Identification of Gene-environment interactions using a marginal robust Bayesian method

Xi Lu[1], Kun Fan[1] and Cen Wu[1,2]

[1] Department of Statistics, Kansas State University, Manhattan, KS

[2] School of Public Health, Yale University, 60 College ST, New Haven, CT 06520

Abstract

In high-throughput cancer studies, an important aim is to identify gene-environment interactions associated with the clinical outcomes. Recently, multiple marginal penalization methods have been developed and shown to be effective in G×E studies. However, within the Bayesian framework, marginal variable selection has not received much attention. In this study, we propose a novel marginal Bayesian variable selection method for G×E studies. In particular, our marginal Bayesian method is robust to data contamination and outliers in the outcome variables. With the incorporation of spike-and-slab priors, the proposed method outperforms a number of alternatives in both identification and prediction in extensive simulation studies. The utility of the marginal robust Bayesian variable selection method has been further demonstrated in the case studies using TCGA data. Some of the identified main and interaction effects from the real data analysis have important biological implications.

Keywords: Gene-environment interaction; marginal analysis; robust Bayesian variable selection; spike-and-slab priors.

1 Introduction

In high-throughput profiling studies, the significance of gene–environment (G×E) interactions in elucidating the genetic basis of complex diseases has been increasingly recognized beyond the main effects. In traditional studies, with Bayesian Lasso method, the shrinkage on the individual level of coefficients can be achieved. Although this method comes with many benefits, the main disadvantage is that they cannot shrink the posterior coefficient estimates to zero exactly. To overcome this difficult, we propose a robust Bayesian approach for variable selections in marginal model. We also incorporate spike-and-slab priors to impose sparsity. The advantages of the proposed method and model performance are evaluated through simulation. In the case study, the proposed Bayesian method is expected to lead to improve prediction and the identification of main and interaction effects with important implications. To facilitate fast computation and reproducible research, we implement the proposed and all alternative methods in C++ for the R package.

2 Data and Model Settings

We use Y to denote a continuous response variable representing the the cancer outcome or disease phenotype. Let $X = (X_1, \ldots, X_p)$ be the p genetic variants, $E = (E_1, \ldots, E_q)$ be the

q environmental factors and $C = (C_1, \ldots, C_m)$ be the m clinical factors. We denote the ith subject with i. Let (Y_i, E_i, C_i, X_i) $(i = 1, \ldots, n)$ be independent and identically distributed random vectors. For the jth gene X_j $(j = 1, \ldots, p)$, define $W_j = (X_j E_1, \ldots, X_j E_q)$, $\eta_j = (\eta_{j1}, \ldots, \eta_{jq})^T$. Denote $\tilde{W} = W_j$, consider the following marginal model:

$$\begin{aligned} Y_i &= \sum_{k=1}^{q} \alpha_k E_{ik} + \sum_{t=1}^{m} \gamma_t C_{it} + \beta_j X_{ij} + \sum_{k=1}^{q} \eta_{jk} X_{ij} E_{ik} + \epsilon_i \\ &= \sum_{k=1}^{q} \alpha_k E_{ik} + \sum_{t=1}^{m} \gamma_t C_{it} + \beta_j X_{ij} + \eta_j \tilde{W}_i + \epsilon_i \end{aligned} \tag{1}$$

where α_k's and γ_t's are the regression coefficients corresponding to effects of environmental and clinical factors, respectively, and β_j's and η_{jk}'s are the regression coefficients of the genetic variants and G×E interactions effects, correspondingly. Denote $\alpha = (\alpha_1, \ldots, \alpha_q)^T$, $\gamma = (\gamma_1, \ldots, \gamma_m)^T$, $\beta = (\beta_1, \ldots, \beta_p)^T$, $\eta = (\eta_1^T, \ldots, \eta_p^T)^T$, $W = (W_1, \ldots, W_p)$. Then model (1) can be written as

$$Y_i = E_i \alpha + C_i \gamma + X_{ij} \beta_j + \tilde{W}_i \eta_j + \epsilon_i. \tag{2}$$

2.1 Bayesian formulation of the LAD regerssion

The least absolute deviation (LAD) regression is well known for its advantages in dealing with heavy-tailed error distributions or outliers in response. In regression analysis, the coefficients can be estimated as the solution to the following minimization problem

$$\min_{\alpha, \gamma, \beta_j, \eta_j} \sum_{i=1}^{n} |Y_i - E_i \alpha - C_i \gamma - X_{ij} \beta_j - \tilde{W}_i \eta_j|.$$

Here, we propose a robust Bayesian variable selection in our marginal model. For a Bayesian formulation of LAD regression, we assume that $\epsilon_i (i = 1, \ldots, n)$ are i.i.d. random variables following the Laplace distribution with density

$$f(\epsilon_i | \tau) = \frac{\tau}{2} \exp(-\tau |\epsilon_i|).$$

The likelihood function can be expressed as

$$f(Y | \alpha, \gamma, \beta_j, \eta_j) = \prod_{i=1}^{n} \frac{\tau}{2} \exp(-\tau |Y_i - E_i \alpha - C_i \gamma - X_{ij} \beta_j - \tilde{W}_i \eta_j|).$$

The Laplace distrubution in Bayesian LAD regression can be treated as a special case of the Laplace distribution in Baysian quantile regression. In Baysian quantile regression, we assume that ϵ_i's are following the skewed Laplace distribution with density

$$f(\epsilon | \tau) = \theta(1 - \theta) \tau \exp(-\tau \rho_\theta(\epsilon)).$$

The random errors can be written as

$$\epsilon_i = \xi_1 v_i + \tau^{-1/2} \xi_2 \sqrt{v_i} z_i,$$

where

$$\xi_1 = \frac{1-2\theta}{\theta(1-\theta)} \quad and \quad \xi_1 = \sqrt{\frac{2}{\theta(1-\theta)}}$$

With $\theta \in (0,1)$, $v_i \sim \exp(\tau^{-1})$, and $z_i \sim \mathrm{N}(0,1)$.

As the Bayesian LAD regression is a special case of Bayesian quantile regression with θ=0.5, so we have $\xi_1 = 0$ and $\xi_2 = \sqrt{8}$. Therefore, the response Y_i can be written as:

$$\begin{aligned} Y_i &= \mu_i + \tau^{-1/2}\xi_2\sqrt{v_i}z_i, \\ v_i|\tau &\overset{iid}{\sim} \tau\exp(-\tau v_i), \\ z_i &\overset{iid}{\sim} \mathrm{N}(0,1). \end{aligned} \tag{3}$$

Where $\mu_i = E_i\alpha + C_i\gamma + X_{ij}\beta_j + \tilde{W}_i\eta_j$.

2.2 Bayesian LAD regression with spike-and-slab priors

In model (1), the coefficients β_j and η_j corresponds to the main and interaction effects with respect to the jth genentic variant, respectively. When $\beta_j = 0$ and $\eta_j = 0$, the genetic variant has no effect on the phenotype. A non-zero β suggests a presence of main effect while a non-zero η suggests an interaction effect. As the traditional method cannot shrink the posterior coefficient estimates to zero exactly, we incorporate spike-and-slab priors to impose sparsity.

For the robust Bayesian marginal model of the jth gene ($j = 1, \ldots, p$), consider the following priors:

$$\begin{aligned} \beta_j|s_1, \pi_1 &\sim (1-\pi_1)\mathrm{N}(0, s_1) + \pi_1\delta_0(\beta_j) \\ s_1|\varphi_1^2 &\sim \frac{\varphi_1^2}{2}\exp(-\frac{\varphi_1^2}{2}s_1) \\ \eta_{jk}|s_{2k}, \pi_2 &\overset{iid}{\sim} (1-\pi_2)\mathrm{N}(0, s_{2k}) + \pi_2\delta_0(\eta_k), (k = 1, \ldots, q) \\ s_{2k}|\varphi_2^2 &\overset{iid}{\sim} \frac{\varphi_2^2}{2}\exp(-\frac{\varphi_2^2}{2}s_{2k}), (k = 1, \ldots, q) \end{aligned} \tag{4}$$

Here, π_1 and π_2 control the sparisity on the main and interaction level, repectively. The prior can be non-informative if π_1 and π_2 are given values with 0.5 as their priors are given the same probability. So we assign $\pi_1 \sim \mathrm{Beta}(r_1, u_1)$ and $\pi_2 \sim \mathrm{Beta}(r_2, u_2)$ with these conjugate beta priors which account for the uncertainty in π_1 and π_2. In this paper, we choose $r_1 = u_1 = r_2 = u_2 = 1$.

