study. After cleaning the data through matching phenotypes and genotypes, removing SNPs with minor allele frequency (MAF) less than 0.05 or deviation from Hardy–Weinberg equilibrium, the working dataset contains 1732 subjects with 35099 SNPs.

For computational convenience prescreening can be conducted to reduce For computational convenience prescreening can be conducted to reduce the feature space to a more attainable size for variable selection. For example, Li et al.[25] and Wu et al.[31] use the single SNP analysis to filter SNPs in a GWA study before downstream analysis. In this study, we use a marginal linear model with weight as the response variable to evaluate the penetrance effect of a variant under the environmental exposure. The marginal linear model uses a group of genetic main effect and G×E interactions corresponding to a SNP as the predictors, and test whether this SNP has any effect, main or G×E interaction. The SNPs with p-values less than a certain cutoff (0.001) for any effect, main or interaction, from the test are kept. 253 SNPs pass the screening.

The proposed approach RBSG-SS identifies 22 main SNP effects and 45 G×E interactions. The detailed estimation results are provided in Table 12 in the Appendix. We observe that the proposed method identifies main and interaction effects of SNPs with important implications in obesity. For example, two important SNPs, rs6482836 and rs10741150, that located within gene DOCK1 are identified. DOCK1 (Dedicator Of Cytokinesis 1) has been reported as a putative candidate for obesity related to adiponectin and triceps skinfold by previous studies[32;33]. RBSG-SS identifies the main effect of rs6482836 and its interaction with the E factor act. Physical activity plays an important role in the prevention of overweight and obese[34]. This result suggests that the expression level of DOCK1 in an individual may influence the effect of physical activity in obesity prevention. RBSG-SS also identifies the interaction between rs10741150 and the E factor chol, suggesting that the effect of cholesterol level can be mediated by DOCK1. Interestingly, a previous study has shown that the expression level of DOCK5, an important paralog of DOCK1, is increased in individuals exposed to a diet high in saturated fatty acids[35]. Our results provide more evidence of the importance of DOCK1 in diet-induced obesity. Another example is the SNP rs11196539, located within gene NRG3. NRG3(Neuregulin 3) has been found to be associated with both the basal metabolic rate (BMR) and body mass index (BMI)[36]. RBSG-SS identifies its interaction with the E factors, gl and alcohol. Both glycemic load and alcohol intake are important dietary variables associated with obesity. The continued intake of high-glycemic load meals leads to an increased risk of obesity[37]. The increasing alcohol consumption is associated with a decline in body mass index in women[38], however, heavy drinking can increase risk of the metabolic syndrome[39]. Our results suggest that further investigation of NRG3

may help explain the mechanism of the effects of glycemic load and alcohol intake on obesity. For the environment main effects, two E factors, chol and gl, have positive coefficients, and the other three, act, ceraf and alcohol, have negative coefficients, which are consistent with findings in the previous literature.

In addition to the proposed approach, we also conduct analysis using the alternatives RBL–SS, BSG–SS and BL–SS. As other alternative methods show inferior performance in simulation, they are not considered in real data analysis. Detailed estimation results are provided in Table 13, 14 and 15 in the Appendix. In Table 3, we provide the numbers of identified main and interaction effects with pairwise overlaps, to show the difference in terms of identification between the proposed method and the others. To further investigate the biological similarity of the identified genes, we conduct the Gene Ontology (GO) analysis. We can find an obvious difference between the proposed RBSG–SS and the three alternatives. The GO analysis results are provided in Figure 3.

With real data, it is difficult to assess the selection accuracy objectively. The prediction performance may provide additional information to the selection results. Following Yan and Huang[40] and Li et al.[25], we refit the models identified by RBSG–SS and RBL–SS using the robust Bayesian Lasso, and refit the models selected by BSG–SS and BL–SS using the Bayesian Lasso. For robust methods, the prediction mean absolute deviations (PMAD) are computed based on the posterior median estimates. The PMADs are 8.64 and 8.88 for RBSG–SS and RBL–SS, respectively. The proposed method outperforms the competitors. For non-robust methods, the prediction mean squared errors, or PMSEs, are 128.39 and 137.77 for BSG–SS and BL–SS, respectively. Overall, the superior performance of RNSG–SS over the alternatives can be observed.

4.2 TCGA skin cutaneous melanoma data

In this case study, we analyze the Cancer Genome Atlas (TCGA) skin cutaneous melanoma (SKCM) data. TCGA is a collaborative effort supported by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), and has published high quality clinical, environmental, as well as multi–omics data. For this study, we use the level-3 gene expression data of SKCM downloaded from the cBio Cancer Genomics Portal[41]. Our goal is to identify genes that have genetic main effect or G×E interaction effects on the Breslow' thickness, an important prognostic variable for SKCM[42]. The log-transformed Breslow's depth is used as the response variable and four E factors are considered, age, AJCC pathologic tumor stage, gender and Clark level. Data are available on 294 subjects and 20,531 gene expressions. We adopt the same screening method used in the first case study to select 109 genes for further analysis.

The proposed approach RBSG–SS identifies 16 main SNP effects and 32 G×E interactions. The detailed estimation results are available from Table 16 in the Appendix. One important gene identified is CXCL6 (C-X-C Motif Chemokine Ligand 6), a chemokine with neutrophil chemotactic and angiogenic activities. It has been reported that CXCL6 plays an important role in melanoma growth and metastasis[43]. RBSG-SS identifies its main effect and its interactions with E factors, stage and Clark level. This suggests that CXCL6 can have different effects at different stages of melanoma. Another important finding is the gene MAGED4, one of member in MAGE(Melanoma-associated antigen) family. MAGE family

contains genes that are highly attractive targets for cancer immunotherapy[44]. MAGED4 has been found to be an potential target for glioma immunotherapy[45]. RBSG-SS identifies the main effect of MAGED4 and its interaction with the E factor tumor stage, suggesting that MAGED4 may also play an important role in SKCM and its effect may change over different tumor stages. For the main effects of the E factors, Clark level and tumor stage have positive coefficients, and age and gender have negative coefficients, which match observations in the literature.

Analysis is also conducted via the three alternative methods, and the results are summarized in Table 3. Detailed estimation results are provided in Table 17, 18 and 19 in the Appendix. Again, the proposed RBSG–SS identifies different sets of main and interaction effects from the rest. We further investigate the biological similarity of the identified genes by GO analysis (Figure 3), which suggests an obvious difference. Prediction performance is also evaluated. The PMADs are 0.69 and 0.83 for RBSG–SS and RBL–SS, respectively. The proposed approach again has better prediction performance than RBL–SS. The PMSEs are 0.93 and 1.05 for BSG-SS and BL-SS, respectively. Combined, the RBSG–SS outperforms the alternatives.

Table 3: The numbers of main G effects and interactions identified by different approaches and their overlaps.

NHS	Main G effects				Interactions			
	RBSG-SS	RBL-SS	BSG-SS	BL-SS	RBSG-SS	RBL-SS	BSG-SS	BL-SS
RBSG-SS	22	20	16	13	45	21	17	10
RBL-SS		29	20	16		39	14	14
BSG-SS			29	25			34	22
BL-SS				27				42
SKCM	Main G effects				Interactions			
	RBSG-SS	RBL-SS	BSG-SS	BL-SS	RBSG-SS	RBL-SS	BSG-SS	BL-SS
RBSG-SS	16	10	14	13	32	11	18	10
RBL-SS		17	12	14		33	15	24
BSG-SS			22	15			29	14
BL-SS				20				33

5 Discussion

In this study, we have developed robust Bayesian variable selection methods for gene-environment interaction studies. The robustness of our methods comes from Bayesian formulation of LAD regression. In G×E studies, the demand for robustness arises in heavy-tailed distribution/ data contamination in both the response and predictors, as well as model

misspecification. We have focused on the first case, which is frequently encountered in practice. Investigations of the robust Bayesian methods accommodating the other two cases are interesting and will be pursued in the future.

Recently, penalization has emerged as a power tool for dissecting G×E interactions[8]. Our literature review suggests that Bayesian variable selection methods, although tightly related to penalization, has not been fully explored for interaction analyses, let alone the robust ones. We are among the first to conduct robust G×E analysis within the Bayesian framework. The proposed Bayesian LAD sparse group LASSO are not only specifically tailored for G×E studies, and but also generally applicable for problems incorporating the bi-level structure in a broader context, such as simultaneously selection of prognostic genes and pathways[46;47]. The spike-and-slab priors have been incorporated to further improve identification and prediction performances. As a byproduct, the Bayesian LAD LASSO and group LASSO, both with spike-and-slab priors, have also been investigated for the first time. The computational feasibility of the Gibbs samplers is guaranteed by the R package roben, with the core modules of the MCMC algorithms developed in C++.

In G×E studies, the form of interaction effects can be linear, nonlinear, and both linear and nonlinear, resulting in parametric[13;48;49], nonparametric[25;50;51] and semiparametric variable selection methods[31;52;53] to dissect G×E interactions, respectively. The proposed study can be potentially generalized to these studies within robust Bayesian framework. For example, variable selection for multiple semiparametric G×E studies can be formulated as a combination of individual and group level selection problem, where the robust Bayesian methods based on sparse group, group and individual level selection are directly applicable. The proposed robust Bayesian framework has paved the way for the future investigations.

References

[1] David J. Hunter. Gene—environment interactions in human diseases. *Nature Reviews Genetics*, 6(4):287–298, 2005. doi: 10.1038/nrg1578.

[2] Naoko I. Simonds, Armen A. Ghazarian, Camilla B. Pimentel, Sheri D. Schully, Gary L. Ellison, Elizabeth M. Gillanders, and Leah E. Mechanic. Review of the gene-environment interaction literature in cancer: What do we know? *Genetic Epidemiology*, 40(5):356–365, 2016. doi: 10.1002/gepi.21967.

[3] Joel Hirschhorn, Kirk Lohmueller, Edward Byrne, and Kurt Hirschhorn. A comprehensive review of genetic association studies. *Genetics in Medicine*, 2(4):45–61, 2002.

[4] Cen Wu, Shaoyu Li, and Yuehua Cui. Genetic association studies: an information content perspective. *Current genomics*, 7(13):566–573, 2012.

[5] Bhramar Mukherjee, Jaeil Ahn, Stephen B. Gruber, and Nilanjan Chatterjee. Testing Gene-Environment Interaction in Large-Scale Case-Control Association Studies: Possible Choices and Comparisons. *American Journal of Epidemiology*, 175(3):177–190, 12 2011. ISSN 0002-9262. doi: 10.1093/aje/kwr367.

[6] Jianqing Fan and Jinchi Lv. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1):101–148, 2010.

[7] Cen Wu and Shuangge Ma. A selective review of robust variable selection with applications in bioinformatics. *Briefings in Bioinformatics*, 16(5):873–883, 12 2014. ISSN 1467-5463. doi: 10.1093/bib/bbu046.

[8] Fei Zhou, Jie Ren, Shuangge Ma, and Cen Wu. Gene–environment interaction: a variable selection perspective. *Epistasis. Methods in Molecular Biology*, 2020.

[9] Centers for Disease Control and Prevention. National diabetes statistics report. 2020. URL `https://www.cdc.gov/diabetes/data/statistics/statistics-report.html`.

[10] Keming Yu and Rana A. Moyeed. Bayesian quantile regression. *Statistics and Probability Letters*, 54(4):437 – 447, 2001. ISSN 0167-7152. doi: https://doi.org/10.1016/S0167-7152(01)00124-9.

[11] Cen Wu, Qingzhao Zhang, Yu Jiang, and Shuangge Ma. Robust network-based analysis of the associations between (epi)genetic measurements. *Journal of Multivariate Analysis*, 168:119 – 130, 2018. ISSN 0047-259X.

[12] Jie Ren, Yinhao Du, Shaoyu Li, Shuangge Ma, Yu Jiang, and Cen Wu. Robust network-based regularization and variable selection for high-dimensional genomic data in cancer prognosis. *Genetic Epidemiology*, 43(3):276–291, 2019. doi: 10.1002/gepi.22194.

[13] Cen Wu, Yu Jiang, Jie Ren, Yuehua Cui, and Shuangge Ma. Dissecting gene–environment interactions: A penalized robust approach accounting for hierarchical structures. *Statistics in Medicine*, 37(3):437–456, 2018. doi: 10.1002/sim.7518.

[14] Qing Li, Ruibin Xi, and Nan Lin. Bayesian regularized quantile regression. *Bayesian Anal.*, 5(3):533–556, 09 2010. doi: 10.1214/10-BA521.

[15] Xiaofan Xu and Malay Ghosh. Bayesian variable selection and estimation for group lasso. *Bayesian Anal.*, 10(4):909–936, 12 2015. doi: 10.1214/14-BA929.

[16] Edward I. George and Robert E. McCulloch. Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993. doi: 10.1080/01621459.1993.10476353.

[17] Veronika Ročková and Edward I. George. The spike-and-slab lasso. *Journal of the American Statistical Association*, 113(521):431–444, 2018. doi: 10.1080/01621459.2016.1260469.

[18] Zaixiang Tang, Yueping Shen, Xinyan Zhang, and Nengjun Yi. The spike-and-slab lasso generalized linear models for prediction and associated genes detection. *Genetics*, 205 (1):77–88, 2017. ISSN 0016-6731. doi: 10.1534/genetics.116.192195.

[19] Patrick Breheny and Jian Huang. Penalized methods for bi-level variable selection. *Statistics and its interface*, 2(3):369–380, 2009. doi: 10.4310/sii.2009.v2.n3.a10.

[20] Jie Ren, Fei Zhou, Xiaoxi Li, and Cen Wu. *roben: Robust Bayesian Variable Selection for Gene-Environment Interactions*, 2020. URL `https://CRAN.R-project.org/package=roben`. R package version 0.1.0.

[21] Hideo Kozumi and Genya Kobayashi. Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81(11):1565–1578, 2011. doi: 10.1080/00949655.2010.496117.

[22] Keming Yu and Jin Zhang. A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics - Theory and Methods*, 34(9-10):1867–1879, 2005. doi: 10.1080/03610920500199018.

[23] Trevor Park and George Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008. doi: 10.1198/016214508000000337.

[24] Maria Maddalena Barbieri and James O. Berger. Optimal predictive model selection. *Ann. Statist.*, 32(3):870–897, 06 2004. doi: 10.1214/009053604000000238. URL `https://doi.org/10.1214/009053604000000238`.

[25] Jiahan Li, Zhong Wang, Runze Li, and Rongling Wu. Bayesian group lasso for non-parametric varying-coefficient models with application to functional genome-wide association studies. *Ann. Appl. Stat.*, 9(2):640–664, 06 2015. doi: 10.1214/15-AOAS808. URL `https://doi.org/10.1214/15-AOAS808`.

