# Robust Bayesian method to sparse high-dimensional regression

by

Xi Lu

M.A., Columbia University, 2017

---

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2024

# Abstract

In high-dimensional regression problems, the demand for robust variable selection arises due to the widely observed outliers and heavy-tailed distributions of the response variable, as well as model misspecifications when structured sparsity is not properly accounted for. Although extensive frequentist-based robust regularization methods have been developed, within the Bayesian framework, the understanding of robust Bayesian analysis is still limited. In this dissertation, we systematically develop a suite of robust Bayesian penalization methods that can efficiently accommodate the data contamination while identifying the complicated underlying sparsity patterns.

In the first project, we propose a marginal robust Bayesian variable selection method for gene-environment (G×E) studies. Our study has been motivated by the difficulty to choosing sensible tuning parameters and lack of inferential power in existing robust marginal penalization methods. In particular, the Laplacian likelihood has been adopted in the Bayesian hierarchical model to accommodate data contamination and outliers. With the incorporation of spike-and-slab priors, we have implemented the Gibbs sampler based on Markov Chain Monte Carlo (MCMC). The proposed method outperforms a number of alternatives in extensive simulation studies. The utility of the proposed method has been further demonstrated using data from the Nurse Health Study (NHS).

In the second project, we develop a novel Bayesian quantile elastic net with spike-and-slab priors, which significantly improves over multiple versions of elastic net regularization methods. The Bayesian formulation of quantile regression in our method distinguishes the proposed one from the least square based elastic net in the presence of long-tailed distributions of the disease phenotype and heteroscedasticity in the regression error. Incorporation of the spike-and-slab priors leads to higher variable selection accuracy over Bayesian methods using the credible interval criterion and the scaled neighborhood criterion. In particular, the

proposed method enables exact statistical inference by providing Bayesian credible intervals with nominal coverage probabilities. The advantage of the proposed method has been fully demonstrated in both the simulation study and a biomedical data with high-dimensional genomics features.

In the third project, we consider the robust Bayesian subgroup analysis on samples extracted from a population with underlying grouping structures, which is an important step toward developing individualized treatment strategies for personalized medicine. Our approach can successfully recover group membership while retaining the robust property in the Bayesian framework. Numerical studies have shown the superiority of the developed method over alternatives when subgroup structure detection has been conducted in the presence of data contaminations. Promising results have also been revealed through a case study using the breast cancer data from The Cancer Genome Atlas (TCGA).

An important component of this dissertation is in developing user-friendly and publicly available software packages for reproducible research and broad dissemination of my work to the scientific community. To facilitate reproducible results and fast computation, we have developed high-speed R packages with C++ implementation for my dissertation projects. The R package *marble* for my first project and *Bayenet* for my second project are available on CRAN as well as on my Github pages. Currently, we are working on the R package of the third project.

# Robust Bayesian method to sparse high-dimensional regression

by

Xi Lu

M.A., Columbia University, 2017

---

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2024

Approved by:

Major Professor
Cen Wu

# Copyright

# Abstract

In high-dimensional regression problems, the demand for robust variable selection arises due to the widely observed outliers and heavy-tailed distributions of the response variable, as well as model misspecifications when structured sparsity is not properly accounted for. Although extensive frequentist-based robust regularization methods have been developed, within the Bayesian framework, the understanding of robust Bayesian analysis is still limited. In this dissertation, we systematically develop a suite of robust Bayesian penalization methods that can efficiently accommodate the data contamination while identifying the complicated underlying sparsity patterns.

In the first project, we propose a marginal robust Bayesian variable selection method for gene-environment (G×E) studies. Our study has been motivated by the difficulty to choosing sensible tuning parameters and lack of inferential power in existing robust marginal penalization methods. In particular, the Laplacian likelihood has been adopted in the Bayesian hierarchical model to accommodate data contamination and outliers. With the incorporation of spike-and-slab priors, we have implemented the Gibbs sampler based on Markov Chain Monte Carlo (MCMC). The proposed method outperforms a number of alternatives in extensive simulation studies. The utility of the proposed method has been further demonstrated using data from the Nurse Health Study (NHS).

In the second project, we develop a novel Bayesian quantile elastic net with spike-and-slab priors, which significantly improves over multiple versions of elastic net regularization methods. The Bayesian formulation of quantile regression in our method distinguishes the proposed one from the least square based elastic net in the presence of long-tailed distributions of the disease phenotype and heteroscedasticity in the regression error. Incorporation of the spike-and-slab priors leads to higher variable selection accuracy over Bayesian methods using the credible interval criterion and the scaled neighborhood criterion. In particular, the

proposed method enables exact statistical inference by providing Bayesian credible intervals with nominal coverage probabilities. The advantage of the proposed method has been fully demonstrated in both the simulation study and a biomedical data with high-dimensional genomics features.

In the third project, we consider the robust Bayesian subgroup analysis on samples extracted from a population with underlying grouping structures, which is an important step toward developing individualized treatment strategies for personalized medicine. Our approach can successfully recover group membership while retaining the robust property in the Bayesian framework. Numerical studies have shown the superiority of the developed method over alternatives when subgroup structure detection has been conducted in the presence of data contaminations. Promising results have also been revealed through a case study using the breast cancer data from The Cancer Genome Atlas (TCGA).

An important component of this dissertation is in developing user-friendly and publicly available software packages for reproducible research and broad dissemination of my work to the scientific community. To facilitate reproducible results and fast computation, we have developed high-speed R packages with C++ implementation for my dissertation projects. The R package *marble* for my first project and *Bayenet* for my second project are available on CRAN as well as on my Github pages. Currently, we are working on the R package of the third project.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgments

First, I would like to express my heartfelt gratitude to my major advisor, Dr. Cen Wu, for his invaluable guidance throughout this project. Having such a supportive and attentive advisor during my graduate years makes my Ph.D. experience productive and successful. Without his patience, constant support and timely encouragement, I would not have achieved the goal of completing this dissertation. Additionally, I would like to thank my committee members Dr. Juan Du, Dr. Michael Higgins and Dr. Weiqun Wang for their supports and guidance through my research progress.

I would like to extend my thanks to the department of Statistics for offering me graduate assistantship so that I could complete my graduate studies. What's more, thanks to the families of Coyne, Lin for providing scholarships to graduate students. Without this support, it would be more difficult to conduct my PhD research. I also want to thank the collaborators and coauthors for their contribution and suggestions, which greatly enhanced the papers included in this dissertation.

Finally, I would like to express my sincere gratitude to my family for their unconditional support and love. To my parents, thanks for supporting me and hoping the best for me. To my son, my cutest little one, thanks for loving me and motivating me to reach greater goals and be a better person. I could not achieve this accomplishment without them.

# Chapter 1

# Introduction

In high-dimensional genetic studies, it's important to identify gene-environment ( G×E ) interactions associated with clinical outcomes. As only a subset of genetic variants is associated with the response, variable selection is commonly used in determining important main and interaction factors. The advantages of penalization methods have be shown in solving problems with data matrix of "large dimensionality, small sample size" nature. Tibshirani (1996) proposed LASSO (least absolute shrinkage and selection operater), which is one of the most popular penalization variable selection method in regression problems. To accommodate correlation among variables, elastic net (Zou and Hastie (2005)) was introduced as a flexible regularized method. To avoid biased estimation in LASSO, plenty of unbiased variable selection methods were developed, such as the smoothly clipped absolute deviation (SCAD)(Fan and Li (2001)) and the minimax concave penalty (MCP) (Zhang (2010)).When dealing with hierarchical models, Bayesian LASSO (Park and Casella (2008)) shows davantages by accessing the posterior distributions of the parameters through a Gibbs sampler.Robust variable selection methods are also popularly used when encountering models with heavy-tailed distributions and ourliers, which is common in analysis of multi-omics data. In this chapter, we give a brief review of variable selection for both marginal and joint models with non-robust and robust methods.

## 1.1 Penalized Variable Selection

Penalization is one of the most important frameworks in selecting vital variables in multi-omics studies. Generally, the penalization model is in the form of "loss function + penalty". We use $X$ to denote the p genetic variants, which can be SNPs or gene expressions. $Y$ denotes the response variable, which can be a continuous disease outcome or categorical disease status. Let $\beta_j = (\beta_1, \ldots, \beta_p)^T$ be the coefficients. The optimization of the penalization regression is defined as

$$\min_{\beta} L(\beta; X, Y) + \sum_{j=1}^{p} P(|\beta_j|; \lambda)$$

Where $L(\beta; X, Y)$ is the loss function such as least square loss function, and $P(\cdot)$ is a penalty function which imposes shrinkage on the coefficients with the tuning paramater $\lambda$. With the increase of $\lambda$, the model complexity decreases and less variables are contained in the model. When $\lambda$ is close to zero, more variables are contained in the model. Thus, choosing a proper tuning parameter is crucial in variable selection.

LASSO (Tibshirani (1996)) was developed as a penalized least square regression with $L_1$ penalty with the form

$$||Y - X\beta||_2^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

The result of LASSO leads to an estimator which is continuous (continuity) to maintain stable in prediction and shrinks small coefficients to zero (sparsity) in estimation. With the property of sparsity, the coefficients of unimportant variables are shrinked to zero when implementing LASSO regression to the data set. However, LASSO doesn't meet the property of unbiasedness. With the development of penalties based on LASSO, plenty of methods are proposed with the proporty of continuity, sparsity and unbiasedness, which overcome the bias in LASSO, including SCAD and MCP. Fan and Li (2001) introduced SCAD with the

penalty as

$$P(\beta_j;\lambda,\gamma)=\begin{cases}\lambda|\beta_j| & \text{if } |\beta_j|\le\lambda\\ -\frac{\beta_j^2-2\gamma\lambda|\beta_j|+\lambda^2}{2(\gamma-1)} & \text{if } \lambda<|\beta_j|\le\gamma\lambda,\\ \frac{1}{2}(\gamma+1)\lambda^2 & \text{if } |\beta_j|\ge\gamma\lambda\end{cases}$$

where $\gamma>0$ and $\lambda>0$.The penalty of MCP is defined by Zhang (2010) as

$$P(\beta_j;\lambda,\gamma)\begin{cases}\lambda|\beta_j|-(2\gamma)^{-1}\beta_j^2 & \text{if } |\beta_j|\le\gamma\lambda\\ \frac{1}{2}\gamma\lambda^2 & \text{if } |\beta_j|\ge\gamma\lambda\end{cases}$$

where $\gamma>1$. When considering more complex data structures such as grouping structure, which is commonly encountered in solving problems with categorical variables, adopting other penalty function is in need. Yuan and Lin (2006) developed the group LASSO regression as the following

$$||Y-X\beta||_2^2+\lambda\sum_{g=1}^{G}||\beta_g||_{K_g},$$

Where $\beta_g$ is a group of coefficients of $\beta=(\beta_1^T,\ldots,\beta_G^T)^T$. $||\beta_g||_k=(\beta_g^TK\beta_g)^{1/2}$ with a symmetric positive definite matrix $K$. When dealing with a group of variable with high pairwise correlation, LASSO tends to select all of them. Ridge regression (Hoerl and Kennard (1970)) was proposed which minimizes the residual sum of squares with the $L_2$ penalty. However, although balancing the interconnection among features, ridge regression cannot do variable selection. The high interconnections among genetic features encourages the use of methods such as elastic net and fused LASSO. To contaminate correlations while producing a sparse solution at the same time, the elastic net (Zou and Hastie (2005)) which uses a combination of $L_1$ and $L_2$ becomes popular. The form of elastic net is given as follows

$$||Y-X\beta||_2^2+\lambda_1\sum_{j=1}^{p}|\beta_j|+\lambda_2\sum_{j=1}^{p}{\beta_j}^2$$

Here $\lambda_1$ and $\lambda_2$ are two tuning parameters. Combining the advantages of LASSO and ridge,

elastic net has a group effect and tends to shrink the coefficients while avoiding high correlation. Fused lasso (Tibshirani et al. (2005)) also performs well when analyzing correlated data.The fused lasso is defined as

$$||Y - X\beta||_2^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} |\beta_j - \beta_{j-1}|$$

With the penalization of $L_1$ of both the coefficients and their successive differences, fused lasso encourages a smooth sparse solution.

## 1.2 Bayesian variable selection

When solving problems with large scale data, Bayesian penalization variable selection is another popular strategy. The ordinary least square is supposed to get the optimal coefficient $\beta$ which minimizes the least square function $||Y - X\beta||^2$. With the penalized penalty like LASSO, important variables are selected through imposing shrinkage on coefficients. From the Bayesian perspective, Assuming $Y = X\beta + \epsilon$ with $\epsilon \sim \mathrm{N}(0, \sigma^2)$, the likelihood function can be expressed as

$$f(Y|\beta, \sigma^2) \propto (\frac{1}{\sigma^2})^{\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}||Y - X\beta||^2\}.$$

When the priors of regression coefficients are independent and identical Laplace priors $\pi(\beta_j) = \frac{\lambda}{2} e^{-\lambda|\beta_j|}$ $(j = 1, \ldots, p)$, the LASSO estimate can be solved as a Bayesian posterior mode estimate (Tibshirani (1996)). With an independent prior such as the improper prior $\pi(\sigma^2) = 1/\sigma^2$ on $\sigma^2$, the posterior distribution can be expressed as

$$\pi(\beta, \sigma^2|y) \propto \pi(\sigma^2)(\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}||Y - X\beta||_2^2 - \lambda \sum_{j=1}^{p} |\beta_j|\}. \tag{1.1}$$

However, with the independent and identical Laplace prior on $\beta$, the posterior (1.1) has more than one mode which causes problems in convergence in computation as well as in

explaining the concepts (Park and Casella (2008)). Park and Casella (2008) proposed the following conditional prior on $\beta_j$

$$\pi(\beta_j|\sigma^2) = \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}},$$

which leads to the posterior distribution with the property of unimodality. Although Bayesian LASSO costs a lot of computations, it's able to accommodate the uncertainty of parameters and provide credible interval estimates which cannot be achieved with LASSO.

A lot of extensions of Bayesian LASSO were developed, such as the Bayesian elastic net, Bayesian fused LASSO and Bayesian group LASSO (Kyung et al. (2010)). For instance, the conditional prior of $\beta$ with fused LASSO can be expressed as

$$\pi(\beta|\sigma^2) \propto \exp\{-\frac{\lambda_1}{\sigma}\sum_{j=1}^{p}|\beta_j| - \frac{\lambda_2}{\sigma}\sum_{j=1}^{p-1}|\beta_{j+1}-\beta_j|\}.$$

And the Bayesian elastic net can be represented with the prior

$$\pi(\beta|\sigma^2) \propto \exp\{-\frac{\lambda_1}{\sigma}\sum_{j=1}^{p}|\beta_j| - \frac{\lambda_2}{2\sigma}\sum_{j=1}^{p}\beta_j^2\}.$$

The main drawback of Bayesian LASSO, as well as its extended methods, is that there is no sparsity in the estimation results. This challenge encourages the incorporation of spike-and-slab priors (Mitchell and Beauchamp (1988)) on the coefficients $\beta$ with a point mass at zero and a normal or uniform distribution elsewhere, which is defined as

$$\beta_j|\phi_j \sim (1-\phi_j)\pi_1(\beta_j) + \phi_j\pi_0(\beta_j),\ j = 1,\ldots,p.$$

$\phi_j$ is an auxiliary indicator variable which can be set as a fixed value such as 0.5 or a conjugate prior such as beta prior. $\pi_1(\cdot)$ denotes a "slab distribution" such as normal distribution which accounts for nonzero effects while $\pi_0(\cdot)$ represents a "spike distribution" like point mass prior $\delta_0(\beta_j)$ at zero. If $\phi_j = 0$, $\beta_j \sim \pi_1(\beta_j)$, indicating that the $j$th genetic variants is

included in the final model while $\phi_j = 1$ implies the exclusion of the $j$th genetic variants in the model. With the widely use of mixture of normal priors(Hastie and Tibshirani (1993)), other mixture of distributions becomes possible. Ročková and George (2018) proposed a mixture of double exponential priors, where $\pi_1(\beta_j) = \frac{\lambda_1}{2}\exp(-\lambda_1|\beta_j|)$ with a small $\lambda_1$ and $\pi_0(\beta_j) = \frac{\lambda_0}{2}\exp(-\lambda_0|\beta_j|)$ with a large $\lambda_0$. Moreover, a group variable selection with spike and slab prior was introduced by Zhang et al. (2014) which allows the selection of grouped variables and within group variables at the same time. They also considered Bayesian fused lasso for interconnections within groups. A bi-level variable selection with a multivariate Laplace distribution was developed by Xu and Ghosh (2015). Variables are selected first at a gourp level and then within a group through this sparse group Bayesian LASSO method.

## 1.3 Robust Variable Selection

The data with heavy-tailed distribution and outliers widely exists, which motivates the incorporation of robust methods in hierarchical models. The most commonly used robust loss function is the check loss funciton in quantile regression. Consider the simplest model $Y = X\beta+\epsilon$. The estimation of quantile regression coefficients for the $\theta$th quantile $(0 < \theta < 1)$ amounts to the following minimization problem

$$\min_{\beta} \sum_{i=1}^{n} \rho_\theta(y_i - x_i^T\beta), \tag{1.2}$$

where $\rho_\theta(\epsilon) = \epsilon\{\theta - I(\epsilon < 0)\}$ is the check loss function. The least absolute deviation (LAD) loss function, as a special case of quantile regression, is defined as $\sum_{i=1}^{n} |y_i - x_i^T\beta|$. Penalization with LAD loss function is adopted in order to accommodate data contamination and improve computaion efficiency. With the framework of "robust loss function + penalty", the advantages of robust penalization have been showed in models with mis-specification, such as heavy-tailed errors and outliers (Wu and Ma (2015)). The regularized quantile

LASSO is given by

$$\min_{\beta} \sum_{i=1}^{n} \rho_\theta(y_i - x_i^T\beta) + \lambda \sum_{j=1}^{p} |\beta_j|. \tag{1.3}$$

Besides LASSO, there is a plenty of choice of penalties such as adaptive LASSO, SCAD and MCP. A comparison among these penalties with LASSO within the quantile regression shows a comparable performance of adaptive LASSO, SCAD and MCP, while they are better than LASSO penalty (Wang et al. (2012b)). In Bayesian quantile regression, for $y_i = x_i^T\beta + \epsilon_i, i = 1, \ldots, n$, error $\epsilon_i$'s are assumed to follow the skewed Laplace distribution with density

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

Then the likelihood function is

$$f(y|X, \beta, \tau) = \theta^n(1-\theta)^n\tau^n \exp(-\tau \sum_{i=1}^{n} \rho_\theta(y_i - x_i^T\beta)). \tag{1.4}$$

So the coefficient $\beta$ which maximizes 1.4 is the same as the one which minimizes 1.2. Moreover, the skewed Laplace distribution can be decomposed with a scaled normal distribution and an exponential distribution (Kozumi and Kobayashi (2009)). Consider quantile regression with the LASSO penalty (1.3), with a Laplace prior $\pi(\beta|\tau, \lambda) = (\frac{\tau\lambda}{2})^p \exp(-\tau\lambda \sum_{j=1}^{p} |\beta_j|)$(Li et al. (2010)), the posterior distribution of $\beta$ is

$$\pi(\beta|\tau, \lambda) \propto \exp\{-\tau \sum_{i=1}^{n} \rho_\theta(y_i - x_i^T\beta) - \tau\lambda \sum_{j=1}^{p} |\beta_j|\}. \tag{1.5}$$

The coefficient $\beta$ which maximizes 1.5 is also the solution of 1.3. The full conditional posterior distribution of all parameters can be obtained through Gibbs samplers. Li et al. (2010) gave a brief summary of Bayesian regularization methods in quantile regression with different penalties. The spike-and-slab priors can be incorporated to impose sparsity within the quantile Bayesian framework.

## 1.4 Marginal and Joint Analysis

Marginal and joint analysis are two important paradigms for G×E studies, especially when incorporating interaction terms.

### 1.4.1 Marginal Analysis

Marginal model is popularly used in high-dimensional G×E studies with its advantages in handling large data sets. Consider the following marginal model:

$$Y \sim Cs + Es + G + G \times (Es) \tag{1.6}$$

Here, we use Y to denote the response variable such as the cancer outcome or disease phenotype. $Cs$ denotes multiple clinical factors and $Es$ denotes multiple environmental factors. $G \times (Es)$ denotes a linear interaciton between one genetic factor and all environmental factors. As the clinical factors and environmental factors are not supposed to be selected, we impose penalty function on $G$ and $G \times (Es)$, which corresponds to main and interactions of one genetic factor. In marginal analysis, only one or a subset of the genetic factors are taken into consideration at one time, which is computationally convenient and stable when the number of features is much larger than that of the observations (Zhou et al. (2021)). For instance, in genome wide association study (GWAS), the interaction between one SNP and all environment factors is fitted at a time when using marginal regression model. The dimensionality of main and interaciton effects at each computation is significantly lower than the sample size(Zhou et al. (2021)).The important main and interaction factors can be selected with penalized regression methods, such as LASSO or Bayesian LASSO. Other statistical methods such as likelihood ratio test which is based on the marginal p-value can also be implemented in accomplishing selection.

Marginal analysis can also be used in models with non-linear interaciton terms by statistical significance based variable selection methods. P-values are computed as evidence for testing nonlinear interaciton effects when comparing the reduced model with the full

model. As the dimensionality of genetic is always high, which leads to millions of p-values, penalization methods are prefered in conducting variable selection in large scale studies. Another limitation of significance based method is that it may result in an unreliable solution when dealing with small sample sizes (Chai et al. (2017)). Penalization method has been widely used in studies with hierarchical structure. For example, Shi et al. (2014) proposed a rank-based penalization method which is robust to heavy-tailed data and model misspecification. Zhang et al. (2020) imposed a bi-level sparse structure between the main effects and interacitons in marginal identificaiton of G×E interactions. Furthermore, Bayesian variable selection methods have also been developed for linear and non-linear interactions in multiple studies.

### 1.4.2 Joint Analysis

Joint analysis is widely develoed on both linear and non-linear effects in G×E studies. While only one genetic feature is considered at a time in mariginal analysis, in joint analysis, all genetic feactures are analyzed in one model with its interaction of environmental factors. The following is a conceptual model for G×E interaction studies:

$$Y \sim Cs + Es + Gs + (Gs) \times (Es) \tag{1.7}$$

It is an extension from the marginal model ( (1.6)). Similarly, significance based methods and penalization methods can be conducted on joint models. Since all features are analyzed at one time, a bi-level variable selection method is highly demanded in a lot of situations. Here, one main genetic variant and its interaciton with all environmental factors are considered as one group as G×(1, Es). The sparse group variable seleciton method determines whether the genetic factor is associated with the phenotypic response on the group level. If the coefficient is zero, then there is no relation between the genetic factor and the response. If the coefficient is non-zero, an analysis is further needed on the individual level to determine the main and ineracton effects.

## 1.5 Works in this dissertation

Robust regularization methods have been widely developed (Wu et al. (2019); Wu and Ma (2015)). However, our literature search suggests that most frequentist robust variable selection methods are focused on regularized estimation under simple model structures and lacks inference procedures. In this dissertation, we have developed fully robust Bayesian analysis with valid statistical inferences. In particular, we have incorporated spike-and-slab priors in all the dissertation projects. The spike-and-slab prior can lead to parameter estimates with exact sparsity and facilitate valid statistical inference. Note that exact sparsity has been initially coined based on the frequentist LASSO estimates under the soft-thresholding rule. Although imposing a user-defined threshold can also lead to exact sparsity even when all initial shrinkage estimates are non-zero (Boyd et al. (2011); Wang et al. (2012a)), it causes issues under models with structured sparsity. For example, applying the cut-off as a judgement call can end up with choosing a sparse group (instead of a group) under a group-level penalty (Zhou et al. (2019, 2022)). Such an issue can be avoided if the developed method can automatically impose exact group-level penalty, in both the frequentist and Bayesian frameworks (Ren et al. (2020); Wu et al. (2018a)).

In chapter two, we propose a novel marginal Bayesian variable selection method for G×E studies. The identification of important G×E interactions plays key roles in dissecting the genetic basis of complex diseases with the main effects of genetic and environmental factors. The limitation in marginal penalized variable selection methods motivates us to consider Bayesian analysis. To contaminate outliers in data set and heavy-tailed distribution, we develop a robust maginal Bayesian method. The proposed robust method is in the form of the Bayesian formulation with the least absolute deviation (LAD) loss function. We consider the Bayesian LAD LASSO for identification of main and interaction effects. As Bayesian LAD LASSO cannot shrink the coefficients the zero, we incorporate the spike-and-slab priors to impose sparsity. The Gibbs sampler has been implemented based on MCMC for computation efficiency. The advantage of the proposed robust Bayesian marginal method has been demonstrated through simulation. We also further explore the utility of the proposed

method in case study with data from the Nurse Health Study (NHS).

In chapter three, we propose a novel robust Bayesian variable selection method with elastic net penalty for quantile regression in genetic analysis. As heavy-tailed error distribution and outliers in the response variable widely exists, models which are robust to data contamination are highly demanded. Although quantile regression is popular in frequentist genetic studies, there are seldom investigations from Bayesian perspective. Here, we consider the Bayesian quantile elastic net for regularized identification of important genetic variants. To impose exact sparsity, the spike-and-slab priors have been incorporated in the adaptive shrinkage framework. A Gibbs sampler based on MCMC has been implemented to obtain the posterior distributions of the coefficients of crucial factors. The advantage of the proposed method has been demonstrated with different quantile level in simulation study. Further findings from case study of the Cancer Genome Atlas(TCGA) skin cutaneous melanoma(SKCM) have been proved with current literature.

In chapter four, motivated by the cancer heterogeneity analysis of the breast cancer data from The Cancer Genome Atlas (TCGA), we have developed robust Bayesian subgroup analysis for high-dimensional gene expression data. A spike-and-slab prior has also been incorpotated to impose sparsity. The proposed method can simultaneously identify subgroups of patients while selecting important mRNA expressions with important genetic characteristics. Our approach can successfully recover group membership while retaining the robust property in the Bayesian framework. Furthermore, the posterior inferences utilizing the posterior samples from MCMC have provided inference procedures for both the identified subgroups and selected omics features, which are not available in frequentist approaches. Numerical studies have shown the advantage of our method over alternatives when subgroup structure detection and variable selection have been conducted in the presence of data contaminations because of cancer heterogeneity. A case study using the BRCA data from TCGA have also shown promising results.

# Chapter 2

# Identifying Gene-environment interactions with robust marginal Bayesian variable selection

## 2.1 Introduction

The risk and progression of complex diseases including cancer, asthma and type 2 diabetes, are associated with the coordinated functioning of genetic factors, the environmental (and clinical) factors, as well as their interactions (Cornelis and Hu (2012); Hunter (2005); Simonds et al. (2016); Von Mutius (2009)). The identification of important Gene-environment(G×E) interactions leads to novel insight in dissecting the genetic basis of complex diseases in addition to the main effects of genetic and environmental factors. In the last two decades, searching for the important G×E interactions has been extensively conducted based on genetic association studies (Cordell and Clayton (2005); Wu et al. (2012)). One representative example is the genome wide association study (GWAS), where the statistical significance of interaction between the environmental exposure and the genetic variant has been marginally assessed one at a time across the whole genome. Important findings are evidenced by genome wide significant p-values after adjusting for multiple comparisons.

Recently, substantial efforts have been devoted to novel penalized variable selection methods for G×E studies (Zhou et al. (2021)). In particular, marginal penalization has achieved very competitive performances with the aforementioned significance based G×E analysis (Chai et al. (2017); Shi et al. (2014); Zhang et al. (2020)). For example, within the framework of maximum rank correlation, Shi et al. (2014) has developed a penalization method robust to outliers and model misspecification in determining important G×E interactions one at a time. Zhang et al. (2020) has imposed hierarchical structure between the main effects and interactions in marginal identification of G×E interactions using regularization. Despite success, these studies have limitations. First, as a common tuning parameter is demanded for all the marginal models, its selection requires pooling all genes together to conduct a joint model based cross validation. While such a strategy is not rare, it seems not in favor of the marginal nature of the proposed G×E studies. Second, a rigorous measure to quantify uncertainty is not available. Zhang et al. (2020) has constructed 95% confidence intervals based on the observed occurrence index (OOI) values (Huang and Ma (2010)), nevertheless, this measure has been used to demonstrate stability of identified effects rather than quantifying uncertainty of penalized estimates.

These limitations have motivated us to consider Bayesian analyses. In literature, Bayesian variable selection methods have been developed for G×E analysis in multiple studies (Zhou et al. (2021)). For example, with indicator model selection, Liu et al. (2015) has imposed hierarchical Bayesian variable selection for linear G×E interactions. Li et al. (2015) has proposed a Bayesian group LASSO to identify non-linear interactions in nonparametric varying coefficient models. Ren et al. (2020) has further incorporated selection of linear and non-linear G×E interactions simultaneously while accounting for structured identification in the Bayesian adaptive shrinkage framework. All these fully Bayesian methods can efficiently provide uncertainty quantification based on the posterior samples from MCMC. Nevertheless, our limited literature mining shows that none of the marginal Bayesian variable selection methods have been proposed for interaction studies so far.

Historically, marginal analysis has prevailed in G×E interaction studies within the framework of genetic association studies. Although recent studies have confirmed the utility of

regularized variable selection in joint G×E analysis, more efforts are needed for marginal penalizations especially through the Bayesian point of view. The step towards marginal Bayesian variable selection is of particular significance in developing a coherent framework of analyzing G×E interactions.

Here, we propose a novel marginal Bayesian variable selection method for the robust identification of G×E interactions. As heavy-tailed distributions and outliers in the response variable have been widely observed, robust modelling is essential for yielding reliable results. Specifically, the robustness of the proposed method is facilitated by the Bayesian formulation of the least absolute deviation (LAD) regression which has been a popular choice in frequentist G×E studies but seldom investigated in a similar context from the Bayesian perspective. We consider the Bayesian LAD LASSO for regularized identification of interaction effects. As Bayesian LAD LASSO does not lead to zero coefficients, the spike-and-slab priors (George and McCulloch (1992); Ishwaran and Rao (2005)) has been incorporated to impose exact sparsity in the adaptive shrinkage framework. The corresponding MCMC algorithm has been developed to accommodate fast computations. We have demonstrated the advantage of the proposed robust Bayesian marginal analysis in simulation. The findings from the case study of the Nurses' Health Study (NHS) with SNP measurements have important biological implications.

## 2.2 Method

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 $i$th subject with $i$. Let $(Y_i, E_i, C_i, X_i)$ $(i = 1, \ldots, n)$ be independent and identically distributed random vectors. For $X_{ij}$ $(j = 1, \ldots, p)$, the measurement of the $j$th genetic factor on the

$i$th subject, consider the following marginal model:

$$\begin{aligned} Y_i &= \sum_{k=1}^{q} \alpha_k E_{ik} + \sum_{t=1}^{m} \gamma_t C_{it} + \beta_j X_{ij} + \sum_{k=1}^{q} \eta_{jk} X_{ij} E_{ik} + \epsilon_i \\ &= \sum_{k=1}^{q} \alpha_k E_{ik} + \sum_{t=1}^{m} \gamma_t C_{it} + \beta_j X_{ij} + \eta_j \tilde{W}_i + \epsilon_i, \end{aligned} \tag{2.1}$$

where $\alpha_k$'s and $\gamma_t$'s are the regression coefficients corresponding to effects of environmental and clinical factors, respectively. For the $j$th gene $X_j$ $(j = 1, \ldots, p)$, the G×E interactions effects are defined with $W_j = (X_j E_1, \ldots, X_j E_q)$, $\eta_j = (\eta_{j1}, \ldots, \eta_{jq})^T$. With a slight abuse of notation, denote $\tilde{W} = W_j$. The $\beta_j$'s and $\eta_{jk}$'s are the regression coefficients of the genetic variants and G×E interactions effects, correspondingly. Denote $\alpha = (\alpha_1, \ldots, \alpha_q)^T$ and $\gamma = (\gamma_1, \ldots, \gamma_m)^T$. Then model (4.1) can be written as

$$Y_i = E_i \alpha + C_i \gamma + X_{ij} \beta_j + \tilde{W}_i \eta_j + \epsilon_i. \tag{2.2}$$

### 2.2.1 Bayesian formulation of the LAD regerssion

The necessity of accounting for robustness in interaction studies has been increasingly recognized (Zhou et al. (2021)). Within the frequentist framework, it is essentially dependent on adopting a robust loss function to quantify lack of fit (Wu and Ma (2015)). Among a variety of popular robust losses, the least absolute deviation (LAD) loss function is well known for its advantages in dealing with heavy-tailed error distributions or outliers in response. The estimation of regression coefficients amounts 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 the robust marginal Bayesian variable selection based on LAD. As the Laplace distribution is equivalent to the mixture of an exponential distribution and a scaled normal distribution(Kozumi and Kobayashi (2011)), 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

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

where $\tau$ is the inverse of the scale parameters from the Laplace density. Then the likelihood function of our marginal G×E model 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 above formulation using Laplace distrubution is a special case of the asymmetric Laplace distribution, which has been widely adopted in Baysian quantile regression(Yu and Moyeed (2001); Yu and Zhang (2005)). In Baysian quantile regression, $\epsilon_i$'s are assumed to follow the skewed Laplace distribution with density

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

where $\rho_\theta(\epsilon) = \epsilon\{\theta - I(\epsilon < 0)\}$ is the check loss function. The random errors can be written as

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

where

$$\xi_1 = \frac{1-2\theta}{\theta(1-\theta)} \quad and \quad \xi_2 = \sqrt{\frac{2}{\theta(1-\theta)}}$$

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

The Bayesian LAD regression is a special case of Bayesian quantile regression(Li et al. (2010)) with $\theta$=0.5, resulting in that $\xi_1 = 0$ and $\xi_2 = \sqrt{8}$. Therefore, the response $Y_i$ can be written as:

$$\begin{aligned} Y_i &= \mu_i + \tau^{-1/2}\xi_2\sqrt{v_i}z_i, \\ v_i|\tau &\overset{iid}{\sim} \tau\exp(-\tau v_i), \\ z_i &\overset{iid}{\sim} \mathrm{N}(0,1), \end{aligned} \tag{2.3}$$

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

### 2.2.2 Bayesian LAD LASSO with spike-and-slab priors

In model (4.1), the coefficients $\beta_j$ and $\eta_j$ corresponds to the main and interaction effects with respect to the $j$th genentic variant, respectively. When $\beta_j = 0$ and $\eta_j = 0$, the genetic variant has no effect on the phenotype. A non-zero $\beta_j$ suggests the presence of main genetic effect. For $\eta_j$, if at least one of its component is not zero, then the G×E interaction effect exist. In the literature, Bayesian quantile LASSO, with Bayesian LAD LASSO as its special case, has been proposed to conduct variable selection (Li et al. (2010)). However, a major limitation is that Bayesian quantile LASSO cannot shrink regression coefficients to 0 exactly, resulting in inaccurate identification and biased estimation. To overcome such an limitation, we incorporate spike-and-slab priors to impose sparsity within Bayesian LAD LASSO framework as follows.