We place normal priors on $\alpha_k (k = 1, \ldots, q)$ and $\gamma_t (t = 1, \ldots, m)$ as

$$\begin{aligned} \alpha_k &\overset{iid}{\sim} \frac{1}{\sqrt{(2\pi\alpha_0)}}\exp(-\frac{\alpha_k^2}{2\alpha_0}), (k = 1, \ldots, q) \\ \gamma_t &\overset{iid}{\sim} \frac{1}{\sqrt{(2\pi\gamma_0)}}\exp(-\frac{\gamma_t^2}{2\gamma_0}), (t = 1, \ldots, m) \end{aligned}$$

We also assume Gamma priors on τ, φ_1^2 and φ_2^2 with

$$\begin{aligned}\tau &\sim \text{Gamma}(a, b),\\ \varphi_1^2 &\sim \text{Gamma}(c_1, d_1),\\ \varphi_2^2 &\sim \text{Gamma}(c_2, d_2).\end{aligned}$$

2.3 Computation

For the jth gene, the joint posterior distribution of all the unknown parameters conditional on data can be expressed as

$$\begin{aligned}&\pi(\alpha, \gamma, \beta_j, \eta_j, v, s_1, s_2, \tau, \varphi_1, \varphi_2, \pi_1, \pi_2, z_i|Y)\\ &\quad\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau^{-1}\xi_2^2 v_i}} \exp\Big\{ -\frac{(y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\}\\ &\quad\times \prod_{i=1}^{n} \tau\exp(-\tau v_i)\tau^{a-1}\exp(-b\tau)\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}z_i^2)\\ &\quad\times \prod_{k=1}^{q} \frac{1}{\sqrt{(2\pi\alpha_0)}}\exp(-\frac{\alpha_k^2}{2\alpha_0})\\ &\quad\times \prod_{t=1}^{m} \frac{1}{\sqrt{(2\pi\gamma_0)}}\exp(-\frac{\gamma_t^2}{2\gamma_0})\\ &\quad\times \Big((1-\pi_1)(2\pi s_1)^{-1/2}\exp(-\frac{\beta_j^2}{2s_1})\mathbf{I}_{\{\beta_j\neq 0\}} + \pi_1\delta_0(\beta_j)\Big)\\ &\quad\times \prod_{k=1}^{q} \Big((1-\pi_2)(2\pi s_{2k})^{-1/2}\exp(-\frac{\eta_{jk}^2}{2s_{2k}})\mathbf{I}_{\{\eta_{jk}\neq 0\}} + \pi_2\delta_0(\eta_{jk})\Big)\\ &\quad\times \frac{\varphi_1^2}{2}\exp(-\frac{\varphi_1^2}{2}s_1)\\ &\quad\times \prod_{k=1}^{q} \frac{\varphi_2^2}{2}\exp(-\frac{\varphi_2^2}{2}s_{2k})\\ &\quad\times (\varphi_1^2)^{c_1-1}\exp(-d_1\varphi_1^2)\\ &\quad\times (\varphi_2^2)^{c_2-1}\exp(-d_2\varphi_2^2)\\ &\quad\times \pi_1^{r_1-1}(1-\pi_1)^{u_1-1}\\ &\quad\times \pi_2^{r_2-1}(1-\pi_2)^{u_2-1}\end{aligned}$$

Let $\mu_{(-\alpha_k)} = E(y_i) - E_{ik}\alpha_k, (i = 1, \ldots, n), (k = 1, \ldots, q)$, representing the mean effect without the contribution of $E_{ik}\alpha_k$. The posterior distribution of α_k conditional on all other

parmeters can be expressed as

$$
\begin{aligned}
&\pi(\alpha_k|\text{rest}) \\
&\quad \propto \pi(\alpha_k)\pi(Y|\cdot) \\
&\quad \propto \exp\Big\{ -\sum_{i=1}^{n} \frac{(y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \times \exp(-\frac{\alpha_k^2}{2\alpha_0}) \\
&\quad \propto \exp\Big\{ -\frac{1}{2}\big[(\sum_{i=1}^{n} \frac{\tau E_{ik}^2}{\xi_2^2 v_i} + \frac{1}{\alpha_0})\alpha_k^2 - 2\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\alpha_k)})E_{ik}}{\xi_2^2 v_i}\alpha_k\big]\Big\}
\end{aligned}
$$

Hence, the full conditional distribution of α_k is normal distribution $N(\mu_{\alpha_k}, \sigma_{\alpha_k}^2)$with mean

$$\mu_{\alpha_k} = \big(\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\alpha_k)})E_{ik}}{\xi_2^2 v_i}\big)\sigma_{\alpha_k}^2,$$

and variance

$$\sigma_{\alpha_k}^2 = \big(\sum_{i=1}^{n} \frac{\tau E_{ik}^2}{\xi_2^2 v_i} + \frac{1}{\alpha_0}\big)^{-1}.$$

The posterior distribution of $\gamma_t (t = 1, \ldots, m)$ conditional on all other parameters can be obtained in similiar way. Let $\mu_{(-\gamma_t)} = E(y_i) - C_{it}\gamma_t$, $i = 1, \ldots, n$, then

$$\gamma_t|\text{rest} \sim N(\mu_{\gamma_k}, \sigma_{\gamma_t}^2)$$

where

$$
\begin{aligned}
\mu_{\gamma_t} &= \big(\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\gamma_t)})C_{it}}{\xi_2^2 v_i}\big)\sigma_{\gamma_t}^2, \\
\sigma_{\gamma_t}^2 &= \big(\sum_{i=1}^{n} \frac{\tau C_{it}^2}{\xi_2^2 v_i} + \frac{1}{\gamma_0}\big)^{-1}.
\end{aligned}
$$

Let $\mu_{(-\beta_j)} = E(y_i) - X_{ij}\beta_j$ and $l_1 = \pi(\beta_j = 0|\text{rest})$, the conditional posterior distribution of β_j is a multivariate spike-and-slab distribution:

$$\beta_j|\text{rest} \sim (1 - l_1)N(\mu_{\beta_j}, \sigma_{\beta_j}^2) + l_1\delta_0(\beta_j) \tag{5}$$

where

$$
\begin{aligned}
\mu_{\beta_j} &= \big(\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\beta_j)})X_{ij}}{\xi_2^2 v_i}\big)\sigma_{\beta_j}^2, \\
\sigma_{\beta_j}^2 &= \big(\sum_{i=1}^{n} \frac{\tau X_{ij}^2}{\xi_2^2 v_i} + \frac{1}{s_1}\big)^{-1}.
\end{aligned}
$$

It's easy to show that

$$l_1 = \frac{\pi_1}{\pi_1 + (1 - \pi_1)s_1^{-1/2}(\sigma_{\beta_j}^2)^{1/2}\exp\{\frac{1}{2}(\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\beta_j)})X_{ij}}{\xi_2^2 v_i})^2\sigma_{\beta_j}^2\}}.$$

The posterior distribution of β_j is a mixture of a multivariate normal distribution and a point mass at 0. That is, at each iteractio of MCMC, β_j is drawn from $N(\mu_{\beta_j}, \sigma^2_{\beta_j})$ with probability $(1 - l_1)$ and is set to 0 with probability l_1.

Similiarly, the posterior distribution of $\eta_{jk}(k = 1, \ldots, q)$ is also a spike-and-slab distribution. Denote $\mu_{(-\eta_{jk})} = E(y_i) - W_{ik}\eta_{jk}$ and $l_{2k} = \pi(\eta_{jk} = 0|\text{rest})$, η_{jk} follows this distribution:

$$\eta_{jk}|\text{rest} \sim (1 - l_{2k})N(\mu_{\eta_{jk}}, \sigma^2_{\eta_{jk}}) + l_{2k}\delta_0(\eta_{jk}) \tag{6}$$

where

$$\mu_{\eta_{jk}} = \big(\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\eta_{jk})})\tilde{W}_{ik}}{\xi_2^2 v_i}\big)\sigma^2_{\eta_{jk}},$$
$$\sigma^2_{\beta_j} = \big(\sum_{i=1}^{n} \frac{\tau\tilde{W}^2_{ik}}{\xi_2^2 v_i} + \frac{1}{s_{2k}}\big)^{-1}.$$

And

$$l_{2k} = \frac{\pi_2}{\pi_2 + (1 - \pi_2)s_{2k}^{-1/2}(\sigma^2_{\eta_{jk}})^{1/2}\exp\{\frac{1}{2}(\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\eta_{jk})})\tilde{W}_{ik}}{\xi_2^2 v_i})^2\sigma^2_{\eta_{jk}}\}}. \tag{7}$$

The full conditional posterior distribution of s_1 is:

$$\begin{aligned} &s_1|\text{rest} \\ &\quad \propto \pi(\beta_j|s_1, \pi_1)\pi(s_1|\varphi_1^2) \\ &\quad \propto \Big((1 - \pi_1)(2\pi s_1)^{-1/2}\exp(-\frac{\beta_j^2}{2s_1})\mathbf{I}_{\{\beta_j \neq 0\}} + \pi_1\delta_0(\beta_j)\Big)\exp(-\frac{\varphi_1^2}{2}s_1) \end{aligned} \tag{8}$$

When $\beta_j = 0$, (8) is proportion to $\exp(-\frac{\varphi_1^2}{2}s_1)$. Therefore, the posterior distribution of s_1 is $\exp(\frac{\varphi_1^2}{2})$.

When $\beta_j \neq 0$, (8) is proportion to

$$\begin{aligned} &\frac{1}{\sqrt{s_1}}\exp(-\frac{\varphi_1^2}{2}s_1)\exp(-\frac{\beta_j^2}{2s_1}) \\ &\qquad \propto \frac{1}{\sqrt{s_1}}\exp\Big\{-\frac{1}{2}[\varphi_1^2 s_1 + \frac{\beta_j^2}{s_1}]\Big\} \end{aligned}$$

Therefore, when $\beta_j \neq 0$, the posterior distribution for s_1^{-1} is Inverse-Gaussian$(\sqrt{\frac{\varphi_1^2}{\beta_j^2}}, \varphi_1^2)$.