[26] Yuehua Cui, Guolian Kang, Kelian Sun, Minping Qian, Roberto Romero, and Wenjiang Fu. Gene–centric genomewide association study via entropy. *Genetics*, 179(1):637–650, 2008. ISSN 0016-6731. doi: 10.1534/genetics.107.082370.

[27] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992. ISSN 08834237.

[28] Stephen P. Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998. doi: 10.1080/10618600.1998.10474787.

[29] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2004.

[30] Frank B. Hu, JoAnn E. Manson, Meir J. Stampfer, Graham Colditz, Simin Liu, Caren G. Solomon, and Walter C. Willett. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *New England Journal of Medicine*, 345(11):790–797, 2001. doi: 10.1056/NEJMoa010492. PMID: 11556298.

[31] Cen Wu, Yuehua Cui, and Shuangge Ma. Integrative analysis of gene–environment interactions under a multi–response partially linear varying coefficient model. *Statistics in Medicine*, 33(28):4988–4998, 2014. doi: 10.1002/sim.6287.

[32] Laura Kelly Vaughan, Howard W. Wiener, Stella Aslibekyan, David B. Allison, Peter J. Havel, Kimber L. Stanhope, Diane M. O'Brien, Scarlett E. Hopkins, Dominick J. Lemas, Bert B. Boyer, and Hemant K. Tiwari. Linkage and association analysis of obesity traits reveals novel loci and interactions with dietary n-3 fatty acids in an alaska native (yup'ik) population. *Metabolism*, 64(6):689 – 697, 2015. ISSN 0026-0495. doi: https://doi.org/10.1016/j.metabol.2015.02.008.

[33] Minjoo Kim, Sarang Jeong, Hye Jin Yoo, Hyoeun An, Sun Ha Jee, and Jong Ho Lee. Newly identified set of obesity-related genotypes and abdominal fat influence the risk of insulin resistance in a korean population. *Clinical Genetics*, 95(4):488–495, 2019. doi: 10.1111/cge.13509.

[34] Nicholas J. Wareham, Esther M. F. van Sluijs, and Ulf Ekelund. Physical activity and obesity prevention: a review of the current evidence. *Proceedings of the Nutrition Society*, 64(2):229–247, 2005. doi: 10.1079/PNS2005423.

[35] Julia S. El-Sayed Moustafa, Hariklia Eleftherohorinou, Adam J. de Smith, Johanna C. Andersson-Assarsson, Alexessander Couto Alves, Eleni Hadjigeorgiou, Robin G. Walters, Julian E. Asher, Leonardo Bottolo, Jessica L. Buxton, Rob Sladek, David Meyre, Christian Dina, Sophie Visvikis-Siest, and Peter Jacobson. Novel association approach for variable number tandem repeats (VNTRs) identifies DOCK5 as a susceptibility gene for severe obesity. *Human Molecular Genetics*, 21(16):3727–3738, 05 2012. ISSN 0964-6906. doi: 10.1093/hmg/dds187.

[36] Myoungsook Lee, Dae Young Kwon, Myung-Sunny Kim, Chong Ran Choi, Mi-Young Park, and Ae-jung Kim. Genome-wide association study for the interaction between BMR and BMI in obese Korean women including overweight. *The Korean Nutrition Society and The Korean Society of Community Nutrition*, 10(1):115–124, 02 2016.

[37] Janette C Brand-Miller, Susanna HA Holt, Dorota B Pawlak, and Joanna McMillan. Glycemic index and obesity. *The American Journal of Clinical Nutrition*, 76(1):281S–285S, 07 2002. ISSN 0002-9165. doi: 10.1093/ajcn/76/1.281S.

[38] Kiran Nanchahal, W David Ashton, and David A Wood. Alcohol consumption, metabolic cardiovascular risk factors and hypertension in women. *International Journal of Epidemiology*, 29(1):57–64, 02 2000. ISSN 0300-5771. doi: 10.1093/ije/29.1.57.

[39] Inkyung Baik and Chol Shin. Prospective study of alcohol consumption and metabolic syndrome. *The American Journal of Clinical Nutrition*, 87(5):1455–1463, 05 2008. ISSN 0002-9165. doi: 10.1093/ajcn/87.5.1455.

[40] Jun Yan and Jian Huang. Model selection for cox models with time-varying coefficients. *Biometrics*, 68(2):419–428, 2012. doi: 10.1111/j.1541-0420.2011.01692.x.

[41] Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, Benjamin E. Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, Caitlin J. Byrne, Michael L. Heuer, Erik Larsson, Yevgeniy Antipin, Boris Reva, Arthur P. Goldberg, Chris Sander, and Nikolaus Schultz. The cbio cancer genomics portal: An open platform for exploring

multidimensional cancer genomics data. *Cancer Discovery*, 2(5):401–404, 2012. doi: 10.1158/2159-8290.CD-12-0095.

[42] Ashfaq A. Marghoob, Karen Koenig, Flavia V. Bittencourt, Alfred W. Kopf, and Robert S. Bart. Breslow thickness and clark level in melanoma. *Cancer*, 88(3):589–595, 2000. doi: 10.1002/(SICI)1097-0142(20000201)88:3⟨589::AID-CNCR15⟩3.0.CO;2-I.

[43] Hannelien Verbeke, Sofie Struyf, Nele Berghmans, Els Van Coillie, Ghislain Opdenakker, Catherine Uyttenhove, Jacques Van Snick, and Jo Van Damme. Isotypic neutralizing antibodies against mouse gcp-2/cxcl6 inhibit melanoma growth and metastasis. *Cancer Letters*, 302(1):54 – 62, 2011. ISSN 0304-3835. doi: https://doi.org/10.1016/j.canlet.2010.12.013.

[44] Qing-Mei Zhang, Ning Shen, Sha Xie, Shui-Qing Bi, Bin Luo, Yong-Da Lin, Jun Fu, Su-Fang Zhou, Guo-Rong Luo, Xiao-Xun Xie, and Shao-Wen Xiao. Maged4 expression in glioma and upregulation in glioma cell lines with 5-aza-2'-deoxycytidine treatment. *Asian Pacific Journal of Cancer Prevention*, 15(8):3495–3501, 2014. ISSN 1513-7368.

[45] Meixiang Sang, Lifang Wang, Chunyan Ding, Xinliang Zhou, Bin Wang, Ling Wang, Yishui Lian, and Baoen Shan. Melanoma-associated antigen genes – an update. *Cancer Letters*, 302(2):85 – 90, 2011. ISSN 0304-3835. doi: https://doi.org/10.1016/j.canlet.2010.10.021.

[46] Yu Jiang, Yuan Huang, Yinhao Du, Yinjun Zhao, Jie Ren, Shuangge Ma, and Cen Wu. Identification of prognostic genes and pathways in lung adenocarcinoma using a bayesian approach. *Cancer Informatics*, 1(7), 2017.

[47] Wei Liu, Songyun Ouyang, Zhigang Zhou, Meng Wang, Tingting Wang, Yu Qi, Chunling Zhao, Kuisheng Chen, and Liping Dai. Identification of genes associated with cancer progression and prognosis in lung adenocarcinoma: Analyses based on microarray from oncomine and the cancer genome atlas databases. *Molecular Genetics & Genomic Medicine*, 7(2):e00528, 2019. doi: 10.1002/mgg3.528.

[48] Yaqing Xu, Tingyan Zhong, Mengyun Wu, and Shuangge Ma. Histopathological imaging–environment interactions in cancer modeling. *Cancers*, 11(4), 2019.

[49] Fei Zhou, Jie Ren, Gengxin Li, Yu Jiang, Xiaoxi Li, Weiqun Wang, and Cen Wu. Penalized variable selection for lipid–environment interactions in a longitudinal lipidomics study. *Genes*, 10(12), 2019. ISSN 2073-4425. doi: 10.3390/genes10121002.

[50] Cen Wu and Yuehua Cui. A novel method for identifying nonlinear gene–environment interactions in case–control association studies. *Human Genetics*, 132(12):1413–1425, 12 2013. ISSN 1432-1203. doi: 10.1007/s00439-013-1350-z.

[51] Cen Wu, Ping-Shou Zhong, and Yuehua Cui. Additive varying–coefficient model for nonlinear gene–environment interactions. *Statistical Applications in Genetics and Molecular Biologye*, 17(2), 2018. doi: 10.1515/sagmb-2017-0008.

[52] Cen Wu, Xingjie Shi, Yuehua Cui, and Shuangge Ma. A penalized robust semiparametric approach for gene–environment interactions. *Statistics in Medicine*, 34(30):4016–4030, 2015. ISSN 0277–6715. doi: 10.1002/sim.6609.

[53] Jie Ren, Fei Zhou, Xiaoxi Li, Qi Chen, Hongmei Zhang, Shuangge Ma, Yu Jiang, and Cen Wu. Semiparametric bayesian variable selection for gene–environment interactions. *Statistics in Medicine*, 39(5):617–638, 2020.

[54] Lin Zhang, Veerabhadran Baladandayuthapani, Bani K. Mallick, Ganiraju C. Manyam, Patricia A. Thompson, Melissa L. Bondy, and Kim-Anh Do. Bayesian hierarchical structured variable selection methods with application to molecular inversion probe studies in breast cancer. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63(4):595–620, 2014. doi: 10.1111/rssc.12053.

[55] Minjung Kyung, Jeff Gill, Malay Ghosh, and George Casella. Penalized regression, standard errors, and bayesian lassos. *Bayesian Anal.*, 5(2):369–411, 06 2010. doi: 10.1214/10-BA607. URL `https://doi.org/10.1214/10-BA607`.

A Summary of methods.

Table 4: Summary of the proposed and alternative methods.

		Methods	Reference
Robust	**RBSG-SS**	Robust Bayesian sparse group selection with spike–and–slab priors	proposed for the first time
	RBG-SS	Robust Bayesian group selection with spike–and–slab priors	proposed for the first time
	RBL-SS	Robust Bayesian Lasso with spike–and–slab priors	proposed for the first time
	RBSG	Robust Bayesian sparse group selection	proposed for the first time
	RBG	Robust Bayesian group Lasso	Li et al. (2010)[14]
	RBL	Robust Bayesian Lasso	Li et al. (2010)[14]
Non-robust	**BSG-SS**	Bayesian sparse group Lasso with spike–and–slab priors	Xu and Ghosh (2015)[15]
	BG-SS	Bayesian group Lasso with spike–and–slab priors	Xu and Ghosh (2015)[15] Zhang et al. (2014)[54]
	BL-SS	Bayesian Lasso with spike–and–slab priors	Xu and Ghosh (2015)[15] Zhang et al. (2014)[54]
	BSG	Bayesian sparse group Lasso	Xu and Ghosh (2015)[15]
	BG	Bayesian group Lasso	Kyung et al. (2010)[55]
	BL	Bayesian Lasso	Park and Casella (2008)[23]

Note: The models in the references are modified to be applicable to G×E settings.

B Hyper-parameters sensitivity analysis

We demonstrate the sensitivity of RBSG-SS for variable selection to the choice of the hyper-parameters for π_0, and π_1. We consider five different Beta priors: (1) Beta(0.5, 0.5) which is a U-shape curve between $(0, 1)$; (2) Beta(1, 1) which is a essentially a uniform prior; (3) Beta(2, 2) which is a quadratic curve; (4) Beta(1, 5) which is highly right-skewed; (5) Beta(5, 1) which is highly left-skewed. As a demonstrating example, we use the same setting of Example 1 to generate data under the Error 2. Table 5 shows the identification performance of the median thresholding model (MPM) with different Beta priors. For all choices of Beta priors, the MPM model is very stable. Also, RBSG-SS correctly identifies most of the true effects with low false positives in all cases. Therefore, we simply use Beta$(1, 1)$ as the prior for π_0, and π_1 in this study.

Table 5: Sensitivity analysis for RBSG-SS using Example 1. mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.

	TP	FP	Pred
Beta(0.5, 0.5)	21.31(1.67)	1.71(1.50)	2.19(0.11)
Beta(1, 1)	21.66(1.72)	1.32(1.33)	2.17(0.10)
Beta(2, 2)	21.13(2.10)	1.47(1.16)	2.18(0.10)
Beta(1, 5)	20.82(1.71)	1.38(1.30)	2.17(0.10)
Beta(5, 1)	21.58(1.75)	2.22(1.52)	2.19(0.09)

C Assessment of the convergence of MCMC chains

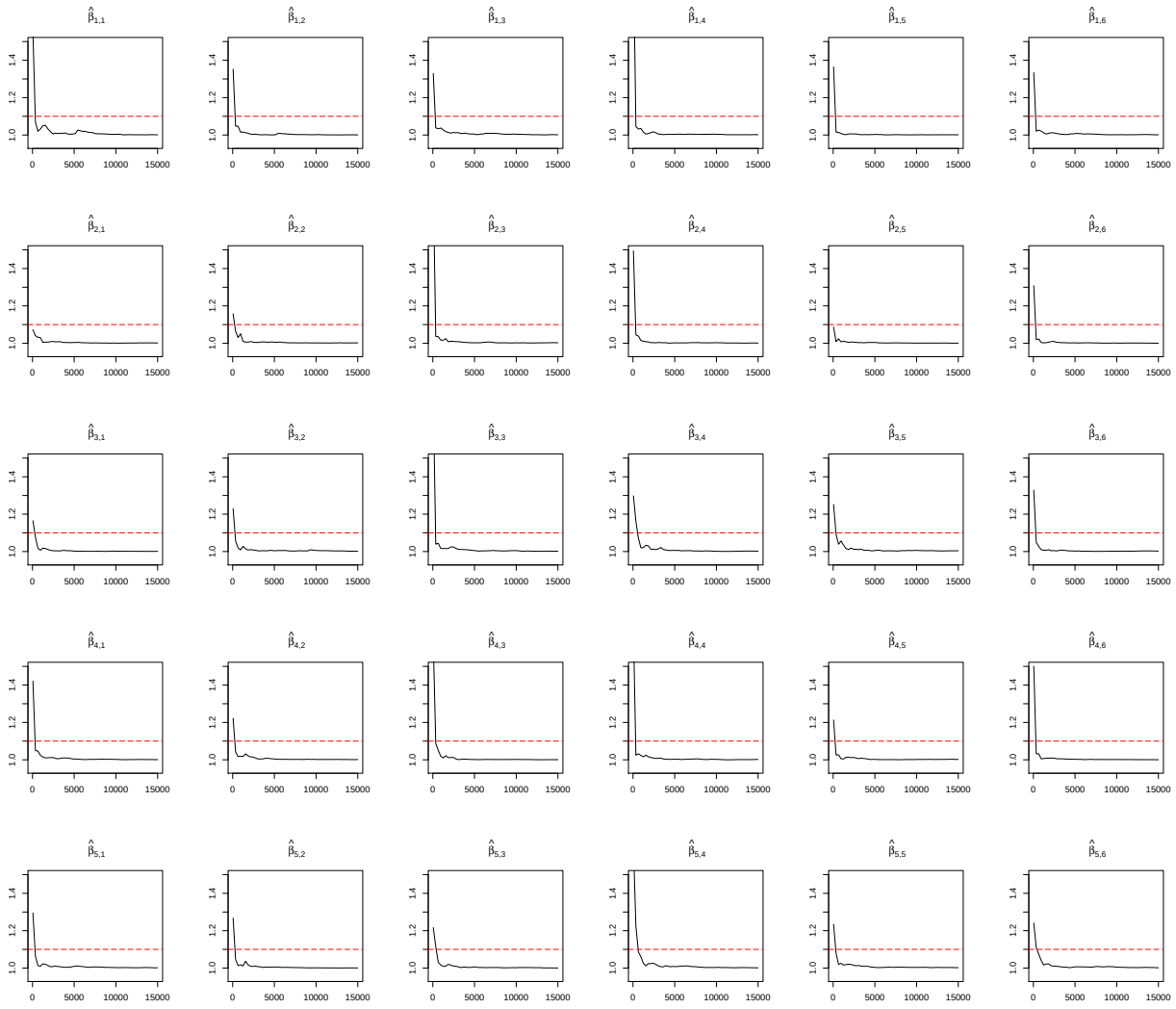

Figure 2: Potential scale reduction factor (PSRF) against iterations for the first five groups of coefficients in Example 1. Black line: the PSRF. Red line: the threshold of 1.1. The $\hat{\beta}_{j1}$ to $\hat{\beta}_{j6}$ represent the six estimated coefficients for the main and interaction effects in the jth group, $(j = 0, \ldots, 5)$, respectively.