For the $j$th gene ($j = 1, \ldots, p$), the marginal LAD LASSO model is given by:

$$\sum_{i=1}^{n} |Y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j| + \lambda_1|\beta_j| + \lambda_2 \sum_{k=1}^{q} |\eta_{jk}|.$$

Let $\varphi_1 = \tau\lambda_1$ and $\varphi_2 = \tau\lambda_2$. Then the conditional Laplace prior on the coefficient of main effect $\beta_j$ can be expressed as scale mixtures of normals:

$$\begin{aligned}\pi(\beta_j|\tau, \lambda_1) &= \frac{\varphi_1}{2}\exp\{-\varphi_1|\beta_j|\} \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi s_1}}\exp(-\frac{\beta_j^2}{2s_1})\frac{\varphi_1^2}{2}\exp(\frac{-\varphi_1^2}{2}s_1)ds_1.\end{aligned}$$

The conditional Laplace prior on the coefficients of interaction effect $\eta_j$ can be written as:

$$\begin{aligned}\pi(\eta_j|\tau, \lambda_2) &= \prod_{k=1}^{q}\frac{\varphi_2}{2}\exp\{-\varphi_2|\eta_{jk}|\} \\ &= \prod_{k=1}^{q}\int_0^\infty \frac{1}{\sqrt{2\pi s_2}}\exp(-\frac{\eta_{jk}^2}{2s_2})\frac{\varphi_2^2}{2}\exp(\frac{-\varphi_2^2}{2}s_2)ds_2.\end{aligned}$$

Therefore, we consider the following hierarchical formulation for the marginal G×E model:

$$\begin{aligned}
\beta_j|s_1,\pi_1 &\sim (1-\pi_1)\mathrm{N}(0,s_1)+\pi_1\delta_0(\beta_j),\\
s_1|\varphi_1^2 &\sim \frac{\varphi_1^2}{2}\exp(-\frac{\varphi_1^2}{2}s_1),\\
\eta_{jk}|s_{2k},\pi_2 &\overset{iid}{\sim} (1-\pi_2)\mathrm{N}(0,s_{2k})+\pi_2\delta_0(\eta_{jk})(k=1,\ldots,q),\\
s_{2k}|\varphi_2^2 &\overset{iid}{\sim} \frac{\varphi_2^2}{2}\exp(-\frac{\varphi_2^2}{2}s_{2k})(k=1,\ldots,q),
\end{aligned} \tag{2.4}$$

where $\delta_0(\beta_j)$ and $\delta_0(\eta_{jk})$ denote the spike at 0, respectively, and the slab distributions are represented by two normal distributions, $\mathrm{N}(0,s_1)$ and $\mathrm{N}(0,s_2k)$. Here, $\pi_1 \in [0,1]$ and $\pi_2 \in [0,1]$. The mixture of the spike and slab components facilitate the selection of main and interaction effects. Instead of setting $\pi_1$ and $\pi_2$ to a fixed value such as 0.5, we assign conjugate beta priors on them as $\pi_1 \sim \mathrm{Beta}(r_1,u_1)$ and $\pi_2 \sim \mathrm{Beta}(r_2,u_2)$ which account for the uncertainty in $\pi_1$ and $\pi_2$. In this paper, we choose $r_1=u_1=r_2=u_2=1$ as it gives a prior mean with 0.5 and it also allows a prior to spread out.

In addition, the normal prior has been placed on the coefficients of environmental factor $\alpha_k(k=1,\ldots,q)$ and clinical factor $\gamma_t(t=1,\ldots,m)$ 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 conjugate Gamma priors on $\tau$, $\varphi_1^2$ and $\varphi_2^2$ with

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

In typical G×E studies, the environmental and clinical factors are of low dimensionality and the selection of them is not of interest. Therefore, the sparsity-inducing priors have not

been adopted for these factors. We consider the Bayesian LAD LASSO type of regularization in the proposed study as published studies have demonstrated that baseline penalty such as MCP and LASSO work well for marginal variable selection (Chai et al. (2017); Shi et al. (2014)).

It is noted that Zhang et al. (2020) has proposed a marginal sparse group MCP to respect the strong hierarchy between main and interaction effects. Their results are promising when long tailed distributions and outliers are not present in the response variable. Although sparse group (or, bi–level) variable selection has been demonstrated as being very effective in multiple G×E studies based on joint models (Zhou et al. (2021)), in our study, there is only one group per each marginal model. The sparse group no longer has significant advantages over individual level selection. Therefore, it has not been considered here.

Our model respects the weak hierarchy of "main effects, interactions". If imposing the strong hierarchy is needed, the genetic factor, once it is not selected given the presence of corresponding interaction effects, can be added back to the identified marginal model for a refit to impose strong hierarchy (Chai et al. (2017)). While such a practice is not uncommon in marginal interaction studies, Shi et al. (2014) has also revealed satisfactory performance when strong hierarchy has not been pursued.

### 2.2.3 The Gibbs sampler for robust marginal G×E analysis

For the $j$th genetic factor, the joint posterior distribution of all the unknown parameters conditional on data can be expressed as

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

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 the coefficient of environ-

mental factor $\alpha_k$ 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 $\alpha_k$ 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 the coefficient of clinical factor $\gamma_t (t = 1, \ldots, m)$ 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 the coefficient of genetic factor $\beta_j$ is a spike-and-slab distribution:

$$\beta_j|\text{rest} \sim (1 - l_1)N(\mu_{\beta_j}, \sigma_{\beta_j}^2) + l_1\delta_0(\beta_j), \tag{2.5}$$

where

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

We can show that

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

The posterior distribution of $\beta_j$ is a mixture of a normal distribution and a point mass at 0. That is, at each iteractio of MCMC, $\beta_j$ 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 the interaction of the $j$th gene and environmental factors $\eta_{jk}(k = 1, \ldots, q)$ is also a spike-and-slab distribution. Denote $\mu_{(-\eta_{jk})} = E(y_i) - W_{ik}\eta_{jk}$ and $l_{2k} = \pi(\eta_{jk} = 0|\text{rest})$, $\eta_{jk}$ follows this distribution:

$$\eta_{jk}|\text{rest} \sim (1 - l_{2k})N(\mu_{\eta_{jk}}, \sigma^2_{\eta_{jk}}) + l_{2k}\delta_0(\eta_{jk}), \tag{2.6}$$

where

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

And

$$l_{2k} = \frac{\pi_2}{\pi_2 + (1-\pi_2)s_{2k}^{-1/2}(\sigma^2_{\eta_{jk}})^{1/2}\exp\{\frac{1}{2}(\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\eta_{jk})})\tilde{W}_{ik}}{\xi_2^2 v_i})^2 \sigma^2_{\eta_{jk}}\}}. \tag{2.7}$$

The full conditional posterior distribution of $s_1$ is:

$$
\begin{aligned}
& s_1|\text{rest} \\
& \propto \pi(\beta_j|s_1,\pi_1)\pi(s_1|\varphi_1^2) \qquad (2.8)\\
& \propto \Big((1-\pi_1)(2\pi s_1)^{-1/2}\exp(-\frac{\beta_j^2}{2s_1})\mathbf{I}_{\{\beta_j\neq 0\}}+\pi_1\delta_0(\beta_j)\Big)\exp(-\frac{\varphi_1^2}{2}s_1).
\end{aligned}
$$

When $\beta_j = 0$, equation(4.6) is proportional to $\exp(-\frac{\varphi_1^2}{2}s_1)$. Therefore, the posterior distribution of $s_1$ is $\exp(\frac{\varphi_1^2}{2})$.

When $\beta_j \neq 0$, equation(4.6) is proportional 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 $\varphi_1^2$:

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

Therefore, the posterior distribution for $\varphi_1^2$ is Gamma$(c_1+1, s_1/2+d_1)$. Similiarly, the posterior distribution for $\varphi_2^2$ is Gamma$(c_2+q, \sum_{k=1}^{q} s_{2k}/2+d_2)$.

The full conditional posterior distribution of $\pi_1$:

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

Then, the posterior distribution for $\pi_1$ 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 $\pi_2$:

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

So, the posterior distribution for $\pi_2$ 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 $\tau$:

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

Therefore, the posterior distribution for $\tau$ 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

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

The spirit of marginal penalization for G×E interactions lies in the usage of a common sparsity cutoff to determine a list of important main and interaction effects. Instead of focusing on a fixed cutoff, varying the cutoff can generate different lists, resulting in a comprehensive view of important findings. The tuning parameter in penalized estimation serves as the cutoff. Therefore, the same tuning parameter has to be adopted for all the sub models (Chai et al. (2017); Shi et al. (2014); Zhang et al. (2020)). To further justify such a common tuning parameter, Zhang et al. (2020) has attempted using the joint model to select the common tuning through cross validation. However, this seems not coherent with the nature of marginal analysis.

Ideally, the tuning parameter should be determined by each model itself to allow for flexibility in controlling sparsity individually, and a common cutoff is still available to examine different lists of important effects. With the Bayesian formulation, we can avoid such a limitation of frequentist marginal penalization methods. In particular, the priors have been placed on regularization parameters to determine the sparsity in a data-driven manner for each sub model. With the spike-and-slab priors, the posterior distributions on the coefficients of main and interaction effects naturally lead to the usage of inclusion probability as a common cutoff to pin down the list of important effects, which is described in detail in the

next section.

## 2.3 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, LAD Bayesian Lasso with spike-and-slab priors (denoted as LADBLSS) with three alternatives, LAD Bayesian Lasso (denoted as LADBL), Bayesian Lasso with spike-and-slab priors (denoted as BLSS) and Bayesian Lasso (denoted as BL). LADBL is similar to the proposed method, except that it does not adopt the spike-and-slab prior. The details of posterior inference are available from the Appendix.

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 distributions with marginal mean 0, marginal variance 1 and AR1 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 AR1 structure with $\rho = 0.5$. Among the $p$ main G effects and $pq$ G×E interactions, 8 and 12 effects are set as being associated with the response, respectively. All the environmental and clinical factors are important with nonzero coefficients, which are randomly generated 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.

In addition, 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 (0,1,2) for genotypes (aa,Aa,AA) respectively. Here, the gene expressions are generated from the first setting.

*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 $\delta$ 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 $\delta = r\sqrt{q_1(1-q_1)q_2(1-q_2)}$.

We collect the posterior samples from the Gibbs Sampler with 10,000 interations and discard the first 5,000 samples as burn-ins. The posterior medians are used to estimate the coefficients. For approaches incorporating spike-and-slab priors, we consider computing the inclusion probability to indicate the importance of predictors. Here we use a binary indicator $\phi$ to denote that the membership of the non-spike distribution. Take the main effect of the $j$th genetic factor, $X_j$, as an example. Suppose we have collected H posterior samples from MCMC after burn-ins. The $j$th G factor is included in the marginal G×E model at the $h$th MCMC iteration if the corresponding indicator is 1, i.e., $\phi_j^{(h)} = 1$. Subsequently, the posterior probability of retaining the $j$th genetic main effect in the final marginal model is defined as the average of all the indicators for the $j$th G factor among the H posterior samples. That is,

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

A larger posterior inclusion probability $p_j$ indicates a stronger empirical evidence that the $j$th genetic main effect has a non-zero coefficient, i.e., a stronger association with the phenotypic trait.

To comprehensively assess the performance of the proposed and alternative methods,

we consider a sequence of probabilities as cutting-offs in inclusion probability for methods with spike-and-slab priors. Given a cutoff probability, the main or interaction is included in the final marginal model if its posterior inclusion probability is larger than the cutoff, and is excluded otherwise. Provided with a sequence of cutting-off probabilities from small to large, we can investigate the set of identified effects and calculate the true/false positive rates (T/FPR) as the ground truth is known in simulation. For the sequence of cut-offs, we are able to compute the area under curve (AUC) as a comprehensive measure. Besides, for methods without spike-and-slab priors, the confidence level of the credible intervals can be adopted as the cut-off to compute TPR and FPRs. Therefore, all the methods under comparison can be evaluated on the same ground.

In addition, we also consider Top100, which is defined as the number of true signals when 100 important main effects (or interactions) are identified. For methods with spik-and-slab priors, 100 main effects or interactions are chosen with the highest inclusion probabilities. For methods without spike-and-slab priors, the indicators of all effects are computed for a sequence of credible levels. The top 100 main effects or interactions are chosen in terms of the highest average identification values.

Simulation results for the gene expression data in the first setting are tabulated in Tables 2.1and 2.2. We can observe that the proposed method has the best performance among all approaches, especially when the response variable has 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 2.1, 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, i.e., 0.8432(sd 0.0115) from LADBL. Also, the AUC of robust methods is much larger than that of non-robust methods, especially in the presence of heavy-tailed errors. For instance, in the first setting under error3, the AUC of LADBLSS is 0.9558 and the AUC of LADBL is 0.8432 while that of BLSS and BL is around 0.5. Similar advantageous performance can also be observed from the identification results with Top100. In Table 2.2 under error 5, LADBLSS identifies 7.80(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, Bayesian LAD method with spike-and-slab priors has the best performance in all identification results. Under error 3, in Table 2.1, 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 2.2, 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, respectively.

**Table 2.1**: *Simulation results of the first setting for BL (Bayesian LASSO), BLSS (Bayesian LASSO with spike-and-slab priors), LADBL (LAD Bayesian LASSO) and LADBLSS (LAD Bayesian LASSO with spike-and-slab priors). 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 |
| N(0,1) | SD | 0.0052 | 0.0021 | 0.0076 | 0.0026 |
| Error 2 | AUC | 0.8332 | 0.9420 | 0.9004 | 0.9841 |
| $t(2)$ | SD | 0.0107 | 0.0235 | 0.0078 | 0.0031 |
| Error 3 | AUC | 0.5343 | 0.5473 | 0.8432 | 0.9558 |
| Lognormal(0,2) | SD | 0.0144 | 0.0576 | 0.0115 | 0.0161 |
| Error 4 | AUC | 0.8221 | 0.9124 | 0.9222 | 0.9895 |
| 90%N(0,1)+10%Cauchy(0,1) | SD | 0.0212 | 0.0410 | 0.0071 | 0.0024 |
| Error 5 | AUC | 0.7507 | 0.8431 | 0.9192 | 0.9904 |
| 80%N(0,1)+20%Cauchy(0,1) | SD | 0.0217 | 0.0633 | 0.0059 | 0.0018 |

Similar patterns can be observed in Table A.1, A.2 for the second setting, and Table A.3, A.4 for the third setting in Appendix. We have also investigated the performance of when n=2000 under setting 1. While the difference among the 4 methods significantly diminishes with such a large sample size, we can still observe the superior performance of LADBLSS by using a shorter list of top ranked effects. The results are provided in the table from supplementary files. Overall, the advantages of conducting robust Bayesian G×E analysis using the proposed approach can be justified based on the results of comprehensive simulation studies. The convergence of the MCMC chains with the potential scale reduction factor (PSRF) (Brooks and Gelman (1998)) has been conducted. In this study, we use PSRF $\leq$ 1.1 (Gelman et al. (2004)) as the cut-off point which indicates that chains converge to a

**Table 2.2**: *Identification results of the first setting with Top100 method for BL (Bayesian LASSO), BLSS (Bayesian LASSO with spike-and-slab priors), LADBL (LAD Bayesian LASSO) and LADBLSS (LAD Bayesian LASSO with spike-and-slab priors). mean(sd) based on 100 replicates. n=200, p=500, q=4 and* $m = 3$.

| | | Main | Interaction | Total |
|---|---|---|---|---|
| Error 1 | BL | 7.60(0.49) | 6.80(1.6) | 14.40(1.73) |
| N(0,1) | BLSS | 7.80(0.41) | 10.80(0.92) | 18.60(1.13) |
| | LADBL | 7.67(0.55) | 6.53(1.85) | 14.20(1.81) |
| | LADBLSS | 7.76(0.5) | 10.53(1.36) | 18.30(1.49) |
| Error 2 | BL | 6.37(1.90) | 3.90(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) | 5.80(1.71) | 13.23(2.01) |
| | LADBLSS | 7.53(0.51) | 9.90(1.56) | 17.43(1.76) |
| Error 3 | BL | 0.90(1.21) | 0.50(0.97) | 1.40(1.45) |
| Lognormal(0,2) | BLSS | 0.73(0.94) | 0.47(0.68) | 1.20(1.35) |
| | LADBL | 6.27(1.55) | 3.67(1.94) | 9.93(2.75) |
| | LADBLSS | 6.10(1.37) | 8.93(2.02) | 15.03(3.09) |
| Error 4 | BL | 5.57(2.99) | 3.63(2.53) | 9.20(5.05) |
| 90%N(0,1) | BLSS | 6.20(2.62) | 8.30(3.98) | 14.50(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) | 3.00(2.49) | 8.07(5.01) |
| 80%N(0,1) | BLSS | 4.60(3.25) | 5.70(4.23) | 10.30(7.27) |
| +20%Cauchy(0,1) | LADBL | 7.57(0.57) | 6.83(1.07) | 14.40(1.83) |
| | LADBLSS | 7.80(0.55) | 10.53(1.36) | 18.33(1.69) |

stationary distribution. The convergence of chains after burn-ins has been checked for all parameters with the value of PSRF less than 1.1. Figure B.1 shows the convergence pattern of PSRF for the main and interaction coefficients of the first genetic factors in Example 1 under Error 3.

In simulation, the hyper parameters for the Gamma priors and Beta priors specified in Section 2.2 are set to 1. In addition, the initial values of the regression parameters are also set to 1. Based on our experiments, the results and convergence of the MCMC algorithm are not sensitive to the choice of these parameters. We have observed satisfactory convergence for all of our simulations. For one simulated dataset under the first setting with n=200, p=500 and standard normal error, the CPU time (in minutes) for fitting all the 500 marginal models

through 10000 MCMC iterations on a laptop with standard configurations are 1.27(BL), 1.75(BLSS), 6.16(LADBL) and 5.95 (LADBLSS) minutes, respectively. The source codes of implementing all the methods under comparison are included in the supplementary files.

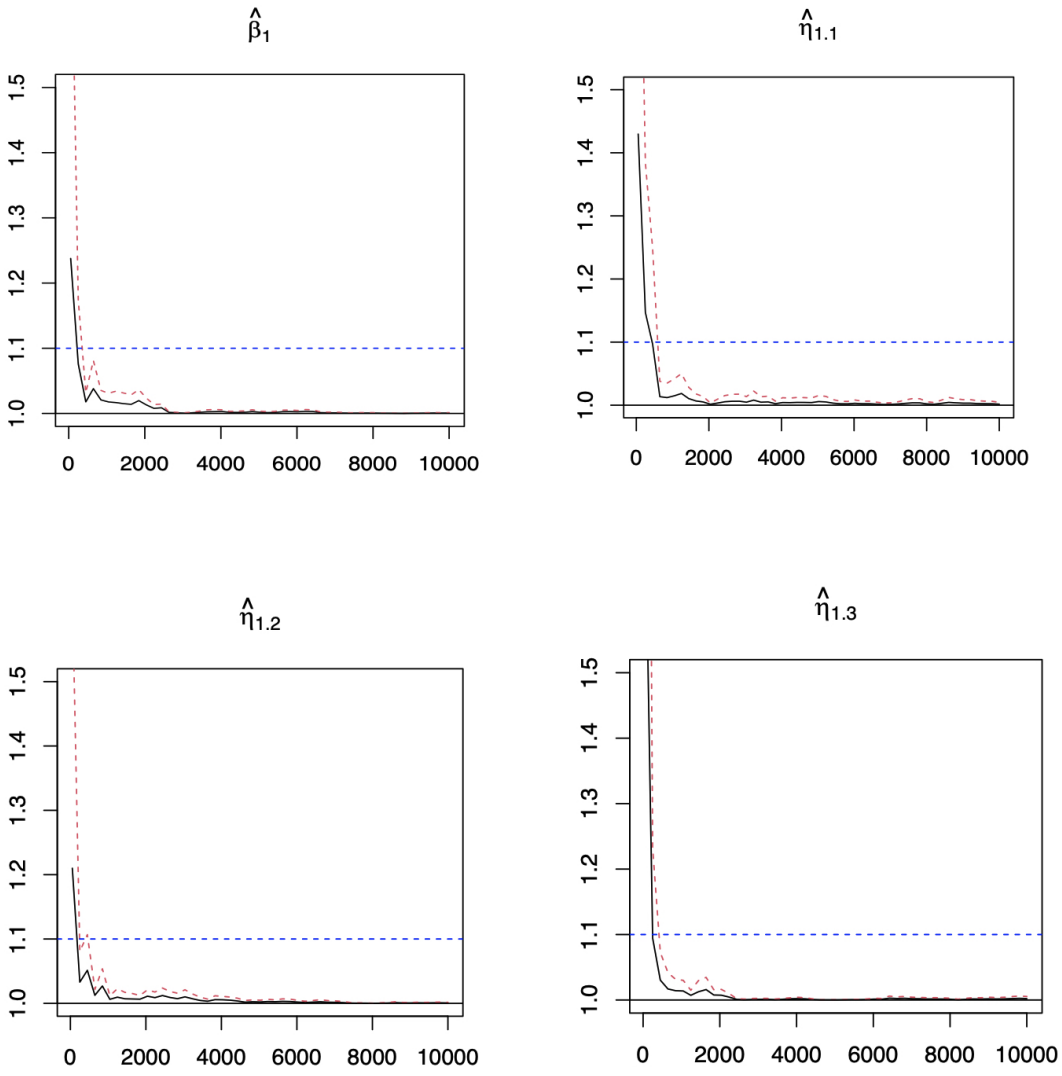


**Figure 2.1**: *Potential scale reduction factor (PSRF) against iterations for the coefficients of the first genetic factors and its interaction with environmental factors in Example 1 under Error 3. Black line: the PSRF. Red dotted line: the upper limits of the 95% confidence interval for the PSRF. Blue dotted line: the threshold of 1.1. The* $\hat{\beta}_1$ *represents the estimated coefficients of the main effects for the first genetic factor. The* $\hat{\eta}_{11}$ *to* $\hat{\eta}_{13}$ *represent the estimated coefficients of the first three interaction effects for the first genetic factor.*

## 2.4 Real Data Analysis

In this study, we analyze the type 2 diabetes (T2D) data from Nurses' Health Study (NHS), which is a well-characterized cohort study of women with high dimensional SNP data, as well as measurements on lifestyle and dietary factors. We consider SNPs on chronmosome 10 to identify main and gene-environment interactions associated with weight, which is an important phenotypic trait related to type 2 diabetes. Here, weight is used as response and five environment factors, age (age), total physical activity (act), trans fat intake (trans), cereal fiber intake (ceraf) and reported high blood cholesterol (chol) are considered. Data are available on 3391 subjects and 17016 gene expressions after cleaning the raw data through matching phenotypes and genotypes and removing SNPs with minor allele frequency (MAF) less than 0.05. A prescreening is done before downstream analysis. We use a marginal linear model with weight as response and age, act, trans, ceraf, chol as environment factors. 10,000 SNPs which have at least two main or interaction effects with p-value less than 0.05 are kept.The scale of working data is generally not a major concern for marginal analysis, as the computation can be done in a highly parallel manner. Here we focus on chromosome 10 which has been reported to harbor interesting genes in existing studies.

We use Top 100 method to identify 100 most important main and interaction effects. The proposed method LADBLSS identifies 20 main SNP effects and 80 gene-environment interactions, which are listed in Table A.5. Our study provides crucial implications in identifying the important main and interactions of SNPs and its associations with weight. For example, three SNPs, rs17011106, rs4838643 and rs17011115, located within gene WDFY4 are identified. WDFY4 has been observed as an influential factor related to weight and obesity (Barclay et al. (2015); Martin et al. (2019)). In addition, SNPs rs10994364, rs10821773 and rs10994308, located within gene ANK3, are identified with interacting environment factors age and chol. There are findings showing an association between ANK3 and higher systolic blood pressure (Ghanbari et al. (2014)). Published studies have also shown that ANK3 is linked to plumonary and renal hypertension (Ghanbari et al. (2014)). Allele risk variants have been identified in ANK3, and these variants explain a proportion of the her-

itability of BD (bipolar disorder), which is associated with higher body mass index (BMI) and increased metabolic comorbidity and the genetic risk for BD relates to common genetic risk with T2D (Winham et al. (2014)). Our proposed method identifies its interaction with chol, the high blood cholesterol. Data from several sources suggest that islet cholesterol metabolism contributes to the pathogenesis of T2D (Brunham et al. (2008)). Furthermore, the SNP rs1244416, corresponding to gene ATP5C1, interacts with the reported high blood cholesterol. This gene has been found to be deregulated in T2D skeletal muscle through pathway-based microanalysis (Morrison et al. (2012)). The interactions between SNP rs10857590 and trans fat intake has also been identified by using the proposed method. The SNP is within gene ARHGAP22 which has been investigated in Huang et al. (2018). As a diabetic retinopathy (DR) susceptibility gene, the expression of ARHGAP22 is positively associated with endothelial progenitor cells (EPC) levels in T2D patients with DR.

Analysis with alternatives BL, BLSS and LADBL has also been conducted. To compare the alternative methods with the proposed method, we provide the numbers of main effects and interactions identified by these methods with pairwise overlaps in Table 2.3. It clearly shows that the proposed one results in a very different set of effects compared to alternatives. We refit the regularized marginal models by LADBL and LADBLSS using robust Bayesian Lasso, and those identified by BL and BLSS using Bayesian Lasso. In addition, the inclusion probabilities of the selected main and interaction effects using LADBLSS are provided in Table A.6. Results from the alternative methods are available from the Supplementary files. The proposed method selects the 100 most important effects with the inclusion probability larger than 0.9, demonstrating its superiority in quantifying uncertain compared to marginal penalization methods (Chai et al. (2017); Shi et al. (2014); Zhang et al. (2020)). We noticed the small magnitude of refitted regression coefficients from LAD based methods compared to those obtained by the non–robust method in the supplementary files. This is due to the difference between the LAD based and least square based loss function for robust and non-robust methods, respectively. The advantage of LADBLSS over the non-robust methods can be clearly observed. First, majority of the top100 important effects identified by BL are main genetic effects. This is less likely to be reasonable as the response variable weight

has been well acknowledged to be also dependent on gene-environment interactions. For BLSS, the inclusion probabilities are low compared to those of the LADBLSS, suggesting lower level of certainty and confidence in the regression coefficients obtained from BLSS. The inferior performance of BL and BLSS further justifies the need of developing robust methods in marginal gene-environment interaction studies. Overall, LADBLSS leads to identification results significantly different from all the alternatives, as well as main and interaction effects of important biological implications that are not discovered by the benchmarks.

**Table 2.3**: *The numbers of main G effects and interactions identified by different approaches and their overlaps for BL (Bayesian LASSO), BLSS (Bayesian LASSO with spike-and-slab priors), LADBL (LAD Bayesian LASSO) and LADBLSS (LAD Bayesian LASSO with spike-and-slab priors).*

| **T2D** | Main | | | | Interaction | | | |
|---|---|---|---|---|---|---|---|---|
| | BL | BLSS | LADBL | LADBLSS | BL | BLSS | LADBL | LADBLSS |
| BL | 86 | 5 | 6 | 8 | 14 | 14 | 4 | 8 |
| BLSS | | 24 | 3 | 6 | | 76 | 20 | 23 |
| LADBL | | | 20 | 12 | | | 80 | 50 |
| LADBLSS | | | | 20 | | | | 80 |

## 2.5 Discussion

In the past, G×E interaction studies have been mainly conducted through marginal hypothesis testing, based on a diversity of study designs utilizing parametric, nonparametric and semiparametric models (Mukherjee et al. (2012); Murcray et al. (2009); Thomas (2010)), which later have been extended to joint analyses driven primarily by the pathway or gene set based association studies (Jiang et al. (2017); Jin et al. (2014); Wu and Cui (2013a)). In addition, published literature has also reported the success of marginal screening studies, including those based on partial correlations (Niu et al. (2018); Xu et al. (2019)). Recently, the effectiveness of regularized variable selection in G×E interaction studies has been increasingly recognized, and a large number of regularization methods have been proposed for joint interaction studies (Zhou et al. (2021)). Marginal penalization has also been demonstrated as promising competitors, although they have only been investigated in a limited number of frequentist studies (Chai et al. (2017); Shi et al. (2014); Zhang et al. (2020)).

Therefore, the proposed marginal robust Bayesian variable selection is of particular importance, since joint and marginal analysis cannot replace each other and marginal Bayesian penalization has not been examined for G×E studies so far. In particular, with the robustness and incorporation of spike- and-slab priors in the adaptive Bayesian shrinkage, the LADBLSS has an analysis framework more coherent with that of the joint robust analysis (Ren et al. (2023)), which significantly facilitates methodological developments for interaction studies.

Nevertheless, the proposed method has limitations. As a fully Bayesian methods based on MCMC algorithms, the computation cost is generally high due to the tradeoff for quantifying uncertainty using posterior samples. Such a drawback can be addressed through conducting the computation in a parallel manner given the marginal nature of the method. Besides, the variable selection conducted in our study is based on the L1 penalty within the Bayesian framework. As this structure ignores the correlation among genetic features, a possible direction for future improvement is to incorporate network or gene set information in the identification of important gene–environment interactions (Wang et al. (2021)).

Furthermore, in our study, the genetic factor is represented by one SNP coded as a triadic factor. A closer look at both the additive and dominant penetrance effects of the SNP will lead to elucidation of the genetic basis using marginal interaction studies on a finer scale. For gene –environment interaction studies, marginal and joint analysis are the two major paradigms, and cannot replace each other (Zhou et al. (2021)). It is always on a safe side to perform marginal analysis in G×E studies in addition to the joint ones, facilitating a more comprehensive understanding on the genetic architecture of complex diseases.

The marginal Bayesian regularization can be extended to different types of response, for example, under binary, categorical, prognostic and multivariate outcomes. Nevertheless, considering robustness in the generalized models with the Bayesian framework is not trivial, especially under the multivariate responses (Wu et al. (2014); Zhou et al. (2019)). We postpone the investigations to the future studies.The interaction between genetic and environmental factors in this study has been modeled as the product of the two corresponding variables, which amounts to "linear" interactions. In practice, the linear interaction assumption has been frequently violated(Ma et al. (2011); Wu and Cui (2013b); Zhao et al. (2019)), which demands accommodation of these nonlinear effects through nonparametic and semiparametric models (Li et al. (2015); Ren et al. (2020); Wu et al. (2015, 2018b)). It is of great interest and importance to migrate the nonlinear G×E studies to marginal cases in the near future.

# Chapter 3

# Bayesian Quantile Elastic Net with Spike-and-Slab Priors

## 3.1 Introduction

In genetic association studies, the central task is to identify the subset of important susceptible variants that is associated with the disease phenotype (Hirschhorn et al. (2002); Uffelmann et al. (2021); Wu et al. (2012)). In literature, a wide array of regularized variable selection methods have been proposed for such an analysis goal, including LASSO(Tibshirani (1996)), SCAD (Fan and Li (2001)) and MCP (Zhang (2010)) among others (Wu and Ma (2015)). To account for the structured sparsity, in particular the strong interconnections among the features, Zou and Hastie (2005) have proposed the elastic net where a weighted L1 penalty imposes sparsity and the ridge penalty accommodates strong correlations in shrinkage estimation. The superiority of elastic net in handling omics data with structured sparsity have been demonstrated in multiple association studies (Cho et al. (2009, 2010); Ogutu et al. (2012); Sokolov et al. (2016); Waldmann et al. (2013)). For example, using Genome wide association studies (GWAS), Waldmann et al. (2013) have demonstrated that elastic net outperforms LASSO when moderate to high linkage disequilibrium between SNPs exists. Cho et al. (2010) have also shown the advantage of applying elastic net in GWAS

since multicollinearity always exists due to linkage disequilibrium among neighboring SNPs.

Meanwhile, a variety of robust elastic net methods have been developed, showing great promise to accommodate data heterogeneity in the genetic association studies. Among those studies, Slawski (2012) has proposed the quantile elastic net along with the solution path algorithm, and shown that the more general structured sparsity can be incorporated in elastic net. Besides, Cohen Freue et al. (2019) have considered robustifying elastic net through regularizing the "S-loss" that rely on robust (squared) scale function of the residuals. In a plasma proteomic biomarkers study, the robust elastic net has yielded a selected model with superior predictive performance and important biomarkers ignored by non-robust methods. More recently, Kepplinger (2023) has further refined the robust elastic net (Cohen Freue et al. (2019)) by adopting an adaptive elastic net penalty, which leads to not only the oracle properties of the regularized estimator, but also a computationally scalable algorithm for model fitting.

In fact, the urgent need to develop robust structured variable selection methods can be further justified by analysis of a representative GWAS data consisting of the single nucleotide polymorphisms (SNPs) and the phenotypic trait, Body Mass Index (BMI), from the Nurse's Health Study (Colditz and Hankinson (2005); Colditz et al. (1997)) conducted in the case study of this paper. The strong correlations among SNPs in terms of high linkage disequilibrium (LD) can be observed through the LD plot in Figure 3.1. Besides, the right panel of Figure 3.1 shows the skewed distribution of BMI and outliers, which is due to the heterogeneity of complex diseases. Therefore, it is necessary to develop regularization approaches that accommodate both the robustness and structured sparsity for genetic association studies, such as the robust elastic net methods.

Despite the success, uncertainty quantification in robust regularization methods still emerges as a common challenge that has not been fully tackled as statistical inference procedures are usually not available due to the difficulty in deriving asymptotic distributions of the robust sparse estimators, especially in the presence of structured sparsity arising from omics features. On the other hand, within the Bayesian framework, methods with the fully Bayesian nature ensures exact statistical inference even when the sample size is limited, thus

overcoming the disadvantage in robust penalization methods. For Bayesian elastic net, multiple Bayesian hierarchical models have been developed, leveraging the strength of sparsity inducing priors including the elastic net prior and the orthant normal priors (Hans (2011); Kyung et al. (2010); Li and Lin (2010)). Furthermore, Li et al. (2010) have proposed the Bayesian quantile elastic net resistant to outliers and data heterogeneity through exploiting the asymmetric Laplace distribution based robust likelihood function (Yu and Moyeed (2001)).

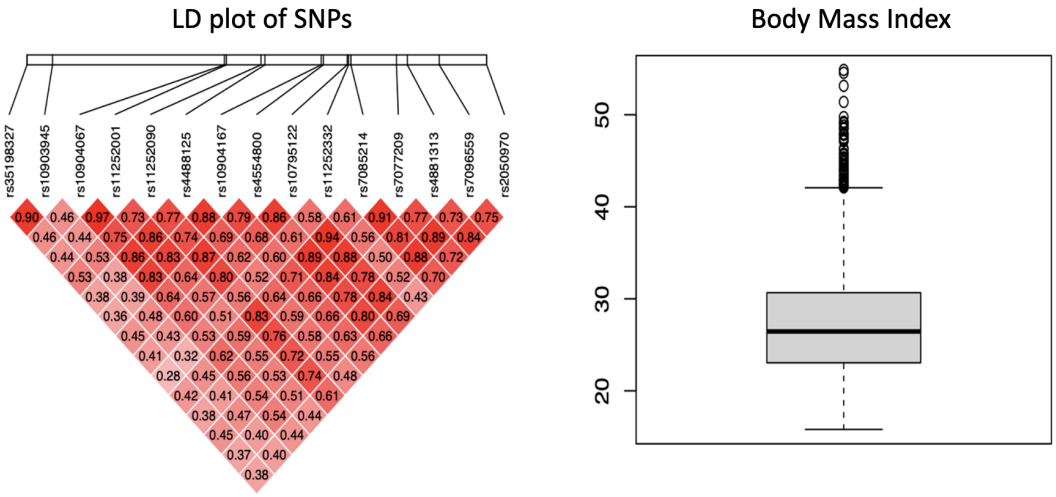


**Figure 3.1**: *LD plot of SNPs and boxplot of Body Mass Index (right) from NHS data.*

Although these fully Bayesian methods can complement frequentist elastic net models, they have suffered from the shortcoming of not producing exact sparsity. Specifically, the elastic net prior examined in Li and Lin (2010) and Li et al. (2010) can be viewed as a compromise between the ridge prior and LASSO prior (Park and Casella (2008)) that cannot lead to exact 0 coefficients in posterior estimates. Therefore, to conduct variable selection, Li et al. (2010) have suggested the credible interval and scaled neighborhood criteria as ad hoc treatments to impose sparsity, which have been well acknowledged to yield inferior selection results. For example, it has been reported in many published studies that the 95% credible intervals are usually too wide and many important predictors are thus discarded (Ren et al. (2023); Xu and Ghosh (2015); Zhou et al. (2023)).

Therefore, to unravel the full potential of robust elastic net in genetic association studies, we have developed a novel Bayesian quantile elastic net with spike-and-slab priors (George and McCulloch (1993); Mitchell and Beauchamp (1988)), which has superior performance over multiple alternatives including the existing family of elastic net methods. First, compared to its direct competitor Bayesian (quantile) elastic net (Li and Lin (2010); Li et al. (2010)), the incorporation of spike-and-slab prior promotes exact sparsity, and therefore bypass the credible intervals and scale neighborhood criteria that results in inferior identification performance, especially under higher dimensions. Second, in contrast to the frequentist robust penalization methods that usually lack procedure for uncertainty quantification, the proposed robust fully Bayesian method enables exact statistical inference even under small sample sizes. Third, elastic net based methods and similar ones are well known to yield large number of false positive findings due to the nature of promoting correlation among genomics features. The proposed one has overcame such a drawback that widely exists with regularization methods accounting for structured sparsity.