Similiarly, for $s_{2k}(k = 1, \ldots, q)$, when $\eta_{jk} = 0$, the posterior distribution of s_{2k} is $\exp(\frac{\varphi_2^2}{2})$. When $\eta_{jk} \neq 0$, the posterior distribution for s_{2k}^{-1} is Inverse-Gaussian$(\sqrt{\frac{\varphi_2^2}{\eta_{jk}^2}}, \varphi_2^2)$.

The full conditional posterior distribution of φ_1^2:

$$\begin{aligned} &\varphi_1^2|\text{rest} \\ &\quad \propto \pi(s_1|\varphi_1^2)\pi(\varphi_1^2) \\ &\quad \propto \frac{\varphi_1^2}{2}\exp(-\frac{\varphi_1^2 s_1}{2})(\varphi_1^2)^{c_1 - 1}\exp(-d_1\varphi_1^2) \\ &\quad \propto (\varphi_1^2)^{c_1}\exp\Big(-\varphi_1^2(s_1/2 + d_1)\Big) \end{aligned}$$

Therefore, the posterior distribution for φ_1^2 is Gamma$(c_1+1, s_1/2+d_1)$. Similiarly, the posterior distribution for φ_2^2 is Gamma$(c_2+q, \sum_{k=1}^{q} s_{2k}/2+d_2)$.

The full conditional posterior distribution of π_1:

$$\begin{aligned}
&\pi_1|\text{rest} \\
&\quad\propto \pi(s_1|\varphi_1^2)\pi(\varphi_1^2) \\
&\quad\propto \pi_1^{r_1-1}(1-\pi_1)^{u_1-1} \\
&\quad\times \Big((1-\pi_1)(2\pi s_1)^{-1/2}\exp(-\frac{\beta_j^2}{2s_1})\mathbf{I}_{\{\beta_j\neq 0\}}+\pi_1\delta_0(\beta_j)\Big)
\end{aligned}$$

Then, the posterior distribution for π_1 is Beta $(1+r_1-\mathbf{I}(\beta_j\neq 0), u_1+\mathbf{I}(\beta_j\neq 0))$.

The full conditional posterior distribution of π_2:

$$\begin{aligned}
&\pi_2|\text{rest} \\
&\quad\propto \pi(s_2|\varphi_2^2)\pi(\varphi_2^2) \\
&\quad\propto \pi_2^{r_2-1}(1-\pi_2)^{u_2-1} \\
&\quad\times \prod_{k=1}^{q}\Big((1-\pi_2)(2\pi s_{2k})^{-1/2}\exp(-\frac{\eta_{jk}^2}{2s_{2k}})\mathbf{I}_{\{\eta_{jk}\neq 0\}}+\pi_2\delta_0(\eta_{jk})\Big)
\end{aligned}$$

So, the posterior distribution for π_2 is Beta $(1+r_1-\sum_{k=1}^{q}\mathbf{I}(\eta_{jk}\neq 0), u_1+\sum_{k=1}^{q}\mathbf{I}(\eta_{jk}\neq 0))$.

The full conditional posterior distribution of τ:

$$\begin{aligned}
&\tau|\text{rest} \\
&\quad\propto \pi(v|\tau)\pi(\tau)\pi(Y|\cdot) \\
&\quad\propto \tau^{n/2}\exp\Big\{-\sum_{i=1}^{n}\frac{(y_i-E_i\alpha-C_i\gamma-X_{ij}\beta_j-\tilde{W}_i\eta_j)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\} \\
&\quad\times \tau^n\exp(-\tau\sum_{i=1}^{n}v_i)\tau^{a-1}\exp(-b\tau) \\
&\quad\propto \tau^{a+\frac{3}{2}n-1}\exp\Big\{-\tau\big[\sum_{i=1}^{n}(\frac{(y_i-E_i\alpha-C_i\gamma-X_{ij}\beta_j-\tilde{W}_i\eta_j)^2}{2\xi_2^2 v_i}+v_i)+b\big]\Big\}
\end{aligned}$$

Therefore, the posterior distribution for τ is Gamma$(a+\frac{3}{2}n, \big[\sum_{i=1}^{n}(\frac{(y_i-E_i\alpha-C_i\gamma-X_{ij}\beta_j-\tilde{W}_i\eta_j)^2}{2\xi_2^2 v_i}+v_i)+b\big])$.

Last, we have The full conditional posterior distribution of v_i:

$$\begin{aligned}
&v_i|\text{rest} \\
&\quad\propto \pi(v|\tau)\pi(Y|\cdot) \\
&\quad\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{(y_i-E_i\alpha-C_i\gamma-X_{ij}\beta_j-\tilde{W}_i\eta_j)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\}\times\exp(-\tau v_i) \\
&\quad\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{1}{2}\big[(2\tau)v_i+\frac{\tau(y_i-E_i\alpha-C_i\gamma-X_{ij}\beta_j-\tilde{W}_i\eta_j)^2}{\xi_2^2 v_i}\big]\Big\}
\end{aligned}$$

It is easy to show that

$$\frac{1}{v_i}|\text{rest} \sim \text{Inverse-Gaussian}(\sqrt{\frac{2\xi_2^2}{(y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2}},\ 2\tau).$$

3 Simulation

To demonstrate the utility of the proposed approach, we evaluate the performance through simulation study. In particular, we compare the performance of the proposed method, robust Bayesian Lasso spike-and-slab variable selection (denoted as LADBLSS) with three alternatives, robust Bayesian Lasso variable selection (denoted as LADBL), Bayesian Lasso with spike-and-slab variable selection (denoted as BLSS) and Bayesian Lasso variable selection (denoted as BL). LADBL is similar to the proposed method, except that it does not adopt the spike-and-slab prior. Comparison of LADBLSS with BLSS demonstrate the importance of accommodating outliers.

Under all settings, the sample size is set as $n = 200$, and the number of G factors is $p = 500$ with $q = 4$, $m = 3$. For environmental factors, we simulate four continuous variables from multivariate normal distributed factors with marginal mean 0, marginal variance 1 and AR (auto-regressive) correlation structure with $\rho = 0.5$. In addition, three clinical factors are generated from a multivariate normal distribution with margianl mean 0 and marginal variance 1 and AR structure with $\rho = 0.5$. Among the p main G effects and $p \times q$ $G \times E$ interactions, 8 and 12 are set as associated with the response, respectively. All environmental factors and clinical factors have important effects. The nonzero coefficients of important effects are randomly generateed from a uniform distribution $Unif[0.1, 0.5]$. The random error are generated from: (1) $N(0, 1)$(Error 1), (2) t-distribution with 2 degress of freedom ($t(2)$)(Error2), (3) LogNormal(0,2)(Error3), (4) $90\%N(0, 1)$+10%Cauchy(0,1)(Error4), (5) $80\%N(0, 1)$+20%Cauchy(0,1)(Error5). All of them are heavy-tailed distribution except the first one.

The genetic factors are simulated in the following four settings.

Setting 1

In simulating continuous genetic variants, we generate multivariate normal distributions with marginal mean 0 and variance 1. The AR structure is considered in computing the correlation of G factors, under which gene j and k have correlation $\rho^{|j-k|}$ with $\rho = 0.5$.

Setting 2

We assess the performance under single-nucleotide polymorphism (SNP) data. The SNPs are obtained by dichotomizing the gene expression values at the 1st and 3rd quartiles, with the 3–level (2,1,0) for genotypes (aa,Aa,AA) respectively. Here, the gene expressions are generated from Example 1.

Setting 3

Consider simulating the SNP data under a pairwise linkage disequilibrium (LD) structure. For the two minor alleles A and B of two adjacent SNPs, let q_1 and q_2 be the minor allele frequencies (MAFs), respectively. The frequencies of four haplotypes are as $p_{AB} = q_1q_2 + \delta$, $p_{ab} = (1 - q_1)(1 - q_2) + \delta$, $p_{Ab} = q_1(1 - q_2) - \delta$, and $p_{aB} = (1 - q_1)q_2 - \delta$, 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 $(q_1^2, 2q_1(1-q_1), (1-q_1)^2)$. Based on the conditional genotype probability matrix, we can simulate the genotypes for locus 2. With MAFs 0.3 and pairwise correlation $r = 0.6$, we have $\delta = r\sqrt{q_1(1-q_1)q_2(1-q_2)}$.

Setting 4

A more practical correlation structure is adopted in this example. We extract the first 500 mRNA gene expressions from SKCM data analyzed in case study, with correlation based on the real data. For each simulation replicate, we randomly sample 200 subjects from the dataset. The same coefficients and error distributions from the first 3 examples are adopted.

Posterior samples are collected from a Gibbs Sampler with 10,000 interations in which the first 5,000 are burn-ins. The posterior medians are used to estimate the coefficients. For approaches incorporating spike-and-slab priors, we consider the median probability model (MPM) to identify significant predictors. Here we use ϕ as the indicator identifying the probability of the posterior distribution of predictor generating from non-spike distribution. Suppose we collect G posterior samples from MCMC after burn-ins. The jth predictor is included in the regression model at gth MCMC iteraction if the indicator of this predictor at this step is 1, that is if $\phi_j^{(g)} = 1$. Then the probability of the posterior distribution including the jth predictor in the final model is defined as the average of all the indicators for the jth predictor among the G posterior samples. That is

$$p_j = \hat{\pi}(\phi_j = 1|y) = \frac{1}{G}\sum_{g=1}^{G} \phi_j^{(g)},\ j = 1, \ldots, p$$

A higher posterior inclusion probability p_j indicates a stronger empirical evidence that the jth predictor has a non-zero coefficient, which also indicates that this predictor has a stronger association with the response variable. Usually, the MPM is defined with the predictors which have posterior inclusion probabiliry no less than $\frac{1}{2}$. For approaches without spike-and-slab priors, the method of credible interval is adopted. A predictor X_j(or $\tilde{W}_k$) is excluded if the credible interval of β_j(or η_{jk}) covers 0 and is retained otherwise.