D Additional simulation results

Table 6: Simulation results in Example 2. $(n, q, k, p) = (500, 2, 5, 100)$. mean(sd) of true positives (TP), false positives (FP)and prediction errors (Pred) based on 100 replicates.

		RBSG-SS	RBG-SS	RBL-SS	BSG-SS	BG-SS	BL-SS
Error 1	TP	24.87(0.35)	25.00(0.00)	24.53(0.51)	24.83(0.38)	25.00(0.00)	24.53(0.51)
	FP	1.63(1.16)	31.40(3.38)	2.30(1.86)	1.13(1.04)	29.20(1.10)	0.60(0.85)
	Pred	0.85(0.03)	0.86(0.03)	0.86(0.03)	1.09(0.06)	1.13(0.06)	1.10(0.07)
Error 2	TP	22.23(1.76)	24.67(0.76)	19.23(1.72)	19.97(1.63)	24.47(0.90)	15.27(1.91)
	FP	1.90(1.30)	35.73(7.83)	2.10(1.40)	2.33(1.42)	34.13(7.44)	1.73(1.39)
	Pred	2.24(0.14)	2.18(0.11)	2.38(0.16)	10.21(1.27)	9.13(0.94)	11.30(1.83)
Error 3	TP	21.50(1.48)	25.00(0.00)	17.43(2.13)	18.73(2.02)	25.00(0.00)	13.10(1.54)
	FP	2.13(1.14)	35.20(6.77)	1.90(1.37)	2.90(1.71)	34.00(6.88)	1.37(0.96)
	Pred	2.39(0.18)	2.29(0.11)	2.52(0.22)	12.46(1.67)	10.40(0.94)	13.04(1.35)
Error 4	TP	23.58(1.49)	25.00(0.00)	21.04(2.29)	15.94(5.34)	23.18(3.50)	12.08(4.60)
	FP	0.80(0.93)	30.32(3.27)	0.78(1.07)	7.46(27.02)	53.50(58.05)	3.56(8.91)
	Pred	1.85(0.16)	1.82(0.13)	1.92(0.17)	25.65(55.13)	25.63(67.60)	30.67(87.77)
Error 5	TP	24.12(1.00)	25.00(0.00)	21.82(1.90)	18.04(3.64)	24.24(1.88)	13.12(2.99)
	FP	0.90(1.02)	29.48(1.64)	0.82(0.90)	2.72(1.75)	36.12(12.21)	1.48(1.25)
	Pred	1.81(0.13)	1.82(0.12)	1.89(0.15)	14.85(6.53)	12.87(5.94)	15.19(6.43)

Table 7: Simulation results in Example 2. $(n, q, k, p) = (500, 2, 5, 100)$. mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.

		RBSG	RBG	RBL	BSG	BG	BL
Error 1	TP	24.20(0.61)	25.00(0.00)	24.23(0.57)	24.33(0.61)	25.00(0.00)	24.30(0.60)
	FP	2.93(1.86)	54.80(17.34)	3.30(1.97)	1.87(1.61)	56.20(10.30)	5.77(2.56)
	Pred	1.14(0.05)	1.32(0.06)	1.15(0.05)	1.74(0.12)	2.13(0.14)	2.01(0.15)
Error 2	TP	14.00(2.27)	22.20(2.33)	13.63(2.66)	13.70(2.29)	23.63(1.61)	14.20(1.97)
	FP	0.60(0.85)	31.40(12.07)	0.83(1.05)	1.33(1.18)	62.77(24.90)	8.80(4.54)
	Pred	2.57(0.13)	2.77(0.14)	2.58(0.14)	12.18(1.15)	14.42(1.40)	14.91(1.43)
Error 3	TP	12.40(2.03)	22.47(1.17)	12.27(1.87)	12.43(1.77)	23.20(1.49)	13.37(2.13)
	FP	0.57(0.77)	29.33(5.54)	0.60(0.93)	1.47(1.31)	59.80(13.17)	8.17(3.04)
	Pred	2.69(0.11)	2.86(0.11)	2.69(0.10)	13.59(1.05)	16.32(1.46)	16.93(1.60)
Error 4	TP	15.98(2.92)	23.04(2.78)	16.10(3.12)	10.20(5.31)	20.52(5.81)	11.08(5.00)
	FP	0.26(0.53)	27.36(6.21)	0.30(0.65)	2.34(3.56)	65.04(30.70)	9.38(6.26)
	Pred	2.19(0.16)	2.35(0.16)	2.21(0.17)	26.27(53.26)	34.08(78.47)	34.95(79.04)
Error 5	TP	16.48(2.69)	23.48(1.58)	16.30(2.63)	11.96(3.66)	22.26(3.70)	12.70(3.50)
	FP	0.32(0.59)	28.72(5.89)	0.34(0.59)	1.40(1.21)	62.34(20.52)	8.34(3.77)
	Pred	2.20(0.14)	2.41(0.14)	2.20(0.13)	15.79(5.97)	18.90(6.61)	19.53(6.65)

Table 8: Simulation results in Example 3. $(n, q, k, p) = (500, 2, 5, 100)$. mean(sd) of true positives (TP), false positives (FP)and prediction errors (Pred) based on 100 replicates.

		RBSG-SS	RBG-SS	RBL-SS	BSG-SS	BG-SS	BL-SS
Error 1	TP	24.00(0.91)	25.00(0.00)	22.13(1.57)	24.33(0.66)	25.00(0.00)	22.83(1.37)
	FP	1.85(1.46)	33.40(5.21)	1.87(1.36)	1.27(1.34)	29.60(1.83)	0.77(1.04)
	Pred	0.86(0.03)	0.86(0.03)	0.89(0.03)	1.11(0.07)	1.13(0.07)	1.17(0.10)
Error 2	TP	17.63(2.37)	24.73(0.69)	14.37(2.54)	15.00(2.32)	23.73(1.46)	11.00(1.95)
	FP	2.50(1.41)	33.27(5.14)	2.67(1.99)	2.60(1.43)	30.67(6.69)	1.87(1.53)
	Pred	2.33(0.12)	2.15(0.10)	2.40(0.17)	10.37(1.01)	9.01(0.81)	10.71(0.94)
Error 3	TP	17.23(1.77)	24.80(0.61)	14.47(2.21)	15.10(2.29)	23.67(1.77)	11.03(1.38)
	FP	2.27(1.78)	32.20(6.94)	1.63(1.43)	2.13(1.70)	30.93(5.48)	1.17(1.34)
	Pred	2.39(0.13)	2.24(0.10)	2.45(0.13)	11.98(1.45)	10.32(1.04)	12.37(1.41)
Error 4	TP	23.63(1.19)	24.67(0.92)	20.13(2.19)	15.07(4.69)	22.67(3.68)	11.40(4.01)
	FP	1.30(1.12)	29.13(2.67)	1.17(0.95)	3.37(1.88)	29.93(9.48)	2.37(1.97)
	Pred	1.48(0.13)	1.45(0.11)	1.55(0.14)	12.66(12.40)	10.10(8.77)	12.75(11.83)
Error 5	TP	24.80(0.48)	25.00(0.00)	23.57(1.43)	20.30(2.83)	24.87(0.51)	15.87(2.32)
	FP	0.33(0.55)	29.60(1.83)	0.40(1.04)	3.00(1.66)	32.93(5.86)	2.33(1.63)
	Pred	1.19(0.10)	1.21(0.10)	1.21(0.11)	6.055(1.77)	5.47(1.73)	6.54(1.75)

Table 9: Simulation results in Example 3. $(n, q, k, p) = (500, 2, 5, 100)$. mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.

		RBSG	RBG	RBL	BSG	BG	BL
Error 1	TP	21.27(1.17)	25.00(0.00)	21.30(1.06)	22.23(0.94)	25.00(0.00)	22.13(1.28)
	FP	1.97(1.56)	45.40(11.68)	2.03(1.56)	1.23(1.33)	43.60(10.41)	3.37(2.03)
	Pred	1.08(0.04)	1.21(0.05)	1.07(0.04)	1.57(0.12)	1.87(0.13)	1.79(0.13)
Error 2	TP	8.40(1.87)	18.67(2.68)	8.27(2.02)	8.07(1.57)	20.73(2.65)	8.73(2.15)
	FP	0.43(0.63)	20.73(4.93)	0.57(0.77)	0.67(0.66)	36.27(11.59)	3.83(2.05)
	Pred	2.44(0.13)	2.58(0.13)	2.44(0.12)	10.97(1.05)	12.78(1.31)	13.02(1.33)
Error 3	TP	8.43(2.18)	16.70(3.29)	8.70(2.00)	7.97(2.04)	18.60(3.07)	8.27(1.78)
	FP	0.33(0.71)	17.70(4.60)	0.43(0.73)	0.60(0.72)	33.60(10.63)	3.70(2.34)
	Pred	2.54(0.11)	2.69(0.12)	2.55(0.11)	12.33(1.15)	14.30(1.40)	14.55(1.40)
Error 4	TP	13.77(2.18)	21.20(2.06)	13.67(2.04)	9.67(3.74)	20.60(4.79)	9.77(3.76)
	FP	0.43(0.63)	22.80(3.98)	0.57(0.63)	1.03(1.13)	38.00(12.21)	4.47(2.50)
	Pred	1.73(0.12)	1.85(0.13)	1.73(0.13)	11.78(9.05)	13.94(11.84)	14.22(12.41)
Error 5	TP	19.10(1.86)	24.87(0.73)	19.10(1.60)	15.27(2.94)	24.07(1.70)	15.07(2.88)
	FP	0.20(0.48)	31.13(5.96)	0.23(0.57)	1.10(1.16)	43.93(11.82)	3.83(2.21)
	Pred	1.45(0.08)	1.61(0.09)	1.46(0.08)	6.13(1.13)	7.19(1.24)	7.16(1.29)

Table 10: Simulation results in Example 4. $(n, q, k, p) = (500, 2, 5, 100)$. mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.

		RBSG-SS	RBG-SS	RBL-SS	BSG-SS	BG-SS	BL-SS
Error 1	TP	24.93(0.37)	25.00(0.00)	24.93(0.25)	25.00(0.00)	25.00(0.00)	24.90(0.31)
	FP	1.33(0.99)	30.60(3.84)	1.47(1.25)	1.00(1.02)	29.20(1.10)	0.33(0.61)
	Pred	0.84(0.02)	0.88(0.02)	0.85(0.03)	1.10(0.04)	1.20(0.06)	1.11(0.05)
Error 2	TP	20.80(2.65)	23.60(1.47)	17.24(2.96)	18.58(3.46)	23.04(1.64)	14.08(3.26)
	FP	1.32(1.22)	30.76(4.93)	1.66(1.29)	1.98(1.53)	27.72(5.49)	1.42(1.25)
	Pred	2.25(0.11)	2.22(0.08)	2.37(0.12)	10.32(1.25)	9.53(0.75)	11.43(1.20)
Error 3	TP	20.56(2.73)	23.69(1.38)	16.53(3.20)	17.56(3.49)	22.80(1.65)	12.67(3.39)
	FP	1.40(1.30)	30.04(5.46)	1.78(1.82)	1.76(1.28)	27.60(5.24)	1.22(1.43)
	Pred	2.38(0.13)	2.35(0.10)	2.51(0.16)	12.04(1.40)	11.12(0.96)	13.32(1.44)
Error 4	TP	24.60(0.93)	24.67(0.92)	23.77(1.57)	20.10(6.38)	22.27(5.10)	15.63(6.69)
	FP	0.40(0.56)	29.13(2.97)	0.47(0.73)	1.83(1.90)	28.13(9.22)	1.17(1.15)
	Pred	1.48(0.09)	1.52(0.09)	1.51(0.11)	11.54(6.94)	11.33(6.95)	12.64(6.74)
Error 5	TP	23.16(1.68)	24.96(0.28)	19.60(2.14)	15.64(3.76)	23.44(1.83)	11.60(2.75)
	FP	1.08(1.16)	29.16(1.33)	0.72(0.83)	2.20(1.83)	30.56(7.43)	1.48(1.43)
	Pred	1.56(0.14)	1.53(0.13)	1.63(0.15)	10.98(5.80)	9.45(5.38)	11.38(6.03)

Table 11: Simulation results in Example 4. $(n, q, k, p) = (500, 2, 5, 100)$. mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.