The robustness of our method roots in the Bayesian formulation of quantile regression as the robust likelihood function, which facilitates the developement of an efficient Gibbs sampler. Consequently, posterior sampling and statistical inference can be readily performed through the MCMC algorithm. Incorporation of the spike-and-slab priors leads to higher variable selection accuracy over Bayesian methods using the credible interval criterion and the scaled neighborhood criterion. The advantage of the proposed method has been demonstrated in both the simulation study and a case study of the high-dimensional SNP data from the Nurse Health Study (NHS). For fast computation and reproducible research, we implement the proposed and alternative methods in C++ and encapsulate them in a publicly available R package Bayenet (Lu and Wu (2023)).

## 3.2 Statistical Models

We denote $(Y_i,\ \mathbf{C}_i,\ \mathbf{X}_i)$ as independent and identically distributed random vectors, where the subscript $i$ refers to the $i$th subject $(i = 1, \ldots, n)$. In the random vector, $Y_i$ is the

disease phenotype, and $\boldsymbol{C}_i = (C_{i1}, \ldots, C_{i(q+1)})^\top$ is the $(q+1)$-dimensional vector for clinical covariates such as age and gender. Without loss of generality, $C_{i1} = 1$ corresponds to the intercept. In addition, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^\top$ represents the $p$ genetic variants such as gene expressions and single nucleotide polymorphisms (SNPs). Consider the following linear quantile regression model at the quantile level $\theta$ ( $0 < \theta < 1$):

$$Y_i = \sum_{k=1}^{q} \gamma_{k,\theta} C_{ik} + \sum_{j=1}^{p} \beta_{j,\theta} X_{ij} + \epsilon_{i,\theta}, \tag{3.1}$$

where $\gamma_{k,\theta}$'s and $\beta_{j,\theta}$'s are the regression coefficients corresponding to the effects of clinical (including the intercept) and genetic factors, respectively. The random error satisfies the condition that $P(\epsilon_{i,\theta} \leq 0 | \boldsymbol{C}_i, \boldsymbol{X}_i) = \theta$. The clinical covariates are of low dimensionality while the high-dimensional genetic factors have structured sparsity. For the rest of the paper, we omit the subscript "$\theta$" for simplicity of notation. Denote $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$. Then model (4.1) can be written as

$$Y_i = \boldsymbol{C}_i^\top \boldsymbol{\gamma} + \boldsymbol{X}_i^\top \boldsymbol{\beta} + \epsilon_i. \tag{3.2}$$

### 3.2.1 Bayesian formulation of the quantile regression

In frequentist quantile regression, estimation of the regression coefficients that depend on quantile level $\theta$ $(0 < \theta < 1)$ amounts to the following minimization problem:

$$\min_{\boldsymbol{\gamma}, \boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\theta (Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta}), \tag{3.3}$$

where the quantile check loss function is defined as $\rho_\theta(\epsilon_i) = \epsilon_i \{\theta - I(\epsilon_i < 0)\}$. In Bayesian quantile regression, the random error $\epsilon_i$'s are assumed to follow the asymmetric Laplace

distribution with density (Yu and Moyeed (2001); Yu and Zhang (2005)):

$$
\begin{aligned}
f(\epsilon_i|\tau) =& \theta(1-\theta)\tau \exp(-\tau\rho_\theta(\epsilon_i)), \\
=& \theta(1-\theta)\tau \begin{cases} e^{-\tau\theta\epsilon_i}, & \text{if } \epsilon_i \geq 0 \\ e^{\tau(1-\theta)\epsilon_i}, & \text{if } \epsilon_i < 0, \end{cases}
\end{aligned}
$$

where the scale parameter $\frac{1}{\tau}$ determines the skewness of the asymmetric Laplace distribution (ALD). Such a probabilistic assumption on model error is critical for formulating Bayesian quantile regression. We can express the likelihood function as:

$$
f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{X}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \tau) = \theta^n(1-\theta)^n\tau^n \exp\Big( -\tau \sum_{i=1}^{n} \rho_\theta(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta})\Big),
$$

where $\boldsymbol{Y} = (Y_1, ..., Y_n)^\top$, $\boldsymbol{C} = (\boldsymbol{C}_1, ..., \boldsymbol{C}_n)^\top$ and $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_n)^\top$. It can be observed that the maximum likelihood estimates of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are corresponding to the minimizer to the quantile check loss in (4.3). In literature, the asymmetric Laplace distribution can be equivalently specified as a mixture of exponential and scaled normal distribution as follows (Kozumi and Kobayashi (2011)). Let $\xi_1 = \frac{1-2\theta}{\theta(1-\theta)}$ and $\xi_2 = \sqrt{\frac{2}{\theta(1-\theta)}}$, given the quantile level $\theta$. Then the error term $\epsilon_i$ that follows an ALD with scale parameter $\frac{1}{\tau}$ can be written as:

$$
\epsilon_i = \tau^{-1}\xi_1\tilde{v}_i + \tau^{-1}\xi_2\sqrt{\tilde{v}_i}z_i,
$$

where $\tilde{v}_i$ and $z_i$ follow standard exponential distribution, Exp(1), and standard normal distribution, N(0,1), respectively. Define $v_i = \tau^{-1}\tilde{v}_i \sim \text{Exp}(\tau^{-1})$ leading to the following hierarchical model:

$$
\begin{aligned}
Y_i &= \boldsymbol{C}_i^\top \boldsymbol{\gamma} + \boldsymbol{X}_i^\top \boldsymbol{\beta} + \xi_1 v_i + \tau^{-1/2}\xi_2\sqrt{v_i}z_i, \\
v_1, ..., v_n &\sim \prod_{i=1}^{n} \tau \exp(-\tau v_i),
\end{aligned}
$$

$$z_1, ..., z_n \sim \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} z_i^2).$$

### 3.2.2 Bayesian quantile elastic net with spike-and-slab priors

The elastic net has been initially proposed in Zou and Hastie (2005) as a regularization method to improve accuracy in penalized estimation, selection and prediction over LASSO in the presence of highly correlated predictors. The penalty function of elastic net consists of the L1 norm and a ridge penalty, which leads to shrinkage estimation while accommodating strong correlations among the predictors. In the Bayesian framework, the Bayesian elastic net has been proposed by adopting a sparsity-inducing prior incorporating both the Gaussian prior and Laplacian prior (Li and Lin (2010), Kyung et al. (2010)). The data heterogeneity and outlying observations widely encountered in genomics studies have further motivated the development of robust elastic net (Li et al. (2010), Kepplinger (2023)). In particular, the Bayesian quantile elastic net proposed in Li et al. (2010) has shown superior performance over frequentist penalization approaches. Given the connection between frequentist and Bayesian shrinkage methods, we first consider the following penalized loss function under the quantile elastic net:

$$\sum_{i=1}^{n} \rho_\theta(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta}) + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2. \tag{3.4}$$

where $\rho_\theta(\cdot)$ is the quantile check loss function, and $\lambda_1$ and $\lambda_2$ are positive tuning parameters. As proposed in Li and Lin (2010) and Kyung et al. (2010), the elastic net prior on regression coefficient $\beta_j$ has the form of $\pi(\beta_j|\lambda_1, \lambda_2, \tau) \propto \lambda_1 \tau \exp\{-\lambda_1 \tau |\beta_j| - \lambda_2 \tau \beta_j^2\}$. With the aforementioned asymmetric Laplace distribution based likelihood function, we can obtain the posterior distribution on $\boldsymbol{\beta}$:

$$f(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \tau, \varphi_1, \varphi_2) \propto \exp\Big\{ -\tau \sum_{i=1}^{n} \rho_\theta(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta}) - \tau\lambda_1 \sum_{j=1}^{p} |\beta_j| - \tau\lambda_2 \sum_{j=1}^{p} \beta_j^2 \Big\}. \tag{3.5}$$

Then maximizing the posterior distribution in (3.5) is equivalent to minimizing the regularized loss function corresponding to the quantile elastic net in (4.5). Let $\varphi_1 = \tau\lambda_1$ and $\varphi_2 = \tau\lambda_2$, then we can set the elastic net prior on $\beta_j$ as:

$$\pi(\beta_j|\varphi_1, \varphi_2) = C(\varphi_1, \varphi_2)\frac{\varphi_1}{2}\exp\{-\varphi_1|\beta_j| - \varphi_2\beta_j^2\},$$

where $C(\varphi_1, \varphi_2)$ is a constant depending on $\varphi_1$ and $\varphi_2$. With the additional re-parametrization $\eta_1 = \frac{\varphi_1^2}{4\varphi_2}$ and $\eta_2 = \varphi_2$, the above elastic net prior, can be further expressed as a scale mixture of normal distribution with truncated gamma mixing density:

$$\pi(\beta_j|\eta_1, \eta_2) = \int_1^\infty \frac{1}{\sqrt{2\pi(t_j-1)/(2\eta_2 t_j)}}\exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1}\beta_j^2\}\Gamma^{-1}(\frac{1}{2}, \eta_1)\eta_1^{\frac{1}{2}}\exp(-\eta_1 t_j)t_j^{-1/2}\,\mathrm{d}t_j. \tag{3.6}$$

where $\beta_j|\eta_1, \eta_2$ is equivalent to a mixture of normal distribution, $\mathrm{N}(0, \frac{t_j-1}{2\eta_2 t_j})$ with $t_j > 1$, with a truncated gamma mixing density on $\frac{t_j-1}{2\eta_2 t_j}$, $(t_j > 1)$. Specifically, the truncated gamma distribution over support $(1, +\infty)$ has a shape parameter $1/2$ and a scale parameter $\eta_1^{-\frac{1}{2}}$. The detail can be found from Section B.5 in the Appendix.The representation directly leads to the following hierarchical specification in Bayesian regularized quantile regression Li et al. (2010):

$$\begin{aligned} \beta_j|t_j, \eta_2 &\overset{ind}{\sim} \mathrm{N}(0, \frac{t_j-1}{2\eta_2 t_j}), \\ t_j|\eta_1 &\overset{ind}{\sim} \Gamma^{-1}(\frac{1}{2}, \eta_1)t_j^{-\frac{1}{2}}\eta_1^{\frac{1}{2}}\exp\{-\eta_1 t_j\}\mathrm{I}(t_j > 1), \end{aligned} \tag{3.7}$$

which has a slight different form in the hierarchical model of Bayesian elastic net (Li and Lin (2010)) based on the normal likelihood. Although the formulation (3.7) ensures a equivalence between the marginal posterior mode of $\beta_j$ and the estimates provided by fitting the frequentist quantile elastic net in (4.5), it cannot yield exact 0 estimate of regression coefficients for variable selection purposes as in frequentist regularization methods. Li et al. (2010) have proposed to adopt the 95% credible intervals and the scaled neighborhood criterion for identification of sparse signals. However, even Li et al. (2010) have acknowledged that the 95% credible intervals are in general too wide to include important signals. On the other

hand, for the scaled neighborhood criterion, a probability threshold is needed to determine whether the predictor is retained or not. Li et al. (2010) have further suggested the usage of the receiver operating characteristic (ROC) and power curves to comprehensively evaluate the performance of the Bayesian elastic net by investigating a sequence of confidence coefficients and the probability thresholds. Such a strategy still cannot automatically lead to a final sparse model, and is feasible only when the ground truth is known a priori, such as in the simulation study.

In literature, to improve variable selection accuracy by promoting the exact sparsity, the spike-and-slab priors have been developed and widely used George and McCulloch (1993); Ishwaran and Rao (2005). For instance, Ren et al. (2023) have shown that adopting the spike-and-slab priors in the Bayesian hierarchical model has significantly boosted the identification and estimation accuracy in Bayesian regularized bi-level selection. Zhou et al. (2023) have also demonstrated the utility of non-parametric Bayesian variable selection in quantile varying coefficient models with the spike-and-slab priors. In this study, we have considered incorporating the spike-and-slab priors in Bayesian quantile elastic net to overcome the aforementioned limitation in Li and Lin (2010) and Li et al. (2010).

Specifically, we consider the following Bayesian hierarchical representation to incorporate the spike-and-slab priors:

$$\begin{aligned} \beta_j|t_j,\pi_1,\eta_2 &\overset{ind}{\sim} (1-\psi_j)\mathrm{N}(0,\frac{t_j-1}{2\eta_2 t_j})+\psi_j\delta_0(\beta_j), \\ \psi_j|\pi_1 &\overset{ind}{\sim} \text{Bernoulli}(\pi_1), \\ t_j|\eta_1 &\overset{ind}{\sim} \Gamma^{-1}(\frac{1}{2},\eta_1)t_j^{-\frac{1}{2}}\eta_1^{\frac{1}{2}}\exp\{-\eta_1 t_j\}\mathrm{I}(t_j>1). \end{aligned} \tag{3.8}$$

where $\delta_0(\beta_j)$ denotes the spike component of the prior corresponding to a point mass at 0 and the slab component is represented by the normal distribution, $\mathrm{N}(0,\frac{t_j-1}{2\eta_2 t_j})$. For each $\beta_j$ that quantifies the genetic contribution of the $j$th genetic feature to the variation in disease phenotype, a latent binary indicator $\psi_j$ has been assigned to facilitate variable selection with exact sparsity. In particular, $\beta_j$ will be shrunk to exact 0 with $\psi_j = 1$, since the spike component is a point mass at 0. Thus the $j$th genetic factor is not associated with the

response. Otherwise, if $\psi_j = 0$, the hierarchical representation in (4.4) reduces to (3.7), and $\beta_j$ is modeled via the normal distribution which does not directly yield a zero estimate. In addition, I$(\cdot)$is the indicator function and $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function. Here, we set $\pi_1 \in [0,1]$. Instead of setting $\pi_1$ with a non-informative prior such as 0.5, a conjugate beta prior as $\pi_1 \sim$Beta$(r_1, u_1)$ is given which accounts for the uncertainty in $\pi_1$. In this paper, we choose $r_1 = u_1 = 1$ as it gives a prior mean with 0.5 and it also allows a prior to spread out.

To sample $\eta_1$, we use a Metropolis-Hastings within Gibbs algorithm. The posterior distribution of $\eta_1$ is:

$$f(\eta_1|t_j) \propto \Gamma^{-p}(\frac{1}{2}, \eta_1)\eta_1^{p/2+c_1-1}\exp\{-\eta_1[d_1 + \sum_{j=1}^{p} t_j]\}.$$

With

$$\lim_{\eta_1 \to \infty} \frac{{\eta_1}^{1/2}\exp(\eta_1)}{\Gamma^{-1}(\frac{1}{2}, \eta_1)} = 1,$$

The proposed distribution is $q(\eta_1|t_j) \propto \eta_1^{p+c_1-1}\exp\{-\eta_1[d_1 + \sum_{j=1}^{p}(t_j - 1)]\}$. We generate samples of $\eta_1$ with the following steps:

1.Generate $\hat{\eta_1}$ from the proposed distribution $q(\eta_1|t_j)$.

2.Generate $U$ from $Uniform(0,1)$.

3. If

$$U \leq \frac{f(\hat{\eta_1})q(\eta_1)}{f(\eta_1)q(\hat{\eta_1})},$$

then set $\eta_1 = \hat{\eta_1}$, which means that $\eta_1$ is generated from the poposed distribution, otherwise $\eta_1$ is genrated from the original posterior distribution.

We place the normal prior on the coefficients of clinical factor $\gamma_k (k = 1, \ldots, q)$ as:

$$\gamma_k \overset{iid}{\sim} \frac{1}{\sqrt{(2\pi\gamma_0)}}\exp(-\frac{\gamma_k^2}{2\gamma_0})(k = 1, \ldots, q),$$

In addition, the conjugate Gamma priors are imposed on $\tau$, $\eta_1$ and $\eta_2$ as:

$$\tau \sim \text{Gamma}(a, b),$$

$$\eta_1 \sim \text{Gamma}(c_1, d_1),$$

$$\eta_2 \sim \text{Gamma}(c_2, d_2).$$

### 3.2.3 The Gibbs sampler for quantile elastic net regression analysis

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

$$\begin{aligned}
&\pi(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{v}, \boldsymbol{t}, \tau, \eta_1, \eta_2, \pi_1 | Y) \\
&\quad \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau^{-1}\xi_2^2 v_i}} \exp\Big\{ - \frac{(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \\
&\quad \times \prod_{i=1}^{n} \tau \exp(-\tau v_i) \tau^{a-1} \exp(-b\tau) \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} z_i^2) \\
&\quad \times \prod_{k=1}^{q} \frac{1}{\sqrt{(2\pi\gamma_0)}} \exp(-\frac{\gamma_k^2}{2\gamma_0}) \\
&\quad \times \prod_{j=1}^{p} \Big( (1-\pi_1)(2\pi(t_j-1)/(2\eta_2 t_j))^{-1/2} \exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1} \beta_j^2\} \mathbf{I}_{\{\boldsymbol{\beta}_j \neq 0\}} + \pi_1 \delta_0(\beta_j) \Big) \\
&\quad \times \prod_{j=1}^{p} \Gamma^{-1}(\frac{1}{2}, \eta_1) t_j^{-\frac{1}{2}} \eta_1^{\frac{1}{2}} \exp\{-\eta_1 t_j\} \mathrm{I}(t_j > 1) \\
&\quad \times (\eta_1)^{c_1 - 1} \exp(-d_1 \eta_1) \\
&\quad \times (\eta_2)^{c_2 - 1} \exp(-d_2 \eta_2) \\
&\quad \times \pi_1^{r_1 - 1} (1-\pi_1)^{u_1 - 1}
\end{aligned}$$

Let $\mu_{(-\gamma_k)} = E(Y_i) - C_{ik}\gamma_k$, $(i = 1, \ldots, n), (k = 1, \ldots, q)$, representing the mean effect without the contribution of $C_{ik}\gamma_k$. The posterior distribution of the coefficient of clinical

factor $\gamma_k$ conditional on all other parameters can be expressed as

$$\begin{aligned}
&\pi(\gamma_k|\text{rest}) \\
&\quad\propto \pi(\gamma_k)\pi(Y|\cdot) \\
&\quad\propto \exp\Big\{-\sum_{i=1}^{n}\frac{(Y_i-\boldsymbol{C}_i^\top\boldsymbol{\gamma}-\boldsymbol{X}_i^\top\boldsymbol{\beta}-\xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\}\times\exp(-\frac{\gamma_k^2}{2\gamma_0}) \\
&\quad\propto \exp\Big\{-\frac{1}{2}\big[(\sum_{i=1}^{n}\frac{\tau C_{ik}^2}{\xi_2^2 v_i}+\frac{1}{\gamma_0})\gamma_k^2-2\sum_{i=1}^{n}\frac{\tau(Y_i-\mu_{(-\gamma_k)})C_{ik}}{\xi_2^2 v_i}\gamma_k\big]\Big\}.
\end{aligned}$$

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

$$\mu_{\gamma_k}=(\sum_{i=1}^{n}\frac{\tau(Y_i-\mu_{(-\gamma_k)}-\xi_1 v_i)C_{ik}}{\xi_2^2 v_i})\sigma_{\gamma_k}^2,$$

and variance

$$\sigma_{\gamma_k}^2=(\sum_{i=1}^{n}\frac{\tau C_{ik}^2}{\xi_2^2 v_i}+\frac{1}{\gamma_0})^{-1}.$$

Let $\mu_{(-\boldsymbol{\beta}_j)}=E(Y_i)-X_{ij}\beta_j$ and $l_j=\pi(\beta_j=0|\text{rest})$, $(j=1,\ldots,p)$the conditional posterior distribution of the coefficient of genetic factor $\beta_j$ is a spike-and-slab distribution:

$$\beta_j|\text{rest}\sim(1-l_j)N(\mu_{\beta_j},\sigma_{\beta_j}^2)+l_j\delta_0(\beta_j), \tag{3.9}$$

where

$$\begin{aligned}
\mu_{\beta_j}&=(\sum_{i=1}^{n}\frac{\tau(Y_i-\mu_{(-\beta_j)}-\xi_1 v_i)X_{ij}}{\xi_2^2 v_i})\sigma_{\beta_j}^2, \\
\sigma_{\beta_j}^2&=(\sum_{i=1}^{n}\frac{\tau X_{ij}^2}{\xi_2^2 v_i}+\frac{2\eta_2 t_j}{t_j-1})^{-1}.
\end{aligned}$$

We can show that

$$l_j=\frac{\pi_1}{\pi_1+(1-\pi_1)(\frac{2\eta_2 t_j}{t_j-1})^{-1/2}(\sigma_{\beta_j}^2)^{1/2}\exp\{\frac{1}{2}(\sum_{i=1}^{n}\frac{\tau(y_i-\mu_{(-\beta_j)}-\xi_1 v_i)X_{ij}}{\xi_2^2 v_i})^2\sigma_{\beta_j}^2\}}.$$

The posterior distribution of $\beta_j$ is a mixture of a normal distribution and a point mass at 0.

That is, at each iteraction of MCMC, $\beta_j$ is drawn from $N(\mu_{\beta_j}, \sigma^2_{\beta_j})$ with probability $(1 - l_j)$ and is set to 0 with probability $l_j$.

The full conditional posterior distribution of $t_j - 1$ is:

$$
\begin{aligned}
& t_j - 1|\text{rest} \\
& \quad \propto \pi(\beta_j|t_j, \pi_1)\pi(t_j|\eta_1) \\
& \quad \propto \Big((1-\pi_1)(2\pi(t_j-1)/(2\eta_2 t_j))^{-1/2}\exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1}\beta_j^2\}\mathbf{I}_{\{\beta_j\neq 0\}} + \pi_1\delta_0(\beta_j)\Big) \\
& \quad \times \Gamma^{-1}(\frac{1}{2}, \eta_1)t_j^{-\frac{1}{2}}\eta_1^{\frac{1}{2}}\exp\{-\eta_1 t_j\}\mathrm{I}(t_j > 1).
\end{aligned}
\tag{3.10}
$$

When $\beta_j = 0$, equation (3.10) is proportional to $\Gamma^{-1}(\frac{1}{2}, \eta_1)t_j^{-\frac{1}{2}}\eta_1^{\frac{1}{2}}\exp\{-\eta_1 t_j\}\mathrm{I}(t_j > 1)$, which is the truncated gamma distribution. Therefore, the posterior distribution of $t_j - 1$ is $\mathrm{TG}(\frac{1}{2}, \eta_1^{-1}, 1)$.

When $\beta_j \neq 0$, equation (3.10) is proportional to

$$
\begin{aligned}
& \frac{1}{\sqrt{t_j-1}}\exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1}\beta_j^2\}\exp(-\eta_1 t_j)\mathrm{I}(t_j > 1) \\
& \qquad \propto \frac{1}{\sqrt{t_j-1}}\exp\Big\{-\frac{1}{2}[2\eta_1(t_j-1) + \frac{2\eta_2\beta_j^2}{t_j-1}]\Big\}\mathrm{I}(t_j > 1).
\end{aligned}
$$

Therefore, when $\beta_j \neq 0$, the posterior distribution for $(t_j - 1)^{-1}$ is Inverse-Gaussian$(\sqrt{\frac{\eta_1}{\eta_2\beta_j^2}}, 2\eta_1)$.

The full conditional posterior distribution of $\eta_1$ is:

$$
\begin{aligned}
& \eta_1|\text{rest} \\
& \quad \propto \pi(t_j|\eta_1)\pi(\eta_1) \\
& \quad \propto \eta_1^{c_1-1}\exp(-d_1\eta_1)\prod_{j=1}^{p}\Gamma^{-1}(\frac{1}{2}, \eta_1)\eta_1^{\frac{1}{2}}\exp\{-\eta_1 t_j\} \\
& \quad \propto \Gamma^{-p}(\frac{1}{2}, \eta_1)\eta_1^{p/2+c_1-1}\exp\{-\eta_1[d_1 + \sum_{j=1}^{p} t_j]\}.
\end{aligned}
$$

We use a Metropolis-Hastings within Gibbs sampling. The proposal distribution for the

Metropolis-Hastings step is $\eta_1^{p+c_1-1}\exp\{-\eta_1[d_1+\sum_{j=1}^{p}(t_j-1)]\}$.

The full conditional posterior distribution of $\eta_2$ is:

$$
\begin{aligned}
&\eta_2|\text{rest}\\
&\quad\propto \pi(\beta_j|\eta_2)\pi(\eta_2)\\
&\quad\propto \Big((1-\pi_1)(2\pi(t_j-1)/(2\eta_2 t_j))^{-1/2}\exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1}\beta_j^2\}\mathbf{I}_{\{\beta_j\neq 0\}}+\pi_1\delta_0(\beta_j)\Big)\\
&\quad\times \eta_2^{c_2-1}\exp(-d_2\eta_2)
\end{aligned}
$$

So, the posterior distribution for $\eta_2$ is Gamma $(c_2+\frac{1}{2}\sum_{j=1}^{p}\mathbf{I}_{\{\beta_j\neq 0\}}, d_2+\sum_{j=1}^{p}\frac{t_j}{t_j-1}\beta_j^2)$.

The full conditional posterior distribution of $\pi_1$:

$$
\begin{aligned}
&\pi_1|\text{rest}\\
&\quad\propto \pi(\beta_j|\pi_1)\pi(\pi_1)\\
&\quad\propto \pi_1^{r_1-1}(1-\pi_1)^{u_1-1}\\
&\quad\times \Big((1-\pi_1)(2\pi(t_j-1)/(2\eta_2 t_j))^{-1/2}\exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1}\beta_j^2\}\mathbf{I}_{\{\beta_j\neq 0\}}+\pi_1\delta_0(\beta_j)\Big).
\end{aligned}
$$

Then, the posterior distribution for $\pi_1$ is Beta $(p+r_1-\sum_{j=1}^{p}\mathbf{I}(\beta_j\neq 0), u_1+\sum_{j=1}^{p}\mathbf{I}(\beta_j\neq 0))$.

The full conditional posterior distribution of $\tau$:

$$
\begin{aligned}
&\tau|\text{rest}\\
&\quad\propto \pi(\boldsymbol{v}|\tau)\pi(\tau)\pi(\boldsymbol{Y}|\cdot)\\
&\quad\propto \tau^{n/2}\exp\Big\{-\sum_{i=1}^{n}\frac{(Y_i-\boldsymbol{C}_i^\top\boldsymbol{\gamma}-\boldsymbol{X}_i^\top\boldsymbol{\beta}-\xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\}\\
&\quad\times \tau^n\exp(-\tau\sum_{i=1}^{n}v_i)\tau^{a-1}\exp(-b\tau)\\
&\quad\propto \tau^{a+\frac{3}{2}n-1}\exp\Big\{-\tau\big[\sum_{i=1}^{n}(\frac{(Y_i-\boldsymbol{C}_i^\top\boldsymbol{\gamma}-\boldsymbol{X}_i^\top\boldsymbol{\beta}-\xi_1 v_i)^2}{2\xi_2^2 v_i}+v_i)+b\big]\Big\}.
\end{aligned}
$$

Therefore, the posterior distribution for $\tau$ is Gamma $(a+\frac{3}{2}n,\ \sum_{i=1}^{n}(\frac{(Y_i-\boldsymbol{C}_i^\top\boldsymbol{\gamma}-\boldsymbol{X}_i^\top\boldsymbol{\beta}-\xi_1 v_i)^2}{2\xi_2^2 v_i}+$

$v_i) + b)$.

Last, we have the full conditional posterior distribution of $v_i$:

$$
\begin{aligned}
& v_i|\text{rest} \\
& \propto \pi(\boldsymbol{v}|\tau)\pi(\boldsymbol{Y}|\cdot) \\
& \propto \frac{1}{\sqrt{v_i}} \exp\Big\{ - \frac{(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \times \exp(-\tau v_i) \\
& \propto \frac{1}{\sqrt{v_i}} \exp\Big\{ - \frac{1}{2}\big[ (\frac{\tau\xi_1^2}{\xi_2^2} + 2\tau) v_i + \frac{\tau(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_i)^2}{\xi_2^2 v_i} \big] \Big\}.
\end{aligned}
$$

It can be found that

$$
\frac{1}{v_i}|\text{rest} \sim \text{Inverse-Gaussian}\Big(\sqrt{\frac{\xi_1^2 + 2\xi_2^2}{(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_i)^2}}, \ \frac{\tau\xi_1^2}{\xi_2^2} + 2\tau\Big).
$$

## 3.3 Simulation

To demonstrate the advantage of the proposed method, we evaluate the performance through simulation study. In particular, we compare the performance of the proposed method, Bayesian quantile elastic net with spike-and-slab priors (denoted as bqenetss) with seven alternatives, Bayesian quantile elastic net (denoted as bqenet), Bayesian elastic net (denoted as benet), Bayesian quantile Lasso (denoted as qbl), Bayesian Lasso with spike-and-slab priors (denoted as blss), Bayesian Lasso (denoted as bl), Elastic net (denoted as enet) and LASSO (denoted as lasso). Among all the methods under comparison, there are six Bayesian methods where bqenetss, bqenet and qbl are robust and benet, blss and bl are non-robust ones. The non-robust frequentist regularization methods enet and lasso have been implemented using R package *glmnet* (Friedman et al., 2010)). The inclusion of clinical covariates in model (4.1) demands tailored Gibbs samplers for all Bayesian methods. The details of posterior inferences are available from the Appendix.

Under all the settings, the sample size is set as $n = 300$, and the number of G factors is $p = 400$ and the number of clinical factors is $q = 3$. Three clinical factors are generated

from a multivariate normal distribution with marginal mean 0, variance 1 and AR(1) correlation structure with $\rho = 0.5$. All the clinical factors have nonzero coefficients, which are randomly generated from a uniform distribution Uniform$[0.8, 1.5]$. The random error are generated from five different distributions: (1)N$(0, 1)$(Error1), (2)$t(2)$(Error2), (3)Lognormal(0,1)(Error3), (4)30%N$(0, 4)$+70%N$(0, 1)$(Error4), (5)20%$t(2)$+80%Laplace$(0, 2)$(Error5). Except Error1, all the rest of errors have heavy-tailed distribution.

### 3.3.1 Independent and identically distributed random errors

For the i.i.d. random errors specified in model (4.1), the genetic factors are simulated in the following two settings.

*Setting 1.* We generate continuous genetic variants from multivariate normal distributions with marginal mean vector 0 and a covariance matrix with auto-regressive (AR-1) correlation. Then the $j$th and $k$th genetic factors have correlation $\rho^{|j-k|}$ with $\rho$ being 0.5. 25 coefficients of genetic effects are randomly chosen to be nonzero and simulated from Uniform$[0.3, 0.9]$.

*Setting 2.* We simulate the genetic variants in a cluster-wise manner with higher correlations so they are more realistic (Ren et al. (2019)). Specifically, the 400 genes are simulated from 80 clusters with 5 gene per cluster which are generated from multivariate normal distributions with mean 0 and auto-regression correlation structure $\rho^{|j-k|}$ with $\rho = 0.8$. 8 out of 80 cluster are randomly selected to have nonzero effects on the disease phenotype, where 3 out of 5 coefficients are randomly chosen to be nonzero from Uniform$[0.6, 1.2]$.

We have collected the posterior samples from the Gibbs Sampler that run with 10,000 iterations. The first 5,000 samples have been discarded as burn-ins. For approaches incorporating spike-and-slab priors, we consider computing the inclusion probability to indicate the importance of predictors with a binary indicator $\phi$ denoting the non-spike distribution. For example, suppose we have collected M posterior samples from MCMC after burn-ins for the $j$th genetic feature, $X_j$. At the $t$th MCMC iteration, if the corresponding indicator is 1, then the $j$th G factor is included in the final model. The posterior probability of retaining the $j$th genetic feature in the final model is defined as the average of all the indicators for

$X_j$ among the M posterior samples. That is,

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

A higher posterior inclusion probability $p_j$ indicates a stronger empirical evidence that the $j$th genetic effect has a nonzero coefficient, which means a stronger association with the response. For Bayesian methods without spike-and-slab priors, the 95% credible interval (95%CI) is adopted for detecting important predictors. The number of true/false positives (TP/FP) are calculated for identifying non-zero effects. Besides identification, we have also evaluated the estimation performance of all approaches. The posterior medians are used to estimate the coefficients for all Bayesian methods. The estimation accuracy (ESM) is computed as $\sum_{j=1}^{p} \left(\beta_j - \hat{\beta}_j\right)^2$. Here we consider three quantile levels with $\theta = 0.3$, 0.5 and 0.7, separately.

Simulation results for the gene expression data in the first setting are tabulated in Tables 4.1. We can observe that the proposed method has the best performance among all approaches in identification and estimation, especially in the presence of heavy-tailed distributions. For example, in Table 4.1 under error 3 and quantile level $\theta = 0.3$, bqenetss identifies 24.50(sd 0.01) out of 25 true positives, which is much larger than the TPs of all the rest of the methods. Meanwhile, it leads to the lowest number of false positives, 0.07(sd 0.25), and smallest estimation error, 0.06(sd 0.02), among all methods under comparison. In Table 4.1, with $\theta = 0.5$, under error 3, bqenetss has the highest TP while the FP is the lowest among all methods. In this setting, bqenetss identifies 24.46(sd 0.97) out of 25 true positive with the lowest false positive as 0.31(0.48), and the estimation is also the smallest (0.23(sd 0.16)) compared to that of the other approaches.