To compare the performance of the four approaches, we consider a sequence of cutting-off probabilities in MPM for methods with spike-and-slab priors and different credible intervals for methods without spike-and-slab priors, compute true-positive-rate (TPR) and false-positive-rate (FPR) values, and use the area under curve (AUC) under the receiver operating characteristic (ROC) framework to compare the identification accuracy. In addition, we also consider Top100, which is defined as the number of true signals when 100 important main effects (or interactions) are identified.

Table 1: Simulation results of the first setting. AUC (mean of AUC), SD (sd of AUC) based on 100 replicates. n=200, p=500, q=4 and $m = 3$.

		BL	BLSS	LADBL	LADBLSS
Error 1	AUC	0.9182	0.9901	0.9258	0.9887
	SD	0.0052	0.0021	0.0076	0.0026
Error 2	AUC	0.8332	0.9420	0.9004	0.9841
	SD	0.0107	0.0235	0.0078	0.0031
Error 3	AUC	0.5343	0.5473	0.8432	0.9558
	SD	0.0144	0.0576	0.0115	0.0161
Error 4	AUC	0.8221	0.9124	0.9222	0.9895
	SD	0.0212	0.0410	0.0071	0.0024
Error 5	AUC	0.7507	0.8431	0.9192	0.9904
	SD	0.0217	0.0633	0.0059	0.0018

Table 2: Identification results of the first setting with Top100 method. mean(sd) based on 100 replicates. n=200, p=500, q=4 and $m = 3$.

		Main	Interaction	Total
Error 1	BL	7.6(0.49)	6.8(1.6)	14.4(1.73)
N(0,1)	BLSS	7.8(0.41)	10.8(0.92)	18.6(1.13)
	LADBL	7.67(0.55)	6.53(1.85)	14.2(1.81)
	LADBLSS	7.76(0.5)	10.53(1.36)	18.3(1.49)
Error 2	BL	6.37(1.90)	3.9(2.07)	10.27(3.19)
$t(2)$	BLSS	6.33(1.63)	8.53(2.46)	14.87(3.71)
	LADBL	7.43(0.94)	5.8(1.71)	13.23(2.01)
	LADBLSS	7.53(0.51)	9.9(1.56)	17.43(1.76)
Error 3	BL	0.9(1.21)	0.5(0.97)	1.4(1.45)
Lognormal(0,2)	BLSS	0.73(0.94)	0.47(0.68)	1.2(1.35)
	LADBL	6.27(1.55)	3.67(1.94)	9.93(2.75)
	LADBLSS	6.1(1.37)	8.93(2.02)	15.03(3.09)
Error 4	BL	5.57(2.99)	3.63(2.53)	9.2(5.05)
90%N(0,1)	BLSS	6.2(2.62)	8.3(3.98)	14.5(6.39)
+10%Cauchy(0,1)	LADBL	7.77(0.43)	7.00(1.93)	14.77(1.81)
	LADBLSS	7.77(0.57)	10.67(1.50)	18.23(1.67)
Error 5	BL	5.07(2.89)	3(2.49)	8.07(5.01)
80%N(0,1)	BLSS	4.6(3.25)	5.7(4.23)	10.3(7.27)
+20%Cauchy(0,1)	LADBL	7.57(0.57)	6.83(1.07)	14.4(1.83)
	LADBLSS	7.8(0.55)	10.53(1.36)	18.33(1.69)

Simulation results for the gene expression data in setting one are tabulated in Tables 1 and 2. We can observe that the proposed method has the best performance among all approaches, especially when dealing with heavy-tailed distributions. First, the performance of methods with spike-and-slab priors is consistently better than methods without spike-and-slab priors. For example in table 1, under error 3, the AUC of LADBLSS is 0.9558(SD 0.0161), which is much larger than that of the robust method without spike-and-slab priors, which is 0.8432(SD 0.0115) from LADBL. Similar advantageous performance can also be observed from the identification results with Top100. In table 2, under error 5, LADBLSS identifies 7.78(SD 0.55) out of the 8 main effects and 10.53(SD 1.36) out of the 12 interaction effects. This is higher than the results of LADBL with 7.57(SD 0.57) of main effects and 6.83(SD 1.07) of interaction effects. Second, among all the methods with spike-and-slab priors, the robust Bayesian method with spike-and-slab priors has the best performance in both AUC and identification results. Under error 3, in Table 1, the AUC of LADBLSS is 0.9558(SD 0.0161) while the AUC of BLSS is 0.5473(SD 0.0576). Under error 4 in table 2, LADBLSS identifies 7.77(SD 0.57) main effects and 10.67(SD 1.50) interaction effects while BLSS identifies 6.2(SD 2.62) main effects and 8.3(SD 3.98) interaction effects.

Similar patterns can be observed in Table 5, 6, 7, 8, 9, 10 for setting 2, 3 and 4, respectively, in Appendix. Overally, the advantages of conducting robust Bayesian selection with spike-and-slab priors can be established besed on the results of comprehensive simulation studies.

4 Real Data Analysis

In this study, we consider skin cutaneous melanoma (SKCM) from the Cancer Genome Atlas (TCGA), which is organized by the National Cancer Institute (NCI) with high quality genetic, clinical and proteomic data. We use the level-3 gene expression data of SKCM from the cBio Cancer Genomics Portal. Messenger RNA (mRNA) gene expressions are used as G factor. For E factor, we consider Age, AJCC pathologic tumor stage, gender and Clark level. The response variable is the log-transformed Breslow's thickness. Data are available on 298 subjects and 18,934 gene expressions among which 10,000 genes with the strongest association with the response variables are selected for $G \times E$ interaction analysis. We are trying to identify important gene expressions that have significant main effect or $G \times E$ interaction effects on the Breslow's thinkness.

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

SKCM	Main				Interaction			
	BL	BLSS	LADBL	LADBLSS	BL	BLSS	LADBL	LADBLSS
BL	8	0	4	2	92	43	17	10
BLSS		2	0	0		98	17	20
LADBL			71	4			29	7
LADBLSS				35				65

Table 4: Analysis of SKCM with proposed method: identified main and interaction effects

		Interactions			
Gene	Main Effects	Clark level	AJCC stage	Age	Gender
EBF2	-0.0061				
KCTD16	-0.0345				
GH2	-0.635		-0.8708		
AVP		0.1131			
LOC105379569		0.1695			-0.3004
MTE		0.8074		-0.0839	-0.6608
UBQLN3			-0.3885		1.341
TDRG1				0.0617	
ACTL9		0.7542			
OR2T12		0.4096			
OR51A7		0.9043			
OR8I2	-0.6614	0.2902			
KCNK10				0.0096	
CRYAA		0.017			
H2A		0.0097			
OR4K14		1.6838	-0.4658	-0.4519	0.7227
HND		0.1046			
ODF1		0.6037			
DKKL1		0.0054			
CCDC166		0.1788	-0.2516		
PTPN4	-0.0025				
CRYBB3		0.1714			
UCN		0.0124			

References

[1] Wu, C. and Ma, S. (2014) A selective review of robust variable selection with applications in bioinformatics. *Brief. Bioinform* 1–11. doi: 10.1093/bib/bbu046

[2] Huang J., Ma S. and Xie H. (2007). Least absolute deviations estimation for the accelerated failure time model. *Statistica Sinica*, **17**: 1533–1548.

[3] Wu et. al. (2015) A robust network–constrained penalization approach for integrative analysis with applications in TCGA data. (Submitted)

A Additional simulation results

Table 5: Simulation results of the second setting. AUC (mean of AUC), SD (sd of AUC) based on 100 replicates. n=200, p=500, q=4 and $m = 3$.

		BL	BLSS	LADBL	LADBLSS
Error 1	AUC	0.9089	0.9881	0.9148	0.9888
	SD	0.0059	0.0019	0.0051	0.0037
Error 2	AUC	0.8187	0.9255	0.8877	0.9769
	SD	0.0142	0.0524	0.0057	0.0048
Error 3	AUC	0.5333	0.5533	0.8239	0.9459
	SD	0.0096	0.0656	0.1045	0.0162
Error 4	AUC	0.8113	0.9122	0.9111	0.9849
	SD	0.0166	0.0502	0.0083	0.0033
Error 5	AUC	0.7425	0.8086	0.9076	0.9856
	SD	0.0241	0.0746	0.0065	0.0024

Table 6: Identification results of the second setting with Top100 method. mean(sd) based on 100 replicates. n=200, p=500, q=4 and $m = 3$.