		RBSG	RBG	RBL	BSG	BG	BL
Error 1	TP	21.47(1.87)	24.40(1.07)	21.67(1.81)	22.70(1.64)	24.87(0.51)	22.53(1.85)
	FP	3.17(2.51)	56.00(20.03)	3.33(2.59)	2.30(1.66)	66.33(14.04)	6.57(2.62)
	Pred	1.26(0.06)	1.44(0.07)	1.27(0.06)	2.40(0.27)	2.83(0.24)	2.75(0.30)
Error 2	TP	9.08(2.54)	19.38(3.12)	9.20(2.60)	9.68(2.41)	20.82(2.93)	10.80(2.65)
	FP	0.78(0.86)	30.30(10.26)	0.84(0.89)	2.18(1.48)	65.94(19.60)	8.62(3.38)
	Pred	2.67(0.08)	2.89(0.09)	2.67(0.09)	13.54(0.87)	16.38(1.19)	16.87(1.27)
Error 3	TP	8.51(2.31)	18.71(3.37)	8.62(2.33)	9.02(2.33)	20.60(2.76)	10.58(2.50)
	FP	0.56(0.69)	25.29(7.94)	0.56(0.66)	1.87(1.36)	56.87(15.55)	7.38(2.91)
	Pred	2.79(0.11)	3.00(0.12)	2.79(0.12)	15.34(1.29)	18.66(1.69)	19.30(1.88)
Error 4	TP	13.30(3.32)	21.93(2.72)	13.47(3.33)	10.93(4.30)	20.97(4.76)	11.97(4.43)
	FP	0.50(0.57)	29.07(9.28)	0.40(0.50)	1.70(1.39)	60.03(20.56)	7.90(3.92)
	Pred	2.03(0.12)	2.22(0.12)	2.03(0.12)	15.20(7.98)	18.61(12.19)	19.41(13.39)
Error 5	TP	14.38(2.64)	22.36(2.22)	14.40(2.70)	10.12(3.53)	20.84(3.74)	10.56(3.39)
	FP	0.30(0.58)	25.16(4.36)	0.22(0.46)	0.88(1.12)	36.52(12.60)	3.70(2.57)
	Pred	1.84(0.15)	1.99(0.17)	1.84(0.15)	11.23(5.54)	13.02(5.80)	13.18(5.90)

E Estimation results for data analysis

Table 12: Analysis of the NHS T2D data using RBSG-SS.

SNP	Gene*		chol	act	gl	ceraf	alcohol
			3.503	-3.447	0.752	-3.364	-2.639
rs10741150	DOCK1		-0.948				
rs10765059	TCERG1L	-0.531					0.877
rs10786611	RF00019	0.668	0.723		0.530		
rs10884466	RNA5SP326	-0.466		0.643			
rs10885423	NRG3				-0.715		
rs10886442	GRK5		0.805				
rs11196539	NRG3				-0.608		-0.801
rs11198590	CACUL1	-0.494				0.994	-0.687
rs11259039	FRMD4A	1.016					
rs1194657	THAP12P3				0.798		
rs1219508	RPS15AP5	-0.742					
rs12265854	SLC16A12	0.397					
rs12414552	TCERG1L	0.667		0.470	0.585		-0.690
rs12767723	SLC25A18P1	0.820				-0.515	
rs12772559	TACR2				0.938	0.510	
rs12774333	LRMDA	-0.599					
rs12775160	FOXI2	-0.651	-0.501				0.647
rs16916794	SLC39A12	-0.552	0.511	0.455			
rs16920092	PLXDC2						-0.843
rs17094114	GFRA1	-0.615					
rs2492664	OR6L1P	0.695			-0.737		
rs2784767	PLAC9	-0.540					
rs2814322	GRID1				-0.830		
rs3740063	ABCC2				-0.966		
rs3763722	LARP4B	0.332	-1.156				0.866
rs4411238	PRKG1	0.537					
rs4578341	CHST15		-0.822			0.602	
rs4747517	ITIH5		-1.468			0.920	
rs4749926	IL2RA	-0.840	-0.815				
rs4917817	PYROXD2	-0.624		0.594			
rs4918904	XRCC6P1					0.997	
rs6482836	DOCK1	-0.957		1.067			
rs7070789	GPAM				-1.245		-0.791
rs7072255	ANTXRLP1		0.800				
rs7077721	SNRPD2P1	0.858		0.774			
rs7896554	NACAP2	0.840		-0.630	-0.565		
rs7897847	LGI1						0.962
rs870753	CFAP58				-0.783		

Continued on the next page

Table 12: Continued from the previous page.

SNP	Gene*	chol	act	gl	ceraf	alcohol
rs881726	GFRA1					1.001

∗ Genes that SNPs belong to or are the closest to.

Table 13: Analysis of the NHS T2D data using RBL-SS.

SNP	Gene*		chol	act	gl	ceraf	alcohol
			1.354	-0.430	-0.778	-2.424	-4.000
rs1041168	PLPP4	0.463					
rs10741150	DOCK1		-1.126				
rs10786611	RF00019	0.632					
rs10794069	ADAM12	0.524					
rs10824802	MBL2	0.553					
rs10884466	RNA5SP326	-0.439		0.503			
rs10885423	NRG3				-1.060		
rs10886047	MIR3663HG						-0.410
rs10886442	GRK5		1.087				
rs10998780	ATP5MC1P7				0.150		
rs11003665	RNA5SP318		0.632				
rs11013740	KIAA1217						0.852
rs11196539	NRG3						-0.624
rs11198590	CACUL1	-0.628					
rs11202221	BMPR1A				0.815		
rs11259039	FRMD4A	1.021					
rs11595123	AKR1E2			1.079			
rs11813505	KIAA1217					1.301	
rs1194657	THAP12P3				0.663		
rs1219508	RPS15AP5	-0.886					
rs12265854	SLC16A12	0.596					
rs12269237	RF00017					0.884	
rs12414552	TCERG1L	0.594					
rs12414627	PNLIPRP1				-0.551		
rs12767723	SLC25A18P1	0.962					
rs12772559	TACR2				0.906		
rs12774333	LRMDA	-0.449					
rs12775160	FOXI2	-0.560					
rs1573137	SORCS3				0.615		
rs16916794	SLC39A12	-0.803	0.528				
rs16920092	PLXDC2						-0.655
rs17094114	GFRA1	-0.563					
rs2291314	PLPP4	0.536					
rs2420979	TACC2					-1.091	

Continued on the next page

Table 13: Continued from the previous page.

SNP	Gene*		chol	act	gl	ceraf	alcohol
rs2492664	OR6L1P	0.655			-0.363		
rs2664339	RNU6-543P	-0.501					
rs2666236	IATPR					0.689	
rs2784767	PLAC9	-0.452	0.730				
rs2814322	GRID1				-0.806		
rs2842129	DYNC1I2P1	-0.662					
rs2900814	SNRPD2P1	-0.643					
rs3740063	ABCC2				-0.885		
rs3763722	LARP4B						1.036
rs4411238	PRKG1	0.399					
rs4578341	CHST15		-0.582			0.479	
rs4747009	LRRC20						0.710
rs4747517	ITIH5		-0.905				
rs4749926	IL2RA	-0.607					
rs4752432	PLPP4		0.725				
rs4917817	PYROXD2	-0.506					
rs4934762	PCAT5	-0.560					
rs6482836	DOCK1	-0.709					
rs7070789	GPAM				-0.820		
rs7072255	ANTXRLP1		0.811				
rs7077721	SNRPD2P1	0.702					
rs7894809	PCGF5	0.501					
rs7896554	NACAP2	0.850		-0.953			
rs7897847	LGI1						0.929
rs7903853	FRMD4A		-1.185				
rs7920351	TCERG1L			-0.713			
rs881726	GFRA1						0.675
rs943213	DOCK1			-0.939			

∗ Genes that SNPs belong to or are the closest to.

Table 14: Analysis of the NHS T2D data using BSG-SS.

SNP	Gene*		chol	act	gl	ceraf	alcohol
			2.045	-2.049	-2.204	-1.796	-4.436
rs1041168	PLPP4	0.638					
rs10765059	TCERG1L			0.709			
rs10786611	RF00019	0.773	0.556				
rs10829671	EBF3	-0.505					
rs10884466	RNA5SP326	-0.563		0.625			
rs10886442	GRK5		1.038				
rs10998780	ATP5MC1P7				0.704		

Continued on the next page

Table 14: Continued from the previous page.

SNP	Gene*		chol	act	gl	ceraf	alcohol
rs11017821	TCERG1L						0.665
rs11198590	CACUL1	-0.698			0.905	0.568	
rs11200996	CCSER2					0.508	
rs11259039	FRMD4A	1.174					
rs1219508	RPS15AP5	-0.787					
rs12265854	SLC16A12	0.681			-0.494		
rs12269237	RF00017					0.684	
rs12414552	TCERG1L	0.480					
rs12764378	ARID5B	-0.420					
rs12767723	SLC25A18P1	0.638				-0.559	
rs12775160	FOXI2	-0.762					0.893
rs1361709	PCDH15			-0.852			
rs1395465	RN7SL63P	0.292					
rs16916794	SLC39A12	-0.614	0.580	0.622			
rs16920092	PLXDC2						-0.692
rs17094114	GFRA1	-0.676					
rs17469499	KIAA1217					-0.527	
rs2472737	RET	0.629					
rs2577356	GFRA1					0.875	
rs2784767	PLAC9	-0.569	0.537				
rs2792708	GPAM	0.488					
rs2900814	SNRPD2P1	-0.460					
rs2926458	RNU6-463P	-0.680					
rs3763722	LARP4B		-1.251				1.186
rs4411238	PRKG1	0.619					
rs4747517	ITIH5		-1.257				
rs4752432	PLPP4		0.956				
rs4917817	PYROXD2	-0.626		0.630			
rs4922535	GDF10			-0.601		-0.649	
rs4934762	PCAT5	-0.640					
rs4934858	NRP1	0.281					
rs6482836	DOCK1	-0.773					
rs7070789	GPAM				-0.642		
rs7072255	ANTXRLP1		0.723				
rs7085788	RHOBTB1	-0.720					
rs7086058	RN7SKP143	-0.507					
rs716168	VTI1A		-0.570				
rs7894809	PCGF5	0.642					
rs7895870	RN7SKP167		-0.867				
rs7896554	NACAP2	1.097		-0.477			
rs7917422	HTR7						0.794

Continued on the next page

Table 14: Continued from the previous page.

SNP	Gene*	chol	act	gl	ceraf	alcohol
rs881726	GFRA1					0.933

∗ Genes that SNPs belong to or are the closest to.

Table 15: Analysis of the NHS T2D data using BL-SS.

SNP	Gene*		chol	act	gl	ceraf	alcohol
			3.095	-2.406	-2.373	-1.716	-3.721
rs1041168	PLPP4	0.670					
rs10508670	KIAA1217		0.773				
rs10765059	TCERG1L			0.445			
rs10829671	EBF3	-0.717					
rs10884466	RNA5SP326	-0.528					
rs10998780	ATP5MC1P7				1.195		
rs11017821	TCERG1L						0.307
rs11198590	CACUL1				1.273		
rs11200996	CCSER2					0.509	
rs11202221	BMPR1A				0.954		
rs11259039	FRMD4A	1.020					
rs11594070	ATE1-AS1				-0.401		
rs1194657	THAP12P3				0.681		
rs12248205	CDH23				-0.938		
rs12256982	ZMIZ1				0.152		
rs12265854	SLC16A12	0.661			-0.830		
rs12269237	RF00017					0.864	
rs12412976	RPLP1P10			0.592	-0.590		
rs12414552	TCERG1L	0.549					
rs12414627	PNLIPRP1				-0.572		
rs12764378	ARID5B	-0.564					
rs12767723	SLC25A18P1	1.062					
rs12775160	FOXI2	-0.636					
rs1361709	PCDH15			-0.729			
rs1395465	RN7SL63P	0.562					
rs1573137	SORCS3				0.869		
rs16916794	SLC39A12	-0.430	0.862				
rs16920092	PLXDC2						-0.508
rs17094114	GFRA1	-0.734					
rs17469499	KIAA1217					-0.680	
rs2384105	SNRPEP8						-0.738
rs2420979	TACC2	-0.629					
rs2472737	RET	0.553					
rs2577356	GFRA1					0.739	

Continued on the next page

Table 15: Continued from the previous page.

SNP	Gene*		chol	act	gl	ceraf	alcohol
rs2784767	PLAC9	-0.593	0.576				
rs2792708	GPAM	0.568					
rs2900814	SNRPD2P1					-0.157	
rs2926458	RNU6-463P	-0.527					
rs3763722	LARP4B		-1.002				1.151
rs4411238	PRKG1	0.461					
rs4747009	LRRC20						1.016
rs4747517	ITIH5		-1.695				
rs4752432	PLPP4						0.787
rs4917817	PYROXD2	-0.637		0.751			
rs4934762	PCAT5	-0.771					
rs4934858	NRP1	0.496					
rs6482836	DOCK1	-0.899					
rs7069001	WDFY4		-0.942				
rs7070789	GPAM				-1.154		-0.771
rs7077718	DNMBP	-0.661					
rs7085788	RHOBTB1	-0.721					
rs7086058	RN7SKP143	-0.872					
rs716168	VTI1A		-0.662				
rs7894809	PCGF5	0.828					
rs7895870	RN7SKP167		-1.295				
rs7896554	NACAP2	0.989					
rs7917422	HTR7				0.663		1.306
rs7920351	TCERG1L				-1.059		
rs809836	LYZL1				1.109		
rs881726	GFRA1						0.922
rs915216	DUSP5		1.102				

∗ Genes that SNPs belong to or are the closest to.

Table 16: Analysis of the TCGA SKCM data using RBSG-SS.

Gene		clark	stage	age	gender
		0.834	0.228	-0.116	-0.183
AHNAKRS	0.107				
ANKRD28	0.134		0.138		
ASH2L		-0.297			
BTD		-0.312			
C1ORF140	-0.002	0.246	-0.083	-0.022	0.092
CD44					0.070
CHP1	0.107	0.045			

Continued on the next page

Table 16: Continued from the previous page.

Gene		clark	stage	age	gender
CXCL6	0.126	-0.120	-0.095		
DLG6	0.113	-0.015	0.067	0.185	-0.142
DOK5				-0.066	
ETNK2	0.152				
FILIP1	-0.030				
JADE1	-0.147				
JPH4		0.115			
KBF2	-0.032	0.182		0.034	-0.026
LRRN2	-0.061				
MAGED4	-0.098		-0.020		
NHSL2	-0.088				
PITPNA	0.151	-0.051	-0.012	-0.033	0.008
SOX8	0.088		-0.212		
TMEM145					0.048
TMEM159	0.160	-0.121	-0.042		0.189
WBSCR27		0.070		0.126	

Table 17: Analysis of the TCGA SKCM data using RBL-SS.