The advantage of incorporating the spike-and-slab priors can also be observed by taking a close look on other simulation scenarios. For example, in Table 4.1, under quantile level $\theta = 0.5$ and error 4, the TP of bqenetss is 22.29(sd 1.62), which is much higher than TP of 11.86(sd 2.61) from bqenet. The estimation error of bqenetss, 0.95(sd 0.45), is much smaller than that of bqenet (4.53(sd 0.54)) at a cost of slightly higher FPs of 1.81(sd 2.27) compared

**Table 3.1**: *Simulation results of the first setting with i.i.d. errors. n=300, p=400. Mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

| | | bqenetss | bqenet | benet | qbl | blss | bl | enet | lasso |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\theta = 0.3$ | | | | |
| **Error1** | TP | 24.90(0.01) | 23.40(1.34) | 23.60(1.14) | 23.40(0.89) | 24.90(0.01) | 24.60(0.55) | 22.00(1.87) | 21.40(1.95) |
| | FP | 3.00(2.00) | 6.20(5.93) | 46.00(6.56) | 3.60(3.21) | 1.20(1.30) | 12.00(2.35) | 45.80(11.52) | 38.40(11.80) |
| | ESM | 0.24(0.06) | 2.28(0.35) | 7.22(1.27) | 2.01(0.17) | 0.14(0.02) | 2.16(0.17) | 3.12(0.72) | 3.00(0.65) |
| **Error2** | TP | 22.33(2.20) | 12.70(4.16) | 15.43(3.73) | 13.40(3.86) | 14.40(6.81) | 14.57(4.47) | 18.00(4.25) | 17.8(4.09) |
| | FP | 1.87(1.76) | 0.57(0.94) | 82.87(8.98) | 0.83(1.02) | 1.57(1.83) | 23.47(6.93) | 43.70(16.22) | 39.2(17.16) |
| | ESM | 0.90(0.40) | 4.43(0.64) | 42.01(32.75) | 4.05(0.78) | 3.89(3.00) | 16.73(11.16) | 5.42(1.49) | 5.47(1.57) |
| **Error3** | TP | 24.50(0.01) | 17.87(3.38) | 18.40(1.92) | 17.90(3.07) | 19.37(3.97) | 18.50(2.97) | 20.17(3.17) | 19.57(3.00) |
| | FP | 0.07(0.25) | 0.43(0.73) | 74.17(6.84) | 0.30(0.53) | 2.13(2.62) | 20.90(5.45) | 55.70(18.43) | 45.73(15.80) |
| | ESM | 0.06(0.02) | 3.11(0.59) | 21.44(5.27) | 2.47(0.49) | 1.76(1.13) | 8.51(2.49) | 4.38(1.01) | 4.32(1.01) |
| **Error4** | TP | 20.90(1.92) | 12.10(1.75) | 17.37(2.58) | 13.03(1.90) | 16.43(2.42) | 16.53(2.11) | 19.93(2.15) | 19.10(2.32) |
| | FP | 4.07(3.31) | 1.20(1.65) | 78.37(6.03) | 1.60(2.66) | 1.27(1.23) | 24.10(7.45) | 48.40(12.30) | 39.70(16.35) |
| | ESM | 1.48(0.61) | 4.64(0.76) | 27.17(5.04) | 4.73(1.01) | 2.57(0.99) | 11.24(1.70) | 4.69(1.00) | 4.74(1.06) |
| **Error5** | TP | 18.50(2.45) | 8.69(2.59) | 15.31(2.48) | 9.76(2.60) | 12.83(4.58) | 13.59(3.41) | 17.45(4.40) | 17.14(4.55) |
| | FP | 4.96(2.49) | 0.48(0.74) | 84.00(7.71) | 1.03(1.43) | 2.66(0.94) | 23.66(8.45) | 42.76(17.32) | 38.72(17.35) |
| | ESM | 2.66(0.94) | 5.21(0.77) | 45.98(50.04) | 5.42(1.17) | 3.87(1.76) | 17.88(17.43) | 5.31(1.22) | 5.27(1.22) |
| | | | | | $\theta = 0.5$ | | | | |
| **Error1** | TP | 24.73(0.52) | 22.60(1.77) | 22.27(1.17) | 22.7(1.73) | 24.93(0.25) | 23.67(1.30) | 20.93(2.18) | 20.77(2.19) |
| | FP | 1.77(1.14) | 5.30(4.51) | 49.47(7.89) | 5.20(3.78) | 1.00(1.17) | 14.67(4.27) | 56.87(16.03) | 51.53(20.16) |
| | ESM | 0.21(0.09) | 2.59(0.42) | 8.05(1.63) | 2.24(0.35) | 0.14(0.06) | 2.51(0.42) | 3.86(0.84) | 3.78(0.79) |
| **Error2** | TP | 23.62(1.71) | 14.38(4.81) | 16.62(3.62) | 14.46(4.89) | 16.38(6.89) | 15.92(4.79) | 19.23(4.32) | 19.08(4.46) |
| | FP | 0.85(0.80) | 0.54(1.39) | 80.23(10.39) | 0.46(0.88) | 2.08(2.56) | 22.77(6.39) | 52.46(21.44) | 45.23(18.57) |
| | ESM | 0.48(0.28) | 4.02(0.90) | 32.34(17.94) | 3.50(0.61) | 3.08(2.45) | 12.99(6.87) | 5.40(1.62) | 5.37(1.70) |
| **Error3** | TP | 24.46(0.97) | 16.23(3.56) | 18.54(3.02) | 16.46(3.95) | 18.46(5.55) | 17.92(3.86) | 19.38(3.12) | 19.23(2.71) |
| | FP | 0.31(0.48) | 0.62(0.96) | 73.54(7.85) | 0.85(1.21) | 1.54(1.27) | 20.62(7.32) | 53.46(23.76) | 49.85(13.83) |
| | ESM | 0.23(0.16) | 3.40(0.65) | 23.63(13.98) | 2.90(0.72) | 2.12(2.27) | 9.15(4.87) | 4.83(1.42) | 4.60(1.28) |
| **Error4** | TP | 22.29(1.62) | 11.86(2.61) | 16.48(2.23) | 13.10(2.10) | 17.9(3.37) | 16.05(2.25) | 19.52(2.73) | 18.71(2.55) |
| | FP | 1.81(2.27) | 0.81(0.98) | 78.10(7.51) | 1.43(1.78) | 2.14(1.56) | 23.71(6.27) | 58.71(21.19) | 49.19(15.56) |
| | ESM | 0.95(0.45) | 4.53(0.54) | 26.87(4.65) | 4.54(0.97) | 2.55(1.08) | 11.53(1.80) | 5.40(0.76) | 5.29(0.72) |
| **Error5** | TP | 17.83(1.72) | 8.50(2.07) | 13.17(3.06) | 9.00(2.19) | 12.50(4.89) | 13.00(2.90) | 16.17(4.17) | 16.17(4.40) |
| | FP | 3.83(2.79) | 0.50(0.55) | 91.33(6.15) | 1.33(0.52) | 1.83(1.83) | 24.00(5.44) | 38.50(20.03) | 30.50(16.36) |
| | ESM | 2.41(0.73) | 5.27(0.76) | 38.40(5.85) | 5.48(0.76) | 4.34(1.93) | 14.63(1.08) | 5.87(1.07) | 5.81(1.14) |
| | | | | | $\theta = 0.7$ | | | | |
| **Error1** | TP | 24.80(0.41) | 22.80(1.30) | 22.17(1.21) | 22.80(1.16) | 24.97(0.18) | 24.03(0.96) | 21.60(2.44) | 21.23(2.37) |
| | FP | 3.97(2.13) | 4.27(3.00) | 48.10(8.63) | 3.73(2.49) | 1.13(1.50) | 13.30(3.73) | 58.03(22.50) | 49.50(18.46) |
| | ESM | 0.26(0.09) | 2.50(0.41) | 7.93(0.85) | 2.13(0.29) | 0.14(0.06) | 2.43(0.25) | 3.84(1.00) | 3.72(0.99) |
| **Error2** | TP | 22.53(1.93) | 14.00(3.12) | 15.80(3.45) | 14.27(3.48) | 15.60(5.96) | 15.07(4.43) | 18.07(4.59) | 17.83(4.49) |
| | FP | 1.23(1.22) | 0.70(0.88) | 81.23(8.84) | 1.13(1.63) | 1.90(1.45) | 21.87(6.17) | 45.50(21.67) | 39.77(15.48) |
| | ESM | 0.82(0.55) | 4.31(0.61) | 73.29(207.38) | 3.99(0.95) | 3.85(4.23) | 22.92(50.75) | 5.32(1.29) | 5.26(1.35) |
| **Error3** | TP | 22.27(1.78) | 15.93(3.61) | 18.53(2.62) | 15.83(3.42) | 20.07(3.50) | 18.87(3.22) | 20.60(2.37) | 0.23(2.21) |
| | FP | 4.60(4.10) | 2.07(2.07) | 73.83(8.89) | 3.10(2.34) | 2.33(1.71) | 23.03(7.36) | 56.23(17.58) | 49.10(15.08) |
| | ESM | 1.13(0.54) | 4.43(0.74) | 20.96(7.28) | 4.57(0.75) | 1.72(1.18) | 8.81(3.19) | 4.72(1.03) | 4.64(1.10) |
| **Error4** | TP | 20.70(2.20) | 13.13(3.21) | 17.47(1.93) | 13.30(2.69) | 17.60(3.15) | 16.97(2.24) | 19.80(3.42) | 19.40(3.16) |
| | FP | 3.90(3.38) | 1.13(1.55) | 78.37(8.19) | 1.47(1.22) | 2.00(2.69) | 23.00(6.94) | 45.73(11.29) | 38.53(13.12) |
| | ESM | 1.55(0.60) | 4.65(0.59) | 27.04(4.54) | 4.88(0.65) | 2.50(0.86) | 11.10(1.83) | 4.60(1.05) | 4.61(1.05) |
| **Error5** | TP | 18.68(2.31) | 9.40(3.25) | 15.97(2.33) | 10.13(3.10) | 13.37(5.02) | 14.57(3.20) | 17.10(4.69) | 15.63(5.25) |
| | FP | 4.85(3.07) | 0.37(0.56) | 82.80(5.42) | 1.03(1.25) | 1.13(1.14) | 23.17(7.60) | 42.10(18.17) | 31.90(19.87) |
| | ESM | 2.80(0.87) | 5.40(0.55) | 51.06(76.62) | 5.47(0.64) | 4.28(1.90) | 19.47(25.35) | 5.67(1.20) | 5.92(1.44) |

to 0.81(sd 0.98) of bqenet. In Table 4.1, at the quantile level $\theta = 0.7$ and the heavy-tailed $t(2)$ distribution (error 2), the proposed method outperforms bqenet with a higher TP of

22.53(sd 1.93) and smaller number of FPs and estimation errors.

Simulation results for the 2nd setting based on highly correlated gene expressions data are provided in Table B.1 in the Appendix. We can again observe the superior performance of the proposed method over the alternatives. For example, with the heavy tailed distribution error4 and quantile level $\theta = 0.7$, bqenetss identifies 22.13(sd 1.68) out of 24 TPs while bqenet identifies 19.67(sd 1.83) TPs. The bqenet leads to a FP of 1.53(sd 1.87), larger than the 3.23(sd 2.21) detected by bqenetss.

When dealing with highly correlated data, the advantages of the proposed method can be clearly observed. In the Table B.1 of the second setting under a heavy tailed distribution error4 when $\theta = 0.7$, bqenetss identifies 22.13(sd 1.68) out of 24 TP while bqenet identifies 19.67(sd 1.83). Meanwhile, the FP of bqenetss (1.53(sd 1.87)) is lower than that of bqenet (3.23(sd 2.21)). A more distinct difference is reflected in the estimation error. An inaccuracy of 2.70(sd 2.08) for bqenetss in estimation is much smaller than the 8.28(sd 1.37) of bqenet. Moreover, a comparison between robust and non-robust methods further shows the necessity of developing Bayesian regularized quantile regression methods. For instance, in Table B.1 in the Appendix, under error 4 and quantile level $\theta = 0.5$, bqenetss identifies 22.87(sd 1.04) TP with a tiny fraction of 0.40(sd 0.56) FPs while bqenet identifies 19.70(sd 1.84) TP with a FP 1.80(sd 1.16). However, the non-robust enet has reported a TP of 17.70(sd 2.48) with a much large FP of 78.00(sd 7.00). The proposed bqenetss has also shown clearly advantageous performance over the other two non-robust Bayesian methods, blss and bl.

It is also worthwhile to observe the advantage of bqenetss among all the Bayesian methods with spike-and-slab priors. For example, in Table 4.1, under error 2 and quantile level $\theta = 0.7$, bqenetss identifies 22.53(sd 1.93) TPs with a small FP(1.23(sd 1.22)) while blss has a much smaller TP as 15.60(sd 5.96), with a FP of 1.90(sd 1.45). Such a difference is due to the fact that compared to blss that ignores the inter-releatedness among gene expressions, bqenetss conducts variable selection while accounting for the correlations among predictors. Further more, the estimation error of bqenetss (0.82(sd 0.55)) is also much smaller than that of blss (3.85(sd 4.23)). The difference in estimation error further increases when the correlation becomes stronger in *Setting 2*. Under the same combination of error distributions

and quantile level in Table B.1, bqenetss yields an estimation error of 1.75(sd 1.30) that is much less than 11.88 (sd 9.34) of blss.

The superior performance of Bayesian methods over frequentist methods (enet and lasso) have also been clearly shown under data contamination and heavily distributed errors in simulation. For instance, under the mixture of normal distributions (error 4) and $\theta = 0.5$ in Table 4.1, the proposed method bqenetss has identified 22.29(sd 1.62) TP with a low number of FPs, 1.81(sd 2.27), and estimation error of 0.95(sd 0.45). When using the frequentist regularization methods, enet identifies 19.52(sd 2.73) TP with a large number of FPs (58.71(sd 21.19)) and estimation error, 5.29(sd 0.72). Also, in Table 4.1 under error 1 with $\theta = 0.3$, the TPs of all approaches are comparable. However, the regularization methods enet and lasso have reported much larger number of FPs of 45.80 (sd 11.52) and 38.40 (sd 11.80), respectively.

### 3.3.2 Heterogeneous random errors

We consider the case where the random errors do not follow $i.i.d.$ distributions. The data are simulated from the following data generating model:

$$Y_i = \boldsymbol{C}_i^\top \boldsymbol{\gamma} + \boldsymbol{X}_i^\top \boldsymbol{\beta} + (1 + x_{ki})\epsilon_i, \quad (i = 1, ..., n),$$

where $x_{ki}$ is a randomly chosen predictor for $k = 1, ..., p$. The same approaches of generating heterogeneous random errors have been adopted in published studies on high dimensional quantile regressions (Li and Lin (2010); Wu and Liu (2009)). Here the genetic factors $\boldsymbol{X}_i$ are generated from the two settings in Section 3.3.1. Simulation results for the gene expression data in the first setting are tabulated in Table 3.2 and simulation results for the 2nd setting based on highly correlated gene expressions data are provided in Table B.2 in the Appendix.

The advantage of bqenetss over all the other methods can be observed given different quantile levels and heterogeneous random errors. The performance of the proposed method outperforms all the other approaches at different quantile levels. For instance, in Table

**Table 3.2**: *Simulation results of the first setting with heterogeneous random errors. n=300, p=400. Mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

| | | bqenetss | bqenet | benet | qbl | blss | bl | enet | lasso |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\theta = 0.3$ | | | | |
| **Error1** | TP | 24.77(0.43) | 19.80(2.01) | 20.70(2.25) | 20.00(1.97) | 23.33(1.30) | 21.53(1.87) | 21.33(2.51) | 20.97(2.59) |
| | FP | 0.87(0.82) | 2.53(2.52) | 63.57(8.06) | 2.40(2.43) | 1.27(1.51) | 18.17(4.79) | 56.33(18.65) | 50.47(19.63) |
| | ESM | 0.28(0.11) | 3.47(0.56) | 13.66(2.31) | 3.26(0.61) | 0.79(0.28) | 5.33(0.87) | 4.20(0.92) | 4.16(0.84) |
| **Error2** | TP | 18.77(5.36) | 7.23(4.68) | 11.93(3.93) | 8.13(4.70) | 8.13(6.92) | 10.20(4.74) | 12.87(6.49) | 12.67(6.52) |
| | FP | 1.40(1.48) | 0.27(0.69) | 87.20(6.30) | 0.33(0.92) | 0.93(1.39) | 24.43(5.73) | 33.23(23.60) | 30.47(22.24) |
| | ESM | 2.21(2.23) | 5.29(1.28) | 26.56(72.5) | 4.72(0.83) | 7.55(6.06) | 93.18(25.55) | 7.01(1.74) | 7.09(1.76) |
| **Error3** | TP | 24.87(0.01) | 14.78(2.92) | 15.50(3.87) | 14.50(2.94) | 13.50(6.37) | 13.89(4.84) | 17.94(2.82) | 17.28(3.58) |
| | FP | 0.06(0.24) | 0.22(0.43) | 81.06(8.75) | 0.28(0.46) | 1.28(1.64) | 24.89(6.67) | 45.39(13.18) | 41.89(17.58) |
| | ESM | 0.17(0.11) | 3.66(0.46) | 45.36(33.03) | 3.22(0.51) | 5.47(3.78) | 19.89(13.74) | 6.00(1.51) | 6.21(1.57) |
| **Error4** | TP | 18.70(2.71) | 7.67(2.67) | 13.97(2.76) | 8.43(2.57) | 8.63(2.95) | 12.07(3.07) | 16.67(2.83) | 16.37(3.01) |
| | FP | 2.67(3.94) | 0.43(0.68) | 85.30(7.64) | 0.67(0.96) | 1.23(1.43) | 24.53(7.56) | 40.83(15.41) | 34.57(13.92) |
| | ESM | 2.26(0.99) | 5.30(0.81) | 52.94(14.95) | 5.69(0.91) | 7.23(2.01) | 21.94(5.99) | 6.05(1.17) | 6.14(1.28) |
| **Error5** | TP | 18.77(1.79) | 16.73(2.64) | 15.60(2.71) | 14.00(3.18) | 15.63(3.26) | 15.00(3.60) | 21.77(3.39) | 21.00(3.86) |
| | FP | 2.47(2.29) | 2.03(1.47) | 83.33(8.05) | 1.07(1.23) | 1.63(1.33) | 9.07(3.86) | 39.17(12.01) | 36.60(15.00) |
| | ESM | 6.30(2.29) | 8.34(1.03) | 82.61(540.69) | 8.06(1.42) | 9.54(3.09) | 31.60(67.20) | 7.16(2.50) | 7.82(2.72) |
| | | | | | $\theta = 0.5$ | | | | |
| **Error1** | TP | 24.40(0.93) | 19.90(2.28) | 20.77(1.92) | 19.80(2.25) | 23.40(1.40) | 21.60(1.59) | 20.80(2.61) | 20.40(2.65) |
| | FP | 1.33(1.63) | 1.83(1.12) | 62.07(7.51) | 2.03(1.81) | 2.07(1.66) | 17.63(5.07) | 55.07(19.10) | 48.63(16.94) |
| | ESM | 0.27(0.21) | 3.21(0.49) | 12.40(1.82) | 2.92(0.49) | 0.58(0.31) | 4.73(0.93) | 4.20(0.84) | 4.14(0.96) |
| **Error2** | TP | 21.47(3.20) | 8.43(4.10) | 13.00(3.97) | 9.27(4.13) | 7.80(6.52) | 10.67(5.29) | 14.30(6.36) | 14.03(5.94) |
| | FP | 0.17(0.46) | 0.07(0.25) | 87.23(7.82) | 0.03(0.18) | 0.80(1.03) | 23.03(8.28) | 30.03(18.35) | 28.43(17.31) |
| | ESM | 0.93(0.91) | 4.73(0.99) | 103.64(19.85) | 3.98(0.78) | 6.55(2.66) | 44.38(10.30) | 6.58(1.57) | 6.55(1.58) |
| **Error3** | TP | 24.10(1.40) | 13.03(3.99) | 14.80(3.70) | 13.01(3.52) | 13.03(6.17) | 13.53(3.86) | 18.77(3.45) | 18.43(3.33) |
| | FP | 0.13(0.35) | 0.20(0.48) | 82.93(7.28) | 0.30(0.84) | 1.17(1.18) | 24.57(5.73) | 52.33(24.36) | 41.07(10.55) |
| | ESM | 0.27(0.26) | 3.93(0.63) | 43.20(22.27) | 3.41(0.55) | 4.76(2.71) | 18.61(9.65) | 5.58(1.18) | 5.44(1.21) |
| **Error4** | TP | 19.60(2.59) | 6.67(3.03) | 13.47(2.85) | 8.50(2.45) | 8.73(3.79) | 11.60(2.53) | 16.63(3.78) | 16.17(3.76) |
| | FP | 1.00(1.20) | 0.10(0.31) | 86.53(5.79) | 0.37(0.56) | 1.37(1.16) | 21.80(6.45) | 42.63(18.75) | 38.77(18.83) |
| | ESM | 1.51(0.76) | 5.18(0.72) | 50.27(12.62) | 4.64(0.69) | 6.04(1.72) | 18.89(3.99) | 6.25(1.21) | 6.35(1.34) |
| **Error5** | TP | 20.17(1.80) | 17.77(1.63) | 16.17(2.00) | 14.17(1.91) | 15.90(2.73) | 15.77(2.31) | 22.63(1.13) | 21.83(1.58) |
| | FP | 1.53(1.17) | 1.93(1.36) | 82.83(6.86) | 1.00(1.11) | 1.93(1.28) | 10.93(4.26) | 44.77(14.32) | 39.90(1.61) |
| | ESM | 4.39(1.98) | 7.91(0.97) | 76.15(14.02) | 7.29(1.19) | 9.87(4.44) | 18.81(3.70) | 6.55(1.80) | 7.25(2.40) |
| | | | | | $\theta = 0.7$ | | | | |
| **Error1** | TP | 24.08(1.02) | 19.19(2.17) | 20.31(1.67) | 19.46(2.20) | 22.81(1.36) | 21.23(1.95) | 19.77(2.29) | 19.27(2.65) |
| | FP | 1.31(1.16) | 2.23(2.29) | 64.38(6.38) | 2.46(1.90) | 1.69(1.41) | 18.69(5.28) | 50.69(15.30) | 42.31(16.33) |
| | ESM | 0.40(0.23) | 3.44(0.53) | 12.48(1.89) | 3.20(0.44) | 0.75(0.26) | 4.88(0.73) | 4.29(1.02) | 4.23(1.09) |
| **Error2** | TP | 19.13(2.61) | 7.53(3.48) | 12.10(3.23) | 8.07(2.86) | 5.73(4.27) | 9.80(3.34) | 13.13(3.97) | 12.23(4.54) |
| | FP | 1.03(1.54) | 0.23(0.43) | 90.80(6.52) | 0.37(0.56) | 0.63(0.89) | 22.93(7.64) | 31.63(15.35) | 26.20(15.78) |
| | ESM | 1.84(0.99) | 5.38(0.69) | 15.58(15.62) | 4.88(0.68) | 7.53(2.17) | 39.48(58.50) | 6.83(1.30) | 6.92(1.40) |
| **Error3** | TP | 19.50(4.30) | 10.09(4.71) | 15.09(4.08) | 10.68(4.11) | 12.00(6.34) | 13.73(4.48) | 16.73(5.78) | 16.64(5.03) |
| | FP | 1.50(2.06) | 0.55(0.91) | 81.59(9.76) | 0.82(1.50) | 1.14(1.42) | 23.64(5.88) | 43.14(17.95) | 39.64(15.90) |
| | ESM | 1.56(1.45) | 4.83(1.04) | 80.91(18.34) | 4.61(1.01) | 4.58(2.54) | 30.43(69.31) | 5.69(1.61) | 5.61(1.41) |
| **Error4** | TP | 17.80(2.46) | 6.97(2.98) | 12.60(3.01) | 7.37(2.34) | 7.40(3.76) | 10.37(2.58) | 15.33(2.91) | 14.67(3.29) |
| | FP | 2.07(2.42) | 0.27(0.58) | 85.73(5.81) | 0.73(1.31) | 0.73(1.23) | 24.40(8.78) | 37.23(17.05) | 29.87(11.93) |
| | ESM | 2.37(1.00) | 5.41(0.83) | 53.81(13.98) | 5.64(1.04) | 6.58(1.79) | 21.08(5.20) | 6.25(1.10) | 6.20(1.17) |
| **Error5** | TP | 19.03(2.25) | 17.37(1.87) | 16.43(2.88) | 15.27(2.08) | 16.00(2.68) | 15.33(2.5) | 22.73(1.14) | 21.90(1.32) |
| | FP | 2.67(1.40) | 2.07(1.62) | 81.50(5.79) | 1.30(1.18) | 1.40(1.19) | 10.93(5.09) | 43.13(13.81) | 33.83(13.16) |
| | ESM | 5.75(2.79) | 8.57(1.11) | 79.43(21.58) | 8.00(1.55) | 9.17(4.13) | 19.21(4.29) | 6.47(1.69) | 7.08(2.06) |

3.2, under the heterogeneous log normal distribution (Error 3) and quantile level $\theta = 0.3$, bqenetss has identified the largest TP of 24.87(sd 0.01), with the smallest FP of 0.06(sd 0.24)

and the lowest estimation error of 0.17(sd 0.11) among all the methods under comparison. Besides, in Table 3.2 under the same error distribution (Error 3), at quantile level $\theta = 0.5$, the TP of bqenetss is 24.10(sd 1.40), which is much larger than the TP of all the rest of the methods. Meanwhile, the FP (0.13(sd 0.35)) and estimation error (0.27(sd 0.26)) of bqenetss are the smallest compared to all the other methods.

To further confirm the superiority of incorporating the spike-and-slab prior under the data heterogeneity, we take a closer look over the performance of bqenet which is the major competitor of the proposed bqenetss. In Table 3.2, under quantile level $\theta = 0.5$ and heterogeneous $t(2)$ error (Error 2), although both methods yield a small fraction of false positive findings, bqenetss has discovered 21.47(sd 3.20) TPs while bqenet has merely detected 8.43(sd 4.10) TPs. In addition, the resulting estimation error by using bqenetss is 0.93(sd 0.91), much smaller than 4.73 (sd 0.99) reported by bqenet. In the Appendix, Table B.2 lists the identification and estimation results according to *Setting 2* where genetic factors are simulated cluster-wisely with strong correlations. When $\theta = 0.5$, under error 5, bqenetss identifies 20.21(sd 2.14) TP out of 25 true signals with a low value in FP (1.41 (sd 1.32)) while bqenet identifies a much lower TP as 18.00(sd 1.83).

Under heterogeneous model errors, Bayesian methods have generally outperformed the frequentist enet and lasso. For example, Table 3.2 shows that under heterogeneous normal error (Error 1) and quantile level $\theta = 0.7$, the numbers of TP and FP corresponding to bqenetss are 24.08(sd 1.02) and 1.31 (sd 1.16), respectively. The bqenetss also leads to a small estimation error of 0.40 (sd 0.23). The TPs of enet and lasso are 19.77(sd 2.29) and 19.27(sd 2.65), correspondingly, which are inferior to the TPs identified by bqenetss. Moreover, both frequentist methods have produced a much higher number of FPs, which are 50.69(sd 15.30) for enet and 42.31(sd 16.33) for lasso, and larger estimation errors 4.29(sd1.02) and 4.23(sd 1.09), respectively. In Table B.2 in the Appendix, although TPs reported by enet and lasso have increased to 23.07(sd0.74) and 22.70(sd 1.18) respectively, which are close to 23.73(sd 0.52) identified by bqenetss, the number of FPs, 43.20(sd 12.88) for enet and 38.33(sd 13.58) for lasso, are still much larger that of bqenetss, 0.60(sd 0.93).

The advantages of conducting robust Bayesian analysis using the proposed approach can

be justified based on the results of comprehensive simulation studies. The convergence of the MCMC chains with the potential scale reduction factor (PSRF) (Brooks and Gelman (1998)) has been conducted. In this study, we use PSRF $\leq 1.1$ (Gelman et al. (2004)) as the cut-off point which indicates that chains converge to a stationary distribution. The convergence of chains after burn-ins has been checked for all parameters with the value of PSRF less than 1.1. Figure B.1 shows the convergence pattern of PSRF for some randomly chosen genetic coefficients in setting 1 under Error 2.

## 3.4 Real Data Analysis

Nurses' Health Study (NHS) is a well-characterized cohort study involving female subjects for whom the lifestyle and dietary phenotypic data, as well as the SNP genotype data are available (Colditz and Hankinson (2005); Colditz et al. (1997)). We focus on SNPs on chromosome 10 to identify important genetic factors that are potentially associated with body mass index, an important phenotypic trait related to obesity. Specifically, we consider the body mass index (BMI) as response variable, and age, total physical activity (act), trans fat intake (trans) and cereal fiber intake (ceraf) as four clinical covariates. Data are available on 3391 subjects and 17016 SNPs after matching phenotypes with genotypes, and removing SNPs that are rare variants with minor allele frequency (MAF) less than 0.05 or deviates from Hardy–Weinberg equilibrium. We use the 1804 healthy subjects for downstream analysis. Following published literature adopting regularization methods (including the elastic net) to analyze GWAS data (Cho et al. (2009); Waldmann et al. (2013); Wu et al. (2014)), we first conduct a pre-screening of SNPs to reduce the size of the working dataset to a reasonable scale. 800 SNPs with p-value less than 0.05 have been obtained by filtering through the marginal model where each SNP has been fitted to BMI individually after accounting for the clinical covariates. The final working data set consists has a sample size of 1804, 800 SNPs and 4 clinical covariates.

All the 8 methods under comparison have been adopted to analyze the data. For quantile regression based methods, we set the quantile level to $\theta = 0.5$. Table 3.3 presents the pairwise

overlap of SNPs, indicating that the proposed bqenetss results in identification of a different set of SNPs compared to alternative methods. Moreover, the frequentist methods, enet and lasso, have reported a much larger number of susceptibility SNPs than their bayesian counterparts. Note that while enet and lasso have only provided penalized estimates of the effects, SNPs discovered by Bayesian methods are all associated with significance measures, such as the posterior inclusion probability and 95% credible intervals for methods with and without incorporating spike-and-slab priors, respectively.

Table B.3 in the Appendix lists important findings corresponding to the proposed bqenetss. Among those SNPs, some have not been found by the alternative approaches. For example, four SNPs, rs983814, rs12255028, rs11815954 and rs7067773, which are located within gene ASAH2, have been identified. ASAH2 has been reported as a crucial survival factor in human immune system and it could reduce accumulation in colon tumors (Kurt et al. (2023); Zhu et al. (2021)). ASAH2 can also increase the stability in certain proteins which may suppress lipid reactive oxygen prodiction(Zhu et al. (2021)). In addition, SNPs rs41408444, rs4748991, rs1886995 and rs11014290, located within gene PRTFDC1, have been observed to be related with ovarian cancer development (Cai et al. (2007)). Furthermore, studies show that PRTFDC1 has a close relation with the unfavorable prognosis of 4 cancers (Xu and Guo (2021)). Also, it has been identified as a potential tumor gene in cancer cells (Cai et al. (2007); Suzuki et al. (2007)). Our proposed method has identified LINC00707. As an important marker for diagnosis, treatment and diseases' prognosis in biological studies, LINC00707 has a close association with clinical manifestations of disease like tumor size, tumor stage, overall survival times, drug sensitivity and other factors (Yao et al. (2022)). LINC00707 has also been reported to be expressed in multiple disease types, especially in multiple types of cancer (Yao et al. (2022)). All these findings indicate that our proposed method is useful in identifying SNPs with sensible biological implications.

In the Appendix, Table B.4 – Table B.8 present the identified SNPs from all the other 5 Bayesian methods. For the methods without incorporating the spike-and-slab priors, that are bqenet, benet, qbl and bl, the SNPs are selected using the 95% credible intervals, in Table B.4, Table B.5, Table B.6 and Table B.8, respectively. Together with Table B.3,

we can observe low similarity of findings between robust and non-robust methods, as well as between methods with and without spike-and-slab priors. The findings from enet and lasso are omitted since simulation study indicates both result in much larger number of false positives, and focus on penalized estimation without uncertainty quantification.

We have evaluated the prediction performance of all approaches using random-split of the data. At each round, the working data has been randomly split into a training and testing set, with three-fourth and one-fourth of the subjects, respectively. The prediction accuracy in mean and standard deviations can then be assessed over 100 splits. For a fair comparison between robust and non-robust methods, both the mean squared error (MSE) and least absolute deviation (LAD) have been computed. Table 3.4 shows the prediction errors for all methods under comparison in both criteria. Here, it can be found that the proposed method bqenetss has the best prediction performance with the smallest MSE (14.91(sd 0.78)) and the LAD (2.88(sd 0.09)), due to its robustness against the heavy-tailed distributions and accommodation of linkage disequilibrium. We can observe that robust methods have a better performance in prediction than non-robust ones. For instance, the MSE of bqenetss is 14.91(0.78) and the MSE of bqenet is 17.69(1.21), which are smaller than that of benet (20.57(1.71)). In addition, the MSE of qbl (19.83(1.78)) is smaller than those of blss (23.67(2.67)) and bl(22.06(1.57)). Similar results can be found when comparing the prediction performance with LAD between quantile-based methods and non-quantile methods. This indicates that developing the quantile Bayesian methods is necessary. It can be found that the prediction performance of quantile elastic methods is better than quantile lasso. The LAD of bqenetss is 2.88(0.09) and the LAD of bqenet is 2.97(0.12), while the LAD of qbl is as large as 3.38(0.13). This is because that the quantile elastic methods can accommodate the correlation among genetic variants.The superiority of the methods incorporating spike-and-slab priors can be clearly observed in the prediction results. The MSE of bqenetss is 14.91(0.78), which is much smaller than that of bqenet (17.69(1.21)). When comparing the performance of Bayesian methods with frequentist methods, it shows that all these frequentist methods have worse performance with high prediction errors. The MSE and LAD of elastic net are 66.01(8.91) and 25.29(0.15), and the MSE and LAD of frequentist

LASSO are 65.97(8.51) and 25.27(0.16), which are much higher than those of bqenetss.

**Table 3.3**: *The number of SNPs identified by different methods and their overlaps.*

| **T2D** | bqenetss | bqenet | benet | qbl | blss | bl | enet | lasso |
|---|---|---|---|---|---|---|---|---|
| bqenetss | 20 | 3 | 3 | 1 | 3 | 4 | 14 | 15 |
| bqenet | | 12 | 6 | 2 | 2 | 5 | 8 | 8 |
| benet | | | 25 | 4 | 3 | 16 | 17 | 17 |
| qbl | | | | 6 | 1 | 5 | 2 | 2 |
| blss | | | | | 4 | 3 | 3 | 3 |
| bl | | | | | | 24 | 12 | 12 |
| enet | | | | | | | 123 | 106 |
| lasso | | | | | | | | 117 |

**Table 3.4**: *Prediction performance in terms of the mean squared error (MSE) and least absolute deviation (LAD).*

| **Criteria** | bqenetss | bqenet | benet | qbl | blss | bl | enet | lasso |
|---|---|---|---|---|---|---|---|---|
| MSE | 14.91(0.78) | 17.69(1.21) | 20.57(1.71) | 19.83(1.78) | 23.67(2.67) | 22.06(1.57) | 66.01(8.91) | 65.97(8.51) |
| LAD | 2.88(0.09) | 2.97(0.12) | 3.51(0.15) | 3.38(0.13) | 3.63(0.14) | 3.66(0.12) | 25.29(0.15) | 25.27(0.16) |

## 3.5 Discussion

In this study, we have developed the Bayesian quantile elastic net with spike-and-slab priors that can accommodate structured sparsity in the presence of outliers and long-tailed distributions in the phenotypic trait. In recent robust sparse Bayesian analysis, elicit priors as a composition of structured shrinkage priors and spike-and-slab priors to simultaneously accommodate underlying patterns in omics data and impose exact sparsity has been demonstrated as an effective strategy to overcome the limitations in lacking exact 0 estimates in existing Bayesian analysis (Ren et al. (2023); Zhou et al. (2023)).

The proposed work can be extended in multiple dimensions. For example, in cancer genomics studies, high-dimensional survival analysis has been extensively conducted to detect important prognostic features (Jiang et al. (2017); Stingo et al. (2011); Tang et al. (2017, 2019)). The proposed method can be generalized to survival traits, as published studies have already shown the advantage of incorporating robustness in survival analysis (Ren et al. (2019); Sha et al. (2006)). Our method can also be extended within the mixed model framework to handle longitudinal responses. We will postpone the investigations to future studies.

# Chapter 4

# Robust Bayesian Subgroup Analysis

## 4.1 Introduction

Subgroup analysis has received intensive attention in biomedical research, especially in detecting group structure in medical studies and making treatment designs. Recently, subgroup analysis through regularizations have attracted much attention (Ma and Huang (2017)) and its robust extension through penalized quantile regression have also been considered (Y. et al. (2019)). Despite success, a major limitation is that these frequentist approaches for subgroup identification cannot handle high-dimensional data, and lack inferential procedures. Bayesian subgroup method has been developed to overcome these difficulties (Li and Zhu (2023)). However, another challenge for exisiting methods comes from the widely observed heavy-tailed distributions and outliers in response varibles. For example, consider the BRCA1 from The Cancer Genome Atlas (TCGA) breast cancer data, the heavy right tails for the response variable can be clearly observed in Frigure 4.1. However, as most of the existing subgroup methods including frequentist methods and Bayesian methods, are based on the least square loss and cannot accommodate heavy tails and outliers, a robust Bayesian subgroup method is in high demand (Cheng et al. (2022)).

In this paper, we propose a robust Bayesian subgroup approach which can conduct variable selection, identify subgroup structures of subjects and estimate the coefficients at the

same time. The approach is under the similar framework as (Ma and Huang (2017)) that detect subgroup structure through a pairwise fusion penalty. The Laplace prior(Andrews and Mallows (1974)) has been adpoted in a robust Bayesian hierarchical (Li et al. (2010)) model to provide full conditional poterior distributions. To improve computation accuracy in shrinkage via inducing exact sparsity, we incorporate the spike-and-slab priors (George and McCulloch (1993), Ishwaran and Rao (2005)) within the robust Bayesian framework through an efficient Gibbs sampling algorithm in Markov Chain Monte Carlo(MCMC). The advantages of the proposed method over the alternatives have been demonstrated through the simulation studies. The analysis of real breast cancer data from The Cancer Genome Atlas (TCGA) has also revealed the superiority of the proposed method.