		Main	Interaction	Total
Error 1	BL	7.5(0.86)	6.7(1.49)	14.2(1.83)
N(0,1)	BLSS	7.6(0.67)	10.2(0.09)	17.8(1.32)
	LADBL	7.67(0.66)	6.83(1.82)	14.5(1.96)
	LADBLSS	7.63(0.56)	9.97(1.54)	17.6(1.67)
Error 2	BL	5.83(2.21)	3.47(1.57)	9.3(2.98)
$t(2)$	BLSS	6.33(2.09)	7.57(3.15)	13.9(4.73)
	LADBL	7.07(0.94)	5.97(1.61)	13.03(1.96)
	LADBLSS	7.4(0.62)	9.2(1.94)	16.6(2.11)
Error 3	BL	0.77(0.86)	0.73(0.94)	1.5(1.11)
Lognormal(0,2)	BLSS	0.57(1.01)	0.67(1.06)	1.23(1.77)
	LADBL	5.9(1.65)	3.5(1.96)	9.4(2.43)
	LADBLSS	5.67(1.73)	9(2.35)	14.67(3.73)
Error 4	BL	6.03(2.19)	4.4(2.44)	10.43(4.17)
90%N(0,1)	BLSS	6.03(2.57)	8(3.33)	14.03(5.76)
+10%Cauchy(0,1)	LADBL	7.27(0.91)	6.87(1.48)	14.13(1.74)
	LADBLSS	7.53(0.63)	10(1.43)	17.53(1.57)
Error 5	BL	5.53(2.45)	3.63(2.19)	9.16(4.13)
80%N(0,1)	BLSS	5.07(2.57)	6.73(3.37)	11.8(5.65)
+20%Cauchy(0,1)	LADBL	7.47(0.97)	5.43(1.77)	12.9(2.04)
	LADBLSS	7.37(0.85)	10.47(1.46)	17.83(1.91)

Table 7: Simulation results of the third setting. AUC (mean of AUC), SD (sd of AUC) based on 100 replicates. n=200, p=500, q=4 and $m = 3$.

		BL	BLSS	LADBL	LADBLSS
Error 1	AUC	0.9158	0.9895	0.9251	0.9878
	SD	0.0041	0.0022	0.0054	0.0028
Error 2	AUC	0.8323	0.9461	0.8972	0.9833
	SD	0.0117	0.0342	0.0062	0.0028
Error 3	AUC	0.5268	0.5531	0.8415	0.9595
	SD	0.0127	0.0590	0.0107	0.0156
Error 4	AUC	0.8261	0.9323	0.9245	0.9889
	SD	0.0191	0.0352	0.0056	0.0034
Error 5	AUC	0.7533	0.8591	0.9204	0.9862
	SD	0.0201	0.0657	0.0067	0.0114

Table 8: Identification results of the third setting with Top100 method. mean(sd) based on 100 replicates. n=200, p=500, q=4 and $m = 3$.

		Main	Interaction	Total
Error 1	BL	7.7(0.47)	6.8(1.63)	14.5(1.79)
N(0,1)	BLSS	7.63(0.72)	10.93(0.98)	18.57(1.22)
	LADBL	7.7(0.75)	7.33(1.95)	15.03(2.14)
	LADBLSS	7.87(0.35)	10.33(1.35)	18.2(1.45)
Error 2	BL	6.57(1.87)	4.47(1.69)	11.03(2.88)
$t(2)$	BLSS	6.6(1.57)	8.4(2.51)	15(3.68)
	LADBL	7.57(0.62)	5.77(1.50)	13.33(1.77)
	LADBLSS	7.43(0.68)	9.3(2.15)	16.73(2.43)
Error 3	BL	0.5(0.73)	0.83(1.02)	1.33(1.47)
Lognormal(0,2)	BLSS	0.7(0.99)	0.4(0.86)	1.1(1.54)
	LADBL	6.13(2.05)	3.8(1.39)	9.93(1.32)
	LADBLSS	6.63(1.16)	10.1(1.73)	16.73(2.52)
Error 4	BL	5.73(2.82)	4.3(2.64)	10.03(5.11)
90%N(0,1)	BLSS	5.73(3.02)	7.67(4.19)	13.4(7.05)
+10%Cauchy(0,1)	LADBL	7.8(0.48)	6.87(1.61)	14.67(1.54)
	LADBLSS	7.83(0.38)	10.5(1.25)	18.33(1.39)
Error 5	BL	5.6(2.61)	2.93(2.23)	8.53(4.27)
80%N(0,1)	BLSS	5.27(2.27)	6.9(3.64)	12.17(5.66)
+20%Cauchy(0,1)	LADBL	7.87(0.35)	6.87(1.45)	14.73(1.46)
	LADBLSS	7.7(0.53)	10.7(1.12)	18.4(1.28)

Table 9: Simulation results of the fourth setting. AUC (mean of AUC), SD (sd of AUC) based on 100 replicates. n=200, p=500, q=4 and $m = 3$.

		BL	BLSS	LADBL	LADBLSS
Error 1	AUC	0.6164	0.8992	0.6577	0.9429
	SD	0.0092	0.0128	0.0088	0.0078
Error 2	AUC	0.5516	0.8672	0.6263	0.9384
	SD	0.0127	0.0235	0.0108	0.0074
Error 3	AUC	0.4052	0.6074	0.6055	0.9201
	SD	0.0155	0.0608	0.0159	0.0132
Error 4	AUC	0.5574	0.8672	0.6582	0.9456
	SD	0.0135	0.0200	0.0112	0.0070
Error 5	AUC	0.5197	0.8183	0.6506	0.9472
	SD	0.0123	0.0524	0.0156	0.0083

Table 10: Identification results of the fourth setting with Top100 method. mean(sd) based on 100 replicates. n=200, p=500, q=4 and $m = 3$.

		Main	Interaction	Total
Error 1	BL	3.3(0.88)	1.63(1.19)	4.93(1.34)
N(0,1)	BLSS	3.07(1.23)	3.53(1.89)	6.6(2.55)
	LADBL	3.5(1.28)	2.27(1.17)	5.77(1.94)
	LADBLSS	4.7(1.37)	5.9(1.84)	10.6(2.31)
Error 2	BL	2.1(1.21)	1(1.05)	3.1(1.60)
$t(2)$	BLSS	3.1(1.94)	3.17(2.59)	6.27(4.05)
	LADBL	2.93(1.11)	1.67(1.18)	4.6(1.79)
	LADBLSS	4.5(1.33)	6.13(1.91)	10.63(2.75)
Error 3	BL	0.27(0.58)	0.33(0.61)	0.6(0.93)
Lognormal(0,2)	BLSS	0.6(0.97)	0.7(0.88)	1.3(1.58)
	LADBL	2.97(1.25)	1(0.74)	3.97(1.33)
	LADBLSS	3.7(1.56)	4.9(2.25)	8.6(3.16)
Error 4	BL	1.97(1.63)	1.43(1.30)	3.4(2.59)
90%N(0,1)	BLSS	3.53(1.63)	3.43(2.28)	6.97(3.44)
+10%Cauchy(0,1)	LADBL	3.5(1.17)	2.6(1.30)	6.1(1.81)
	LADBLSS	4.7(1.32)	6.03(2.03)	10.73(2.64)
Error 5	BL	2.03(1.47)	1.13(1.28)	3.17(2.34)
80%N(0,1)	BLSS	2.9(1.81)	2.6(2.09)	5.5(3.54)
+20%Cauchy(0,1)	LADBL	3.43(1.41)	2.23(1.07)	5.67(1.47)
	LADBLSS	5.53(1.50)	7(2.07)	12.53(3.09)