Gene		clark	stage	age	gender
		0.926	-0.062	-0.011	0.388
AHNAKRS	0.084				
ANKRD28	0.191		0.207		
ASH2L		-0.258			
BAIAP2	0.043				
BTD		-0.309		-0.255	
C1ORF140		0.129			
C1ORF54					-0.102
CHP1	0.081				-0.111
CPXM1				0.005	
CSNK2A2	-0.003				
CYP1B1-AS1			0.104		
DAP	0.036			-0.116	
DLG6				0.242	
ETNK2	0.109				
FHL5		0.220			
FILIP1			-0.016		
GAMT				0.082	
IL11RA	-0.087				
IQCK					-0.090

Continued on the next page

Table 17: Continued from the previous page.

Gene		clark	stage	age	gender
JADE1	-0.161				
JPH4		0.159			
KDM6B			-0.142		
LRFN2		0.096			
MAGED4	-0.130				
MAPE					-0.191
MPD1	-0.078				
NHSL2	-0.144	-0.306			
PAX1	0.171		0.217		
PBX2	0.141			0.130	
PITPNA	0.161			-0.056	
RNPEPL1			0.052		
SLC12A5					-0.081
SOX8	0.140		-0.091		
STPG1			0.184		
TMEM145					0.222
TMEM159	0.123				
TNFAIP1				0.283	
TP53TG1		0.102			-0.063
WBSCR27		0.090		0.126	

Table 18: Analysis of the TCGA SKCM data using BSG-SS.

Gene		clark	stage	age	gender
		0.487	0.163	0.048	0.087
AHNAKRS	0.120				
ANKRD28	0.138				
ARMC9	0.008				
ASH2L	0.019	-0.194		-0.107	
BTD		-0.303		-0.138	
C14ORF2		0.251			
C1ORF140		0.100	0.024		0.029
CD44					0.125
CHP1	0.123				
CPXM1	-0.047				
CXCL6	0.032				
DLG6		0.093		0.204	-0.061
DOK5			-0.052		
ETNK2	0.094				
FILIP1	-0.049				

Continued on the next page

Table 18: Continued from the previous page.

Gene		clark	stage	age	gender
GAMT		-0.004			
IL11RA	-0.045				
JADE1	-0.149				
JPH4		0.110			
KBF2	-0.077				
LRRN2	-0.073				
MAGED4	-0.122				
MAPE					-0.217
NHSL2	-0.026				
PBX2	0.133			0.155	
PHP1B	-0.076				
PITPNA	0.150	0.077		0.038	-0.039
SOX8	0.103		-0.148		
STPG1			0.197		
TMEM145	0.015	-0.045			0.147
TMEM159	0.140				0.113
TNFRSF4		0.077			
TP53TG1		0.072			
WBSCR27		0.015		0.092	
ZFP62	-0.010				

Table 19: Analysis of the TCGA SKCM data using BL-SS.

Gene		clark	stage	age	gender
		0.545	0.308	0.080	0.047
AHNAKRS	0.102				
ANKRD28	0.180		0.134		
ASH2L				-0.185	
BTD		-0.386			
C14ORF2		0.126			
C1ORF140		0.199			
CELSR2			0.112		
CHP1	0.080				
CPXM1	-0.067				
CSNK2A2	-0.026				
CYP1B1-AS1			0.104		
DAP				-0.139	
DLG6	0.088			0.236	
ETNK2	0.206				-0.089
FHL5		0.076			

Continued on the next page

Table 19: Continued from the previous page.

Gene		clark	stage	age	gender
FILIP1			-0.062		
GAMT				0.058	
IL11RA	-0.056				
IQCK					-0.098
JADE1	-0.203				
JPH4		0.101			
KBF2	-0.089				
KDM6B			-0.173		
LRFN2		0.109			
LRRN2	-0.091				
MAGED4	-0.113				
MAPE					-0.114
MPD1	-0.100				
NHSL2	-0.035				
PAX1			0.050		
PBX2	0.126			0.072	
PHP1B					-0.054
PIP4K2C			-0.101		
PITPNA	0.193				
PTP4A3			-0.138		
RNPEPL1			0.171		
SAA2	0.021			-0.058	
SLC12A5					-0.112
SOX8	0.132		-0.084		
TIE1	-0.093				
TMEM145					0.188
TMEM159	0.174				0.181
TP53TG1		0.156			-0.030
WBSCR27		0.048		0.105	

F Biological similarity analysis

We carried out an examination of the Gene Ontology (GO) biological processes which provide us with a deeper insight on the differences of the markers identified by different methods. We totally identified 77 unique genes using our proposed method along with three other methods for the NHS data. We conducted the GO enrichment analysis using the R package GOSim and found these genes involve in a total of 158 GO biological processes, the p-values of which are smaller than 0.1 in the GO enrichment analysis. Then we divided the 158 processes into four categories: positive regulation (P), negative regulation (N), regulation (R, without a well-defined "direction") and other (O). We computed the proportions of genes that involve in the four categories of processes for each of the four methods. Similarly, for the TCGA SKCM data, 109 genes were identified by our method along with three other alternative

methods. GO enrichment analysis showed that they involve in 183 biological processes, with p-values smaller than 0.1. The results for NHS and TCGA SKCM are provided in Figure 3, which shows an obvious difference between our proposed method and the three alternatives in both datasets.

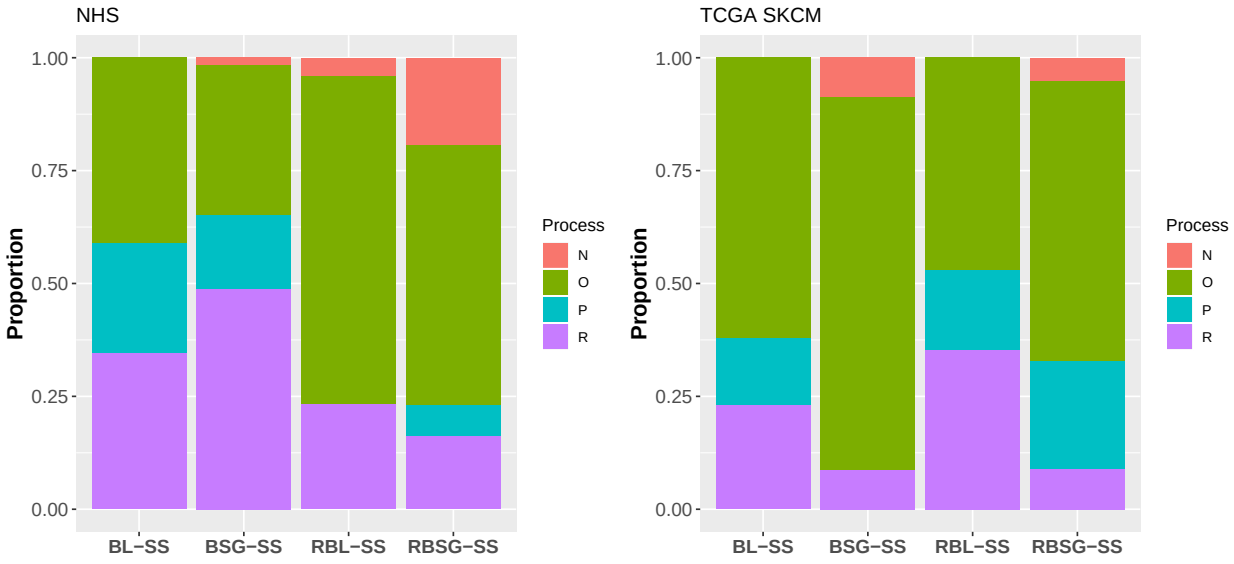

Figure 3: Gene Ontology (GO) analysis: proportions of genes that have the four categories of processes with different approaches. Left: NHS data. Right: TCGA SKCM data.

G Posterior inference

G.1 RBG-SS

G.1.1 Hierarchical model specification

$$y_i = W_i^\top \alpha + E_i^\top \theta + U_i^\top \beta + \nu^{-\frac{1}{2}} \kappa \sqrt{u_i} z_i \quad i = 1, \ldots, n$$

$$u_i | \nu \overset{ind}{\sim} \nu \exp\left(-\nu u_i\right) \quad i = 1, \ldots, n$$

$$z_i \overset{ind}{\sim} \mathrm{N}(0,\ 1) \quad i = 1, \ldots, n$$

$$\nu \sim \mathrm{Gamma}\left(c_1,\ c_2\right)$$

$$\alpha \sim \mathrm{N}_q(0,\ \Sigma_{\alpha 0})$$

$$\theta \sim \mathrm{N}_k(0,\ \Sigma_{\theta 0})$$

$$\beta_j | \phi_j, s_j \overset{ind}{\sim} \phi_j \, \mathrm{N}_L\left(0,\ s_j \mathbf{I}_L\right) + (1 - \phi_j)\delta_0(\beta_j) \quad j = 1, \ldots, p$$

$$\phi_j | \pi_0 \overset{ind}{\sim} \mathrm{Bernoulli}(\pi_0) \quad j = 1, \ldots, p$$

$$\pi_0 \sim \mathrm{Beta}\left(a_0,\ b_0\right)$$

$$s_j | \eta \sim \mathrm{Gamma}\left(\frac{L+1}{2}, \frac{\eta}{2}\right) \quad j = 1, \ldots, p$$

$$\eta \sim \mathrm{Gamma}\left(d_1,\ d_2\right)$$

G.1.2 Gibbs Sampler

- $u_i | \text{rest} \sim \text{Inverse-Gaussian}(\mu_{u_i},\ \lambda_{u_i})$, where the shape parameter $\lambda_{u_i} = 2\nu$, mean parameter $\mu_{u_i} = \sqrt{\frac{2\kappa^2}{(y_i - \tilde{y}_i)^2}}$ and $\tilde{y}_i = y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta$.
- $\nu | \text{rest} \sim \mathrm{Gamma}\left(s_\nu,\ r_\nu\right)$, where the shape parameter $s_\nu = c_1 + \frac{3n}{2}$ and the rate parameter $r_\nu = c_2 + \sum_{i=1}^n u_i + (2\kappa^2)^{-1} \sum_{i=1}^n u_i^{-1} \tilde{y}_i^2$.
- $\alpha | \text{rest} \sim \mathrm{N}(\mu_\alpha,\ \Sigma_\alpha)$, where

$$\mu_\alpha = \Sigma_\alpha \nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} W_i (y_i - E_i^\top \theta - U_i^\top \beta)$$

$$\Sigma_\alpha = \left(\nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} W_i W_i^\top + \Sigma_{\alpha 0}^{-1} \right)^{-1}$$

- $\theta|\text{rest} \sim \text{N}(\mu_\theta,\, \Sigma_\theta)$, where

$$\mu_\theta = \Sigma_\theta \nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} E_i (y_i - W_i^\top \alpha - U_i^\top \beta)$$

$$\Sigma_\theta = \left(\nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} E_i E_i^\top + \Sigma_{\theta 0}^{-1} \right)^{-1}$$

- $\beta_j|\text{rest} \sim l_j \text{N}(\mu_{\beta_j},\, \Sigma_{\beta_j}) + (1 - l_j)\delta_0(\beta_j)$ where

$$\mu_{\beta_j} = \Sigma_{\beta_j} \nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} U_{ij} \tilde{y}_{ij}$$

$$\Sigma_{\beta_j} = \left(\nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} U_{ij} U_{ij}^\top + \frac{1}{s_j} \mathbf{I}_L \right)^{-1}$$

$$l_j = \frac{\pi_0}{\pi_0 + (1 - \pi_0) s_j^{\frac{L}{2}} |\Sigma_{\beta_j}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \| \Sigma_{\beta_j}^{\frac{1}{2}} \nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} U_{ij} \tilde{y}_{ij} \|_2^2 \right\}}$$

  and $\tilde{y}_{ij}$ is defined as $\tilde{y}_{ij} = y_i - W_i^\top \alpha - E_i^\top \theta - U_{(j)}^\top \beta_{(j)}$.

- The posterior of s_j is

$$s_j^{-1}|\text{rest} \sim \begin{cases} \text{Inverse-Gamma}(\frac{L+1}{2},\ \frac{\eta}{2}) & \text{if } \beta_j = 0 \\ \text{Inverse-Gaussian}(\eta,\ \sqrt{\frac{\eta}{\|\beta_j\|_2^2}}) & \text{if } \beta_j \neq 0 \end{cases}$$

- $\pi_0|\text{rest} \sim \text{Beta}\left(a_0 + \sum_{j=1}^{p} \mathbf{I}_{\{\beta_j \neq 0\}},\ b_0 + \sum_{j=1}^{p} \mathbf{I}_{\{\beta_j = 0\}} \right)$

- $\eta|\text{rest} \sim \text{Gamma}\,(s_\eta,\, r_\eta)$, where $s_\eta = \frac{p + p \times L}{2} + d_1$ and the rate parameter $r_\eta = \frac{\sum_{j=1}^{p} s_j}{2} + d_2$.