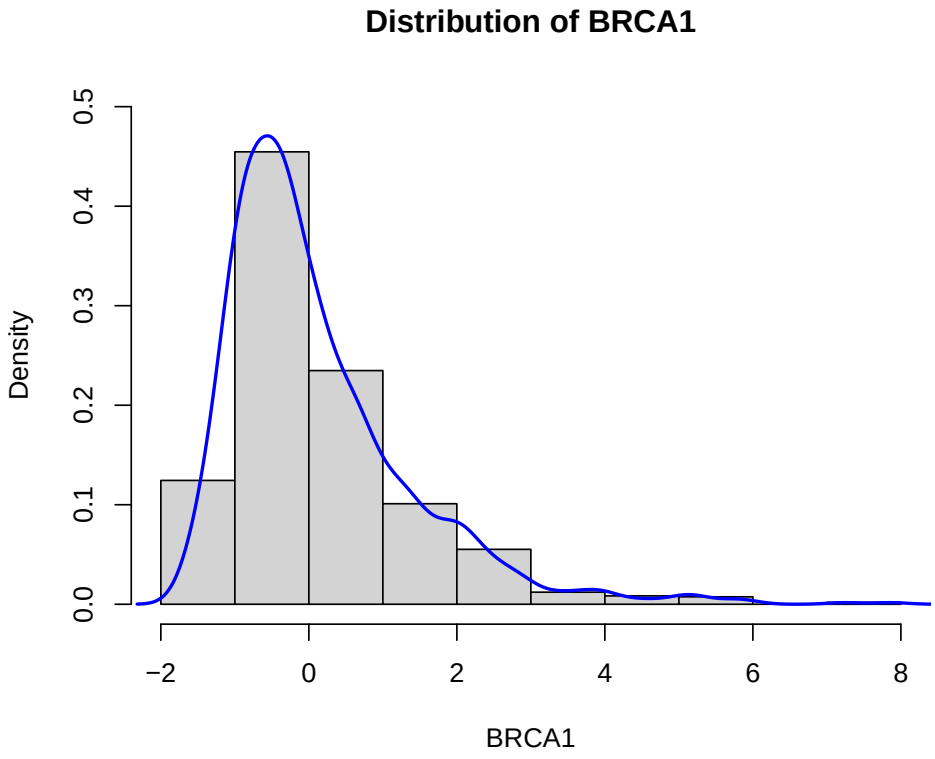


**Figure 4.1**: *Distribution of BRCA1 from NHS data*

## 4.2 Model Settings

We denote $(Y_i, \boldsymbol{\mu}_i, \mathbf{X}_i)$ as independent and identically distributed random vectors, where the subscript $i$ refers to the $i$th subject $(i = 1, \ldots, n)$. In the random vector, $Y_i$ is a continuous response variable representing the cancer outcome or disease phenotype and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ is the $n$ unknown subject-specific intercepts. Additionally, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^\top$ represents the $p$ genetic variants such as gene expressions and single nucleotide polymorphisms (SNPs).Consider the following model:

$$Y_i = \mu_i + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i, \tag{4.1}$$

where $\beta_j$'s are the regression coefficients of the genetic variants. Denote $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)^T$. Then model (4.1) can be written as

$$Y_i = \mu_i + \mathbf{X}_i^\top \boldsymbol{\beta} + \epsilon_i. \tag{4.2}$$

### 4.2.1 Bayesian formulation of the quantile regression

In frequentist quantile regression, minimizing the following problem leads to the estimation of the regression coefficients with quantile level $\theta$ $(0 < \theta < 1)$:

$$\min_{\boldsymbol{\gamma}, \boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\theta(Y_i - \mu_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}), \tag{4.3}$$

where the quantile check loss function is defined as $\rho_\theta(\epsilon_i) = \epsilon_i\{\theta - I(\epsilon_i < 0)\}$. In Bayesian quantile regression, with the assumption that the random error $\epsilon_i$'s following the asymmetric

Laplace distribution with density (Yu and Moyeed (2001); Yu and Zhang (2005)):

$$\begin{aligned} f(\epsilon_i|\tau) =& \theta(1-\theta)\tau \exp(-\tau\rho_\theta(\epsilon_i)), \\ =& \theta(1-\theta)\tau \begin{cases} e^{-\tau\theta\epsilon_i}, & \text{if } \epsilon_i \geq 0 \\ e^{\tau(1-\theta)\epsilon_i}, & \text{if } \epsilon_i < 0, \end{cases} \end{aligned}$$

where the scale parameter $\frac{1}{\tau}$ determines the skewness of the asymmetric Laplace distribution (ALD), the likelihood function can be obtained as:

$$f(\boldsymbol{Y}|\mu_i, \boldsymbol{X}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \tau) = \theta^n(1-\theta)^n\tau^n \exp\big(-\tau\sum_{i=1}^{n}\rho_\theta(Y_i - \mu_i - \boldsymbol{X}_i^\top\boldsymbol{\beta})\big),$$

where $\boldsymbol{Y} = (Y_1, ..., Y_n)^\top$, $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)^\top$ and $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_n)^\top$. Then maximizing the likelihood function to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ are corresponding to minimizing the quantile check loss in (4.3). In literature, the asymmetric Laplace distribution can be equivalently specified as a mixture of exponential and scaled normal distribution as follows (Kozumi and Kobayashi (2011)):

$$\epsilon_i = \tau^{-1}\xi_1\tilde{v}_i + \tau^{-1}\xi_2\sqrt{\tilde{v}_i}z_i,$$

where $\xi_1 = \frac{1-2\theta}{\theta(1-\theta)}$, $\xi_2 = \sqrt{\frac{2}{\theta(1-\theta)}}$, $v_i = \tau^{-1}\tilde{v}_i \sim \text{Exp}(\tau^{-1})$ and $z_i \sim N(0,1)$. Then we can have the following hierarchical model:

$$Y_i = \mu_i + \boldsymbol{X}_i^\top\boldsymbol{\beta} + \xi_1 v_i + \tau^{-1/2}\xi_2\sqrt{v_i}z_i,$$

$$v_1, ..., v_n \sim \prod_{i=1}^{n}\tau\exp(-\tau v_i),$$

$$z_1, ..., z_n \sim \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}z_i^2).$$

### 4.2.2 Bayesian quantile subgroup with spike-and-slab priors

Subgroup analysis with a pairwire fusion structure in frequentist approach is implemented by Ma and Huang (2017) to identify subgroups of heterogeneous population for individualized treatment. Through imposing a concave pairwise fusion penalty function $p(\cdot, \lambda)$ on the subject-specific intercepts $\mu_i$'s from the following objective function:

$$Q_n(\mu, \beta; \lambda) = \frac{1}{2}\sum_{i=1}^{n}(Y_i - \mu_i - \boldsymbol{X}_i^\top \boldsymbol{\beta})^2 + \sum_{1\le i<j\le n} p(|\mu_i - \mu_j|, \lambda),$$

dividing the subjects into cetain subgroups can be obtained. Caused by various factors such as data errors, mis-reporting, mistakes from sampling, false assumption and others (Osborne and Overbay (2004)), heavy-tailed distributions and outliers in the response variable have been commonly observed in heterogeneous data analysis in practical cancer studies. To overcome this difficulty and conduct subgroup identification and variable selection simultaneously, Cheng et al. (2022) propose the following robust penalized objective function:

$$\sum_{i=1}^{n}\rho(Y_i - \mu_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}) + \lambda_1 \sum_{1\le i<j\le n} |\mu_i - \mu_j| + \lambda_2 \sum_{j=1}^{p} |\beta_j|,$$

where $\rho(\cdot)$ is the robust loss function, and $\lambda_1$ and $\lambda_2$ are positive tuning parameters.

As inference procedure is not always available for penalization methods, Bayesian subgroup analysis of the heterogeneous population is introduced by Li and Zhu (2023). To deal with the heavy-tailed distributions and outliers of the response variable within the Bayesian framework, here we propose a robust Bayesian subgroup function as below:

$$\sum_{i=1}^{n}\rho_\theta(Y_i - \mu_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}) + \lambda_1 \sum_{1\le i<j\le n} \omega_{ij}|\mu_i - \mu_j| + \lambda_2 \sum_{j=1}^{p} |\beta_j|,$$

where $\rho(\cdot)_\theta$ is the check loss function and $\theta$ is the quantile level. A conditional Laplace prior

is imposed on the intercept $\boldsymbol{\mu}$ as:

$$\pi(\boldsymbol{\mu}|\sigma^2) \propto \exp\{-\tau\lambda_1 \sum_{1\leq i<j\leq n} \omega_{ij}|\mu_i - \mu_j|\},$$

where $\lambda_1$ is the hyperparameter and $\omega_{ij}$'s are the weights. The weights $\omega_{ij}$ is defined as: $\omega_{ij} = \frac{2}{n-1}\exp\{-\phi(\hat{\mu_i} - \hat{\mu_j})\}$ (Ma and Huang (2017)), where $\phi$ is a nonnegative constant with $\phi = 0$ when $\mu_i = \mu_j$. $\hat{\mu_i}$ is estimated with $y_i - \boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the estimated value with smoothly clipped absolute deviation penalty (SCAD). Let $\eta_1 = \tau\lambda_1$, a conditional Laplace prior on $\boldsymbol{\mu}$ can be expressed as a scale mixture of normal distribution and an exponential distribution as below:

$$\begin{aligned}
\boldsymbol{\mu}|\varphi_{ij}^2 &\sim \mathrm{N}_n(0, \Sigma_\mu), \\
\varphi_{ij}^2 &\sim \frac{\eta_1^2}{2}\exp(-\frac{\eta_1^2}{2}\varphi_{ij}^2), \\
\eta_1^2 &\sim \mathrm{Gamma}(c_1, d_1),
\end{aligned}$$

where

$$\Sigma_\mu^{-1} = \begin{cases} \sum_{1\leq k<i} \frac{\omega_{ki}^2}{\varphi_{ki}^2} + \sum_{i<k\leq n} \frac{\omega_{ik}^2}{\varphi_{ik}^2}, & \text{if } i = j. \\ -\frac{\omega_{ij}^2}{\varphi_{ij}^2}, & \text{if } i < j. \\ -\frac{\omega_{ji}^2}{\varphi_{ji}^2}, & \text{if } i > j. \end{cases}$$

Li and Zhu (2023) use the noninformative marginal prior $\pi(\beta) = 1$ on $\beta$ which leads to a multivariate normal posterior distribution of $\beta$. However, it cannot yield exact 0 of estimating regression coefficients. Li et al. (2010) have proposed to adopt the 95% credible intervals and the scaled neighborhood criterion for identification of sparse signals. However, the 95% credible intervals has been found too wide in general to include important signals (Li et al. (2010)) and a probability threshold is required to determine whether the predictor is included or not for the scaled neighborhood criterion.

To improve variable selection accuracy by imposing the exact sparsity, George and McCulloch (1993); Ishwaran and Rao (2005) have developed and widely used the spike-and-slab priors. In this study, we have considered incorporating the spike-and-slab priors in Bayesian

quantile subgroup and Bayesian subgroup to overcome the limitations in Li and Zhu (2023).

Specifically, let $\eta_2 = \tau\lambda_2$, we consider the following Bayesian hierarchical model to incorporate the spike-and-slab priors:

$$\begin{aligned}
\beta_j|s_j, \pi_j &\overset{iid}{\sim} (1-\pi_j)\mathrm{N}(0, s_j) + \pi_j\delta_0(\beta_j) \quad j = 1, \ldots, p \\
\pi_j|\psi_j &\overset{iid}{\sim} \text{Bernoulli}(\psi_j), \\
s_j|\eta_2^2 &\overset{iid}{\sim} \frac{\eta_2^2}{2}\exp(-\frac{\eta_2^2}{2}s_j), \\
\eta_2^2 &\sim \text{Gamma}(c_2, d_2),
\end{aligned} \tag{4.4}$$

where $\delta_0(\beta_j)$ denotes the spike component of the prior with a point mass at 0 and the normal distribution, $\mathrm{N}(0, s_j)$ is representing the slab component. For each $\beta_j$ measuring the genetic contribution of the $j$th genetic feature to the variation in disease phenotype, a latent binary indicator $\psi_j$ is used in variable selection with exact sparsity. In particular, as the spike component is a point mass at 0, $\beta_j$ will be shrunk to exact 0 with $\psi_j = 1$ and the $j$th genetic factor is not included in the final model. Otherwise, if $\psi_j = 0$, $\beta_j$ is generated from non-zero slab componenet with the normal distribution which does not directly yield a zero estimate, meaning the $j$th genetic factor is associated with the response. Here, we set $\pi_j \in [0, 1]$. Instead of setting $\pi_j$ with a non-informative prior such as 0.5, a conjugate beta prior as $\pi_j \sim \text{Beta}(r_1, u_1)$ is given which accounts for the uncertainty in $\pi_j$. In this paper, we choose $r_1 = u_1 = 1$ as it gives a prior mean with 0.5 and it also allows a prior to spread out.

### 4.2.3 The Gibbs sampler for quantile subgroup with spike-and-slab priors (qsubgroupss)

$$\sum_{i=1}^{n} \rho_\theta(Y_i - \mu_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}) + \lambda_1 \sum_{1 \le i < j \le n} \omega_{ij}|\mu_i - \mu_j| + \lambda_2 \sum_{j=1}^{p} |\beta_j|. \tag{4.5}$$

**Hierarchical model specification**

$$Y_i = \mu_i + \boldsymbol{X}_i^\top \boldsymbol{\beta} + \xi_1 v_i + \tau^{-1/2}\xi_2\sqrt{v_i}z_i \quad i = 1, \ldots, n$$

$$v_i|\tau \overset{iid}{\sim} \tau\exp(-\tau v_i) \quad i = 1, \ldots, n$$

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

$$\beta_j|s_j, \pi_1 \overset{iid}{\sim} (1-\pi_1)\mathrm{N}(0, s_j) + \pi_1\delta_0(\beta_j) \quad j = 1, \ldots, p$$

$$s_j|\eta_2^2 \overset{iid}{\sim} \frac{\eta_2^2}{2}\exp(-\frac{\eta_2^2}{2}s_j) \quad j = 1, \ldots, p$$

$$\tau \sim \mathrm{Gamma}(a, b)$$

$$\eta_2^2 \sim \mathrm{Gamma}(c_2, d_2)$$

$$\pi_1 \sim \mathrm{Beta}(r_1, u_1)$$

$$\boldsymbol{\mu}|\varphi_{ij}^2 \sim \mathrm{N}_n(0, \Sigma_\mu)$$

$$\varphi_{ij}^2 \sim \frac{\eta_1^2}{2}\exp(-\frac{\eta_1^2}{2}\varphi_{ij}^2)$$

$$\eta_1^2 \sim \mathrm{Gamma}(c_1, d_1)$$

where

$$\Sigma_\mu^{-1} = \begin{cases} \sum_{1\le k<i} \frac{\omega_{ki}^2}{\varphi_{ki}^2} + \sum_{i<k\le n} \frac{\omega_{ik}^2}{\varphi_{ik}^2}, & \text{if } i = j. \\ -\frac{\omega_{ij}^2}{\varphi_{ij}^2}, & \text{if } i < j. \\ -\frac{\omega_{ji}^2}{\varphi_{ji}^2}, & \text{if } i > j. \end{cases}$$

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

be expressed as

$$
\begin{aligned}
&\pi(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{v}, \tau, \eta_1^2, \eta_2^2, \varphi_{ij}^2, \pi_1 | Y) \\
&\quad \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau^{-1}\xi_2^2 v_i}} \exp\Big\{ - \frac{(y_i - \mu_i - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \\
&\qquad \times \prod_{i=1}^{n} \tau \exp(-\tau v_i) \tau^{a-1} \exp(-b\tau) \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} z_i^2) \\
&\qquad \times \prod_{j=1}^{p} \Big( (1-\pi_1)(2\pi s_j)^{-1/2} \exp\{-\frac{\beta_j^2}{2s_j}\} \mathbf{I}_{\{\beta_j \neq 0\}} + \pi_1 \delta_0(\beta_j) \Big) \\
&\qquad \times (\eta_1^2)^{c_1 - 1} \exp(-d_1 \eta_1) \times (\eta_2^2)^{c_2 - 1} \exp(-d_2 \eta_2) \times \pi_1^{r_1 - 1} (1-\pi_1)^{u_1 - 1} \\
&\qquad \times |\Sigma_\mu|^{-\frac{1}{2}} \exp\{-\frac{1}{2} \boldsymbol{\mu}^\top (\Sigma_\mu)^{-1} \boldsymbol{\mu}\} \times \prod_{1 \leq i < j \leq n} \frac{\eta_1^2}{2} \exp(-\frac{\eta_1^2}{2} \varphi_{ij}^2)
\end{aligned}
$$

Let $\mu_{(-\boldsymbol{\beta}_j)} = \mu_i + \boldsymbol{X}_i^\top \boldsymbol{\beta} + \xi_1 v_i - X_{ij}\beta_j$ and $l_j = \pi(\beta_j = 0 | \text{rest})$, the conditional posterior distribution of the coefficient of genetic factor $\beta_j$ is a spike-and-slab distribution:

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

where

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

We can show that

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

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

The full conditional posterior distribution of $s_j$ is:

$$
\begin{aligned}
& s_j|\text{rest} \\
& \quad \propto \pi(\beta_j|s_j,\pi_1)\pi(s_j|\eta_2^2) \\
& \quad \propto \Big((1-\pi_1)(2\pi s_j)^{-1/2}\exp(-\frac{\beta_j^2}{2s_j})\mathbf{I}_{\{\beta_j\neq 0\}} + \pi_1\delta_0(\beta_j)\Big)\exp(-\frac{\eta_2^2}{2}s_j).
\end{aligned} \tag{4.6}
$$

When $\beta_j = 0$, equation(4.6) is proportional to $\exp(-\frac{\eta_2^2}{2}s_j)$. Therefore, the posterior distribution of $s_j$ is $\exp(\frac{\eta_2^2}{2})$.

When $\beta_j \neq 0$, equation(4.6) is proportional to

$$
\begin{aligned}
& \frac{1}{\sqrt{s_j}}\exp(-\frac{\eta_2^2}{2}s_j)\exp(-\frac{\beta_j^2}{2s_j}) \\
& \qquad \propto \frac{1}{\sqrt{s_j}}\exp\Big\{-\frac{1}{2}[\eta_2^2 s_j + \frac{\beta_j^2}{s_j}]\Big\}.
\end{aligned}
$$

Therefore, when $\beta_j \neq 0$, the posterior distribution for $s_j^{-1}$ is Inverse-Gaussian$(\sqrt{\frac{\eta_2^2}{\beta_j^2}}, \eta_2^2)$.

The full conditional posterior distribution of $\eta_2^2$:

$$
\begin{aligned}
& \eta_2^2|\text{rest} \\
& \quad \propto \pi(s_1|\eta_2^2)\pi(\eta_2^2) \\
& \quad \propto \prod_{j=1}^{p}\frac{\eta_2^2}{2}\exp(-\frac{\eta_2^2 s_j}{2})(\eta_2^2)^{c_2-1}\exp(-d_2\eta_2^2) \\
& \quad \propto (\eta_2^2)^{p+c_2-1}\exp\Big(-\eta_2^2(\sum_{j=1}^{p}\frac{s_j}{2}+d_2)\Big).
\end{aligned}
$$

Therefore, the posterior distribution for $\eta_2^2$ is Gamma$(p+c_2, \sum_{j=1}^{p}\frac{s_j}{2}+d_2)$.

The full conditional posterior distribution of $\tau$:

$$
\begin{aligned}
&\tau|\text{rest} \\
&\quad\propto \pi(\boldsymbol{v}|\tau)\pi(\tau)\pi(\boldsymbol{Y}|\cdot) \\
&\quad\propto \tau^{n/2}\exp\Big\{-\sum_{i=1}^{n}\frac{(y_i-\mu_i-X_{ij}\beta_j-\xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\} \\
&\quad\times \tau^n\exp(-\tau\sum_{i=1}^{n}v_i)\tau^{a-1}\exp(-b\tau) \\
&\quad\propto \tau^{a+\frac{3}{2}n-1}\exp\Big\{-\tau\big[\sum_{i=1}^{n}(\frac{(y_i-\mu_i-X_{ij}\beta_j-\xi_1 v_i}{2\xi_2^2 v_i}+v_i)+b\big]\Big\}.
\end{aligned}
$$

Therefore, the posterior distribution for $\tau$ is Gamma$(a+\frac{3}{2}n,\ \big[\sum_{i=1}^{n}(\frac{(y_i-\mu_i-X_{ij}\boldsymbol{\beta}_j-\xi_1 v_i)^2}{2\xi_2^2 v_i}+v_i)+b\big])$.

The full conditional posterior distribution of $\pi_1$:

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

So, the posterior distribution for $\pi_1$ is Beta $(1+r_1-\sum_{j=1}^{p}\mathbf{I}(\beta_j\neq 0), u_1+\sum_{j=1}^{p}\mathbf{I}(\beta_j\neq 0))$.

We have the full conditional posterior distribution of $v_i$:

$$
\begin{aligned}
&v_i|\text{rest} \\
&\quad\propto \pi(v|\tau)\pi(Y|\cdot) \\
&\quad\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{(y_i-\mu_i-\boldsymbol{X}_i^\top\boldsymbol{\beta})^2}{2\tau^{-1}\xi_2^2 v_i}\Big\}\times\exp(-\tau v_i) \\
&\quad\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{1}{2}\big[(\frac{\tau\xi_1^2}{\xi_2^2}+2\tau)v_i+\frac{\tau(y_i-\mu_i-\boldsymbol{X}_i^\top\boldsymbol{\beta})^2}{\xi_2^2 v_i}\big]\Big\}.
\end{aligned}
$$

It can be found that

$$\frac{1}{v_i}|\text{rest} \sim \text{ Inverse-Gaussian}(\sqrt{\frac{\xi_1^2+2\xi_2^2}{(y_i-\mu_i-\boldsymbol{X}_i^\top\boldsymbol{\beta})^2}},\ \frac{\tau\xi_1^2}{\xi_2^2}+2\tau).$$

The full conditional posterior distribution of $\boldsymbol{\mu}$:

$$\begin{aligned}
&\boldsymbol{\mu}|\text{rest}\\
&\propto \exp\{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v)^\top\Sigma_2^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v)\}\\
&\times|\Sigma_\mu|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\mu}^\top\Sigma_\mu^{-1}\boldsymbol{\mu}\}\\
&\propto \exp\{-\frac{1}{2}[\boldsymbol{\mu}^\top({\Sigma_2}^{-1}+{\Sigma_\mu}^{-1})\boldsymbol{\mu}-2\boldsymbol{\mu}^\top{\Sigma_2}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v)]\}
\end{aligned}$$

Where $\mu_2^{-1}=diag(\frac{1}{\tau^{-1}\xi_2^2 v})$. It can be found that the posterior distribution of $\boldsymbol{\mu}$ is $N(\mu_\mu,\sigma_\mu^2)$, where

$$\begin{aligned}
\mu_\mu &= \sigma_\mu^2\Sigma_2^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v),\\
\sigma_\mu^2 &= (\Sigma_2^{-1}+{\Sigma_\mu}^{-1})^{-1}.
\end{aligned}$$

The full conditional posterior distribution of $\varphi_{ij}$:

$$\begin{aligned}
&\varphi_{ij}^2|\text{rest}\\
&\propto \frac{1}{\sqrt{2\pi\varphi_{ij}^2}}\exp\{-\frac{(\mu_i-\mu_j)^2}{2\varphi_{ij}^2}\}\exp(-\frac{\lambda^2\varphi_{ij}^2}{2})\\
&= \frac{1}{\sqrt{2\pi\varphi_{ij}^2}}\exp\{-\frac{1}{2}[\frac{1}{\varphi_{ij}^2}(\mu_i-\mu_j)^2+\eta_1^2\varphi_{ij}^2]\}
\end{aligned}$$

Then,

$$\frac{1}{\varphi_{ij}^2}|\text{rest} \sim \text{ Inverse-Gaussian}(\sqrt{\frac{\eta_1^2}{(\mu_i-\mu_j)^2}},\ \eta_1^2), 1\le i<j\le n.$$

The full conditional posterior distribution of $\eta_1^2$:

$$\begin{aligned}
&\eta_1^2|\text{rest} \\
&\propto \prod_{1\le i<j\le n} \frac{\eta_1^2}{2}\exp(-\frac{\eta_1^2}{2}\varphi_{ij}^2) \times (\eta_1^2)^{c_1-1}\exp(-d_1\eta_1) \\
&= (\eta_1^2)^{\frac{n(n-1)}{2}+c_1-1}\exp\{-\eta_1^2(\sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+d_1)\}.
\end{aligned}$$

Therefore, the posterior distribution for $\eta_1^2$ is Gamma($\frac{n(n-1)}{2}+c_1$, $\sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+d_1$).

## 4.3 Simulation

To demonstrate the advantage of the proposed method, we evaluate the performance through simulation study. We compare the performance of the proposed method, Bayesian quantile subgroup with spike-and-slab priors (denoted as bqsgss) with several alternatives, Bayesian quantile subgroup (denoted as bqsg), Bayesian subgroup with spike-and-slab priors (denoted as bsubgroupss), Bayesian subgroup (denoted as bsubgroup), subgroup with $L_1$ loss (denoted as LASSO), subgroup with minimax concave penalty (denoted as MCP), subgroup with smoothly clipped absolute deviation penalty (denoted as SCAD), subgroup with truncated $L_1$ penalty (denoted as TL). Among all the methods under comparison, there are four Bayesian methods where bqsgss and bqsg are robust methods, bsubgroupss and bsubgroup are non-robust methods. The other four methods are frequentist methods.

To evaluate the performance of all the methods for each scenario, we consider comparing the $\bar{K}$ which is the average values of the estimated number of subgroups over 100 replicates and $\tilde{K}$ which is the median values of the estimated number of subgroups over 100 replicates. Moreover, the square root of the mean squared error (RMSE) for the estimated values of the intercepts and covariate effects are used for comparison, which are defined as $\text{RMSE}_\mu = ||\hat{\mu_i} - \mu_i||/\sqrt{n}$ and $\text{RMSE}_\beta = ||\hat{\beta}_i - \beta_i||/\sqrt{p}$. The mean absolute errors for the intercepts ($\text{MAE}_\mu$) and the mean absolute errors for the covariate effects ($\text{MAE}_\beta$) are evaluated as well, given by $\text{MAE}_\mu = \frac{1}{n}\sum_{i=1}^n |\hat{\mu_i} - \mu_i|$ and $\text{MAE}_\beta = \frac{1}{p}\sum_{i=1}^p |\hat{\beta}_i - \beta_i|$.

### 4.3.1 Subgroup analysis for low-dimensional cases

We first consider $n \in \{200, 400\}$ and $p = 5$ to compare the performance of the proposed method under situations with low dimensionality. Specifically, we consider $K = 2$ with center $1, -1$. $\mu_i$ 's are independently generated from a multinomial distribution with equal probability at each group center. $x_i$'s are generated independently from the standard normal distribution with margianl mean vector 0 and a convariate matrix with auto-regressive (AR-1) correlation. The corresponding coefficients are set to be $\beta = \mathbf{I}_5$. The error terms are generated from two different distributions: (1)N$(0, 0.5^2)$(Error1), (2)$0.5 \times t(5)$(Error2).

### 4.3.2 Subgroup analysis for high-dimensional cases

We also consider high-dimensional cases where $n \in \{200, 400\}$ and $p \in \{50, 100, 500\}$ with $\beta = (\mathbf{I}_5, \mathbf{0}_(p-5))$. The generation of the covariates $\mu_i$ 's, $x_i$'s and errors are the same as those in 4.3.1.Simulation results for the gene expression data in the 4.3.1 and 4.3.2 are tabulated in Table 4.1 and Table B.1.

We can observe that the proposed robust method with spike-and-slab priors has the best performance among all approaches in both identification and estimation, especially when the dimensionality is high. For example, in Table 4.1, when $n = 200, p = 500$, the median of the estimated subgroup number of the robust Bayesian subgroup with spike-and-slab priors and Bayesian subgroup with spike-and-slab priors is 2, which are more accurate than other methods.

The advantages of the methods incorporating spike-and-slab priors over the methods without spike-and-slab methods can also be found from the results. For instance, when $n = 200, p = 50$ in Table 4.1, Bayesian subgroup with spike-and-slab priors performs better than Baysian subgroup as well as the frequentist methods. For Bayesian subgroup with spike-and-slab priors method, the $\text{RMSE}_\mu$ is 0.393(0.115) and the $\text{MAE}_\mu$ is 0.144(0.144), which is the smallest among all methods. Also, when comparing the performance with the frequentist methods, we can also observe that the mean (1.992(0.413)) of the estimated subgroup number with Bayesian subgroup with spike-and-slab priors method is more accurate as it's close to

2, which is the true value of the group number.

Moreover, a comparison between the Bayesian methods and frequentist methods shows the necessity of developing Bayesian methods, especially when the dimensionality is high. For example, in Table B.1, when $n = 400, p = 100$, the mean and median of the estimated subgroup number is around the true value 2, while those of the frequentist methods are far from 2. What's more, the frequentist methods don;t work in high dimensional data, which indicates the advantages of developing robust Bayesian methods in subgroup analysis.

Similar patterns can be found from data with heavy-tailed distribution and outliers in Table B.1. The proposed robust methods have more stable results than the other methods, both in low-dimensional cases and high-dimensional cases. For example, in Table B.1, when $n = 400, p = 5$, the median of the subgroups identified by the robut Bayesian methods is the same as the true group number 2, while that value of the frequentist methods is far from 2. What's more, the average value of the subgroup number estimated by the proposed method is also close to the true value 2.

The proposed method also performs better than the other methods in high-dimensional data. When $n = 400, p = 100$ in Table B.1, robust Bayesian methods outperform the other methods with a more accurate estimation of subgroups while the frequentist methods lead to a biased estimation in high-dimensional data. When increasing the dimensionality with $n = 200, p = 500$ in Table B.1, the Bayesian methods also have a stable results in estimating the subgroups. However, the frequentist methods cannot give reasonable results when dealing with high-dimensional data.

**Table 4.1**: *Simulation results with the mean (sd) of the setting with i.i.d. errors $N(0, 0.5^2)$ over 100 replicates.*

| | | bqsgss | bqsg | bsubgroupss | bsubgroup | LASSO | MCP | SCAD | TL |
|---|---|---|---|---|---|---|---|---|---|
| **n=200,** | $\bar{K}$ | 1.032(0.179) | 1.032(0.179) | 2.030(0.490) | 2.030(0.490) | 2.530(0.860) | 2.200(0.760) | 2.230(0.770) | 2.400(0.500) |
| **p=5** | $\tilde{K}$ | 1.000 | 1.000 | 2.000 | 2.000 | 2.000 | 2.000 | 2.000 | 2.000 |
| | $\text{RMSE}_\mu$ | 0.964(0.011) | 0.941(0.013) | 0.423(0.111) | 0.423(0.111) | 0.761(0.234) | 0.322(0.092) | 0.322(0.089) | 0.333(0.081) |
| | $\text{RMSE}_\beta$ | 0.245(0.020) | 0.185(0.061) | 0.055(0.033) | 0.055(0.033) | 0.068(0.032) | 0.049(0.023) | 0.049(0.023) | 0.048(0.022) |
| | $\text{MAE}_\mu$ | 0.959(0.011) | 0.937(0.013) | 0.344(0.141) | 0.344(0.141) | 0.741(0.255) | 0.126(0.067) | 0.126(0.065) | 0.138(0.056) |
| | $\text{MAE}_\beta$ | 0.025(0.000) | 0.004(0.001) | 0.001(0.001) | 0.001(0.001) | 0.058(0.028) | 0.042(0.022) | 0.042(0.022) | 0.040(0.020) |
| **n=400,** | $\bar{K}$ | 1.857(0.355) | 1.750(0.438) | 1.530(0.571) | 1.500(0.512) | 3.071(1.362) | 2.800(1.300) | 3.000(1.391) | 2.800(1.032) |
| **p=5** | $\tilde{K}$ | 2.000 | 2.000 | 1.500 | 1.500 | 3.000 | 2.000 | 2.000 | 3.000 |
| | $\text{RMSE}_\mu$ | 0.996(0.006) | 0.986(0.006) | 0.637(0.159) | 0.637(0.160) | 0.790(0.197) | 0.341(0.065) | 0.350(0.059) | 0.318(0.050) |
| | $\text{RMSE}_\beta$ | 0.257(0.010) | 0.289(0.059) | 0.054(0.021) | 0.053(0.021) | 0.053(0.021) | 0.039(0.015) | 0.041(0.015) | 0.035(0.011) |
| | $\text{MAE}_\mu$ | 0.990(0.006) | 0.980(0.006) | 0.614(0.180) | 0.615(0.181) | 0.779(0.212) | 0.121(0.056) | 0.131(0.058) | 0.110(0.039) |
| | $\text{MAE}_\beta$ | 0.013(0.000) | 0.003(0.001) | 0.001(0.000) | 0.001(0.000) | 0.044(0.018) | 0.031(0.012) | 0.033(0.013) | 0.029(0.010) |
| **n=200,** | $\bar{K}$ | 1.643(0.427) | 1.231(0.497) | 1.992(0.413) | 1.972(1.591) | 2.472(0.970) | 4.471(1.140) | 4.372(1.104) | 3.370(1.131) |
| **p=50** | $\tilde{K}$ | 2.000 | 2.000 | 2.000 | 2.000 | 2.500 | 5.000 | 4.500 | 3.000 |
| | $\text{RMSE}_\mu$ | 0.957(0.009) | 0.978(0.010) | 0.393(0.115) | 0.451(0.142) | 0.999(0.005) | 0.672(0.083) | 0.668(0.059) | 0.921(0.159) |
| | $\text{RMSE}_\beta$ | 0.316(0.000) | 0.292(0.039) | 0.019(0.007) | 0.056(0.012) | 0.108(0.012) | 0.104(0.014) | 0.104(0.014) | 0.117(0.019) |
| | $\text{MAE}_\mu$ | 0.952(0.009) | 0.975(0.010) | 0.144(0.144) | 0.381(0.175) | 0.994(0.006) | 0.484(0.132) | 0.464(0.098) | 0.877(0.217) |
| | $\text{MAE}_\beta$ | 0.025(0.000) | 0.058(0.008) | 0.001(0.001) | 0.011(0.003) | 0.087(0.010) | 0.084(0.012) | 0.084(0.013) | 0.093(0.015) |
| **n=400,** | $\bar{K}$ | 2.000(0.000) | 2.000(0.000) | 1.432(0.631) | 1.371(0.410) | 2.700(0.920) | 4.800(1.350) | 4.600(1.191) | 4.20(01.522) |
| **p=50** | $\tilde{K}$ | 2.000 | 2.000 | 1.000 | 1.430 | 2.000 | 5.000 | 5.000 | 4.000 |
| | $\text{RMSE}_\mu$ | 0.984(0.005) | 0.986(0.004) | 0.665(0.153) | 0.722(0.144) | 0.999(0.003) | 0.541(0.057) | 0.541(0.048) | 0.674(0.224) |
| | $\text{RMSE}_\beta$ | 0.316(0.000) | 0.277(0.059) | 0.016(0.005) | 0.052(0.010) | 0.071(0.007) | 0.061(0.010) | 0.061(0.009) | 0.064(0.013) |
| | $\text{MAE}_\mu$ | 0.98(0.005) | 0.983(0.004) | 0.647(0.166) | 0.709(0.152) | 0.997(0.005) | 0.326(0.087) | 0.320(0.078) | 0.016(0.090) |
| | $\text{MAE}_\beta$ | 0.013(0.000) | 0.028(0.006) | 0.001(0.000) | 0.005(0.001) | 0.056(0.006) | 0.048(0.008) | 0.048(0.007) | 0.002(0.011) |
| **n=200,** | $\bar{K}$ | 1.840(0.374) | 1.840(0.374) | 2.130(0.460) | 1.501(3.422) | 2.300(0.991) | 4.270(1.080) | 4.230(1.041) | 2.930(0.832) |
| **p=100** | $\tilde{K}$ | 2.000 | 2.000 | 2.000 | 1.000 | 2.000 | 4.000 | 4.000 | 3.000 |
| | $\text{RMSE}_\mu$ | 0.963(0.010) | 0.989(0.005) | 0.385(0.091) | 0.586(0.158) | 1.000(0.009) | 0.863(0.055) | 0.863(0.057) | 1.021(0.023) |
| | $\text{RMSE}_\beta$ | 0.224(0.000) | 0.462(0.046) | 0.014(0.005) | 0.083(0.017) | 0.135(0.014) | 0.141(0.014) | 0.140(0.014) | 0.150(0.013) |
| | $\text{MAE}_\mu$ | 0.959(0.010) | 0.986(0.005) | 0.289(0.122) | 0.550(0.189) | 0.992(0.009) | 0.731(0.096) | 0.732(0.097) | 1.004(0.015) |
| | $\text{MAE}_\beta$ | 0.025(0.000) | 0.190(0.018) | 0.001(0.001) | 0.033(0.007) | 0.107(0.012) | 0.112(0.012) | 0.111(0.012) | 0.119(0.011) |
| **n=400,** | $\bar{K}$ | 1.500(0.001) | 1.500(0.001) | 1.830(0.500) | 1.071(0.432) | 2.930(0.981) | 5.130(1.070) | 5.030(0.891) | 3.070(0.742) |
| **p=100** | $\tilde{K}$ | 1.000 | 1.000 | 1.430 | 1.000 | 3.000 | 5.000 | 5.000 | 3.000 |
| | $\text{RMSE}_\mu$ | 0.987(0.002) | 0.993(0.009) | 0.705(0.141) | 0.818(0.101) | 0.999(0.002) | 0.662(0.038) | 0.659(0.037) | 1.014(0.010) |
| | $\text{RMSE}_\beta$ | 0.224(0.000) | 0.341(0.029) | 0.012(0.004) | 0.064(0.007) | 0.076(0.006) | 0.073(0.007) | 0.073(0.007) | 0.088(0.009) |
| | $\text{MAE}_\mu$ | 0.982(0.002) | 0.989(0.009) | 0.691(0.155) | 0.811(0.106) | 0.992(0.009) | 0.731(0.096) | 0.732(0.097) | 1.004(0.015) |
| | $\text{MAE}_\beta$ | 0.013(0.000) | 0.067(0.007) | 0.001(0.001) | 0.013(0.001) | 0.107(0.012) | 0.112(0.012) | 0.111(0.012) | 0.119(0.011) |
| **n=200,** | $\bar{K}$ | 1.667(0.500) | 1.462(0.519) | 1.871(0.350) | 1.000(0.010) | — | — | — | — |
| **p=500** | $\tilde{K}$ | 2.000 | 1.000 | 2.000 | 1.000 | — | — | — | — |
| | $\text{RMSE}_\mu$ | 0.991(0.008) | 1.004(0.008) | 0.402(0.107) | 1.456(0.666) | — | — | — | — |
| | $\text{RMSE}_\beta$ | 0.100(0.000) | 0.536(0.019) | 0.006(0.003) | 0.104(0.020) | 0.0231(0.0044) | 0.0111(0.0038) | 0.0121(0.0072) | — |
| | $\text{MAE}_\mu$ | 0.988(0.008) | 1.000(0.008) | 0.301(0.149) | 1.266(0.601) | — | — | — | — |
| | $\text{MAE}_\beta$ | 0.025(0.000) | 1.068(0.044) | 0.001(0.001) | 0.183(0.045) | 0.0038(0.0014) | 0.0013(0.0006) | 0.0012(0.0010) | — |
| **n=400,** | $\bar{K}$ | 1.000(0.000) | 1.000(0.000) | 2.100(1.172) | 1.000(0.000) | — | — | — | — |
| **p=500** | $\tilde{K}$ | 1.000 | 1.000 | 1.000 | 1.000 | — | — | — | — |
| | $\text{RMSE}_\mu$ | 0.989(0.006) | 1.005(0.008) | 0.657(0.160) | 1.075(0.114) | — | — | — | — |
| | $\text{RMSE}_\beta$ | 0.100(0.012) | 0.571(0.020) | 0.005(0.002) | 0.133(0.013) | 0.0157(0.0029) | 0.0059(0.0015) | 0.0063(0.0029) | — |
| | $\text{MAE}_\mu$ | 0.985(0.022) | 1.001(0.025) | 0.639(0.176) | 1.000(0.010) | — | — | — | — |
| | $\text{MAE}_\beta$ | 0.013(0.014) | 0.562(0.011) | 0.001(0.000) | 0.133(0.013) | 0.0024(0.0008) | 0.0006(0.0002) | 0.0006(0.0004) | — |