A.1 The ROC curves in simulation

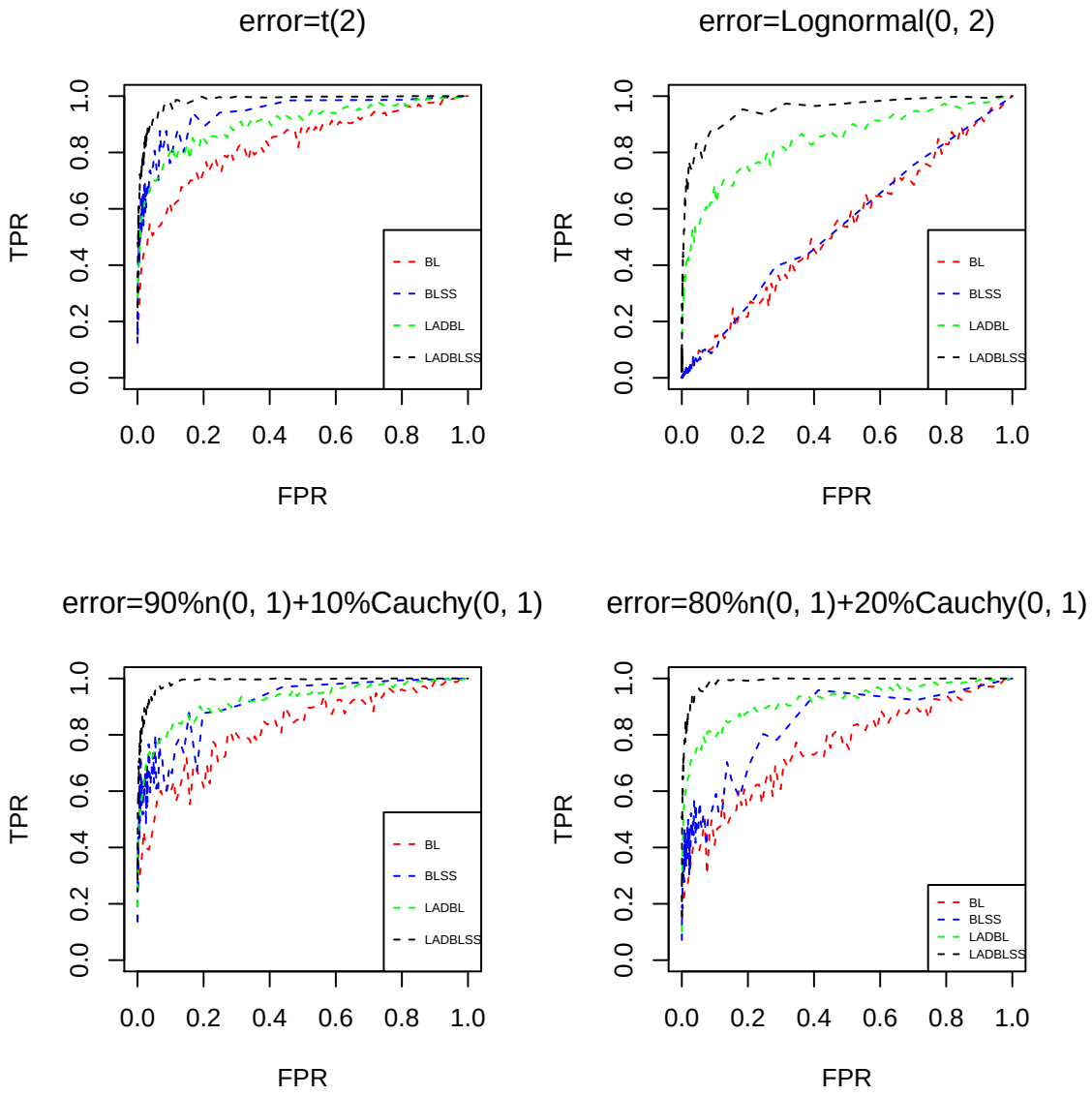

Figure 1: ROC curves of the first setting in simulation

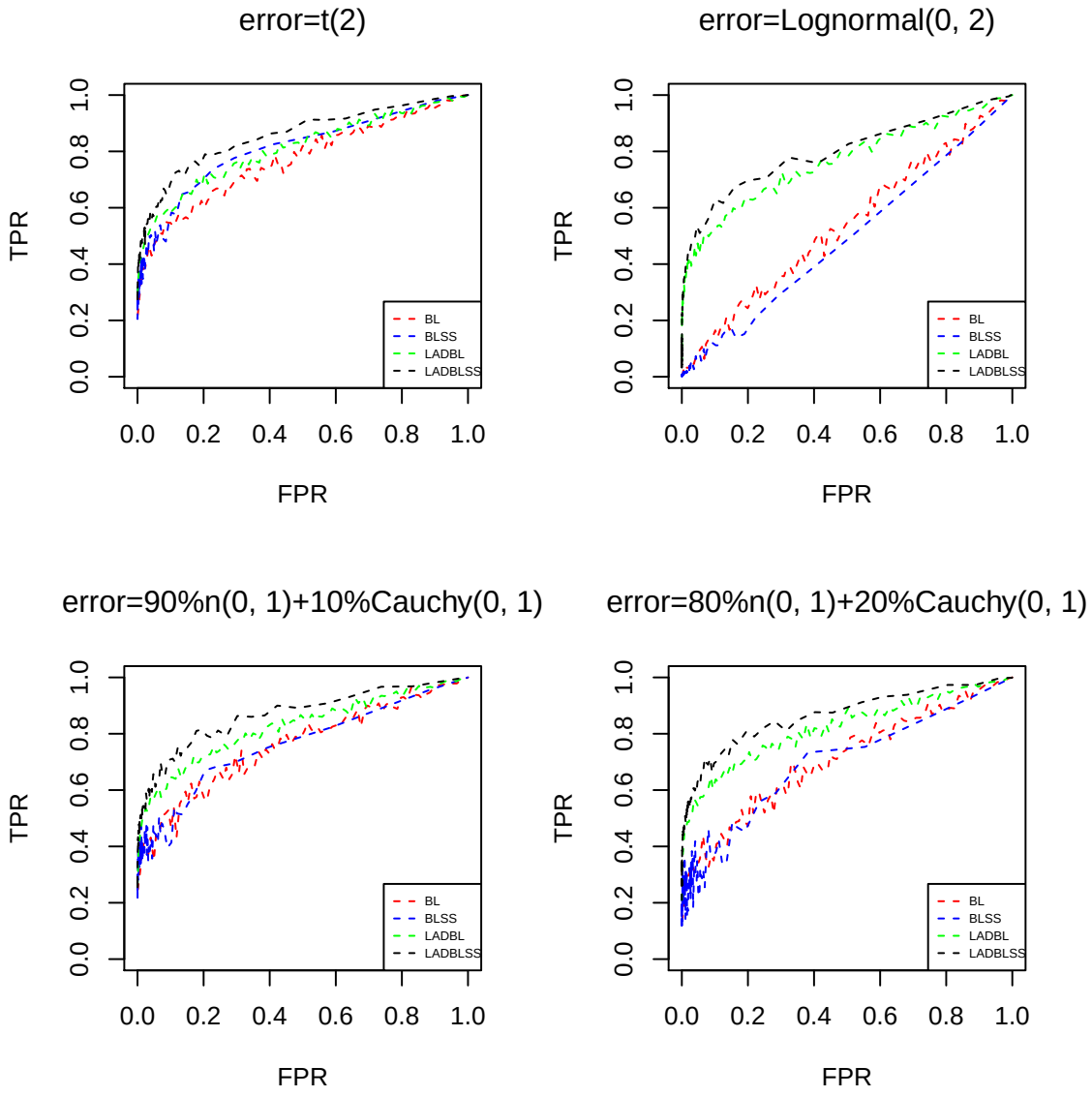

Figure 2: ROC curves of the second setting in simulation

B Posterior inference

B.1 LADBL

B.1.1 Hierarchical model specification

$$
\begin{gathered}
Y_i = E_i\alpha + C_i\gamma + X_{ij}\beta_j + \tilde{W}_i\eta_j + \tau^{-1/2}\xi_2\sqrt{v_i}z_i \quad i = 1,\ldots,n \\
v_i|\tau \overset{iid}{\sim} \tau\exp(-\tau v_i) \quad i = 1,\ldots,n \\
z_i \overset{iid}{\sim} \mathrm{N}(0,1) \quad i = 1,\ldots,n \\
\beta_j|s_1 \sim \frac{1}{\sqrt{2\pi s_1}}\exp(-\frac{\beta_j^2}{2s_1}) \\
s_1|\varphi_1^2 \sim \frac{\varphi_1^2}{2}\exp(-\frac{\varphi_1^2}{2}s_1) \\
\eta_{jk}|s_{2k} \overset{iid}{\sim} \frac{1}{\sqrt{2\pi s_{2k}}}\exp(-\frac{\eta_{jk}^2}{2s_{2k}}) \quad k = 1,\ldots,q \\
s_{2k}|\varphi_2^2 \overset{iid}{\sim} \frac{\varphi_2^2}{2}\exp(-\frac{\varphi_2^2}{2}s_{2k}) \quad k = 1,\ldots,q \\
\alpha_k \overset{iid}{\sim} \frac{1}{\sqrt{(2\pi\alpha_0)}}\exp(-\frac{\alpha_k^2}{2\alpha_0}) \quad k = 1,\ldots,q \\
\gamma_t \overset{iid}{\sim} \frac{1}{\sqrt{(2\pi\gamma_0)}}\exp(-\frac{\gamma_t^2}{2\gamma_0}) \quad t = 1,\ldots,m \\
\tau \sim \mathrm{Gamma}(a,b) \\
\varphi_1^2 \sim \mathrm{Gamma}(c_1,d_1) \\
\varphi_2^2 \sim \mathrm{Gamma}(c_2,d_2)
\end{gathered}
$$

B.1.2 Gibbs Sampler

Let $\mu_{(-\alpha_k)} = E(y_i) - E_{ik}\alpha_k$, then

$$
\begin{aligned}
&\pi(\alpha_k|\text{rest}) \\
&\quad \propto \pi(Y|\cdot)\pi(\alpha_k) \\
&\quad \propto \exp\Big\{-\sum_{i=1}^{n}\frac{(y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\} \times \exp(-\frac{\alpha_k^2}{2\alpha_0}) \\
&\quad \propto \exp\Big\{-\frac{1}{2}\big[(\sum_{i=1}^{n}\frac{\tau E_{ik}^2}{\xi_2^2 v_i} + \frac{1}{\alpha_0})\alpha_k^2 - 2\sum_{i=1}^{n}\frac{\tau(y_i - \mu_{(-\alpha_k)})E_{ik}}{\xi_2^2 v_i}\alpha_k\big]\Big\}
\end{aligned}
$$

Hence, $\alpha_k|\text{rest}\sim N(\mu_{\alpha_k}, \sigma^2_{\alpha_k})$ where

$$\mu_{\alpha_k} = \big(\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\alpha_k)})E_{ik}}{\xi_2^2 v_i}\big)\sigma^2_{\alpha_k},$$
$$\sigma^2_{\alpha_k} = \big(\sum_{i=1}^{n} \frac{\tau E_{ik}^2}{\xi_2^2 v_i} + \frac{1}{\alpha_0}\big)^{-1}.$$

Let $\mu_{(-\gamma_t)} = E(y_i) - C_{it}\gamma_t$, So $\gamma_t|\text{rest}\sim N(\mu_{\gamma_k}, \sigma^2_{\gamma_t})$ where

$$\mu_{\gamma_t} = \big(\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\gamma_t)})C_{it}}{\xi_2^2 v_i}\big)\sigma^2_{\gamma_t},$$
$$\sigma^2_{\gamma_t} = \big(\sum_{i=1}^{n} \frac{\tau C_{it}^2}{\xi_2^2 v_i} + \frac{1}{\gamma_0}\big)^{-1}.$$