G.2 RBL-SS

G.2.1 Hierarchical model specification

$$
\begin{gathered}
y_i = W_i^\top \alpha + E_i^\top \theta + U_i^\top \beta + \nu^{-\frac{1}{2}} \kappa \sqrt{u_i} z_i \\
u_i | \nu \overset{ind}{\sim} \nu \exp\left(-\nu u_i\right) \\
z_i \overset{ind}{\sim} \mathrm{N}(0,\ 1) \\
\nu \sim \mathrm{Gamma}\left(c_1,\ c_2\right) \\
\alpha \sim \mathrm{N}_q(0,\ \Sigma_{\alpha 0}) \\
\theta \sim \mathrm{N}_k(0,\ \Sigma_{\theta 0}) \\
\beta_{jl} | \phi_{jl}, s_{jl} \overset{ind}{\sim} \phi_{jl}\, \mathrm{N}\left(0,\ s_{jl}\right) + (1 - \phi_{jl}) \delta_0(\beta_{jl}) \quad j = 1, \ldots, p;\ l = 1, \ldots, L \\
\phi_{jl} | \pi_1 \overset{ind}{\sim} \mathrm{Bernoulli}(\pi_1) \quad j = 1, \ldots, p;\ l = 1, \ldots, L \\
s_{jl} | \eta \sim \mathrm{Gamma}\left(1, \frac{\eta}{2}\right) \quad j = 1, \ldots, p;\ l = 1, \ldots, L \\
\pi_1 \sim \mathrm{Beta}\left(a_1,\ b_1\right) \\
\eta \sim \mathrm{Gamma}\left(d_1,\ d_2\right)
\end{gathered}
$$

G.2.2 Gibbs Sampler

- $u_i|\text{rest} \sim \text{Inverse-Gaussian}(\mu_{u_i},\ \lambda_{u_i})$, where the shape parameter $\lambda_{u_i} = 2\nu$, mean parameter $\mu_{u_i} = \sqrt{\frac{2\kappa^2}{(Y_i - \tilde{y}_i)^2}}$ and $\tilde{y}_i = y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta$.
- $\nu|\text{rest} \sim \mathrm{Gamma}\left(s_\nu,\ r_\nu\right)$, where the shape parameter $s_\nu = c_1 + \frac{3n}{2}$ and the rate parameter $r_\nu = c_2 + \sum_{i=1}^n u_i + (2\kappa^2)^{-1} \sum_{i=1}^n u_i^{-1} \tilde{y}_i{}^2$.
- $\alpha|\text{rest} \sim \mathrm{N}(\mu_\alpha,\ \Sigma_\alpha)$, where

$$
\begin{aligned}
\mu_\alpha &= \Sigma_\alpha \nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} W_i (Y_i - E_i^\top \theta - U_i^\top \beta) \\
\Sigma_\alpha &= \left(\nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} W_i W_i^\top + \Sigma_{\alpha 0}^{-1} \right)^{-1}
\end{aligned}
$$

- $\theta|\text{rest} \sim \mathrm{N}(\mu_\theta,\ \Sigma_\theta)$, where

$$
\begin{aligned}
\mu_\theta &= \Sigma_\theta \nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} E_i (Y_i - W_i^\top \alpha - U_i^\top \beta) \\
\Sigma_\theta &= \left(\nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} E_i E_i^\top + \Sigma_{\theta 0}^{-1} \right)^{-1}
\end{aligned}
$$

- $\beta_{jl}|\text{rest} \sim l_{jl}\text{N}(\mu_{\beta_{jl}},\, \sigma^2_{\beta_{jl}}) + (1 - l_{jl})\delta_0(\beta_{jl})$ where

$$\mu_{\beta_{jl}} = \sigma^2_{\beta_{jl}} \nu\kappa^{-2} \sum_{i=1}^{n} u_i^{-1} U_{ijl} \tilde{y}_{ijl}$$

$$\sigma^2_{\beta_{jl}} = \left(\nu\kappa^{-2} \sum_{i=1}^{n} u_i^{-1} U^2_{ijl} + \frac{1}{s_{jl}} \right)^{-1}$$

$$l_{jl} = \frac{\pi_1}{\pi_1 + (1-\pi_1) s_{jl}^{\frac{1}{2}} (\sigma^2_{\beta_{jl}})^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}\sigma^2_{\beta_{jl}} (\nu\kappa^{-2} \sum_{i=1}^{n} u_i^{-1} U_{ijl} \tilde{y}_{ijl})^2 \right\}}$$

  and $\tilde{y}_{ijl}$ is defined as $\tilde{y}_{ijl} = y_i - W_i^\top \alpha - E_i^\top \theta - U_{(jl)}^\top \beta_{(jl)}$.

- The posterior of s_{jl} is

$$s_{jl}^{-1}|\text{rest} \sim \begin{cases} \text{Inverse-Gamma}(1,\ \frac{\eta}{2}) & \text{if } \beta_{jl} = 0 \\ \text{Inverse-Gaussian}(\eta,\ \sqrt{\frac{\eta}{\beta^2_{jl}}}) & \text{if } \beta_{jl} \neq 0 \end{cases}$$

- $\pi_1|\text{rest} \sim \text{Beta}\left(a_1 + \sum_{j,l} \mathbf{I}_{\{\beta_{jl} \neq 0\}},\ b_1 + \sum_{j,l} \mathbf{I}_{\{\beta_{jl}=0\}}\right)$

- $\eta|\text{rest} \sim \text{Gamma}\,(s_\eta,\, r_\eta)$, where $s_\eta = p \times L + d_1$ and the rate parameter $r_\eta = \frac{\sum_{j,l} s_{jl}}{2} + d_2$.

G.3 RBSG

G.3.1 Hierarchical model specification

$$y_i = W_i^\top \alpha + E_i^\top \theta + U_i^\top \beta + \nu^{-\frac{1}{2}} \kappa \sqrt{u_i} z_i$$

$$u_i|\nu \overset{ind}{\sim} \nu \exp\left(-\nu u_i\right)$$

$$z_i \overset{ind}{\sim} \text{N}(0,\ 1)$$

$$\nu \sim \text{Gamma}\,(c_1,\ c_2)$$

$$\alpha \sim \text{N}_q(0,\, \Sigma_{\alpha 0})$$

$$\theta \sim \text{N}_k(0,\, \Sigma_{\theta 0})$$

$$\beta_j|r_j, \omega_{jl} \overset{ind}{\sim} \text{N}_L(0,\, V_j),\ \text{where } V_j = \text{diag}\left\{ \left(\frac{1}{r_j} + \frac{1}{\omega_{jl}}\right)^{-1},\ l = 1, 2, \ldots, L \right\}$$

$$r_j, \omega_{j1}, \ldots, \omega_{jL}|\eta_1, \eta_2 \propto \prod_{l=1}^{L} \left[(\omega_{jl})^{-\frac{1}{2}} \left(\frac{1}{r_j} + \frac{1}{\omega_{jl}} \right)^{-\frac{1}{2}} \right] (r_j)^{-\frac{1}{2}} \exp\left(-\frac{\eta_1}{2} r_j - \frac{\eta_2}{2} \sum_{l=1}^{L} \omega_{jl} \right)$$

$$\eta_1, \eta_2 \propto \eta_1^{\frac{p}{2}} \eta_2^{pL} \exp\left\{ -d_1\eta_1 - d_2\eta_2 \right\}$$

$$\sigma^2 \sim 1/\sigma^2$$

G.3.2 Gibbs Sampler

- $u_i|\text{rest} \sim$ Inverse-Gaussian$(\mu_{u_i},\, \lambda_{u_i})$, where the shape parameter $\lambda_{u_i} = 2\nu$, mean parameter $\mu_{u_i} = \sqrt{\frac{2\kappa^2}{(Y_i - \tilde{y}_i)^2}}$ and $\tilde{y}_i = y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta$.

- $\nu|\text{rest} \sim \text{Gamma}\,(s_\nu,\, r_\nu)$, where the shape parameter $s_\nu = c_1 + \frac{3n}{2}$ and the rate parameter $r_\nu = c_2 + \sum_{i=1}^n u_i + (2\kappa^2)^{-1} \sum_{i=1}^n u_i^{-1} \tilde{y}_i^{\,2}$.

- $\alpha|\text{rest} \sim \text{N}(\mu_\alpha,\, \Sigma_\alpha)$, where

$$\mu_\alpha = \Sigma_\alpha \nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} W_i (Y_i - E_i^\top \theta - U_i^\top \beta)$$

$$\Sigma_\alpha = \left(\nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} W_i W_i^\top + \Sigma_{\alpha 0}^{-1} \right)^{-1}$$

- $\theta|\text{rest} \sim \text{N}(\mu_\theta,\, \Sigma_\theta)$, where

$$\mu_\theta = \Sigma_\theta \nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} E_i (Y_i - W_i^\top \alpha - U_i^\top \beta)$$

$$\Sigma_\theta = \left(\nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} E_i E_i^\top + \Sigma_{\theta 0}^{-1} \right)^{-1}$$

- $\beta_j|\text{rest} \sim \text{N}(\mu_{\beta_j},\, \Sigma_{\beta_j})$ where

$$\mu_{\beta_j} = \Sigma_{\beta_j} \nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} U_{ij} \tilde{y}_{ij}$$

$$\Sigma_{\beta_j} = \left(\nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} U_{ij} U_{ij}^\top + V_j^{-1} \right)^{-1}$$

  and $\tilde{y}_{ij}$ is defined as $\tilde{y}_{ij} = y_i - W_i^\top \alpha - E_i^\top \theta - U_{(j)}^\top \beta_{(j)}$.

- $r_j^{-1}|\text{rest} \sim$ Inv-Gaussian$(\eta_1,\, \sqrt{\frac{\eta_1 \sigma^2}{\|\beta_j\|_2^2}})$

- $\omega_{jl}^{-1}|\text{rest} \sim$ Inv-Gaussian$(\eta_2,\, \sqrt{\frac{\eta_2 \sigma^2}{\beta_{jl}^2}})$

- $\eta_1|\text{rest} \sim \text{Gamma}\,(s_{\eta_1},\, r_{\eta_1})$, where $s_{\eta_1} = \frac{p}{2} + 1$ and the rate parameter $r_{\eta_1} = \frac{\sum_{j=1}^p r_j}{2} + d_1$.

- $\eta_2|\text{rest} \sim \text{Gamma}\,(s_{\eta_2},\, r_{\eta_2})$, where $s_{\eta_2} = p \times L + 1$ and the rate parameter $r_{\eta_2} = \frac{\sum_{j,l} \omega_{jl}}{2} + d_2$.

G.4 RBG

G.4.1 Hierarchical model specification

$$
\begin{aligned}
y_i &= W_i^\top \alpha + E_i^\top \theta + U_i^\top \beta + \nu^{-\frac{1}{2}} \kappa \sqrt{u_i} z_i \quad i = 1, \ldots, n \\
u_i|\nu &\overset{ind}{\sim} \nu \exp\left(-\nu u_i\right) \quad i = 1, \ldots, n \\
z_i &\overset{ind}{\sim} \mathrm{N}(0,\ 1) \quad i = 1, \ldots, n \\
\nu &\sim \mathrm{Gamma}\left(c_1,\ c_2\right) \\
\alpha &\sim \mathrm{N}_q(0,\ \Sigma_{\alpha 0}) \\
\theta &\sim \mathrm{N}_k(0,\ \Sigma_{\theta 0}) \\
\beta_j|s_j &\overset{ind}{\sim} \mathrm{N}_L\left(0,\ s_j \mathbf{I}_L\right) \quad j = 1, \ldots, p \\
s_j|\eta &\sim \mathrm{Gamma}\left(\frac{L+1}{2}, \frac{\eta}{2}\right) \quad j = 1, \ldots, p \\
\eta &\sim \mathrm{Gamma}\left(d_1,\ d_2\right)
\end{aligned}
$$

G.4.2 Gibbs Sampler

- $u_i|\text{rest} \sim$ Inverse-Gaussian$(\mu_{u_i},\ \lambda_{u_i})$, where the shape parameter $\lambda_{u_i} = 2\nu$, mean parameter $\mu_{u_i} = \sqrt{\frac{2\kappa^2}{(y_i - \tilde{y}_i)^2}}$ and $\tilde{y}_i = y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta$.
- $\nu|\text{rest} \sim \mathrm{Gamma}\left(s_\nu,\ r_\nu\right)$, where the shape parameter $s_\nu = c_1 + \frac{3n}{2}$ and the rate parameter $r_\nu = c_2 + \sum_{i=1}^n u_i + (2\kappa^2)^{-1} \sum_{i=1}^n u_i^{-1} \tilde{y}_i^2$.
- $\alpha|\text{rest} \sim \mathrm{N}(\mu_\alpha,\ \Sigma_\alpha)$, where

$$
\mu_\alpha = \Sigma_\alpha \nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} W_i (y_i - E_i^\top \theta - U_i^\top \beta)
$$

$$
\Sigma_\alpha = \left(\nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} W_i W_i^\top + \Sigma_{\alpha 0}^{-1} \right)^{-1}
$$

- $\theta|\text{rest} \sim \mathrm{N}(\mu_\theta,\ \Sigma_\theta)$, where

$$
\mu_\theta = \Sigma_\theta \nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} E_i (y_i - W_i^\top \alpha - U_i^\top \beta)
$$

$$
\Sigma_\theta = \left(\nu \kappa^{-2} \sum_{i=1}^n u_i^{-1} E_i E_i^\top + \Sigma_{\theta 0}^{-1} \right)^{-1}
$$

- $\beta_j|\text{rest} \sim \text{N}(\mu_{\beta_j},\, \Sigma_{\beta_j})$ where

$$\mu_{\beta_j} = \Sigma_{\beta_j} \nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} U_{ij} \tilde{y}_{ij}$$

$$\Sigma_{\beta_j} = \left(\nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} U_{ij} U_{ij}^\top + \frac{1}{s_j} \mathbf{I}_L \right)^{-1}$$

  and $\tilde{y}_{ij}$ is defined as $\tilde{y}_{ij} = y_i - W_i^\top \alpha - E_i^\top \theta - U_{(j)}^\top \beta_{(j)}$.

- $s_j^{-1}|\text{rest} \sim \text{Inverse-Gaussian}(\eta,\, \sqrt{\frac{\eta}{\|\beta_j\|_2^2}})$

- $\eta|\text{rest} \sim \text{Gamma}\,(s_\eta,\, r_\eta)$, where $s_\eta = \frac{p + p \times L}{2} + d_1$ and the rate parameter $r_\eta = \frac{\sum_{j=1}^{p} s_j}{2} + d_2$.

G.5 RBL

G.5.1 Hierarchical model specification

$$\begin{gathered}
y_i = W_i^\top \alpha + E_i^\top \theta + U_i^\top \beta + \nu^{-\frac{1}{2}} \kappa \sqrt{u_i} z_i \\
u_i|\nu \overset{ind}{\sim} \nu \exp\left(-\nu u_i\right) \\
z_i \overset{ind}{\sim} \text{N}(0,\, 1) \\
\nu \sim \text{Gamma}\,(c_1,\, c_2) \\
\alpha \sim \text{N}_q(0,\, \Sigma_{\alpha 0}) \\
\theta \sim \text{N}_k(0,\, \Sigma_{\theta 0}) \\
\beta_{jl}|s_{jl} \overset{ind}{\sim} \text{N}\,(0,\, s_{jl}) \quad j = 1, \ldots, p;\; l = 1, \ldots, L \\
s_{jl}|\eta \sim \text{Gamma}\left(1, \frac{\eta}{2}\right) \quad j = 1, \ldots, p;\; l = 1, \ldots, L \\
\eta \sim \text{Gamma}\,(d_1,\, d_2)
\end{gathered}$$

G.5.2 Gibbs Sampler

- $u_i|\text{rest} \sim \text{Inverse-Gaussian}(\mu_{u_i},\, \lambda_{u_i})$, where the shape parameter $\lambda_{u_i} = 2\nu$, mean parameter $\mu_{u_i} = \sqrt{\frac{2\kappa^2}{(Y_i - \tilde{y}_i)^2}}$ and $\tilde{y}_i = y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta$.

- $\nu|\text{rest} \sim \text{Gamma}\,(s_\nu,\, r_\nu)$, where the shape parameter $s_\nu = c_1 + \frac{3n}{2}$ and the rate parameter $r_\nu = c_2 + \sum_{i=1}^{n} u_i + (2\kappa^2)^{-1} \sum_{i=1}^{n} u_i^{-1} \tilde{y}_i^{\,2}$.