**Table 4.2**: *Simulation results with the mean (sd) of the setting with i.i.d. errors* $0.5 \times t(5)$ *over 100 replicates.*

| | | bqsgss | bqsg | bsubgroupss | bsubgroup | LASSO | MCP | SCAD | TL |
|---|---|---|---|---|---|---|---|---|---|
| **n=200,** | $\bar{K}$ | 1.583(0.515) | 1.500(0.519) | 2.400(0.621) | 2.433(0.626) | 4.030(1.100) | 3.630(1.220) | 3.970(1.330) | 3.970(1.330) |
| **p=5** | $\tilde{K}$ | 2.000 | 1.500 | 2.400 | 2.433 | 4.000 | 3.000 | 4.000 | 4.00 0 |
| | $\text{RMSE}_\mu$ | 0.962(0.017) | 0.943(0.011) | 0.563(0.088) | 0.563(0.088) | 0.960(0.110) | 0.547(0.080) | 0.557(0.075) | 0.543(0.067) |
| | $\text{RMSE}_\beta$ | 1.000(0.000) | 0.179(0.055) | 0.071(0.026) | 0.071(0.025) | 0.081(0.023) | 0.059(0.018) | 0.060(0.019) | 0.056(0.018) |
| | $\text{MAE}_\mu$ | 0.956(0.017) | 0.938(0.011) | 0.401(0.136) | 0.401(0.137) | 0.945(0.120) | 0.231(0.074) | 0.250(0.088) | 0.225(0.074) |
| | $\text{MAE}_\beta$ | 0.025(0.000) | 0.004(0.001) | 0.060(0.022) | 0.060(0.022) | 0.068(0.019) | 0.051(0.016) | 0.052(0.017) | 0.048(0.017) |
| **n=400,** | $\bar{K}$ | 1.857(0.355) | 1.811(0.397) | 2.500(1.225) | 2.433(1.006) | 4.83(1.42) | 5.30(1.80) | 1.63(2.71) | 2.00(2.62) |
| **p=5** | $\tilde{K}$ | 2.000 | 2.000 | 2.000 | 2.000 | 5.000 | 5.00 | 3.00 | 3.00 |
| | $\text{RMSE}_\mu$ | 0.995(0.007) | 0.985(0.006) | 0.747(0.148) | 0.7480.148) | 0.971(0.105) | 0.557(0.051) | 0.163(0.254) | 0.204(0.266) |
| | $\text{RMSE}_\beta$ | 1.000(0.000) | 0.262(0.064) | 0.060(0.018) | 0.060(0.018) | 0.063(0.022) | 0.044(0.020) | 0.011(0.018) | 0.015(0.022) |
| | $\text{MAE}_\mu$ | 0.989(0.007) | 0.979(0.006) | 0.704(0.178) | 0.704(0.178) | 0.960(0.110) | 0.252(0.090) | 0.076(0.120) | 0.051(0.107) |
| | $\text{MAE}_\beta$ | 1.000(0.000) | 0.219(0.058) | 0.050(0.016) | 0.050(0.016) | 0.054(0.021) | 0.037(0.018) | 0.010(0.016) | 0.007(0.016) |
| **n=200,** | $\bar{K}$ | 1.000(0.000) | 1.000(0.000) | 2.667(0.547) | 2.933(1.172) | 3.130(1.200) | 4.830(1.180) | 4.770(1.220) | 3.630(1.250) |
| **p=50** | $\tilde{K}$ | 1.000 | 1.000 | 3.000 | 3.000 | 3.000 | 5.000 | 5.000 | 3.000 |
| | $\text{RMSE}_\mu$ | 0.955(0.014) | 0.975(0.008) | 0.564(0.084) | 0.602(0.111) | 1.006(0.011) | 0.757(0.075) | 0.758(0.073) | 0.961(0.139) |
| | $\text{RMSE}_\beta$ | 0.316(0.000) | 0.288(0.049) | 0.021(0.07) | 0.059(0.011) | 0.115(0.015) | 0.113(0.013) | 0.113(0.012) | 0.126(0.020) |
| | $\text{MAE}_\mu$ | 0.950(0.015) | 0.972(0.008) | 0.386(0.127) | 0.472(0.156) | 0.997(0.010) | 0.549(0.136) | 0.546(0.137) | 0.894(0.229) |
| | $\text{MAE}_\beta$ | 0.100(0.000) | 0.231(0.041) | 0.005(0.002) | 0.047(0.008) | 0.092(0.012) | 0.090(0.011) | 0.090(0.011) | 0.102(0.017) |
| **n=400,** | $\bar{K}$ | 1.811(0.397) | 1.750(0.439) | 2.767(1.775) | 2.733(1.617) | 4.93(1.68) | 5.87(1.43) | 0.43(1.33) | 1.20(1.93) |
| **p=50** | $\tilde{K}$ | 2.000 | 2.000 | 3.000 | 3.000 | 4.000 | 6.000 | 5.500 | 3.000 |
| | $\text{RMSE}_\mu$ | 0.981(0.006) | 0.986(0.006) | 0.773(0.131) | 0.816(0.121) | 1.001(0.044) | 0.673(0.046) | 0.070(0.213) | 0.245(0.405) |
| | $\text{RMSE}_\beta$ | 0.316(0.000) | 0.283(0.034) | 0.018(0.006) | 0.059(0.011) | 0.075(0.008) | 0.068(0.010) | 0.007(0.022) | 0.023(0.039) |
| | $\text{MAE}_\mu$ | 0.976(0.006) | 0.982(0.006) | 0.731(0.163) | 0.782(0.145) | 0.993(0.047) | 0.333(0.188) | 0.414(0.083) | 0.282(0.422) |
| | $\text{MAE}_\beta$ | 0.100(0.000) | 0.226(0.030) | 0.005(0.002) | 0.047(0.009) | 0.060(0.006) | 0.044(0.024) | 0.057(0.007) | 0.026(0.034) |
| **n=200,** | $\bar{K}$ | 1.632(0.496) | 1.545(0.509) | 2.967(2.533) | 2.033(1.033) | 2.670(1.270) | 5.070(0.940) | 5.070(0.870) | 3.330(0.880) |
| **p=100** | $\tilde{K}$ | 2.000 | 2.000 | 2.500 | 2.000 | 2.500 | 5.000 | 5.000 | 3.000 |
| | $\text{RMSE}_\mu$ | 0.964(0.009) | 0.995(0.007) | 0.574(0.119) | 0.715(0.151) | 1.004(0.015) | 0.881(0.041) | 0.880(0.042) | 1.038(0.031) |
| | $\text{RMSE}_\beta$ | 0.224(0.000) | 0.429(0.036) | 0.016(0.005) | 0.091(0.016) | 0.145(0.012) | 0.148(0.011) | 0.148(0.011) | 0.162(0.015) |
| | $\text{MAE}_\mu$ | 0.960(0.009) | 0.992(0.007) | 0.418(0.160) | 0.653(0.183) | 0.995(0.010) | 0.713(0.066) | 0.712(0.057) | 1.017(0.022) |
| | $\text{MAE}_\beta$ | 0.050(0.000) | 0.342(0.027) | 0.003(0.001) | 0.073(0.013) | 0.116(0.010) | 0.118(0.010) | 0.118(0.010) | 0.129(0.013) |
| **n=400,** | $\bar{K}$ | 1.828(0.384) | 1.706(0.463) | 2.567(1.040) | 2.467(0.937) | 4.800(1.480) | 6.300(1.160) | 6.100(1.200) | 2.900(1.600) |
| **p=100** | $\tilde{K}$ | 2.000 | 2.000 | 2.000 | 2.500 | 5.000 | 6.500 | 6.000 | 3.500 |
| | $\text{RMSE}_\mu$ | 0.985(0.007) | 0.993(0.005) | 0.777(0.126) | 0.862(0.101) | 1.007(0.011) | 0.747(0.020) | 0.748(0.022) | 0.823(0.434) |
| | $\text{RMSE}_\beta$ | 0.224(0.000) | 0.343(0.051) | 0.013(0.006) | 0.067(0.010) | 0.079(0.009) | 0.077(0.008) | 0.077(0.008) | 0.073(0.039) |
| | $\text{MAE}_\mu$ | 0.981(0.006) | 0.989(0.006) | 0.733(0.150) | 0.835(0.108) | 1.001(0.006) | 0.549(0.071) | 0.554(0.057) | 1.018(0.014) |
| | $\text{MAE}_\beta$ | 0.050(0.000) | 0.273(0.041) | 0.002(0.001) | 0.054(0.007) | 0.063(0.008) | 0.062(0.006) | 0.062(0.006) | 0.072(0.008) |
| **n=200,** | $\bar{K}$ | 1.727(0.467) | 1.571(0.514) | 2.667(1.516) | 1.633(0.718) | — | — | — | — |
| **p=500** | $\tilde{K}$ | 2.000 | 2.000 | 2.000 | 1.500 | — | — | — | — |
| | $\text{RMSE}_\mu$ | 0.993(0.009) | 1.001(0.008) | 0.548(0.097) | 1.450(0.410) | — | — | — | — |
| | $\text{RMSE}_\beta$ | 0.100(0.000) | 0.530(0.017) | 0.007(0.002) | 0.104(0.010) | 0.0234(0.0036) | 0.0090(0.0031) | 0.0091(0.0032) | — |
| | $\text{MAE}_\mu$ | 0.989(0.009) | 0.996(0.008) | 0.383(1.541) | 1.091(0.145) | — | — | — | — |
| | $\text{MAE}_\beta$ | 0.010(0.000) | 0.424(0.014) | 0.001(0.000) | 0.072(0.009) | 0.0037(0.0007) | 0.0009(0.0005) | 0.0008(0.0004) | — |
| **n=400,** | $\bar{K}$ | 1.000(0.000) | 1.000(0.000) | 2.667(1.470) | 1.667(0.884) | — | — | — | — |
| **p=500** | $\tilde{K}$ | 1.000 | 1.000 | 2.000 | 1.000 | — | — | — | — |
| | $\text{RMSE}_\mu$ | 0.996(0.005) | 1.006(0.006) | 0.807(0.104) | 1.175(0.174) | — | — | — | — |
| | $\text{RMSE}_\beta$ | 0.100(0.000) | 0.570(0.030) | 0.007(0.002) | 0.144(0.014) | 0.0157(0.0023) | 0.0053(0.0017) | 0.0055(0.0020) | — |
| | $\text{MAE}_\mu$ | 0.992(0.005) | 1.002(0.006) | 0.772(0.121) | 1.027(0.048) | — | — | — | — |
| | $\text{MAE}_\beta$ | 0.010(0.000) | 0.457(0.025) | 0.001(0.000) | 0.115(0.012) | 0.0025(0.0006) | 0.0005(0.0002) | 0.0005(0.0002) | — |

## 4.4 Real Data Analysis

In this case study, we analyze the Breast Invasive Carcinoma data from The Cancer Genome Atlas (TCGA), which is a well-characterized genomics program for cancer research. For high-dimensional genetic factors, we focus on the mRNA gene expression data, which is analyzed and downloaded from cBioPortal. As a well-studied tumor suppressor for BRCA, the gene expression BRCA1 is used as the response variable. Its mRNA expression has been shown to be able to predict the breast cancer risk and patients' response to certain treatment. BRCA1 is skewed distributed with a heavy tail, which can be observed in Figure4.1, indicating that a robust method is in demand. Data are available on 1069 female subjects and 20440 gene expressions after matching. A marginal screening has been done based on the linear regression and 800 gene expressions with p-value less than 0.001 is obtained for the downstream analysis.

Table B.3 shows important genetic findings using one of our proposed methods bsubgroupss, which can be proved with existing literature. For example, gene PSME3 has been identified by our method that is highly related with the response variable, which has also been reported as a significant factor in tumor progression, playing an important role in affecting the immune infiltration and immune cells in the tumor microenvironment (Dong et al. (2024)). PSME3 has been found to be differentially associated with all examined cancer forms as a predictive biomarker (P. et al. (2022)). In addition, gene HMMR, which has a high posterior inclusion probability with our methods bsubgorupss, has been found to be associated with risk of breast cancer in BRCA1 mutation carriers (I. et al. (2015)), as it can regulate the progression of basal-like breast cancer cells (Liu et al. (2016)). Furthermore, the neighbor to BRCA1 gene, NBR1, which has been revealed to be positively expressed in human breast cancer and highly associated with the survival rates (Marsh and Debnath (2020)), has also been identified by our method with a high posterior probability.Marsh and Debnath (2020) have also found that the gene NBR1 independently is sufficient to drive metastatic outgrowth, even in autophagy-competent tumor cells and preventing NBR1 accumulation can reverse the inhibition on experimental metastasis. We can find that identifying signifi-

cant genetic factors in breast cancer studies is crucial in diagnosis, treatment and diseases' prognosis. Another gene DTL, which is found with our method with a poster inclusion probability 0.9584, has been observed to play an important role in the pathophysiology of breast and lung tumors (Perez-Peña et al. (2017)).All these findings indicate that our proposed method is useful in identifying SNPs with sensible biological implications.

**Table 4.3**: *The analysis results of the TCGA data using bsubgroupss.*

| Gene | Posterior Inclusion Probability | Coefficients |
|---|---|---|
| PSME3 | 1.000 | 0.1389 |
| WDR76 | 1.000 | 0.1849 |
| CENPK | 1.000 | 0.1747 |
| HROB | 1.000 | 0.1684 |
| CKAP2L | 1.000 | 0.1996 |
| DSN1 | 1.000 | 0.0953 |
| HMMR | 1.000 | 0.1618 |
| NBR1 | 1.000 | 0.1785 |
| XRCC2 | 1.000 | 0.1431 |
| SGOL1 | 1.000 | -0.1871 |
| UTP20 | 1.000 | 0.1193 |
| RFC1 | 1.000 | -0.1351 |
| PBRM1 | 1.000 | 0.1565 |
| PRC1 | 1.000 | -0.1242 |
| SPDL1 | 0.9954 | -0.1234 |
| UBXN2B | 0.991 | 0.0511 |
| DTL | 0.9584 | 0.1474 |
| CPT1A | 0.9492 | 0.0476 |
| SLX4 | 0.9390 | 0.0649 |
| ACLY | 0.9104 | 0.0781 |
| RUNDC1 | 0.8946 | 0.1084 |
| KIF21A | 0.8866 | -0.0788 |
| DNAJC7 | 0.833 0 | 0.0685 |
| KIF20B | 0.8232 | 0.1331 |
| SMC1A | 0.6747 | 0.0686 |
| PDLIM2 | 0.6543 | 0.0845 |
| SENP1 | 0.5297 | -0.056 |

## 4.5 Discussion

The Bayesian quantile subgroup variable selection methods have shown the advantages of detecting subgroup structure while selecting important factors simutaneously with large scale datasets. Although there are some limitations with the proposed method, such as the high computation which is commonly exists in Bayesian methods, the superior performance of the proposed Bayesian method cannot be ignored. Moreover, the Bayesian quantile regularization can be extended to other types of response, such as categorical and multivariate outcomes. Also, besides the linear model, the proposed method can be applied with nonlinear effects through nonparametric and semiparametric models.

# Bibliography

David F Andrews and Colin L Mallows. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(1):99–102, 1974.

S.F. Barclay, C.M. Rand, L.A. Borch, L. Nguyen, P.A. Gray, W.T. Gibson, P.M. Wilson, R.J.and Gordon, Z. Aung, E.M. Berry-Kravis, D. Ize-Ludlow, D.E. Weese-Mayer, and N.T. Bech-Hansen. Rapid-onset obesity with hypothalamic dysfunction, hypoventilation, and autonomic dysregulation (rohhad): exome sequencing of trios, monozygotic twins and tumours. *Orphanet Journal of Rare Diseases*, 10(103), 2015.

S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine learning*, 3(1):1–122, 2011.

S.P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434—455, 1998.

L.R. Brunham, J.K. Kruit, C.B. Verchere, and M.R. Hayden. Cholesterol in islet dysfunction and type 2 diabetes. *J. Clin. Invest*, 118:403–408, 2008. doi: 10.1172/JCI33296.

L. Cai, M. Abe, S. Izumi, M. Imura, T. Yasugi, and T. Ushijima. Identification of prtfdc1 silencing and aberrant promoter methylation of gpr150, itga8 and hoxd11 in ovarian cancers. *Life Sciences*, 80:1458–1465, 2007.

H. Chai, Q. Zhang, Y. Jiang, G. Wang, S. Zhang, S. E. Ahmed, and S. Ma. Identifying gene-environment interactions for prognosis using a robust approach. *Econometrics and Statistics*, 4:105–120, 2017.

Chao Cheng, Xingdong Feng, Xiaoguang Li, and Mengyun Wu. Robust analysis of cancer

heterogeneity for high-dimensional data. *Stat Med*, 41(27):5448–5462, 2022. doi: 10.1002/sim.9578.

Seoae Cho, Haseong Kim, Sohee Oh, Kyunga Kim, and Taesung Park. Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis. In *BMC proceedings*, volume 3, pages 1–6. BioMed Central, 2009.

Seoae Cho, Kyunga Kim, Young Jin Kim, Jong-Keuk Lee, Yoon Shin Cho, Jong-Young Lee, Bok-Ghee Han, Heebal Kim, Jurg Ott, and Taesung Park. Joint identification of multiple genetic variants via elastic-net variable selection in a genome-wide association analysis. *Annals of human genetics*, 74(5):416–428, 2010.

Gabriela V Cohen Freue, David Kepplinger, Matías Salibián-Barrera, and Ezequiel Smucler. Robust elastic net estimators for variable selection and identification of proteomic biomarkers. *The Annals of Applied Statistics*, 13(4):2065–2090, 2019.

Graham A Colditz and Susan E Hankinson. The nurses' health study: lifestyle and health among women. *Nature Reviews Cancer*, 5(5):388–396, 2005.

Graham A Colditz, Joann E Manson, and Susan E Hankinson. The nurses' health study: 20-year contribution to the understanding of health among women. *Journal of women's health*, 6(1):49–62, 1997.

H. J. Cordell and D. G. Clayton. Genetic association studies. *The Lancet*, 366(9491):1121–1131, 2005.

M. C. Cornelis and F. B. Hu. Gene-environment interactions in the development of type 2 diabetes: recent progress and continuing challenges. *Annual review of nutrition*, 32:245–259, 2012.

Chengyuan Dong, Yadong Guo, Yanrong Yang, and Xin Ge. Comprehensive analysis of psme3: from pan-cancer analysis to experimental validation. *Frontiers in Immunology*, 15 (1295693), 2024. doi: 10.3389/fimmu.2024.1295693.

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 94(456):1348–1360, 2001. doi: 10.1198/016214501753382273.

Jerome Friedman, Trevor Hastie, Rob Tibshirani, and Balasubramanian Narasimhan. *Lasso and Elastic-Net Regularized Generalized Linear Models*. R Package version 4.1-4.

Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2004.

E. I. George and R. E. McCulloch. Semiparametric bayesian variable selection for gene-environment interactions. *Variable selection via Gibbs sampling*, 88(423):881–889, 1992.

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

M. Ghanbari, P.S. de Vries, H. de Looper, M.J. Peters, C. Schurmann, H. Yaghootkar, M. Dörr, T.M. Frayling, A.G. Uitterlinden, A. Hofman, van Meurs, S.J. J.B, Erkeland, O.H. France, and A. Dehghan. A genetic variant in the seed region of mir-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease. *Hum Mutat*, 35:1524–1531, 2014.

Chris Hans. Elastic net regression modeling with the orthant normal prior. *Journal of the American Statistical Association*, 106(496):1383–1393, 2011.

T. Hastie and R. Tibshirani. Varying-coefficient models. *Journal of the Royal Sta- tistical Society: Series B (Methodological)*, 55(4):757–779, 1993.

J. N. Hirschhorn, K. Lohmueller, E. Byrne, and K. Hirschhorn. A comprehensive review of genetic association studies. *Genetics in medicine*, 4(2):45–61, 2002.

E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970. URL https://www.jstor.org/stable/1267351.

J. Huang and S. Ma. Variable selection in the accelerated failure time model via the bridge method. *Lifetime data analysis*, 16(2):176–195, 2010.

Y. C. Huang, W. L. Liao, J. M. Lin, C. C. Chen, S. P. Liu, S. Y. Chen, Y.N. Lin, Y.J. Lei, H.T. Liu, Y. J. Chen, and F. J. Tsai. High levels of circulating endothelial progenitor cells in patients with diabetic retinopathy are positively associated with arhgap22 expression. *Oncotarget*, 9(25):17858, 2018.

D. J. Hunter. Gene–environment interactions in human diseases. *Nature Reviews Genetics*, 6(4):287–298, 2005.

Blanco I., Kuchenbaecker K., Cuadras D., Wang X., Barrowdale D., and de Garibay GR. et al. Assessing associations between the aurka-hmmr- tpx2-tubg1 functional module and breast cancer risk in brca1/2 mutation carriers. *PLoS ONE*, 10(4), 2015. doi: 10.1371/journal.pone.0120020.

H. Ishwaran and J. S. Rao. Spike and slab variable selection: frequentist and bayesian strategies. *The Annals of Statistics*, 33(2):730–773, 2005.

Y. Jiang, Y. Huang, Y. Du, Y. Zhao, J. Ren, S. Ma, and C. Wu. Identification of prognostic genes and pathways in lung adenocarcinoma using a bayesian approach. *Cancer Informatics*, 1(17), 2017.

L. Jin, X. Zuo, W. Su, X. Zhao, M. Yuan, L. Han, X. Zhao, Y. Chen, and S. Rao. Pathway-based analysis tools for complex diseases: a review. *Genomics, Proteomics and Bioinformatics*, 12(5):210–220, 2014.

David Kepplinger. Robust variable selection and estimation via adaptive elastic net s-estimators for linear regression. *Computational Statistics & Data Analysis*, 183:107730, 2023.

H. Kozumi and G. Kobayashi. Gibbs sampling methods for bayesian quantile regression. *Technical report, Graduate School of Business Administration, Kobe University.*, pages 534–535, 2009.

H. Kozumi and G. Kobayashi. Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81(11):1565–1578, 2011.

F. Kurt, S. Lasser, I. Arkhypov, J. Utikal, and V. Umansky. Enhancing immunotherapy response in melanoma: myeloid-derived suppressor cells as a therapeutic target. *The Journal of Clinical Investigation*, 133(13), 2023.

Minjung Kyung, Jeff Gill, Malay Ghosh, and George Casella. Penalized regression, standard errors, and bayesian lassos. *Bayesian Analysis*, 5(2):369–411, 2010. doi: 10.1214/10-BA607.

J. Li, Z. Wang, R. Li, and R. Wu. Bayesian group lasso for nonparametric varying-coefficient models with application to functional genome-wide association studies. *The annals of applied statistics*, 9(2):640, 2015.

Nan Li and Wensheng Zhu. A bayesian approach for subgroup analysis. *Biometrical Journal*, 65(2200231), 2023. doi: 10.1002/bimj.202200231.

Qing Li and Nan Lin. The bayesian elastic net. *Bayesian Analysis*, 5(1):151–170, 2010.

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

C. Liu, J. Ma, and C. I. Amos. Bayesian variable selection for hierarchical gene–environment and gene–gene interactions. *Human genetics*, 134(1):23–36, 2015.

Wei Liu, Jun Ma, Yong Cheng, Hongbo Zhang, Wengguang Luo, and Hongyan Zhang. Hmmr antisense rna 1, a novel long noncoding rna, regulates the progression of basal-like breast cancer cells. *Breast Cancer: Targets and Therapy*, pages 223–229, 2016. doi: 10.2147/BCTT.S119997.

Xi Lu and Cen Wu. *Bayenet: Bayenet: Bayesian Quantile Elastic Net for Genetic Study*, 2023. URL https://CRAN.R-project.org/package=Bayenet. R package version 0.1.0.

S. Ma and J. Huang. A concave pairwise fusion approach to subgroup analysis. *Journal of the American Statistical Association*, 112(517):410–423, 2017. doi: 10.1080/01621459.2016.1148039.

S. Ma, L. Yang, R. Romero, and Y. Cui. Varying coefficient model for gene–environment interaction: a non-linear look. *Bioinformatics*, 27(15):2119–2126, 2011.

Timothy Marsh and Jayanta Debnath. Autophagy suppresses breast cancer metastasis by degrading nbr1. *Autophagy*, 16(6):1164–1165, 2020. doi: 10.1080/15548627.2020.1753001.

C.L. Martin, D. Jima, G.C. Sharp, L.E. McCullough, S.S. Park, K.M. Gowdy, D. Skaar, M. Cowley, R.L. Maguire, B. Fuemmeler, D. Collier, C.L. Relton, S.K. Murphy, and C. Hoyo. Maternal pre-pregnancy obesity, offspring cord blood dna methylation, and offspring cardiometabolic health in early childhood: an epigenome-wide association study. *Epigenetics*, 14(4):325–340, 2019.

T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988.

F. Morrison, K. Johnstone, A. Murray, J. Locke, and L. W. Harries. Oxidative metabolism genes are not responsive to oxidative stress in rodent beta cell lines. *Experimental diabetes research*, page 793783, 2012.

B. Mukherjee, J. Ahn, S. B. Gruber, and N. Chatterjee. Testing gene-environment interaction in large-scale case-control association studies: possible choices and comparisons. *American journal of epidemiology*, 175(3):177–190, 2012.

C. E. Murcray, J. P. Lewinger, and W. J. Gauderman. Gene-environment interaction in genome-wide association studies. *American journal of epidemiology*, 169(2):219–226, 2009.

Y. S. Niu, N. Hao, and H. H. Zhang. Interaction screening by partial correlation. *Statistics and Its Interface*, 11(2):317–325, 2018.

Joseph O Ogutu, Torben Schulz-Streeck, and Hans-Peter Piepho. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. In *BMC proceedings*, volume 6, pages 1–6. Springer, 2012.

J. W. Osborne and A. Overbay. The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, 9, 2004.

Larsson P., Pettersson D., Engqvist H., Rönnerman E., Forssell-Aronsson E., Kovács A., Karlsson P.and Helou k., and Parris T. Pan-cancer analysis of genomic and transcriptomic data reveals the prognostic relevance of human proteasome genes in different cancer types. *BMC Cancer*, 22(993), 2022. doi: 10.1186/s12885-022-10079-4.

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

Javier Perez-Peña, Verónica Corrales-Sánchez, Eitan Amir, Atanasio Pandiella3, and Alberto Ocana1. Ubiquitin-conjugating enzyme e2t(ube2t) and denticleless protein homolog (dtl) are linked to poor outcome in breast and lung cancers. *SCIENTIfIC REPORTS*, 7(17530), 2017. doi: 10.1038/s41598-017-17836-7.

J. Ren, Y. Du, S. Li, S. Ma, Y. Jiang, and C. Wu. Robust network-based regularization and variable selection for high-dimensional genomic data in cancer prognosis. *Genetic epidemiology*, 43(3):276–291, 2019.

J. Ren, F. Zhou, X. Li, Q. Chen, H. Zhang, S. Ma, Y. Jiang, and C. Wu. Semiparametric bayesian variable selection for gene-environment interactions. *Statistics in Medicine*, 39 (5):617–638, 2020.

J. Ren, F. Zhou, X. Li, S. Ma, Y. Jiang, and C. Wu. Robust bayesian variable selection for gene-environment interactions. *Biometrics*, 79(2):684–694, 2023.

V. Ročková and E. I. George. The spike-and-slab lasso. *Journal of the American Statistical Association*, 113(521):431–444, 2018.

Naijun Sha, Mahlet G Tadesse, and Marina Vannucci. Bayesian variable selection for the analysis of microarray data with censored outcomes. *Bioinformatics*, 22(18):2262–2268, 2006.

X. Shi, J. Liu, J. Huang, Y. Zhou, Y. Xie, and S. Ma. A penalized robust method for identifying gene–environment interactions. *Genetic epidemiology*, 38(3):220–230, 2014.

N. I. Simonds, A. A. Ghazarian, C. B. Pimentel, S. D. Schully, G. L. Ellison, E. M. Gillanders, and L. E. Mechanic. Review of the gene-environment interaction literature in cancer: what do we know? *Genetic epidemiology*, 40(5):356–365, 2016.

Martin Slawski. The structured elastic net for quantile regression and support vector classification. *Statistics and Computing*, 22:153–168, 2012.

Artem Sokolov, Daniel E Carlin, Evan O Paull, Robert Baertsch, and Joshua M Stuart. Pathway-based genomics prediction using generalized elastic net. *PLoS computational biology*, 12(3):e1004790, 2016.

Francesco C Stingo, Yian A Chen, Mahlet G Tadesse, and Marina Vannucci. Incorporating biological information into linear models: A bayesian approach to the selection of pathways and genes. *The annals of applied statistics*, 5(3), 2011.

E. Suzuki, I. Imoto, and A. Pimkhaokham. Prtfdc1, a possible tumor-suppressor gene, is frequently silenced in oral squamous- cell carcinomas by aberrant promoter hypermethylation. *Oncogene*, 26:7921–32, 2007.

Zaixiang Tang, Yueping Shen, Xinyan Zhang, and Nengjun Yi. The spike-and-slab lasso cox model for survival prediction and associated genes detection. *Bioinformatics*, 33(18): 2799–2807, 2017.

Zaixiang Tang, Shufeng Lei, Xinyan Zhang, Zixuan Yi, Boyi Guo, Jake Y Chen, Yueping Shen, and Nengjun Yi. Gsslasso cox: a bayesian hierarchical model for predicting survival and detecting associated genes by incorporating pathway information. *BMC bioinformatics*, 20(1):1–15, 2019.

D. Thomas. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. *Annual review of public health*, 31:21–36, 2010.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *J. R. Statist. Soc. B*, 67(1):91–108, 2005.

E. Uffelmann, Q. Q. Huang, N. S. Munung, J. De Vries, Y. Okada, A. R. Martin, ..., and D. Posthuma. Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1): 1–21, 2021.

E. Von Mutius. Gene-environment interactions in asthma. *Journal of Allergy and Clinical Immunology*, 123(1):3–11, 2009.

Patrik Waldmann, Gábor Mészáros, Birgit Gredler, Christian Fuerst, and Johann Sölkner. Evaluation of the lasso and the elastic net in genome-wide association studies. *Frontiers in genetics*, 4:270, 2013.

H. Wang, M. Ye, Y. Fu, A. Dong, M. Zhang, L. Feng, X. Zhu, W. Bo, L. Jiang, C. Griffin, D. Liang, and R. Wu. Modeling genome-wide by environment interactions through omnigenic interactome networks. *Cell Reports*, 35(6):109114, 2021.

L. Wang, J. Zhou, and A. Qu. Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics*, 68(2):353–360, 2012a.

Lan Wang, Yichao Wu, and Runze Li. Quantile regression for analyzing heterogeneity in

ultra-high dimension. *Journal of the American Statistical Association*, 107(497):214–222, 2012b. doi: 10.1080/01621459.2012.656014.

S.J. Winham, A.B. Cuellar-Barboza, A. Oliveros, S.L. McElroy, S. Crow, C. Colby, Chauhan M. Frye M. Choi, D.S., and J.M. Biernacka. Genome-wide association study of bipolar disorder accounting for effect of body mass index identifies a new risk allele in tcf7l2. *Molecular Psychiatry*, 19:1010–1016, 2014.

C. Wu and Y. Cui. Boosting signals in gene-based association studies via efficient snp selection. *Briefings in Bioinformatics*, 15(2):279–291, 2013a.

C. Wu and Y. Cui. A novel method for identifying nonlinear gene–environment interactions in case–control association studies. *Human genetics*, 132(12):1413–1425, 2013b.

C. Wu, S. Li, and Y. Cui. Genetic association studies: an information content perspective. *Current genomics*, 13(7):566–573, 2012.

C. Wu, Y. Cui, and S. Ma. Integrative analysis of gene–environment interactions under a multi-response partially linear varying coefficient model. *Statistics in medicine*, 33(28): 4988–4998, 2014.

C. Wu, X. Shi, Y. Cui, and S. Ma. A penalized robust semiparametric approach for gene–environment interactions. *Statistics in medicine*, 34(30):4016–4030, 2015.

C. Wu, Y. Jiang, J. Ren, Y. Cui, and S. Ma. Dissecting gene-environment interactions: A penalized robust approach accounting for hierarchical structures. *Statistics in medicine*, 37(3):437–456, 2018a.

C. Wu, P. S. Zhong, and Y. Cui. Additive varying-coefficient model for nonlinear gene-environment interactions. *Statistical applications in genetics and molecular biology*, 17(2), 2018b.

C. Wu, F. Zhou, J. Ren, X. Li, Y. Jiang, and S. Ma. A selective review of multi-level omics data integration using variable selection. *High-throughput*, 8(1):4, 2019.

Cen Wu and Shuangge Ma. A selective review of robust variable selection with applications in bioinformatics. *Briefings in Bioinformatics*, 16(5):873–883, 2015. doi: 10.1093/bib/bbu046.

Yichao Wu and Yufeng Liu. Variable selection in quantile regression. *Statistica Sinica*, pages 801–817, 2009.

J. Xu and Y. Guo. Identification of gene loci that overlap between mental disorders and poor prognosis of cancers. *Frontiers in Psychiatry*, 12:678943, 2021.

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

Y. Xu, M. Wu, Q. Zhang, and S. Ma. Robust identification of gene-environment interactions for prognosis using a quantile partial correlation approach. *Genomics*, 111(5):1115–1123, 2019.