Let $\mu_{(-\beta_j)} = E(y_i) - X_{ij}\beta_j$

$$\begin{aligned}
&\pi(\beta_j|\text{rest}) \\
&\quad\propto \pi(y|\cdot)\pi(\beta_j|s_1) \\
&\quad\propto \exp\Big\{-\sum_{i=1}^{n} \frac{(y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\} \times \exp(-\frac{\beta_j^2}{2s_1}) \\
&\quad\propto \exp\Big\{-\frac{1}{2}\big[(\sum_{i=1}^{n} \frac{\tau X_{ij}^2}{\xi_2^2 v_i} + \frac{1}{s_1})\beta_j^2 - 2\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\beta_j)})X_{ij}}{\xi_2^2 v_i}\beta_j\big]\Big\}
\end{aligned}$$

So, $\beta_j|\text{rest}\sim N(\mu_{\beta_j}, \sigma^2_{\beta_j})$ with

$$\mu_{\beta_j} = \big(\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\beta_j)})X_{ij}}{\xi_2^2 v_i}\big)\sigma^2_{\beta_j},$$
$$\sigma^2_{\beta_j} = \big(\sum_{i=1}^{n} \frac{\tau X_{ij}^2}{\xi_2^2 v_i} + \frac{1}{s_1}\big)^{-1}.$$

Let $\mu_{(-\eta_{jk})} = E(y_i) - W_{ik}\eta_{jk}$, then $\eta_{jk}|\text{rest}\sim N(\mu_{\eta_{jk}}, \sigma^2_{\eta_{jk}})$, where

$$\mu_{\eta_{jk}} = \big(\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\eta_{jk})})\tilde{W}_{ik}}{\xi_2^2 v_i}\big)\sigma^2_{\eta_{jk}},$$
$$\sigma^2_{\beta_j} = \big(\sum_{i=1}^{n} \frac{\tau \tilde{W}_{ik}^2}{\xi_2^2 v_i} + \frac{1}{s_{2k}}\big)^{-1}.$$

The full conditional posterior distribution of s_1 is:

$$\begin{aligned} & s_1|\text{rest} \\ & \quad \propto \pi(\beta_j|s_1)\pi(s_1|\varphi_1^2) \\ & \quad \propto \frac{1}{\sqrt{s_1}}\exp(-\frac{\varphi_1^2}{2}s_1)\exp(-\frac{\beta_j^2}{2s_1}) \\ & \quad \propto \frac{1}{\sqrt{s_1}}\exp\Big\{-\frac{1}{2}[\varphi_1^2 s_1+\frac{\beta_j^2}{s_1}]\Big\} \end{aligned}$$

Therefore, $s_1^{-1}|\text{rest}\sim$ Inverse-Gaussian$(\sqrt{\frac{\varphi_1^2}{\beta_j^2}},\varphi_1^2)$.

Similiarly, for $s_{2k}(k=1,\ldots,q)$, the posterior distribution for is $s_{2k}^{-1}|\text{rest}\sim$ Inverse-Gaussian$(\sqrt{\frac{\varphi_2^2}{\eta_{jk}^2}},\varphi_2^2)$.

The full conditional posterior distribution of φ_1^2 is:

$$\begin{aligned} & \varphi_1^2|\text{rest} \\ & \quad \propto \pi(s_1|\varphi_1^2)\pi(\varphi_1^2) \\ & \quad \propto \frac{\varphi_1^2}{2}\exp(-\frac{\varphi_1^2 s_1}{2})(\varphi_1^2)^{c_1-1}\exp(-d_1\varphi_1^2) \\ & \quad \propto (\varphi_1^2)^{c_1}\exp\Big(-\varphi_1^2(s_1/2+d_1)\Big) \end{aligned}$$

Therefore, the posterior distribution for φ_1^2 is Gamma$(c_1+1, s_1/2+d_1)$.

The full conditional posterior distribution of φ_2^2 is:

$$\begin{aligned} & \varphi_2^2|\text{rest} \\ & \quad \propto \pi(s_2|\varphi_2^2)\pi(\varphi_2^2) \\ & \quad \propto \prod_{k=1}^{q}\frac{\varphi_2^2}{2}\exp(-\frac{\varphi_2^2 s_{2k}}{2})(\varphi_2^2)^{c_2-1}\exp(-d_2\varphi_2^2) \\ & \quad \propto (\varphi_2^2)^{q+c_2-1}\exp\Big(-\varphi_2^2(\sum_{k=1}^{q}\frac{s_{2k}}{2}+d_2)\Big) \end{aligned}$$

The posterior distribution for φ_2^2 is Gamma$(c_2+q, \sum_{k=1}^{q} s_{2k}/2+d_2)$.

The full conditional posterior distribution of τ:

$$\begin{aligned} & \tau|\text{rest} \\ & \quad \propto \pi(v|\tau)\pi(\tau)\pi(Y|\cdot) \\ & \quad \propto \tau^{n/2}\exp\Big\{-\sum_{i=1}^{n}\frac{(y_i-E_i\alpha-C_i\gamma-X_{ij}\beta_j-\tilde{W}_i\eta_j)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\} \\ & \quad \times \tau^n\exp(-\tau\sum_{i=1}^{n}v_i)\tau^{a-1}\exp(-b\tau) \\ & \quad \propto \tau^{a+\frac{3}{2}n-1}\exp\Big\{-\tau\big[\sum_{i=1}^{n}(\frac{(y_i-E_i\alpha-C_i\gamma-X_{ij}\beta_j-\tilde{W}_i\eta_j)^2}{2\xi_2^2 v_i}+v_i)+b\big]\Big\} \end{aligned}$$

Therefore, $\tau|\text{rest} \sim \text{Gamma}(a + \frac{3}{2}n,\ \big[\sum_{i=1}^{n}(\frac{(y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2}{2\xi_2^2 v_i} + v_i) + b\big])$.

The full conditional posterior distribution of v_i:

$$\begin{aligned} & v_i|\text{rest} \\ & \propto \pi(v|\tau)\pi(y|\cdot) \\ & \propto \frac{1}{\sqrt{v_i}} \exp\Big\{ -\frac{(y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \times \exp(-\tau v_i) \\ & \propto \frac{1}{\sqrt{v_i}} \exp\Big\{ -\frac{1}{2}\big[(2\tau)v_i + \frac{\tau(y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2}{\xi_2^2 v_i}\big] \Big\} \end{aligned}$$

Therefore,

$$\frac{1}{v_i}|\text{rest} \sim \text{ Inverse-Gaussian}(\sqrt{\frac{2\xi_2^2}{(y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2}},\ 2\tau).$$

B.2 BLSS

B.2.1 Hierarchical model specification

$$\begin{gathered} Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2 \right\} \\ \alpha \sim \text{N}_q(0,\ \Sigma_{\alpha 0}) \\ \gamma \sim \text{N}_m(0,\ \Sigma_{\gamma 0}) \\ \beta_j|\pi_c, \tau_c^2, \sigma^2 \sim (1 - \pi_c)\,\text{N}\left(0,\ \sigma^2\tau_c^2\right) + \pi_c\,\delta_0(\beta_j) \quad j = 1, \ldots, p \\ \eta_{jk}|\pi_e, \tau_{ek}^2, \sigma^2 \overset{iid}{\sim} (1 - \pi_e)\,\text{N}\left(0,\ \sigma^2\tau_{ek}^2\right) + \pi_e\,\delta_0(\eta_{jk}) \quad j = 1, \ldots, p,\ k = 1, \ldots, q \\ \tau_c^2|\lambda_c^2 \sim \text{Gamma}(1, \frac{\lambda_c^2}{2}) \\ \tau_{ek}^2|\lambda_e^2 \overset{iid}{\sim} \text{Gamma}(1, \frac{\lambda_e^2}{2}) \quad k = 1, \ldots, q \\ \pi_c \sim \text{Beta}\,(r_c,\ u_c) \\ \pi_e \sim \text{Beta}\,(r_e,\ u_e) \\ \lambda_c^2 \sim \text{Gamma}\,(a_c,\ b_c) \\ \lambda_e^2 \sim \text{Gamma}\,(a_e,\ b_e) \\ \sigma^2 \sim \text{Inverse-Gamma}\,(s,\ h) \end{gathered}$$

B.2.2 Gibbs Sampler

Denote $\mu_{(-\alpha)} = \text{E}(Y) - E\alpha$, then $\alpha|\text{rest} \sim \text{N}(\mu_\alpha,\ \Sigma_\alpha)$, where

$$\begin{aligned} \mu_\alpha &= \Sigma_\alpha(\frac{1}{\sigma^2}(Y - \mu_{(-\alpha)})^\top E)^\top \\ \Sigma_\alpha &= \left(\frac{1}{\sigma^2}E^\top E + \Sigma_{\alpha 0}^{-1}\right)^{-1} \end{aligned}$$

Denote $\mu_{(-\gamma)} = \mathrm{E}(Y) - C\gamma$, then $\gamma|\text{rest} \sim \mathrm{N}(\mu_\gamma,\, \Sigma_\gamma)$, where

$$\mu_\gamma = \Sigma_\gamma(\frac{1}{\sigma^2}(Y - \mu_{(-\gamma)})^\top C)^\top$$
$$\Sigma_\gamma = \left(\frac{1}{\sigma^2}C^\top C + \Sigma_{\gamma 0}^{-1}\right)^{-1}$$