- $\alpha|\text{rest} \sim \text{N}(\mu_\alpha,\, \Sigma_\alpha)$, where

$$\mu_\alpha = \Sigma_\alpha \nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} W_i (Y_i - E_i^\top \theta - U_i^\top \beta)$$

$$\Sigma_\alpha = \left(\nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} W_i W_i^\top + \Sigma_{\alpha 0}^{-1} \right)^{-1}$$

- $\theta|\text{rest} \sim \text{N}(\mu_\theta,\, \Sigma_\theta)$, where

$$\mu_\theta = \Sigma_\theta \nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} E_i (Y_i - W_i^\top \alpha - U_i^\top \beta)$$

$$\Sigma_\theta = \left(\nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} E_i E_i^\top + \Sigma_{\theta 0}^{-1} \right)^{-1}$$

- $\beta_{jl}|\text{rest} \sim \text{N}(\mu_{\beta_{jl}},\, \sigma^2_{\beta_{jl}})$ where

$$\mu_{\beta_{jl}} = \sigma^2_{\beta_{jl}} \nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} U_{ijl} \tilde{y}_{ijl}$$

$$\sigma^2_{\beta_{jl}} = \left(\nu \kappa^{-2} \sum_{i=1}^{n} u_i^{-1} U_{ijl}^2 + \frac{1}{s_{jl}} \right)^{-1}$$

  and $\tilde{y}_{ijl}$ is defined as $\tilde{y}_{ijl} = y_i - W_i^\top \alpha - E_i^\top \theta - U_{(jl)}^\top \beta_{(jl)}$.

- $s_j^{-1}|\text{rest} \sim \text{Inverse-Gaussian}(\eta,\, \sqrt{\frac{\eta}{\beta_{jl}^2}})$

- $\eta|\text{rest} \sim \text{Gamma}\,(s_\eta,\, r_\eta)$, where $s_\eta = p \times L + d_1$ and the rate parameter $r_\eta = \frac{\sum_{j,l} s_{jl}}{2} + d_2$.

G.6 BSG-SS

G.6.1 Hierarchical model specification

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta)^2\right\}$$

$$\alpha \sim \mathrm{N}_q(0,\, \Sigma_{\alpha 0})$$

$$\theta \sim \mathrm{N}_k(0,\, \Sigma_{\theta 0})$$

$$\beta_j = V_j^{\frac{1}{2}} b_j, \quad \text{where } V_j^{\frac{1}{2}} = \mathrm{diag}\{\omega_{j1}, \ldots, \omega_{jL}\}$$

$$b_j|\phi_j^b \overset{ind}{\sim} \phi_j^b \,\mathrm{N}_L\,(0,\, \mathbf{I}_L) + (1-\phi_j^b)\delta_0(b_j)$$

$$\phi_j^b|\pi_0 \overset{ind}{\sim} \text{Bernoulli}(\pi_0)$$

$$\pi_0 \sim \text{Beta}\,(a_0,\, b_0)$$

$$\omega_{jl}|\phi_{jl}^w \overset{ind}{\sim} \phi_{jl}^w \,\mathrm{N}^+\left(0,\, s^2\right) + (1-\phi_{jl}^w)\delta_0(\omega_{jl})$$

$$\phi_{jl}^w|\pi_1 \overset{ind}{\sim} \text{Bernoulli}(\pi_1)$$

$$\pi_1 \sim \text{Beta}\,(a_1,\, b_1)$$

$$s^2 \sim \text{Inverse-Gamma}\,(1, \eta)$$

$$\sigma^2 \sim 1/\sigma^2$$

G.6.2 Gibbs Sampler

- $\alpha|\text{rest} \sim \mathrm{N}(\mu_\alpha,\, \Sigma_\alpha)$, where

$$\mu_\alpha = \Sigma_\alpha(\sigma^2)^{-1}\sum_{i=1}^{n} W_i(y_i - E_i^\top \theta - U_i^\top \beta)$$

$$\Sigma_\alpha = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} W_i W_i^\top + \Sigma_{\alpha 0}^{-1}\right)^{-1}$$

- $\theta|\text{rest} \sim \mathrm{N}(\mu_\theta,\, \Sigma_\theta)$, where

$$\mu_\theta = \Sigma_\theta(\sigma^2)^{-1}\sum_{i=1}^{n} E_i(y_i - W_i^\top \alpha - U_i^\top \beta)$$

$$\Sigma_\theta = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} E_i E_i^\top + \Sigma_{\theta 0}^{-1}\right)^{-1}$$

- $b_j|\text{rest} \sim l_j \text{N}(\mu_{b_j}, \Sigma_{b_j}) + (1 - l_j)\delta_0(b_j)$ where

$$\mu_{b_j} = \Sigma_{b_j}(\sigma^2)^{-1} \sum_{i=1}^{n} V_j^{\frac{1}{2}} U_{ij} \tilde{y}_{ij}$$

$$\Sigma_{b_j} = \left(\frac{1}{\sigma^2} \sum_{i=1}^{n} V_j^{\frac{1}{2}} U_{ij} U_{ij}^\top V_j^{\frac{1}{2}} + \mathbf{I}_L \right)^{-1}$$

$$l_j^b = \frac{\pi_0}{\pi_0 + (1-\pi_0)|\Sigma_{b_j}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^4} \| \Sigma_{b_j}^{\frac{1}{2}} \sum_{i=1}^{n} V_j^{\frac{1}{2}} U_{ij} \tilde{y}_{ij} \|_2^2 \right\}}$$

  and $\tilde{y}_{ij}$ is defined as $\tilde{y}_{ij} = y_i - W_i^\top \alpha - E_i^\top \theta - U_{(j)}^\top \beta_{(j)}$.

- $\omega_{jl}|\text{rest} \sim l_{jl}^w \, \text{N}^+(\mu_{\omega_{jl}}, \sigma^2_{\omega_{jl}}) + (1 - l_{jl}^w)\delta_0(\omega_{jl})$ where

$$\mu_{\omega_{jl}} = \sigma^2_{\omega_{jl}} (\sigma^2)^{-1} \sum_{i=1}^{n} b_{jl} U_{ijl} \tilde{y}_{ijl}$$

$$\sigma^2_{\omega_{jl}} = \left(\frac{1}{\sigma^2} \sum_{i=1}^{n} U_{ijl}^2 b_{jl}^2 + \frac{1}{s^2} \right)^{-1}$$

$$l_{jl}^w = \frac{\pi_1}{\pi_1 + (1-\pi_1)\frac{1}{2} s (\sigma^2_{\omega_{jl}})^{-\frac{1}{2}} \exp\left\{ -\frac{\sigma^2_{\omega_{jl}}}{2\sigma^4} \left(\sum_{i=1}^{n} b_{jl} U_{ijl} \tilde{y}_{ijl} \right)^2 \right\} \left[\Phi\left(\frac{\mu_{\omega_{jl}}}{\sigma_{\omega_{jl}}} \right) \right]^{-1}}$$

  and $\tilde{y}_{ijl}$ is defined as $\tilde{y}_{ijl} = y_i - W_i^\top \alpha - E_i^\top \theta - U_{(jl)}^\top \beta_{(jl)}$.

- $s^2|\text{rest} \sim \text{Inv-Gamma}\left(1 + \frac{1}{2}\sum_{j,l} \mathbf{I}_{\{\omega_{jl} \neq 0\}},\ \eta + \frac{1}{2}\sum_{j,l} \omega_{jl}^2\right)$

- $\pi_0|\text{rest} \sim \text{Beta}\left(a_0 + \sum_{j=1}^{p} \mathbf{I}_{\{\beta_j \neq 0\}},\ b_0 + \sum_{j=1}^{p} \mathbf{I}_{\{\beta_j = 0\}}\right)$

- $\pi_1|\text{rest} \sim \text{Beta}\left(a_1 + \sum_{j,l} \mathbf{I}_{\{\omega_{jl} \neq 0\}},\ b_1 + \sum_{j,l} \mathbf{I}_{\{\omega_{jl} = 0\}}\right)$

- η is estimated with the EM approach used in the proposed method. For the gth EM update $\eta^{(g)} = \frac{1}{E_{\eta^{(g-1)}}\left[\frac{1}{s^2}|Y\right]}$.

- $\sigma^2|\text{rest} \sim \text{Inv-Gamma}(\frac{n}{2},\ \frac{\sum_{i=1}^{n} \tilde{y_i}^2}{2})$, where $\tilde{y_i} = y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta$.

G.7 BGL-SS

G.7.1 Hierarchical model specification

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta)^2\right\}$$

$$\alpha \sim \mathrm{N}_q(0,\, \Sigma_{\alpha 0})$$

$$\theta \sim \mathrm{N}_k(0,\, \Sigma_{\theta 0})$$

$$\beta_j|\phi_j, \sigma^2, s_j \overset{ind}{\sim} \phi_j \,\mathrm{N}_L\left(0,\, \sigma^2 s_j \mathbf{I}_L\right) + (1-\phi_j)\delta_0(\beta_j) \quad j = 1, \ldots, p$$

$$\phi_j|\pi_0 \overset{ind}{\sim} \text{Bernoulli}(\pi_0) \quad j = 1, \ldots, p$$

$$\pi_0 \sim \text{Beta}\,(a_0,\, b_0)$$

$$s_j|\eta \overset{ind}{\sim} \text{Gamma}\left(\frac{L+1}{2}, \frac{\eta}{2}\right) \quad j = 1, \ldots, p$$

$$\eta \sim \text{Gamma}\,(d_1,\ d_2)$$

$$\sigma^2 \sim 1/\sigma^2$$

G.7.2 Gibbs Sampler

- $\alpha|\text{rest} \sim \mathrm{N}(\mu_\alpha,\, \Sigma_\alpha)$, where

$$\mu_\alpha = \Sigma_\alpha(\sigma^2)^{-1}\sum_{i=1}^{n} W_i(y_i - E_i^\top \theta - U_i^\top \beta)$$

$$\Sigma_\alpha = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} W_i W_i^\top + \Sigma_{\alpha 0}^{-1}\right)^{-1}$$

- $\theta|\text{rest} \sim \mathrm{N}(\mu_\theta,\, \Sigma_\theta)$, where

$$\mu_\theta = \Sigma_\theta(\sigma^2)^{-1}\sum_{i=1}^{n} E_i(y_i - W_i^\top \alpha - U_i^\top \beta)$$

$$\Sigma_\theta = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} E_i E_i^\top + \Sigma_{\theta 0}^{-1}\right)^{-1}$$

- $\beta_j|\text{rest} \sim l_j \text{N}(\mu_{\beta_j},\, \sigma^2\Sigma_{\beta_j}) + (1-l_j)\delta_0(\beta_j)$ where

$$\mu_{\beta_j} = \Sigma_{\beta_j} \sum_{i=1}^{n} U_{ij}\tilde{y}_{ij}$$

$$\Sigma_{\beta_j} = \left(\sum_{i=1}^{n} U_{ij}U_{ij}^\top + \frac{1}{s_j}\mathbf{I}_L \right)^{-1}$$

$$l_j = \frac{\pi_0}{\pi_0 + (1-\pi_0) s_j^{\frac{L}{2}} |\Sigma_{\beta_j}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \| \Sigma_{\beta_j}^{\frac{1}{2}} \sum_{i=1}^{n} U_{ij}\tilde{y}_{ij} \|_2^2 \right\}}$$

  and $\tilde{y}_{ij}$ is defined as $\tilde{y}_{ij} = y_i - W_i^\top \alpha - E_i^\top \theta - U_{(j)}^\top \beta_{(j)}$.

- The posterior of s_j is

$$s_j^{-1}|\text{rest} \sim \begin{cases} \text{Inverse-Gamma}(\frac{L+1}{2},\ \frac{\eta}{2}) & \text{if } \beta_j = 0 \\ \text{Inverse-Gaussian}(\eta,\ \sqrt{\frac{\eta\sigma^2}{\|\beta_j\|_2^2}}) & \text{if } \beta_j \neq 0 \end{cases}$$

- $\pi_0|\text{rest} \sim \text{Beta}\left(a_0 + \sum_{j=1}^{p} \mathbf{I}_{\{\beta_j \neq 0\}},\ b_0 + \sum_{j=1}^{p} \mathbf{I}_{\{\beta_j = 0\}} \right)$

- $\eta|\text{rest} \sim \text{Gamma}\,(s_\eta,\, r_\eta)$, where $s_\eta = \frac{p+p\times L}{2} + d_1$ and the rate parameter $r_\eta = \frac{\sum_{j=1}^{p} s_j}{2} + d_2$.

- $\sigma^2|\text{rest} \sim \text{Inv-Gamma}(\frac{n+L\sum_{j=1}^{p} \mathbf{I}_{\{\beta_j \neq 0\}}}{2},\ \frac{\sum_{i=1}^{n} \tilde{y_i}^2 + \sum_{j=1}^{p} (s_j)^{-1} \beta_j^\top \beta_j}{2})$, where $\tilde{y_i} = y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta$.