Zhang Y., Wang H. J., and Z. Zhu. Robust subgroup identification. *Statistica Sinica*, 29: 1873–1889, 2019.

Q. Yao, Z. Li, and D. Chen. Review of linc00707: A novel lncrna and promising biomarker for human diseases. *Frontiers in Cell and Developmental Biology*, 10:813963, 2022.

K. Yu and R.A. Moyeed. Bayesian quantile regression. *Statistics and Probability Letters*, 54 (4):437–447, 2001.

K. Yu and J. Zhang. A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics - Theory and Methods*, 34(9-10):1867—1879, 2005.

Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *J. R. Statist. Soc. B*, 68(1):49–67, 2006.

CunHui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, 2010. doi: 10.1214/09-AOS729.

Lin Zhang, Veerabhadran Baladandayuthapani, Ganiraju C. Manyam Bani K. Mallick, Pa tricia 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.

S. Zhang, Y. Xue, Q. Zhang, C. Ma, M. Wu, and S. Ma. Identification of gene–environment interactions with marginal penalization. *Genetic epidemiology*, 44(2):159–196, 2020.

N. Zhao, H. Zhang, J. J. Clark, A. Maity, and M. C. Wu. Composite kernel machine regression based on likelihood ratio test for joint testing of genetic and gene–environment interaction effect. *Biometrics*, 75(2):625–637, 2019.

F. Zhou, J. Ren, G. Li, Y. Jiang, X. Li, W. Wang, and C. Wu. Penalized variable selection for lipid–environment interactions in a longitudinal lipidomics study. *Genes*, 10(12):1002, 2019.

Fei Zhou, Jie Ren, Xi Lu, Shuangge Ma, and Cen Wu. Gene–environment interaction: A variable selection perspective. *Epistasis: Methods and Protocols*, pages 191–223, 2021.

Fei Zhou, Xi Lu, Jie Ren, Kun Fan, Shuangge Ma, and Cen Wu. Sparse group variable selection for gene–environment interactions in the longitudinal study. *Genetic epidemiology*, pages 1–24, 2022. doi: 10.1198/016214508000000337.

Fei Zhou, Jie Ren, Shuangge Ma, and Cen Wu. The bayesian regularized quantile varying coefficient model. *Computational Statistics & Data Analysis*, page 107808, 2023.

H. Zhu, J. Klement, C. Lu, P. Redd, D. Yang, A. Smith, D. B. Poschel, J. Zou, D. Liu, P. Wang, D. Ostrov, N. Coant, Y. A. Hannun, A. H. Colby, Grinstaff M. W., and K. Liu. Asah2 represses the p53–hmox1 axis to protect myeloid-derived suppressor cells from ferroptosis. *J Immunol*, 206(6):1395–1404, 2021.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B*, 67(2):301–320, 2005.

# Appendix A

# Appendices for chapter 2

## A.1 Additional simulation results

### A.1.1 Identification results in simulation

**Table A.1**: *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 |
| N(0,1) | SD | 0.0059 | 0.0019 | 0.0051 | 0.0037 |
| Error 2 | AUC | 0.8187 | 0.9255 | 0.8877 | 0.9769 |
| $t(2)$ | SD | 0.0142 | 0.0524 | 0.0057 | 0.0048 |
| Error 3 | AUC | 0.5333 | 0.5533 | 0.8239 | 0.9459 |
| Lognormal(0,2) | SD | 0.0096 | 0.0656 | 0.1045 | 0.0162 |
| Error 4 | AUC | 0.8113 | 0.9122 | 0.9111 | 0.9849 |
| 90%N(0,1)+10%Cauchy(0,1) | SD | 0.0166 | 0.0502 | 0.0083 | 0.0033 |
| Error 5 | AUC | 0.7425 | 0.8086 | 0.9076 | 0.9856 |
| 80%N(0,1)+20%Cauchy(0,1) | SD | 0.0241 | 0.0746 | 0.0065 | 0.0024 |

**Table A.2**: *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 | 7.50(0.86) | 6.70(1.49) | 14.20(1.83) |
| N(0,1) | BLSS | 7.60(0.67) | 10.20(0.09) | 17.80(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) | 9.30(2.98) |
| $t(2)$ | BLSS | 6.33(2.09) | 7.57(3.15) | 13.90(4.73) |
| | LADBL | 7.07(0.94) | 5.97(1.61) | 13.03(1.96) |
| | LADBLSS | 7.40(0.62) | 9.20(1.94) | 16.60(2.11) |
| Error 3 | BL | 0.77(0.86) | 0.73(0.94) | 1.50(1.11) |
| Lognormal(0,2) | BLSS | 0.57(1.01) | 0.67(1.06) | 1.23(1.77) |
| | LADBL | 5.90(1.65) | 3.50(1.96) | 9.40(2.43) |
| | LADBLSS | 5.67(1.73) | 9.00(2.35) | 14.67(3.73) |
| Error 4 | BL | 6.03(2.19) | 4.40(2.44) | 10.43(4.17) |
| 90%N(0,1) | BLSS | 6.03(2.57) | 8.00(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) | 10.00(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) | 11.80(5.65) |
| +20%Cauchy(0,1) | LADBL | 7.47(0.97) | 5.43(1.77) | 12.90(2.04) |
| | LADBLSS | 7.37(0.85) | 10.47(1.46) | 17.83(1.91) |

**Table A.3**: *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 |
| N(0,1) | SD | 0.0041 | 0.0022 | 0.0054 | 0.0028 |
| Error 2 | AUC | 0.8323 | 0.9461 | 0.8972 | 0.9833 |
| $t(2)$ | SD | 0.0117 | 0.0342 | 0.0062 | 0.0028 |
| Error 3 | AUC | 0.5268 | 0.5531 | 0.8415 | 0.9595 |
| Lognormal(0,2) | SD | 0.0127 | 0.0590 | 0.0107 | 0.0156 |
| Error 4 | AUC | 0.8261 | 0.9323 | 0.9245 | 0.9889 |
| 90%N(0,1)+10%Cauchy(0,1) | SD | 0.0191 | 0.0352 | 0.0056 | 0.0034 |
| Error 5 | AUC | 0.7533 | 0.8591 | 0.9204 | 0.9862 |
| 80%N(0,1)+20%Cauchy(0,1) | SD | 0.0201 | 0.0657 | 0.0067 | 0.0114 |

**Table A.4**: *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 | 7.70(0.47) | 6.80(1.63) | 14.50(1.79) |
| N(0,1) | BLSS | 7.63(0.72) | 10.93(0.98) | 18.57(1.22) |
| | LADBL | 7.70(0.75) | 7.33(1.95) | 15.03(2.14) |
| | LADBLSS | 7.87(0.35) | 10.33(1.35) | 18.20(1.45) |
| Error 2 | BL | 6.57(1.87) | 4.47(1.69) | 11.03(2.88) |
| $t(2)$ | BLSS | 6.60(1.57) | 8.40(2.51) | 15.00(3.68) |
| | LADBL | 7.57(0.62) | 5.77(1.50) | 13.33(1.77) |
| | LADBLSS | 7.43(0.68) | 9.30(2.15) | 16.73(2.43) |
| Error 3 | BL | 0.50(0.73) | 0.83(1.02) | 1.33(1.47) |
| Lognormal(0,2) | BLSS | 0.70(0.99) | 0.40(0.86) | 1.10(1.54) |
| | LADBL | 6.13(2.05) | 3.80(1.39) | 9.93(1.32) |
| | LADBLSS | 6.63(1.16) | 10.10(1.73) | 16.73(2.52) |
| Error 4 | BL | 5.73(2.82) | 4.30(2.64) | 10.03(5.11) |
| 90%N(0,1) | BLSS | 5.73(3.02) | 7.67(4.19) | 13.40(7.05) |
| +10%Cauchy(0,1) | LADBL | 7.80(0.48) | 6.87(1.61) | 14.67(1.54) |
| | LADBLSS | 7.83(0.38) | 10.50(1.25) | 18.33(1.39) |
| Error 5 | BL | 5.60(2.61) | 2.93(2.23) | 8.53(4.27) |
| 80%N(0,1) | BLSS | 5.27(2.27) | 6.90(3.64) | 12.17(5.66) |
| +20%Cauchy(0,1) | LADBL | 7.87(0.35) | 6.87(1.45) | 14.73(1.46) |
| | LADBLSS | 7.70(0.53) | 10.70(1.12) | 18.40(1.28) |

## A.2 Estimation results for data analysis

Table A.5: Analysis of the NHS T2D data using LAD-BLSS.

| | | | Interactions | | | | |
|---|---|---|---|---|---|---|---|
| SNP | Gene* | Main Effects | age | act | trans | ceraf | chol |
| rs17011106 | WDFY4 | -0.024 | | | | | |
| rs7077294 | KIAA1217 | | | | | | -0.0491 |
| rs7093682 | RP11-170M17.1 | | | | -0.1239 | | |
| rs17011106 | WDFY4 | -0.0953 | | | | | |
| rs10826028 | MIR3924 | | | | | -0.0524 | |
| rs4748996 | THNSL1 | | | | | | 0.0064 |
| rs2646392 | KRT8P37 | 0.0148 | | | | | |
| rs7904629 | RP11-170M17.1 | | | | -0.0592 | | |
| rs1244416 | ATP5C1 | | | | | | 0.0851 |
| rs4838643 | WDFY4 | -0.0051 | | | | | |
| rs1916458 | RP11-170M17.1 | | | | -0.0264 | | |
| rs1537615 | RP11-526P5.2 | 0.0477 | | | | | |
| rs2765398 | KRT8P37 | -0.0157 | | | | | |
| rs4317891 | CELF2 | | 0.0647 | | | | |
| rs7922793 | LINC00845 | -0.0345 | | | | | |
| rs1916412 | RP11-170M17.1 | | | | -0.0614 | | |
| rs1916411 | RP11-170M17.1 | | | | -0.0448 | | |
| rs4747800 | KRT8P37 | 0.0036 | | | | | |
| rs11258040 | CAMK1D | | | | | -0.0983 | |
| rs1984275 | RP11-319F12.2 | | | | | 0.0065 | |

Continued on the next page

Table A.5: Continued from the previous page.

| | | | Interactions | | | | |
|---|---|---|---|---|---|---|---|
| SNP | Gene* | Main Effects | age | act | trans | ceraf | chol |
| rs17432763 | MIR5100 | | | | | | -0.0677 |
| rs10796113 | FRMD4A | -0.0931 | | | | | |
| rs224765 | RP11-490O24.2 | | -0.0521 | | | | |
| rs6482387 | KIAA1217 | | | | | | 0.011 |
| rs1492608 | ENKUR | | | | | | -0.0287 |
| rs11257323 | ECHDC3 | | | | | | 0.0084 |
| rs4434904 | KIAA1217 | | | | | | -0.0374 |
| rs10994364 | ANK3 | | 0.1086 | | | | |
| rs12220246 | KIAA1462 | | | | 0.0371 | | |
| rs11010390 | RP11-309N24.1 | | 0.0271 | | | | |
| rs10828584 | KIAA1217 | 0.087 | | | | | |
| rs10857590 | ARHGAP22 | | | | -0.1379 | | |
| rs1537616 | RP11-526P5.2 | 0.086 | | | | | |
| rs17295031 | KIAA1462 | | | | -0.0468 | | |
| rs10905778 | RP11-271F18.4 | | | | | | 0.0055 |
| rs7093161 | SNRPEP8 | | -0.0577 | | | | 0.0107 |
| rs2377872 | CHAT | | | | | -0.0213 | |
| rs1916409 | RP11-170M17.1 | | | | -0.0642 | | |
| rs2245456 | MALRD1 | | -0.0042 | | | | |
| rs787116 | RP11-478H13.1 | | 0.0259 | | | | |
| rs2817825 | RP11-492M23.2 | | 0.0278 | | | | |
| rs11255338 | KIN | | | | | | -0.0401 |

Continued on the next page

Table A.5: Continued from the previous page.

| | | | Interactions | | | | |
|---|---|---|---|---|---|---|---|
| SNP | Gene* | Main Effects | age | act | trans | ceraf | chol |
| rs17011115 | WDFY4 | -0.0045 | | | | | |
| rs11010821 | Y-RNA | -0.025 | | | | | |
| rs2532760 | RP11-492M23.2 | | 0.0272 | | | | |
| rs10821773 | ANK3 | | -0.0107 | | | | |
| rs17454012 | CELF2 | | -0.1076 | | | | |
| rs4372368 | RP11-478B11.2 | | -0.0449 | | | | |
| rs1916420 | RP11-170M17.1 | | | | 0.0885 | | |
| rs2446588 | FRMD4A | | -0.0142 | | | | |
| rs10995687 | RP11-170M17.1 | | | | 0.1065 | | |
| rs161279 | RP11-192P3.5 | | | | 0.0333 | | |
| rs161279 | ZEB1 | | | | 0.0333 | | |
| rs161258 | ZEB1 | | | | 0.0362 | | |
| rs10509149 | TMEM26 | | | | | | -0.0428 |
| rs3740000 | LINC00837 | -0.106 | | | | | |
| rs17314489 | ZNF365 | | 0.1543 | | | | |
| rs17453876 | CELF2 | | 0.0518 | | | | |
| rs10793451 | ZNF485 | | | | -0.1028 | | |
| rs4749527 | KIAA1462 | | | | 0.0093 | | |
| rs12570207 | SEPHS1 | | -0.0329 | | | | |
| rs902904 | THNSL1 | | | | | | -0.0885 |
| rs7921813 | CAMK1D | | 0.0063 | | | | |
| rs10218945 | SNRPEP8 | | | | | | -0.0351 |

Continued on the next page

Table A.5: Continued from the previous page.

| | | | Interactions | | | | |
|---|---|---|---|---|---|---|---|
| SNP | Gene* | Main Effects | age | act | trans | ceraf | chol |
| rs2804551 | RP11-492M23.2 | 0.0616 | | | | | |
| rs12266433 | CELF2 | | -0.0301 | | | | |
| rs16919385 | PLXDC2 | | | | | | 0.0112 |
| rs4750039 | CELF2 | | 0.0333 | | | | |
| rs12249964 | KIAA1217 | | | | | | 0.0607 |
| rs4745829 | RP11-170M17.1 | | | | -0.0252 | | |
| rs11257932 | CAMK1D | | 0.0256 | | | | |
| rs10827602 | RP11-810B23.1 | | | 0.0167 | | | |
| rs7081466 | RP11-526P5.2 | | | | | 0.0267 | |
| rs12256642 | THNSL1 | | | | | | -0.0258 |
| rs2796304 | RP11-492M23.2 | | 0.0875 | | | | |
| rs10826964 | ZEB1 | | | | 0.0316 | | |
| rs11257933 | CAMK1D | | 0.0547 | | | | |
| rs17432532 | MIR5100 | | | | | | 0.017 |
| rs10826899 | UBE2V2P1 | | | | | -0.0234 | |
| rs11592473 | UBE2V2P1 | | | | | -0.0642 | |
| rs12764778 | OR13A1 | | 0.0263 | | | | |
| rs12762462 | GPR158 | | | | -0.0153 | | |
| rs1011763 | MIR3924 | | | | | 0.1871 | |
| rs1916450 | RP11-170M17.1 | | | | -0.1767 | | |
| rs1917814 | CHAT | | | | | -0.0194 | |
| rs6602809 | DCLRE1CP1 | 0.0043 | | | | | |

Continued on the next page

Table A.5: Continued from the previous page.

| | | | Interactions | | | | |
|---|---|---|---|---|---|---|---|
| SNP | Gene* | Main Effects | age | act | trans | ceraf | chol |
| rs923757 | THNSL1 | | | | | | 0.0226 |
| rs7092368 | RP11-526P5.2 | | | | | -0.0531 | |
| rs6602806 | DCLRE1CP1 | 0.0527 | | | | | |
| rs6602806 | ACBD7 | 0.0527 | | | | | |
| rs10994308 | ANK3 | | | | | | -0.0124 |
| rs224699 | RP11-490O24.2 | | -0.0351 | | | | |
| rs7083349 | KIAA1217 | -0.0651 | | | | | |
| rs10828905 | RNU6-632P | | -0.0799 | | | | |
| rs10764441 | KIAA1217 | -0.0447 | | | | | |
| rs10752217 | CELF2 | | 0.02 | | | | |
| rs17566968 | CDC123 | | | | | | -0.057 |
| rs7093183 | KIAA1217 | | | | | | -0.0422 |
| rs2887230 | RP11-478H13.3 | 0.0864 | | | | | |
| rs1761379 | ZEB1 | | | | 0.1187 | | |
| rs7097429 | ALOX5 | | 0.0789 | | | | |

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

Table A.6: Inclusion probability of the NHS T2D data using LADBLSS.

| SNP | Gene* | Main Effects | age | act | trans | ceraf | chol |
|---|---|---|---|---|---|---|---|
| rs17011106 | WDFY4 | 0.9930 | | | | | |
| rs7077294 | KIAA1217 | | | | | | 0.9736 |

Continued on the next page

Table A.6: Continued from the previous page.

| SNP | Gene* | Main Effects | age | act | trans | ceraf | chol |
|---|---|---|---|---|---|---|---|
| rs7093682 | RP11-170M17.1 | | | | 0.9938 | | |
| rs17011106 | WDFY4 | 0.9900 | | | | | |
| rs10826028 | MIR3924 | | | | | 0.9612 | |
| rs4748996 | THNSL1 | | | | | | 0.9834 |
| rs2646392 | KRT8P37 | 0.9818 | | | | | |
| rs7904629 | RP11-170M17.1 | | | | 0.9646 | | |
| rs1244416 | ATP5C1 | | | | | | 0.9656 |
| rs4838643 | WDFY4 | 0.9768 | | | | | |
| rs1916458 | RP11-170M17.1 | | | | 0.9832 | | |
| rs1537615 | RP11-526P5.2 | 0.9956 | | | | | |
| rs2765398 | KRT8P37 | 0.9756 | | | | | |
| rs4317891 | CELF2 | | 1.000 | | | | |
| rs7922793 | LINC00845 | 0.9744 | | | | | |
| rs1916412 | RP11-170M17.1 | | | | 0.9774 | | |
| rs1916411 | RP11-170M17.1 | | | | 0.9700 | | |
| rs4747800 | KRT8P37 | 0.9840 | | | | | |
| rs11258040 | CAMK1D | | | | | 0.9738 | |
| rs1984275 | RP11-319F12.2 | | | | | 0.9952 | |
| rs17432763 | MIR5100 | | | | | | 0.9862 |
| rs10796113 | FRMD4A | 0.9636 | | | | | |
| rs224765 | RP11-490O24.2 | | 0.9710 | | | | |
| rs6482387 | KIAA1217 | | | | | | 0.9638 |
| rs1492608 | ENKUR | | | | | | 0.9680 |

Continued on the next page

Table A.6: Continued from the previous page.

| SNP | Gene* | Main Effects | age | act | trans | ceraf | chol |
|---|---|---|---|---|---|---|---|
| rs11257323 | ECHDC3 | | | | | | 0.9892 |
| rs4434904 | KIAA1217 | | | | | | 0.9716 |
| rs10994364 | ANK3 | | 0.9942 | | | | |
| rs12220246 | KIAA1462 | | | | 0.9610 | | |
| rs11010390 | RP11-309N24.1 | | 0.9820 | | | | |
| rs10828584 | KIAA1217 | 0.9752 | | | | | |
| rs10857590 | ARHGAP22 | | | | 0.9848 | | |
| rs1537616 | RP11-526P5.2 | 0.9944 | | | | | |
| rs17295031 | KIAA1462 | | | | 0.9816 | | |
| rs10905778 | RP11-271F18.4 | | | | | | 0.9988 |
| rs7093161 | SNRPEP8 | | 0.9542 | | | | 0.9902 |
| rs2377872 | CHAT | | | | | 0.9728 | |
| rs1916409 | RP11-170M17.1 | | | | 0.9612 | | |
| rs2245456 | MALRD1 | | 0.9630 | | | | |
| rs787116 | RP11-478H13.1 | | 0.9638 | | | | |
| rs2817825 | RP11-492M23.2 | | 0.9550 | | | | |
| rs11255338 | KIN | | | | | | 0.9964 |
| rs17011115 | WDFY4 | 0.9712 | | | | | |
| rs11010821 | Y-RNA | 0.9916 | | | | | |
| rs2532760 | RP11-492M23.2 | | 0.9720 | | | | |
| rs10821773 | ANK3 | | 0.9586 | | | | |
| rs17454012 | CELF2 | | 0.9998 | | | | |
| rs4372368 | RP11-478B11.2 | | 0.9618 | | | | |

Continued on the next page

Table A.6: Continued from the previous page.

| SNP | Gene* | Main Effects | age | act | trans | ceraf | chol |
|---|---|---|---|---|---|---|---|
| rs1916420 | RP11-170M17.1 | | | | 0.9672 | | |
| rs2446588 | FRMD4A | | 0.9724 | | | | |
| rs10995687 | RP11-170M17.1 | | | | 0.9588 | | |
| rs161279 | RP11-192P3.5 | | | | 0.9770 | | |
| rs161279 | ZEB1 | | | | 0.9770 | | |
| rs161258 | ZEB1 | | | | 0.9876 | | |
| rs10509149 | TMEM26 | | | | | | 0.9726 |
| rs3740000 | LINC00837 | 0.9964 | | | | | |
| rs17314489 | ZNF365 | | 0.9866 | | | | |
| rs17453876 | CELF2 | | 0.9952 | | | | |
| rs10793451 | ZNF485 | | | | 0.9604 | | |
| rs4749527 | KIAA1462 | | | | 0.9794 | | |
| rs12570207 | SEPHS1 | | 0.9698 | | | | |
| rs902904 | THNSL1 | | | | | | 0.9884 |
| rs7921813 | CAMK1D | | 0.9998 | | | | |
| rs10218945 | SNRPEP8 | | | | | | 0.9612 |
| rs2804551 | RP11-492M23.2 | 0.9848 | | | | | |
| rs12266433 | CELF2 | | 0.9618 | | | | |
| rs16919385 | PLXDC2 | | | | | | 0.9806 |
| rs4750039 | CELF2 | | 0.9910 | | | | |
| rs12249964 | KIAA1217 | | | | | | 0.9558 |
| rs4745829 | RP11-170M17.1 | | | | 0.9940 | | |
| rs11257932 | CAMK1D | | 0.9826 | | | | |

Continued on the next page

Table A.6: Continued from the previous page.

| SNP | Gene* | Main Effects | age | act | trans | ceraf | chol |
|---|---|---|---|---|---|---|---|
| rs10827602 | RP11-810B23.1 | | | 0.9728 | | | |
| rs7081466 | RP11-526P5.2 | | | | | 0.9714 | |
| rs12256642 | THNSL1 | | | | | | 0.9616 |
| rs2796304 | RP11-492M23.2 | | 0.9928 | | | | |
| rs10826964 | ZEB1 | | | | 0.9592 | | |
| rs11257933 | CAMK1D | | 0.9726 | | | | |
| rs17432532 | MIR5100 | | | | | | 0.9834 |
| rs10826899 | UBE2V2P1 | | | | | 0.9784 | |
| rs11592473 | UBE2V2P1 | | | | | 0.9864 | |
| rs12764778 | OR13A1 | | 0.9894 | | | | |
| rs12762462 | GPR158 | | | | 0.9636 | | |
| rs1011763 | MIR3924 | | | | | 0.9954 | |
| rs1916450 | RP11-170M17.1 | | | | 0.9820 | | |
| rs1917814 | CHAT | | | | | 0.9670 | |
| rs6602809 | DCLRE1CP1 | 0.9614 | | | | | |
| rs923757 | THNSL1 | | | | | | 0.9964 |
| rs7092368 | RP11-526P5.2 | | | | | 0.9868 | |
| rs6602806 | DCLRE1CP1 | 0.9912 | | | | | |
| rs6602806 | ACBD7 | 0.9912 | | | | | |
| rs10994308 | ANK3 | | | | | | 0.9542 |
| rs224699 | RP11-490O24.2 | | 0.9768 | | | | |
| rs7083349 | KIAA1217 | 0.9986 | | | | | |
| rs10828905 | RNU6-632P | | 0.9626 | | | | |

Continued on the next page

Table A.6: Continued from the previous page.

| SNP | Gene* | Main Effects | age | act | trans | ceraf | chol |
|---|---|---|---|---|---|---|---|
| rs10764441 | KIAA1217 | 0.9896 | | | | | |
| rs10752217 | CELF2 | | 0.9762 | | | | |
| rs17566968 | CDC123 | | | | | | 0.9838 |
| rs7093183 | KIAA1217 | | | | | | 0.9954 |
| rs2887230 | RP11-478H13.3 | 0.9588 | | | | | |
| rs1761379 | ZEB1 | | | | 0.9682 | | |
| rs7097429 | ALOX5 | | 0.9974 | | | | |

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

## A.3 Posterior inference

### A.3.1 LADBL

#### Hierarchical model specification

$$Y_i = E_i\alpha + C_i\gamma + X_{ij}\beta_j + \tilde{W}_i\eta_j + \tau^{-1/2}\xi_2\sqrt{v_i}z_i \quad i = 1, \ldots, n$$

$$v_i|\tau \overset{iid}{\sim} \tau\exp(-\tau v_i) \quad i = 1, \ldots, n$$

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

$$\beta_j|s_1 \sim \frac{1}{\sqrt{2\pi s_1}}\exp(-\frac{\beta_j^2}{2s_1})$$

$$s_1|\varphi_1^2 \sim \frac{\varphi_1^2}{2}\exp(-\frac{\varphi_1^2}{2}s_1)$$

$$\eta_{jk}|s_{2k} \overset{iid}{\sim} \frac{1}{\sqrt{2\pi s_{2k}}}\exp(-\frac{\eta_{jk}^2}{2s_{2k}}) \quad k = 1, \ldots, q$$

$$s_{2k}|\varphi_2^2 \overset{iid}{\sim} \frac{\varphi_2^2}{2}\exp(-\frac{\varphi_2^2}{2}s_{2k}) \quad k = 1, \ldots, q$$

$$\alpha_k \overset{iid}{\sim} \frac{1}{\sqrt{(2\pi\alpha_0)}}\exp(-\frac{\alpha_k^2}{2\alpha_0}) \quad k = 1, \ldots, q$$

$$\gamma_t \overset{iid}{\sim} \frac{1}{\sqrt{(2\pi\gamma_0)}}\exp(-\frac{\gamma_t^2}{2\gamma_0}) \quad t = 1, \ldots, m$$

$$\tau \sim \mathrm{Gamma}(a, b)$$

$$\varphi_1^2 \sim \mathrm{Gamma}(c_1, d_1)$$

$$\varphi_2^2 \sim \mathrm{Gamma}(c_2, d_2)$$

**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

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

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

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

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

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

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:

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

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 $\varphi_1^2$ 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 $\varphi_1^2$ is Gamma$(c_1+1, s_1/2+d_1)$.

The full conditional posterior distribution of $\varphi_2^2$ 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 $\varphi_2^2$ is Gamma$(c_2+q, \sum_{k=1}^{q} s_{2k}/2+d_2)$.

The full conditional posterior distribution of $\tau$:

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

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 $v_i$ is:

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

Therefore,

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

### A.3.2 BLSS

**Hierarchical model specification**

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2\right\}$$

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

$$\gamma \sim \mathrm{N}_m(0,\, \Sigma_{\gamma 0})$$

$$\beta_j|\pi_c, \tau_c^2, \sigma^2 \sim (1-\pi_c)\,\mathrm{N}\left(0,\, \sigma^2\tau_c^2\right) + \pi_c\,\delta_0(\beta_j) \quad j = 1, \ldots, p$$

$$\eta_{jk}|\pi_e, \tau_{ek}^2, \sigma^2 \overset{iid}{\sim} (1-\pi_e)\,\mathrm{N}\left(0,\, \sigma^2\tau_{ek}^2\right) + \pi_e\,\delta_0(\eta_{jk}) \quad j = 1, \ldots, p,\, k = 1, \ldots, q$$

$$\tau_c^2|\lambda_c^2 \sim \text{Gamma}(1, \frac{\lambda_c^2}{2})$$

$$\tau_{ek}^2|\lambda_e^2 \overset{iid}{\sim} \text{Gamma}(1, \frac{\lambda_e^2}{2}) \quad k = 1, \ldots, q$$

$$\pi_c \sim \text{Beta}\,(r_c,\, u_c)$$

$$\pi_e \sim \text{Beta}\,(r_e,\, u_e)$$

$$\lambda_c^2 \sim \text{Gamma}\,(a_c,\, b_c)$$

$$\lambda_e^2 \sim \text{Gamma}\,(a_e,\, b_e)$$

$$\sigma^2 \sim \text{Inverse-Gamma}\,(s,\, h)$$

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

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

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

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

The posterior of $\tau_c^2$ 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 $\tau_{ek}^2$ is:

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

$\lambda_c^2$ and $\lambda_e^2$ have Gamma posterior distributions:

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

$\pi_c$ and $\pi_e$ 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

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

### A.3.3 BL

**Hierarchical model specification**

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - E_i\alpha - C_i\gamma - X_{ij}\beta_j - \tilde{W}_i\eta_j)^2\right\}$$

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

$$\gamma \sim \mathrm{N}_m(0,\, \Sigma_{\gamma 0})$$

$$\beta_j|\tau_c^2, \sigma^2 \sim \mathrm{N}\left(0,\, \sigma^2\tau_c^2\right) \quad j = 1, \ldots, p$$

$$\eta_{jk}|\tau_{ek}^2, \sigma^2 \overset{iid}{\sim} \mathrm{N}\left(0,\, \sigma^2\tau_{ek}^2\right) \quad j = 1, \ldots, p,\, k = 1, \ldots, q$$

$$\tau_c^2|\lambda_c^2 \sim \exp(\frac{\lambda_c^2}{2})$$

$$\tau_{ek}^2|\lambda_e^2 \overset{iid}{\sim} \exp(\frac{\lambda_e^2}{2}) \quad k = 1, \ldots, q$$

$$\lambda_c^2 \sim \mathrm{Gamma}\,(a_c,\, b_c)$$

$$\lambda_e^2 \sim \mathrm{Gamma}\,(a_e,\, b_e)$$

$$\sigma^2 \propto \frac{1}{\sigma^2}$$

**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

$$\begin{aligned}\mu_{\beta_j} &= \Sigma_{\beta_j} X_j^\top (Y - \mu_{(-\beta_j)}),\\ \Sigma_{\beta_j} &= \left(X_j^\top X_j + \frac{1}{\tau_c^2}\right)^{-1}.\end{aligned}$$

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

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

The posterior of $\tau_c^2$ 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 $\tau_{ek}^2$ is:

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

$\lambda_c^2$ and $\lambda_e^2$ have Gamma posterior distributions:

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

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

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

# Appendix B

# Appendices for chapter 3

## B.1 Additional simulation results

### B.1.1 Identification results in simulation

**Table B.1**: *Simulation results of the second setting with i.i.d. errors. n=300, p=400. Mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

| | | bqenetss | bqenet | benet | qbl | blss | bl | enet | lasso |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\theta = 0.3$ | | | | |
| **Error1** | TP | 24.00(0.01) | 23.70(0.53) | 23.47(0.63) | 23.43(0.94) | 24.00(0.01) | 23.70(0.65) | 23.40(0.89) | 22.93(1.39) |
| | FP | 1.57(1.45) | 8.90(5.23) | 43.60(6.84) | 3.73(2.38) | 0.27(0.52) | 7.77(3.26) | 44.23(11.93) | 37.20(11.55) |
| | ESM | 0.36(0.19) | 5.06(1.02) | 15.18(2.44) | 3.47(0.73) | 0.22(0.09) | 3.78(0.67) | 4.83(1.74) | 4.89(1.88) |
| **Error2** | TP | 23.17(1.09) | 19.03(2.19) | 16.97(3.06) | 16.70(2.76) | 16.43(4.42) | 15.77(3.69) | 22.27(1.11) | 21.57(1.59) |
| | FP | 0.63(0.89) | 2.13(1.43) | 82.03(7.18) | 0.70(0.79) | 1.70(1.39) | 11.17(5.47) | 41.03(11.27) | 35.50(10.96) |
| | ESM | 1.39(1.20) | 6.97(1.00) | 82.68(52.15) | 6.03(1.01) | 9.50(4.96) | 20.21(8.35) | 6.82(1.14) | 7.39(1.38) |
| **Error3** | TP | 24.00(0.01) | 22.40(1.28) | 19.57(2.58) | 21.60(1.69) | 119.93(3.52) | 19.33(3.07) | 22.77(1.33) | 22.23(1.59) |
| | FP | 0.02(0.01) | 2.33(1.71) | 71.87(12.31) | 0.60(0.81) | 1.50(2.05) | 9.87(3.57) | 41.30(13.00) | 36.13(14.43) |
| | ESM | 0.11(0.05) | 4.83(0.86) | 49.26(22.65) | 3.35(0.79) | 4.71(3.82) | 12.4(4.74) | 5.71(1.81) | 6.15(2.06) |
| **Error4** | TP | 21.70(1.93) | 18.53(2.06) | 17.13(2.05) | 6.53(2.18) | 18.43(2.30) | 17.43(2.39) | 22.80(0.96) | 22.30(1.29) |
| | FP | 1.57(1.72) | 2.23(1.74) | 77.73(6.33) | 1.33(1.32) | 1.73(1.66) | 10.37(4.45) | 47.47(11.06) | 40.80(10.96) |
| | ESM | 3.07(2.19) | 7.78(1.13) | 60.74(12.00) | 7.12(1.39) | 6.93(2.83) | 15.2(2.69) | 6.36(1.70) | 6.90(2.08) |
| **Error5** | TP | 19.57(2.08) | 17.73(1.93) | 16.60(2.59) | 14.87(2.39) | 16.57(2.70) | 16.20(1.97) | 22.37(1.07) | 21.77(1.30) |
| | FP | 2.60(2.09) | 2.23(1.52) | 83.23(5.41) | 1.20(1.03) | 1.27(1.05) | 9.53(3.75) | 43.00(18.07) | 34.37(12.94) |
| | ESM | 5.61(2.33) | 8.34(0.99) | 81.08(18.82) | 7.86(1.28) | 9.07(3.40) | 17.92(4.34) | 6.92(1.52) | 7.42(1.77) |
| | | | | | $\theta = 0.5$ | | | | |
| **Error1** | TP | 23.97(0.18) | 23.70(0.47) | 23.13(0.97) | 23.73(0.45) | 24.00(0.01) | 23.80(0.41) | 23.47(0.86) | 23.17(1.12) |
| | FP | 0.87(1.01) | 6.03(4.27) | 42.33(6.60) | 3.20(3.26) | 0.37(0.76) | 6.73(3.07) | 40.43(12.49) | 37.50(15.05) |
| | ESM | 0.30(0.13) | 4.62(0.76) | 15.52(2.12) | 3.21(0.60) | 0.24(0.08) | 3.57(0.54) | 4.34(1.33) | 4.58(1.53) |
| **Error2** | TP | 23.50(0.78) | 20.43(2.05) | 17.87(2.70) | 18.67(2.37) | 18.47(3.22) | 17.50(2.94) | 22.77(1.33) | 22.13(1.57) |
| | FP | 0.30(0.60) | 2.23(1.30) | 78.33(10.05) | 0.77(1.07) | 1.87(1.38) | 10.60(4.21) | 44.67(11.05) | 37.33(11.15) |
| | ESM | 0.82(0.55) | 6.21(0.91) | 59.74(23.71) | 5.13(1.07) | 6.59(4.31) | 15.55(5.97) | 6.09(1.46) | 6.69(1.67) |
| **Error3** | TP | 23.97(0.18) | 22.00(1.31) | 19.50(2.32) | 20.73(1.53) | 20.07(3.42) | 19.37(2.68) | 23.27(0.78) | 22.77(0.97) |
| | FP | 0.17(0.38) | 2.50(1.91) | 71.97(9.21) | 0.57(0.82) | 1.30(1.29) | 11.03(4.04) | 42.20(12.76) | 34.13(11.98) |
| | ESM | 0.29(0.16) | 5.66(1.06) | 47.69(24.23) | 4.32(1.21) | 5.09(4.41) | 12.91(5.51) | 5.73(1.80) | 6.13(2.11) |
| **Error4** | TP | 22.87(1.04) | 19.70(1.84) | 17.70(2.48) | 16.97(2.67) | 18.90(2.47) | 17.63(2.55) | 22.87(1.01) | 22.33(1.24) |
| | FP | 0.40(0.56) | 1.80(1.16) | 78.00(7.00) | 0.90(0.96) | 1.47(1.28) | 8.07(3.34) | 44.80(14.36) | 37.43(14.65) |
| | ESM | 1.53(0.86) | 7.01(0.94) | 58.16(10.71) | 6.30(1.25) | 5.63(2.53) | 14.01(2.74) | 6.07(1.31) | 6.64(1.66) |
| **Error5** | TP | 19.87(1.91) | 17.27(2.29) | 17.20(2.25) | 14.43(2.34) | 16.33(3.74) | 15.40(3.19) | 21.83(3.45) | 21.17(3.39) |
| | FP | 1.87(1.50) | 2.27(1.64) | 80.37(6.05) | 0.90(1.16) | 1.83(1.68) | 11.33(5.28) | 38.83(11.50) | 35.33(12.08) |
| | ESM | 4.98(2.01) | 7.99(1.18) | 50.12(42.87) | 7.35(1.34) | 9.36(4.11) | 33.10(80.31) | 7.33(2.87) | 7.85(3.17) |
| | | | | | $\theta = 0.7$ | | | | |
| **Error1** | TP | 23.95(0.22) | 23.76(0.44) | 23.10(1.00) | 23.57(0.75) | 24.00(0.01) | 23.86(0.36) | 23.43(0.68) | 23.14(0.85) |
| | FP | 1.38(1.24) | 11.00(11.02) | 42.52(6.18) | 4.19(4.34) | 0.19(0.40) | 6.71(3.04) | 47.43(12.82) | 42.05(8.02) |
| | ESM | 0.34(0.15) | 5.25(1.19) | 14.66(2.41) | 3.54(0.68) | 0.23(0.09) | 3.57(0.55) | 4.43(1.33) | 4.56(1.47) |
| **Error2** | TP | 22.73(1.34) | 19.57(2.42) | 15.10(4.62) | 16.50(3.63) | 15.13(6.03) | 13.57(5.99) | 21.20(3.01) | 20.50(3.30) |
| | FP | 0.53(1.04) | 2.33(1.83) | 83.30(9.81) | 0.93(1.11) | 1.47(0.82) | 11.17(3.27) | 38.07(10.28) | 33.00(11.25) |
| | ESM | 1.75(1.30) | 6.80(1.14) | 161.11(24.45) | 5.88(1.55) | 11.88(9.34) | 32.67(37.73) | 8.14(3.95) | 8.69(4.17) |
| **Error3** | TP | 22.23(1.94) | 20.57(2.37) | 18.23(3.52) | 18.17(3.31) | 19.00(4.02) | 18.13(3.98) | 23.00(0.91) | 22.27(1.48) |
| | FP | 1.37(1.43) | 2.97(2.40) | 76.77(10.78) | 1.53(1.48) | 1.43(1.43) | 9.47(3.74) | 39.50(11.59) | 32.17(7.56) |
| | ESM | 2.11(1.59) | 7.22(1.02) | 57.28(34.19) | 6.33(1.41) | 6.20(5.60) | 13.98(7.00) | 5.97(1.81) | 6.40(2.11) |
| **Error4** | TP | 22.13(1.68) | 19.67(1.83) | 17.90(2.25) | 17.20(2.30) | 18.80(2.77) | 17.97(1.81) | 22.93(1.08) | 22.53(1.14) |
| | FP | 1.53(1.87) | 3.23(2.21) | 80.10(3.83) | 1.93(1.51) | 1.60(1.38) | 10.43(3.88) | 44.00(11.18) | 38.33(11.07) |
| | ESM | 2.70(2.08) | 8.28(1.37) | 57.64(9.24) | 7.55(1.69) | 6.40(3.06) | 14.64(2.69) | 6.05(1.68) | 6.48(2.07) |
| **Error5** | TP | 18.80(2.31) | 17.30(2.18) | 16.03(2.30) | 14.57(2.47) | 16.30(2.09) | 16.07(2.35) | 22.27(1.20) | 21.60(1.38) |
| | FP | 3.30(2.22) | 1.80(1.35) | 81.63(5.54) | 1.17(1.29) | 2.60(1.38) | 11.70(4.97) | 41.40(10.81) | 39.23(11.52) |
| | ESM | 6.42(2.76) | 8.73(1.49) | 76.34(17.30) | 8.36(1.72) | 9.39(2.88) | 19.25(4.17) | 6.83(1.72) | 7.44(1.96) |