Denote $\mu_{(-\beta_j)} = \mathrm{E}(Y) - X_j\beta_j$, then $\beta_j|\text{rest} \sim (1 - l_c)\mathrm{N}(\mu_{\beta_j},\, \sigma^2\Sigma_{\beta_j}) + l_c\delta_0(\beta_j)$ where

$$\mu_{\beta_j} = \Sigma_{\beta_j}X_j^\top(Y - \mu_{(-\beta_j)})$$
$$\Sigma_{\beta_j} = \left(X_j^\top X_j + \frac{1}{\tau_c^2}\right)^{-1}$$
$$l_c = \frac{\pi_c}{\pi_c + (1-\pi_c)(\tau_c^2)^{-1/2}|\Sigma_{\beta_j}|^{1/2}\exp\{\frac{1}{2\sigma^2}\Sigma_{\beta_j}||X_j^\top(Y - \mu_{(-\beta_j)})||_2^2\}}$$

Denote $\mu_{(-\eta_{jk})} = \mathrm{E}(Y) - \tilde{W}_k\eta_{jk}$, then $\eta_{jk}|\text{rest} \sim (1 - l_{ek})\mathrm{N}(\mu_{\eta_{jk}},\, \sigma^2\Sigma_{\eta_{jk}}) + l_e\delta_0(\eta_{jk})$ where

$$\mu_{\eta_{jk}} = \Sigma_{\eta_{jk}}\tilde{W}_k^\top(Y - \mu_{(-\eta_{jk})})$$
$$\Sigma_{\eta_{jk}} = \left(\tilde{W}_k^\top\tilde{W}_k + \frac{1}{\tau_{ek}^2}\right)^{-1}$$
$$l_e = \frac{\pi_e}{\pi_e + (1-\pi_e)(\tau_{ek}^2)^{-1/2}|\Sigma_{\eta_{jk}}|^{1/2}\exp\{\frac{1}{2\sigma^2}\Sigma_{\eta_{jk}}||\tilde{W}_k^\top(Y - \mu_{(-\eta_{jk})})||_2^2\}}$$

The posterior of τ_c^2 is:

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

The posterior of τ_{ek}^2 is:

$$\frac{1}{\tau_{ek}^2}|\text{rest} \sim \begin{cases}\text{Inverse-Gamma}(1, \frac{\lambda_e^2}{2}) & \text{if } \eta_{jk} = 0\\ \text{Inverse-Gaussian}(\sqrt{\frac{\sigma^2}{\eta_{jk}^2}\lambda_e^2},\, \lambda_e^2) & \text{if } \eta_{jk} \neq 0\end{cases}$$

λ_c^2 and λ_e^2 have Gamma posterior distributions

$$\lambda_c^2|\text{rest} \sim \text{Gamma}(a_c + 1,\, \frac{\tau_c^2}{2} + b_c)$$
$$\lambda_e^2|\text{rest} \sim \text{Gamma}(a_e + q,\, \sum_{k=1}^{q}\frac{\tau_{ek}^2}{2} + b_e)$$

π_c and π_e have Gamma posterior distributions

$$\pi_c|\text{rest} \sim \text{Beta}(r_c - \mathbf{I}_{\{\beta_j \neq 0\}} + 1,\, u_c + \mathbf{I}_{\{\beta_j \neq 0\}})$$
$$\pi_e|\text{rest} \sim \text{Beta}(r_e - \sum_{k=1}^{q}\mathbf{I}_{\{\eta_{jk} \neq 0\}} + q,\, u_e + \sum_{k=1}^{q}\mathbf{I}_{\{\eta_{jk} \neq 0\}})$$

$\sigma^2 \sim$ Inverse-Gamma($\mu_{\sigma^2},\ \Sigma_{\sigma^2}$) where

$$\mu_{\sigma^2} = s + \frac{n + \mathbf{I}_{\{\beta_j \neq 0\}} + \sum_{k=1}^{q} \mathbf{I}_{\{\eta_{jk} \neq 0\}}}{2}$$
$$\Sigma_{\sigma^2} = h + \frac{(Y-\mu)^\top (Y-\mu) + (\tau_c^2)^{-1}\beta_j^2 + \sum_{k=1}^{q} (\tau_{ek}^2)^{-1} \eta_j^\top \eta_j}{2}$$

B.3 BL

B.3.1 Hierarchical model specification

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2 \right\}$$
$$\alpha \sim \mathrm{N}_q(0,\ \Sigma_{\alpha 0})$$
$$\gamma \sim \mathrm{N}_m(0,\ \Sigma_{\gamma 0})$$
$$\beta_j | \tau_c^2, \sigma^2 \sim \mathrm{N}\left(0,\ \sigma^2\tau_c^2\right) \quad j = 1, \ldots, p$$
$$\eta_{jk} | \tau_{ek}^2, \sigma^2 \overset{iid}{\sim} \mathrm{N}\left(0,\ \sigma^2\tau_{ek}^2\right) \quad j = 1, \ldots, p,\ k = 1, \ldots, q$$
$$\tau_c^2 | \lambda_c^2 \sim \exp(\frac{\lambda_c^2}{2})$$
$$\tau_{ek}^2 | \lambda_e^2 \overset{iid}{\sim} \exp(\frac{\lambda_e^2}{2}) \quad k = 1, \ldots, q$$
$$\lambda_c^2 \sim \mathrm{Gamma}\,(a_c,\ b_c)$$
$$\lambda_e^2 \sim \mathrm{Gamma}\,(a_e,\ b_e)$$
$$\sigma^2 \propto \frac{1}{\sigma^2}$$

B.3.2 Gibbs Sampler

Denote $\mu_{(-\alpha)} = \mathrm{E}(Y) - E\alpha$, then $\alpha|\text{rest} \sim \mathrm{N}(\mu_\alpha,\ \Sigma_\alpha)$, where

$$\mu_\alpha = \Sigma_\alpha (\frac{1}{\sigma^2}(Y - \mu_{(-\alpha)})^\top E)^\top$$
$$\Sigma_\alpha = \left(\frac{1}{\sigma^2} E^\top E + \Sigma_{\alpha 0}^{-1}\right)^{-1}$$

Denote $\mu_{(-\gamma)} = \mathrm{E}(Y) - C\gamma$, then $\gamma|\text{rest} \sim \mathrm{N}(\mu_\gamma,\ \Sigma_\gamma)$, where

$$\mu_\gamma = \Sigma_\gamma (\frac{1}{\sigma^2}(Y - \mu_{(-\gamma)})^\top C)^\top$$
$$\Sigma_\gamma = \left(\frac{1}{\sigma^2} C^\top C + \Sigma_{\gamma 0}^{-1}\right)^{-1}$$

Denote $\mu_{(-\beta_j)} = \mathrm{E}(Y) - X_j\beta_j$, then $\beta_j|\text{rest} \sim \mathrm{N}(\mu_{\beta_j},\, \sigma^2\Sigma_{\beta_j})$ where

$$\mu_{\beta_j} = \Sigma_{\beta_j} X_j^\top (Y - \mu_{(-\beta_j)})$$
$$\Sigma_{\beta_j} = \left(X_j^\top X_j + \frac{1}{\tau_c^2} \right)^{-1}$$

Denote $\mu_{(-\eta_{jk})} = \mathrm{E}(Y) - \tilde{W}_k\eta_{jk}$, then $\eta_{jk}|\text{rest} \sim \mathrm{N}(\mu_{\eta_{jk}},\, \sigma^2\Sigma_{\eta_{jk}})$ where

$$\mu_{\eta_{jk}} = \Sigma_{\eta_{jk}} \tilde{W}_k^\top (Y - \mu_{(-\eta_{jk})})$$
$$\Sigma_{\eta_{jk}} = \left(\tilde{W}_k^\top \tilde{W}_k + \frac{1}{\tau_{ek}^2} \right)^{-1}$$

The posterior of τ_c^2 is:

$$\frac{1}{\tau_c^2}|\text{rest} \sim \text{Inverse-Gaussian}(\sqrt{\frac{\sigma^2}{\beta_j^2}\lambda_c^2}, \lambda_c^2)$$

The posterior of τ_{ek}^2 is:

$$\frac{1}{\tau_{ek}^2}|\text{rest} \sim \text{Inverse-Gaussian}(\sqrt{\frac{\sigma^2}{\eta_{jk}^2}\lambda_e^2}, \lambda_e^2)$$

λ_c^2 and λ_e^2 have Gamma posterior distributions

$$\lambda_c^2|\text{rest} \sim \text{Gamma}(a_c + 1,\ \frac{\tau_c^2}{2} + b_c)$$
$$\lambda_e^2|\text{rest} \sim \text{Gamma}(a_e + q,\ \sum_{k=1}^{q} \frac{\tau_{ek}^2}{2} + b_e)$$

$\sigma^2 \sim \text{Inverse-Gamma}(\mu_{\sigma^2},\ \Sigma_{\sigma^2})$ where

$$\mu_{\sigma^2} = \frac{n + 1 + q}{2}$$
$$\Sigma_{\sigma^2} = \frac{(Y - \mu)^\top (Y - \mu) + (\tau_c^2)^{-1}\beta_j^2 + \sum_{k=1}^{q} (\tau_{ek}^2)^{-1}\eta_j^\top \eta_j}{2}$$