G.8 BL-SS

G.8.1 Hierarchical model specification

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta)^2\right\}$$

$$\alpha \sim \mathrm{N}_q(0,\, \Sigma_{\alpha 0})$$

$$\theta \sim \mathrm{N}_k(0,\, \Sigma_{\theta 0})$$

$$\beta_{jl}|\phi_{jl}, \sigma^2, s_{jl} \overset{ind}{\sim} \phi_{jl}\,\mathrm{N}\left(0,\, \sigma^2 s_{jl}\right) + (1-\phi_{jl})\delta_0(\beta_{jl}) \quad j = 1, \ldots, p;\ l = 1, \ldots, L$$

$$\phi_{jl}|\pi_1 \overset{ind}{\sim} \text{Bernoulli}(\pi_1) \quad j = 1, \ldots, p;\ l = 1, \ldots, L$$

$$s_{jl}|\eta \overset{ind}{\sim} \text{Gamma}\left(1, \frac{\eta}{2}\right) \quad j = 1, \ldots, p;\ l = 1, \ldots, L$$

$$\pi_1 \sim \text{Beta}\,(a_1,\, b_1)$$

$$\eta \sim \text{Gamma}\,(d_1,\, d_2)$$

$$\sigma^2 \sim 1/\sigma^2$$

G.8.2 Gibbs Sampler

- $\alpha|\text{rest} \sim \mathrm{N}(\mu_\alpha,\, \Sigma_\alpha)$, where

$$\mu_\alpha = \Sigma_\alpha(\sigma^2)^{-1}\sum_{i=1}^{n} W_i(y_i - E_i^\top \theta - U_i^\top \beta)$$

$$\Sigma_\alpha = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} W_i W_i^\top + \Sigma_{\alpha 0}^{-1}\right)^{-1}$$

- $\theta|\text{rest} \sim \mathrm{N}(\mu_\theta,\, \Sigma_\theta)$, where

$$\mu_\theta = \Sigma_\theta(\sigma^2)^{-1}\sum_{i=1}^{n} E_i(y_i - W_i^\top \alpha - U_i^\top \beta)$$

$$\Sigma_\theta = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} E_i E_i^\top + \Sigma_{\theta 0}^{-1}\right)^{-1}$$

- $\beta_{jl}|\text{rest} \sim l_{jl}\text{N}(\mu_{\beta_{jl}},\, \sigma^2_{\beta_{jl}}) + (1 - l_{jl})\delta_0(\beta_{jl})$ where

$$\mu_{\beta_{jl}} = \sigma^2_{\beta_{jl}}(\sigma^2)^{-1}\sum_{i=1}^{n} U_{ijl}\tilde{y}_{ijl}$$

$$\sigma^2_{\beta_{jl}} = \sigma^2\left(\sum_{i=1}^{n} U^2_{ijl} + \frac{1}{s_{jl}}\right)^{-1}$$

$$l_{jl} = \frac{\pi_0}{\pi_0 + (1-\pi_0)s_{jl}^{\frac{1}{2}}(\sigma^2_{\beta_{jl}})^{-\frac{1}{2}}(\sigma^2)^{-\frac{1}{2}}\exp\left\{-\frac{\sigma^2_{\beta_{jl}}}{2\sigma^4}\left(\sum_{i=1}^{n} U_{ijl}\tilde{y}_{ijl}\right)^2\right\}}$$

  and $\tilde{y}_{ijl}$ is defined as $\tilde{y}_{ijl} = y_i - W_i^\top\alpha - E_i^\top\theta - U_{(jl)}^\top\beta_{(jl)}$.

- The posterior of s_{jl} is

$$s_{jl}^{-1}|\text{rest} \sim \begin{cases} \text{Inverse-Gamma}(1,\ \frac{\eta}{2}) & \text{if } \beta_{jl} = 0 \\ \text{Inverse-Gaussian}(\eta,\ \sqrt{\frac{\eta\sigma^2}{\beta^2_{jl}}}) & \text{if } \beta_{jl} \neq 0 \end{cases}$$

- $\pi_1|\text{rest} \sim \text{Beta}\left(a_1 + \sum_{j,l}\mathbf{I}_{\{\beta_{jl}\neq 0\}},\ b_1 + \sum_{j,l}\mathbf{I}_{\{\beta_{jl}=0\}}\right)$

- $\eta|\text{rest} \sim \text{Gamma}\,(s_\eta,\, r_\eta)$, where $s_\eta = p\times L + d_1$ and the rate parameter $r_\eta = \frac{\sum_{j,l} s_{jl}}{2} + d_2$.

- $\sigma^2|\text{rest} \sim \text{Inv-Gamma}(\frac{n+\sum_{j,l}\mathbf{I}_{\{\beta_{jl}\neq 0\}}}{2},\ \frac{\sum_{i=1}^{n}\tilde{y_i}^2 + \sum_{j,l}(s_{jl}^{-1})\beta^2_{jl}}{2})$, where $\tilde{y}_i = y_i - W_i^\top\alpha - E_i^\top\theta - U_i^\top\beta$.

G.9 BSG

G.9.1 Hierarchical model specification

$$Y|\alpha,\theta,\beta,\sigma^2 \propto (\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - W_i^\top\alpha - E_i^\top\theta - U_i^\top\beta)^2\right\}$$

$$\alpha \sim \text{N}_q(0,\, \Sigma_{\alpha 0})$$

$$\theta \sim \text{N}_k(0,\, \Sigma_{\theta 0})$$

$$\beta_j|\omega_{jl}, r_j \overset{ind}{\sim} \text{N}_L(0,\, \sigma^2 V_j),\ \text{where } V_j = \text{diag}\left\{\left(\frac{1}{r_j} + \frac{1}{\omega_{jl}}\right)^{-1},\ l = 1, 2, \ldots, L\right\}$$

$$r_j, \omega_{j1}, \ldots, \omega_{jL}|\eta_1,\eta_2 \propto \prod_{l=1}^{L}\left[(\omega_{jl})^{-\frac{1}{2}}\left(\frac{1}{r_j} + \frac{1}{\omega_{jl}}\right)^{-\frac{1}{2}}\right](r_j)^{-\frac{1}{2}}\exp\left(-\frac{\eta_1}{2}r_j - \frac{\eta_2}{2}\sum_{l=1}^{L}\omega_{jl}\right)$$

$$\eta_1,\eta_2 \propto \eta_1^{\frac{p}{2}}\eta_2^{pL}\exp\{-d_1\eta_1 - d_2\eta_2\}$$

$$\sigma^2 \sim 1/\sigma^2$$

G.9.2 Gibbs Sampler

- $\alpha|\text{rest} \sim \text{N}(\mu_\alpha, \Sigma_\alpha)$, where

$$\mu_\alpha = \Sigma_\alpha(\sigma^2)^{-1}\sum_{i=1}^{n} W_i(y_i - E_i^\top\theta - U_i^\top\beta)$$

$$\Sigma_\alpha = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} W_i W_i^\top + \Sigma_{\alpha 0}^{-1}\right)^{-1}$$

- $\theta|\text{rest} \sim \text{N}(\mu_\theta, \Sigma_\theta)$, where

$$\mu_\theta = \Sigma_\theta(\sigma^2)^{-1}\sum_{i=1}^{n} E_i(y_i - W_i^\top\alpha - U_i^\top\beta)$$

$$\Sigma_\theta = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} E_i E_i^\top + \Sigma_{\theta 0}^{-1}\right)^{-1}$$

- $\beta_j|\text{rest} \sim \text{N}(\mu_{\beta_j}, \Sigma_{\beta_j})$ where

$$\mu_{\beta_j} = \Sigma_{\beta_j}(\sigma^2)^{-1}\sum_{i=1}^{n} U_{ij}\tilde{y}_{ij}$$

$$\Sigma_{\beta_j} = \sigma^2\left(\sum_{i=1}^{n} U_{ij}U_{ij}^\top + V_j^{-1}\right)^{-1}$$

  and $\tilde{y}_{ij}$ is defined as $\tilde{y}_{ij} = y_i - W_i^\top\alpha - E_i^\top\theta - U_{(j)}^\top\beta_{(j)}$.

- $r_j^{-1}|\text{rest} \sim \text{Inv-Gaussian}(\eta_1, \sqrt{\frac{\eta_1\sigma^2}{\|\beta_j\|_2^2}})$

- $\omega_{jl}^{-1}|\text{rest} \sim \text{Inv-Gaussian}(\eta_2, \sqrt{\frac{\eta_2\sigma^2}{\beta_{jl}^2}})$

- $\eta_1|\text{rest} \sim \text{Gamma}\,(s_{\eta_1}, r_{\eta_1})$, where $s_{\eta_1} = \frac{p}{2}+1$ and the rate parameter $r_{\eta_1} = \frac{\sum_{j=1}^{p} r_j}{2} + d_1$.

- $\eta_2|\text{rest} \sim \text{Gamma}\,(s_{\eta_2}, r_{\eta_2})$, where $s_{\eta_2} = p \times L + 1$ and the rate parameter $r_{\eta_2} = \frac{\sum_{j,l}\omega_j}{2} + d_2$.

- $\sigma^2|\text{rest} \sim \text{Inv-Gamma}(\frac{n+p\times L}{2}, \frac{\sum_{i=1}^{n}\tilde{y_i}^2 + \sum_{j=1}^{p}\beta_j^\top V_j^{-1}\beta_j}{2})$, where $\tilde{y_i} = y_i - W_i^\top\alpha - E_i^\top\theta - U_i^\top\beta$.

G.10 BGL

G.10.1 Hierarchical model specification

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta)^2\right\}$$

$$\alpha \sim \mathrm{N}_q(0,\, \Sigma_{\alpha 0})$$

$$\theta \sim \mathrm{N}_k(0,\, \Sigma_{\theta 0})$$

$$\beta_j|\sigma^2, s_j \overset{ind}{\sim} \mathrm{N}_L\left(0,\, \sigma^2 s_j \mathbf{I}_L\right) \quad j = 1, \ldots, p$$

$$s_j|\eta \overset{ind}{\sim} \text{Gamma}\left(\frac{L+1}{2}, \frac{\eta}{2}\right) \quad j = 1, \ldots, p$$

$$\eta \sim \text{Gamma}\,(d_1,\ d_2)$$

$$\sigma^2 \sim 1/\sigma^2$$

G.10.2 Gibbs Sampler

- $\alpha|\text{rest} \sim \mathrm{N}(\mu_\alpha,\, \Sigma_\alpha)$, where

$$\mu_\alpha = \Sigma_\alpha(\sigma^2)^{-1}\sum_{i=1}^{n} W_i(y_i - E_i^\top \theta - U_i^\top \beta)$$

$$\Sigma_\alpha = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} W_i W_i^\top + \Sigma_{\alpha 0}^{-1}\right)^{-1}$$

- $\theta|\text{rest} \sim \mathrm{N}(\mu_\theta,\, \Sigma_\theta)$, where

$$\mu_\theta = \Sigma_\theta(\sigma^2)^{-1}\sum_{i=1}^{n} E_i(y_i - W_i^\top \alpha - U_i^\top \beta)$$

$$\Sigma_\theta = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} E_i E_i^\top + \Sigma_{\theta 0}^{-1}\right)^{-1}$$

- $\beta_j|\text{rest} \sim \mathrm{N}(\mu_{\beta_j},\, \sigma^2\Sigma_{\beta_j})$ where

$$\mu_{\beta_j} = \Sigma_{\beta_j}\sum_{i=1}^{n} U_{ij}\tilde{y}_{ij}$$

$$\Sigma_{\beta_j} = \left(\sum_{i=1}^{n} U_{ij}U_{ij}^\top + \frac{1}{s_j}\mathbf{I}_L\right)^{-1}$$

  and $\tilde{y}_{ij}$ is defined as $\tilde{y}_{ij} = y_i - W_i^\top \alpha - E_i^\top \theta - U_{(j)}^\top \beta_{(j)}$.

- $s_j^{-1}|\text{rest} \sim \text{Inverse-Gaussian}(\eta, \sqrt{\frac{\eta\sigma^2}{\|\beta_j\|_2^2}})$
- $\eta|\text{rest} \sim \text{Gamma}\,(s_\eta,\, r_\eta)$, where $s_\eta = \frac{p+p\times L}{2} + d_1$ and the rate parameter $r_\eta = \frac{\sum_{j=1}^p s_j}{2} + d_2$.
- $\sigma^2|\text{rest} \sim \text{Inv-Gamma}(\frac{n+p\times L}{2},\ \frac{\sum_{i=1}^n \tilde{y_i}^2 + \sum_{j=1}^p (s_j)^{-1}\beta_j^\top \beta_j}{2})$, where $\tilde{y_i} = y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta$.

G.11 BL

G.11.1 Hierarchical model specification

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta)^2\right\}$$

$$\alpha \sim \text{N}_q(0,\, \Sigma_{\alpha 0})$$

$$\theta \sim \text{N}_k(0,\, \Sigma_{\theta 0})$$

$$\beta_{jl}|\sigma^2, s_{jl} \overset{ind}{\sim} \text{N}\left(0,\, \sigma^2 s_{jl}\right) \quad j = 1, \ldots, p;\ l = 1, \ldots, L$$

$$s_{jl}|\eta \overset{ind}{\sim} \text{Gamma}\left(1, \frac{\eta}{2}\right) \quad j = 1, \ldots, p;\ l = 1, \ldots, L$$

$$\eta \sim \text{Gamma}\,(d_1,\ d_2)$$

$$\sigma^2 \sim 1/\sigma^2$$

G.11.2 Gibbs Sampler

- $\alpha|\text{rest} \sim \text{N}(\mu_\alpha,\, \Sigma_\alpha)$, where

$$\mu_\alpha = \Sigma_\alpha (\sigma^2)^{-1} \sum_{i=1}^n W_i (y_i - E_i^\top \theta - U_i^\top \beta)$$

$$\Sigma_\alpha = \left(\frac{1}{\sigma^2}\sum_{i=1}^n W_i W_i^\top + \Sigma_{\alpha 0}^{-1}\right)^{-1}$$

- $\theta|\text{rest} \sim \text{N}(\mu_\theta,\, \Sigma_\theta)$, where

$$\mu_\theta = \Sigma_\theta (\sigma^2)^{-1} \sum_{i=1}^n E_i (y_i - W_i^\top \alpha - U_i^\top \beta)$$

$$\Sigma_\theta = \left(\frac{1}{\sigma^2}\sum_{i=1}^n E_i E_i^\top + \Sigma_{\theta 0}^{-1}\right)^{-1}$$

- $\beta_{jl}|\text{rest} \sim \text{N}(\mu_{\beta_{jl}}, \sigma^2_{\beta_{jl}})$ where

$$\mu_{\beta_{jl}} = \sigma^2_{\beta_{jl}}(\sigma^2)^{-1}\sum_{i=1}^{n} U_{ijl}\tilde{y}_{ijl}$$

$$\sigma^2_{\beta_{jl}} = \sigma^2\left(\sum_{i=1}^{n} U^2_{ijl} + \frac{1}{s_{jl}}\right)^{-1}$$

  and $\tilde{y}_{ijl}$ is defined as $\tilde{y}_{ijl} = y_i - W_i^\top\alpha - E_i^\top\theta - U_{(jl)}^\top\beta_{(jl)}$.

- $s_j^{-1}|\text{rest} \sim \text{Inverse-Gaussian}(\eta, \sqrt{\frac{\eta\sigma^2}{\beta^2_{jl}}})$

- $\eta|\text{rest} \sim \text{Gamma}\,(s_\eta, r_\eta)$, where $s_\eta = p\times L + d_1$ and the rate parameter $r_\eta = \frac{\sum_{j,l} s_{jl}}{2} + d_2$.

- $\sigma^2|\text{rest} \sim \text{Inv-Gamma}(\frac{n+p\times L}{2}, \frac{\sum_{i=1}^{n} \tilde{y_i}^2 + \sum_{j,l} s_{jl}^{-1}\beta^2_{jl}}{2})$, where $\tilde{y_i} = y_i - W_i^\top\alpha - E_i^\top\theta - U_i^\top\beta$.