**Table B.2**: *Simulation results of the second setting with heterogeneous random errors. n=300, p=400. Mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

| | | bqenetss | bqenet | benet | qbl | blss | bl | enet | lasso |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\theta = 0.3$ | | | | |
| **Error1** | TP | 24.00(0.01) | 22.93(1.01) | 21.80(1.49) | 22.43(1.45) | 23.67(0.66) | 22.87(1.07) | 23.37(0.56) | 22.97(0.81) |
| | FP | 0.47(0.73) | 4.87(4.68) | 55.47(7.18) | 2.30(1.60) | 0.53(0.63) | 8.47(3.60) | 41.40(11.21) | 36.53(12.80) |
| | ESM | 0.35(0.15) | 5.76(1.09) | 24.38(4.19) | 4.54(0.91) | 0.96(0.63) | 6.55(1.39) | 4.57(1.06) | 4.82(1.27) |
| **Error2** | TP | 21.93(1.44) | 17.13(2.56) | 14.27(3.66) | 14.00(3.38) | 13.03(3.91) | 13.17(4.29) | 21.33(2.12) | 20.53(2.67) |
| | FP | 0.53(1.01) | 1.27(1.17) | 84.20(6.84) | 0.50(0.78) | 1.50(1.11) | 9.53(3.56) | 41.23(9.72) | 37.77(11.49) |
| | ESM | 2.26(1.45) | 7.29(1.07) | 16.36(25.99) | 6.41(1.39) | 14.93(7.15) | 30.93(28.35) | 8.81(3.99) | 9.74(3.73) |
| **Error3** | TP | 24.00(0.01) | 21.47(1.74) | 16.33(2.84) | 19.27(2.43) | 16.73(4.16) | 15.10(4.08) | 22.43(1.63) | 21.90(2.11) |
| | FP | 0.01(0.02) | 1.93(1.26) | 82.27(11.21) | 0.47(0.63) | 1.63(1.40) | 9.97(3.70) | 39.70(10.00) | 33.63(10.59) |
| | ESM | 0.16(0.07) | 5.49(0.81) | 89.91(56.83) | 4.39(0.98) | 10.76(7.15) | 22.41(12.54) | 6.97(2.50) | 7.72(3.00) |
| **Error4** | TP | 21.20(2.35) | 16.10(2.70) | 14.17(2.78) | 13.10(2.71) | 13.33(2.52) | 13.20(2.81) | 22.13(1.33) | 21.40(1.67) |
| | FP | 0.83(0.87) | 1.43(1.07) | 87.17(6.42) | 0.53(0.86) | 1.63(1.63) | 8.93(4.18) | 39.13(14.74) | 33.37(12.08) |
| | ESM | 3.64(2.52) | 8.45(1.49) | 121.43(32.16) | 7.90(1.71) | 14.65(4.57) | 25.34(4.96) | 7.53(1.68) | 8.17(1.96) |
| **Error5** | TP | 19.33(1.83) | 17.17(2.26) | 16.57(2.87) | 14.67(2.71) | 15.67(3.20) | 15.60(2.92) | 22.53(1.17) | 22.00(1.66) |
| | FP | 2.10(1.54) | 1.63(1.40) | 84.67(8.58) | 0.93(0.98) | 1.20(1.19) | 9.57(3.37) | 39.80(12.27) | 34.23(13.72) |
| | ESM | 5.91(2.56) | 8.41(1.36) | 93.09(58.5) | 7.77(1.38) | 10.28(5.16) | 20.90(11.77) | 6.70(1.84) | 7.27(2.19) |
| | | | | | $\theta = 0.5$ | | | | |
| **Error1** | TP | 23.97(0.18) | 23.33(1.06) | 22.23(1.33) | 22.60(0.93) | 23.73(0.45) | 22.73(0.98) | 23.53(0.63) | 23.27(0.69) |
| | FP | 0.17(0.59) | 5.10(3.83) | 56.43(9.15) | 2.83(2.18) | 0.77(0.86) | 9.73(4.03) | 42.93(10.30) | 42.03(17.37) |
| | ESM | 0.22(0.16) | 5.75(1.04) | 23.05(5.08) | 4.49(0.94) | 0.68(0.36) | 6.44(1.03) | 4.46(1.07) | 4.62(1.18) |
| **Error2** | TP | 22.97(1.19) | 17.23(2.28) | 12.83(3.01) | 14.07(2.85) | 10.57(4.16) | 10.63(4.32) | 21.00(3.05) | 19.63(3.36) |
| | FP | 0.13(0.43) | 1.33(1.15) | 86.97(7.83) | 0.27(0.58) | 1.83(1.15) | 12.30(4.03) | 37.80(11.76) | 30.87(11.76) |
| | ESM | 1.27(1.14) | 6.72(0.59) | 19.26(19.48) | 5.73(0.92) | 18.35(5.87) | 42.29(30.29) | 9.25(2.96) | 10.35(3.43) |
| **Error3** | TP | 23.97(0.18) | 20.27(1.60) | 17.30(2.41) | 18.37(2.51) | 17.33(2.82) | 16.33(2.81) | 22.90(1.06) | 22.30(1.21) |
| | FP | 0.03(0.17) | 2.00(1.58) | 79.83(6.47) | 0.77(1.10) | 1.43(0.94) | 10.37(5.18) | 43.67(11.02) | 38.23(9.82) |
| | ESM | 0.28(0.18) | 5.92(0.74) | 73.21(25.39) | 4.83(0.85) | 8.30(4.05) | 17.72(4.73) | 6.29(1.43) | 6.86(1.83) |
| **Error4** | TP | 22.00(2.08) | 15.97(2.13) | 14.20(1.73) | 13.20(1.95) | 13.23(2.24) | 12.33(2.58) | 21.57(1.28) | 20.77(1.43) |
| | FP | 0.33(0.66) | 1.43(1.14) | 87.23(5.83) | 0.57(0.86) | 2.30(2.04) | 11.13(3.54) | 42.03(12.14) | 37.33(12.07) |
| | ESM | 2.20(1.88) | 7.46(1.02) | 114.01(36.62) | 6.91(1.42) | 14.44(4.57) | 26.43(5.80) | 8.02(2.17) | 8.87(2.75) |
| **Error5** | TP | 20.21(2.14) | 18.00(1.83) | 16.62(2.57) | 15.03(2.11) | 15.72(2.66) | 15.59(1.99) | 22.72(1.16) | 21.69(1.58) |
| | FP | 1.41(1.32) | 1.59(1.27) | 83.79(4.60) | 0.62(0.78) | 1.69(1.14) | 8.90(4.73) | 40.07(12.13) | 32.38(10.85) |
| | ESM | 4.36(2.40) | 7.63(1.00) | 82.04(21.45) | 6.90(1.35) | 10.14(4.1) | 18.90(5.75) | 6.34(1.62) | 6.88(1.79) |
| | | | | | $\theta = 0.7$ | | | | |
| **Error1** | TP | 23.73(0.52) | 22.43(0.97) | 21.83(0.79) | 21.83(1.37) | 23.23(0.57) | 22.13(1.01) | 23.07(0.74) | 22.70(1.18) |
| | FP | 0.60(0.93) | 4.50(3.32) | 57.40(6.95) | 2.20(1.94) | 0.63(1.13) | 8.77(3.71) | 43.20(12.88) | 38.33(13.58) |
| | ESM | 0.54(0.34) | 5.89(0.71) | 23.46(4.18) | 4.77(0.77) | 1.04(0.47) | 6.54(1.03) | 4.76(1.24) | 4.92(1.36) |
| **Error2** | TP | 22.03(1.79) | 17.20(2.61) | 14.23(3.02) | 14.47(2.81) | 13.57(3.48) | 12.27(4.09) | 21.57(1.48) | 20.37(1.87) |
| | FP | 0.70(0.79) | 1.30(1.32) | 87.57(6.78) | 0.43(0.86) | 1.23(1.38) | 11.37(4.33) | 41.30(14.76) | 35.77(11.25) |
| | ESM | 2.27(1.87) | 7.51(1.02) | 127.16(62.14) | 6.65(1.31) | 13.58(6.26) | 28.75(12.68) | 7.98(2.12) | 8.54(2.36) |
| **Error3** | TP | 21.93(2.15) | 17.40(2.88) | 16.13(3.09) | 15.37(3.40) | 15.03(5.05) | 14.87(3.83) | 21.73(1.55) | 21.30(1.74) |
| | FP | 0.57(0.90) | 2.13(1.63) | 82.33(8.61) | 0.90(1.16) | 1.53(1.48) | 12.33(5.03) | 39.77(10.05) | 34.53(9.73) |
| | ESM | 2.50(2.15) | 7.62(1.17) | 99.27(60.75) | 6.96(1.35) | 11.66(7.27) | 23.70(10.61) | 7.29(2.08) | 7.87(2.44) |
| **Error4** | TP | 20.50(2.21) | 15.67(2.37) | 15.17(2.69) | 13.17(2.41) | 14.10(2.83) | 13.13(2.99) | 21.77(1.68) | 21.13(1.85) |
| | FP | 0.73(0.74) | 1.63(1.33) | 86.23(6.67) | 0.83(0.87) | 1.60(1.79) | 10.93(4.96) | 45.17(14.29) | 35.70(10.95) |
| | ESM | 3.75(2.28) | 8.61(1.04) | 98.14(22.40) | 8.13(1.34) | 12.01(4.34) | 23.57(5.86) | 7.71(1.79) | 8.26(1.94) |
| **Error5** | TP | 20.00(1.53) | 17.67(2.02) | 15.70(3.06) | 15.00(2.03) | 16.33(3.24) | 15.43(2.97) | 22.53(1.63) | 21.77(1.61) |
| | FP | 3.00(2.52) | 2.13(1.59) | 83.07(6.78) | 1.30(1.18) | 1.60(1.10) | 10.23(3.46) | 46.10(12.09) | 38.30(10.58) |
| | ESM | 5.09(1.92) | 8.35(0.89) | 87.38(54.62) | 7.84(0.99) | 9.13(5.03) | 19.80(8.01) | 6.98(1.38) | 7.42(1.58) |

## B.2 Assessment of the convergence of MCMC chains

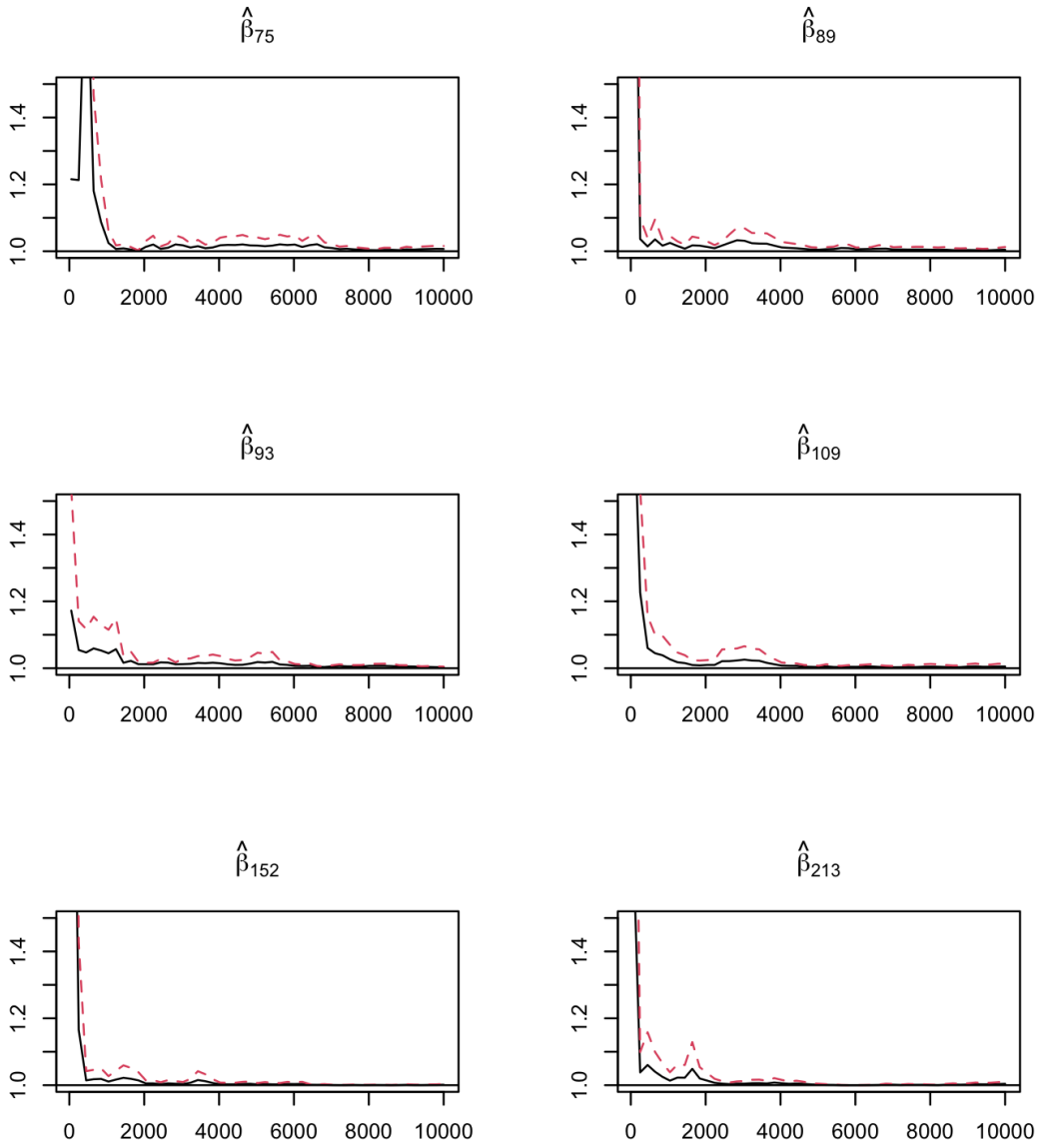


**Figure B.1**: *Potential scale reduction factor (PSRF) against iterations for the coefficients of the selected sixed genetic factors in setting 1 under error 2. Black line: the PSRF. Red dotted line: the upper limits of the 95% confidence interval for the PSRF. The* $\hat{\beta}_{75}, \hat{\beta}_{89}, \hat{\beta}_{93}, \hat{\beta}_{109}, \hat{\beta}_{152}, \hat{\beta}_{213}$ *represents six estimated coefficients of the non-zero main genetic effects.*

## B.3 Estimation results for data analysis

**Table B.3**: *The analysis results of the NHS data using bqenetss.*

| SNP | Gene | Coefficients | Posterior Inclusion Probability |
|---|---|---|---|
| rs2368945 | RP11-492M23.2 | -0.2341 | 0.7534 |
| rs983814 | ASAH2 | -0.0090 | 0.5314 |
| rs11256106 | RP11-428L9.2 | 0.2322 | 0.8130 |
| rs3006258 | RNU6-632P | 0.0001 | 0.7441 |
| rs41408444 | PRTFDC1 | 0.1984 | 0.7481 |
| rs12255028 | ASAH2 | 0.0848 | 0.5896 |
| rs11010821 | Y_RNA | -0.1395 | 0.6615 |
| rs12411923 | MPP7 | 0.2351 | 0.8252 |
| rs3740000 | LINC00837 | 0.3386 | 0.9566 |
| rs4749924 | IL2RA | -0.2294 | 0.7948 |
| rs1556397 | KIAA1217 | 0.2278 | 0.7928 |
| rs2804551 | RP11-492M23.2 | 0.3619 | 0.9640 |
| rs4748991 | PRTFDC1 | 0.2266 | 0.8032 |
| rs4948293 | ARID5B | 0.2738 | 0.8614 |
| rs11815954 | ASAH2 | -0.0414 | 0.5563 |
| rs1886995 | PRTFDC1 | 0.2468 | 0.8300 |
| rs11014290 | PRTFDC1 | 0.2957 | 0.9024 |
| rs7067773 | ASAH2 | 0.3498 | 0.9674 |
| rs7083059 | LINC00707 | 0.2014 | 0.7544 |
| rs7096851 | PRKG1 | 0.1301 | 0.6348 |

**Table B.4**: *The analysis results of the NHS data using bqenet.*

| SNP | Gene | Coefficients |
|---|---|---|
| rs780844 | CUBN | 0.1062 |
| rs16928572 | MPP7 | 0.1158 |
| rs11256531 | TCEB1P3 | 0.1204 |
| rs1539445 | RP11-336A10.5 | 0.0992 |
| rs4749165 | RP11-128B16.3 | 0.1083 |
| rs11259039 | FRMD4A | 0.0906 |
| rs17226417 | RP11-482E14.2 | 0.1044 |
| rs11010821 | Y_RNA | 0.0993 |
| rs12411923 | MPP7 | 0.1333 |
| rs3740000 | LINC00837 | 0.0892 |
| rs2642192 | ZNF438 | 0.0952 |
| rs4922535 | GDF10 | 0.1166 |

**Table B.5**: *The analysis results of the NHS data using benet.*

| SNP | Gene | Coefficients |
|---|---|---|
| rs1937653 | PRKG1 | -0.1932 |
| rs7085788 | CDK1 | -0.1785 |
| rs2153122 | LINC00709 | -0.1383 |
| rs11258614 | FRMD4A | -0.1538 |
| rs11008324 | UBE2V2P1 | -0.1475 |
| rs7100461 | PRKG1-AS1 | -0.1643 |
| rs10796113 | FRMD4A | -0.1507 |
| rs6481648 | RP11-224P11.1 | 0.1387 |
| rs11256531 | TCEB1P3 | 0.2126 |
| rs4749165 | RP11-128B16.3 | 0.1768 |
| rs7913773 | PLXDC2 | -0.1837 |
| rs7921651 | FAM107B | 0.1655 |
| rs17226417 | RP11-482E14.2 | 0.1928 |
| rs9423936 | FAM208B | 0.1622 |
| rs11010821 | Y_RNA | 0.2245 |
| rs2266374 | FAM107B | 0.1866 |
| rs1194657 | RP11-346D6.4 | -0.2129 |
| rs12411923 | MPP7 | 0.1994 |
| rs7101305 | KRT8P37 | -0.1587 |
| rs4922535 | GDF10 | 0.1619 |
| rs2451918 | MKX | -0.1617 |
| rs2049587 | RNU6-15P | 0.1429 |
| rs17140609 | ST8SIA6 | -0.1652 |
| rs7083059 | LINC00707 | 0.2424 |
| rs2884505 | PLXDC2 | 0.2166 |

**Table B.6**: *The analysis results of the NHS data using qbl.*

| SNP | Gene | Coefficients |
|---|---|---|
| rs2764351 | RP11-733D4.1 | -0.1807 |
| rs17415597 | RP11-271F18.4 | -0.2228 |
| rs11256531 | TCEB1P3 | 0.2093 |
| rs7913773 | PLXDC2 | -0.1846 |
| rs12411923 | MPP7 | 0.2213 |
| rs17140609 | ST8SIA6 | -0.1798 |

**Table B.7**: *The analysis results of the NHS data using blss.*

| SNP | Gene | Coefficients | Posterior Inclusion Probability |
|---|---|---|---|
| rs10764477 | KIAA1217 | 0.0184 | 0.5031 |
| rs11010821 | Y_RNA | 0.3419 | 0.7886 |
| rs12411923 | MPP7 | 0.2544 | 0.6349 |
| rs7083059 | LINC00707 | 0.2554 | 0.6055 |

**Table B.8**: *The analysis results of the NHS data using bl.*

| SNP | Gene | Coefficients |
|---|---|---|
| rrs1937653 | PRKG1 | -0.2442 |
| rs7085788 | CDK1 | -0.2065 |
| rs11258614 | FRMD4A | -0.2004 |
| rs17504713 | KIAA1217 | 0.2339 |
| rs17415597 | RP11-271F18.4 | -0.3222 |
| rs7100461 | PRKG1-AS1 | -0.1904 |
| rs2224371 | TACC1P1 | 0.1963 |
| rs10828230 | LUZP4P1 | 0.2325 |
| rs11256531 | TCEB1P3 | 0.2555 |
| rs4749165 | RP11-128B16.3 | 0.21305 |
| rs7913773 | PLXDC2 | -0.2333 |
| rs7921651 | FAM107B | 0.1982 |
| rs17226417 | RP11-482E14.2 | 0.2092 |
| rs11010821 | Y_RNA | 0.3445 |
| rs2266374 | FAM107B | 0.2746 |
| rs2496722 | PARD3 | 0.2307 |
| rs1194657 | RP11-346D6.4 | -0.2744 |
| rs12411923 | MPP7 | 0.2407 |
| rs1556397 | KIAA1217 | -0.3671 |
| rs3861046 | RP11-478B11.2 | -0.2463 |
| rs12146414 | CUBN | -0.2687 |
| rs17140609 | ST8SIA6 | -0.1898 |
| rs7083059 | LINC00707 | 0.5830 |
| rs2884505 | PLXDC2 | 0.2550 |

## B.4 Posterior inference

### B.4.1 Bayesian quantile elastic net

#### Hierarchical model specification

$$
\begin{aligned}
Y_i &= \boldsymbol{C}_i^\top \boldsymbol{\gamma} + \boldsymbol{X}_i^\top \boldsymbol{\beta} + \xi_1 v_1 + \tau^{-1/2}\xi_2\sqrt{v_i}z_i \quad i = 1, \ldots, n \\
v_i|\tau &\overset{iid}{\sim} \tau \exp(-\tau v_i) \quad i = 1, \ldots, n \\
z_i &\overset{iid}{\sim} \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2}) \quad i = 1, \ldots, n \\
\beta_j|t_j, \pi_1, \eta_2 &\overset{iid}{\sim} \sqrt{\frac{\eta_2 t_j}{\pi(t_j - 1)}}\exp\{-\frac{1}{2}(\frac{t_j - 1}{2\eta_2 t_j})^{-1}\beta_j^2\} \quad j = 1, \ldots, p \\
t_j|\eta_1 &\overset{iid}{\sim} \Gamma^{-1}(\frac{1}{2}, \eta_1) t_j^{-\frac{1}{2}} \eta_1^{\frac{1}{2}} \exp\{-\eta_1 t_j\}\mathrm{I}(t_j > 1) \quad j = 1, \ldots, p \\
\gamma_k &\overset{iid}{\sim} \frac{1}{\sqrt{(2\pi\gamma_0)}}\exp(-\frac{\gamma_k^2}{2\gamma_0}) \quad k = 1, \ldots, q \\
\tau &\sim \text{Gamma}(a, b) \\
\eta_1 &\sim \text{Gamma}(c_1, d_1) \\
\eta_2 &\sim \text{Gamma}(c_2, d_2)
\end{aligned}
$$

#### Gibbs Sampler

Let $\mu_{(-\gamma_k)} = \boldsymbol{C}_i^\top \boldsymbol{\gamma} + \boldsymbol{X}_i^\top \boldsymbol{\beta} - C_{ik}\gamma_k$, $(i = 1, \ldots, n), (k = 1, \ldots, q)$, then

$$
\begin{aligned}
&\pi(\gamma_k|\text{rest}) \\
&\quad \propto \pi(\gamma_k)\pi(\boldsymbol{Y}|\cdot) \\
&\quad \propto \exp\Big\{ -\sum_{i=1}^{n} \frac{(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_1)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \times \exp(-\frac{\gamma_k^2}{2\gamma_0}) \\
&\quad \propto \exp\Big\{ -\frac{1}{2}\big[(\sum_{i=1}^{n} \frac{\tau C_{ik}^2}{\xi_2^2 v_i} + \frac{1}{\gamma_0})\gamma_k^2 - 2\sum_{i=1}^{n} \frac{\tau(Y_i - \mu_{(-\gamma_k)} - \xi_1 v_1)C_{ik}}{\xi_2^2 v_i}\gamma_k\big] \Big\}.
\end{aligned}
$$

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

$$\mu_{\gamma_k} = \Big(\sum_{i=1}^{n} \frac{\tau(Y_i - \mu_{(-\gamma_k)} - \xi_1 v_1)C_{ik}}{\xi_2^2 v_i}\Big)\sigma^2_{\gamma_k},$$
$$\sigma^2_{\gamma_k} = \Big(\sum_{i=1}^{n} \frac{\tau C_{ik}^2}{\xi_2^2 v_i} + \frac{1}{\gamma_0}\Big)^{-1}.$$

Let $\mu_{(-\beta_j)} = \boldsymbol{C}_i^\top \boldsymbol{\gamma} + \boldsymbol{X}_i^\top \boldsymbol{\beta} - X_{ij}\beta_j$, $(j = 1, \ldots, p)$, then

$\pi(\boldsymbol{\beta}_j|\text{rest})$

$$\propto \pi(y|\cdot)\pi(\boldsymbol{\beta}_j|t_j, \eta_2)$$
$$\propto \exp\Big\{-\sum_{i=1}^{n} \frac{(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_1)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\} \times \exp\{-\frac{1}{2}(\frac{t_j - 1}{2\eta_2 t_j})^{-1}\boldsymbol{\beta}_j^2\}$$
$$\propto \exp\Big\{-\frac{1}{2}\big[(\sum_{i=1}^{n} \frac{\tau X_{ij}^2}{\xi_2^2 v_i} + \frac{2\eta_2 t_j}{t_j - 1})\boldsymbol{\beta}_j^2 - 2\sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\beta_j)} - \xi_1 v_1)X_{ij}}{\xi_2^2 v_i}\beta_j\big]\Big\}.$$

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

$$\mu_{\beta_j} = \Big(\sum_{i=1}^{n} \frac{\tau(Y_i - \mu_{(-\beta_j)} - \xi_1 v_1)X_{ij}}{\xi_2^2 v_i}\Big)\sigma^2_{\beta_j},$$
$$\sigma^2_{\beta_j} = \Big(\sum_{i=1}^{n} \frac{\tau X_{ij}^2}{\xi_2^2 v_i} + \frac{2\eta_2 t_j}{t_j - 1}\Big)^{-1}.$$

The full conditional posterior distribution of $t_j - 1$ is:

$$
\begin{aligned}
& t_j - 1|\text{rest} \\
& \quad \propto \pi(\beta_j|t_j, \pi_1)\pi(t_j|\eta_1) \\
& \quad \propto (1-\pi_1)(2\pi(t_j-1)/(2\eta_2 t_j))^{-1/2}\exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1}\beta_j^2\} \\
& \quad \times \Gamma^{-1}(\frac{1}{2}, \eta_1) t_j^{-\frac{1}{2}} \eta_1^{\frac{1}{2}} \exp\{-\eta_1 t_j\} \mathrm{I}(t_j > 1) \\
& \quad \propto \frac{1}{\sqrt{t_j - 1}} \exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1}\beta_j^2\} \exp(-\eta_1 t_j) \mathrm{I}(t_j > 1) \\
& \quad \propto \frac{1}{\sqrt{t_j - 1}} \exp\Big\{ -\frac{1}{2}[2\eta_1(t_j - 1) + \frac{2\eta_2\beta_j^2}{t_j - 1}]\} \mathrm{I}(t_j > 1).
\end{aligned}
$$

Therefore, the posterior distribution for $(t_j - 1)^{-1}$ is Inverse-Gaussian$(\sqrt{\frac{\eta_1}{\eta_2\beta_j^2}}, 2\eta_1)$.

The full conditional posterior distribution of $\eta_1$ is:

$$
\begin{aligned}
& \eta_1|\text{rest} \\
& \quad \propto \pi(t_j|\eta_1)\pi(\eta_1) \\
& \quad \propto \eta_1^{c_1 - 1}\exp(-d_1\eta_1) \prod_{j=1}^{p} \Gamma^{-1}(\frac{1}{2}, \eta_1) \eta_1^{\frac{1}{2}} \exp\{-\eta_1 t_j\} \\
& \quad \propto \Gamma^{-p}(\frac{1}{2}, \eta_1) \eta_1^{p/2 + c_1 - 1} \exp\{-\eta_1[d_1 + \sum_{j=1}^{p} t_j]\}.
\end{aligned}
$$

We use a Metropolis-Hastings within Gibbs sampling. The proposal distribution for the Metropolis-Hastings step is $\eta_1^{p+c_1-1}\exp\{-\eta_1[d_1 + \sum_{j=1}^{p}(t_j - 1)]\}$.

The full conditional posterior distribution of $\eta_2$ is:

$$
\begin{aligned}
& \eta_2|\text{rest} \\
& \quad \propto \pi(\beta_j|\eta_2)\pi(\eta_2) \\
& \quad \propto (2\pi(t_j-1)/(2\eta_2 t_j))^{-1/2}\exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1}\beta_j^2\} \\
& \quad \times \eta_2^{c_2 - 1}\exp(-d_2\eta_2)
\end{aligned}
$$

So, the posterior distribution for $\eta_2$ is Gamma$(c_2 + \frac{p}{2}, d_2 + \sum_{j=1}^{p} \frac{t_j}{t_j - 1}\beta_j^2)$.

The full conditional posterior distribution of $\tau$ is:

$$
\begin{aligned}
&\tau|\text{rest} \\
&\propto \pi(\boldsymbol{v}|\tau)\pi(\tau)\pi(\boldsymbol{Y}|\cdot) \\
&\propto \tau^{n/2} \exp\Big\{ -\sum_{i=1}^{n} \frac{(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_1)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \\
&\times \tau^n \exp(-\tau \sum_{i=1}^{n} v_i) \tau^{a-1} \exp(-b\tau) \\
&\propto \tau^{a+\frac{3}{2}n-1} \exp\Big\{ -\tau\big[ \sum_{i=1}^{n} (\frac{(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_1)^2}{2\xi_2^2 v_i} + v_i) + b \big] \Big\}.
\end{aligned}
$$

Therefore, the posterior distribution for $\tau$ is Gamma$(a + \frac{3}{2}n,\ \sum_{i=1}^{n} (\frac{(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_1)^2}{2\xi_2^2 v_i} + v_i) + b)$.

The full conditional posterior distribution of $v_i$ is:

$$
\begin{aligned}
&v_i|\text{rest} \\
&\propto \pi(\boldsymbol{v}|\tau)\pi(\boldsymbol{Y}|\cdot) \\
&\propto \frac{1}{\sqrt{v_i}} \exp\Big\{ -\frac{(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_1)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \times \exp(-\tau v_i) \\
&\propto \frac{1}{\sqrt{v_i}} \exp\Big\{ -\frac{1}{2}\big[ (\frac{\tau\xi_1^2}{\xi_2^2} + 2\tau) v_i + \frac{\tau (Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_1)^2}{\xi_2^2 v_i} \big] \Big\}.
\end{aligned}
$$

It can be found that

$$
\frac{1}{v_i}|\text{rest} \sim \text{ Inverse-Gaussian}\Big( \sqrt{\frac{\xi_1^2 + 2\xi_2^2}{(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_1)^2}},\ \frac{\tau\xi_1^2}{\xi_2^2} + 2\tau \Big).
$$

### B.4.2 Bayesian elastic net

#### Hierarchical model specification

$$\boldsymbol{Y} \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \boldsymbol{C}_i^\top\boldsymbol{\gamma} - \boldsymbol{X}_i^\top\boldsymbol{\beta})\right\}$$

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

$$\beta_j|\tau_j, \sigma^2, \eta_2 \overset{iid}{\sim} \mathrm{N}\left(0,\, \frac{\sigma^2}{\eta_2}\frac{\tau_j - 1}{\tau_j}\right) \quad j = 1, \ldots, p$$

$$\tau_j|\sigma^2, \eta_1 \overset{iid}{\sim} \mathrm{TG}(\frac{1}{2}, \frac{2\sigma^2}{\eta_1}, 1) \quad j = 1, \ldots, p$$

$$\eta_1 \sim \mathrm{Gamma}\,(c_1,\, d_1)$$

$$\eta_2 \sim \mathrm{Gamma}\,(c_2,\, d_2)$$

$$\sigma^2 \sim \frac{1}{\sigma^2}$$

#### Gibbs Sampler

Denote $\mu_{(-\gamma)} = \boldsymbol{C}\boldsymbol{\gamma} + \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{C}\boldsymbol{\gamma}$, then $\gamma|\text{rest} \sim \mathrm{N}(\mu_{\boldsymbol{\gamma}},\, \Sigma_{\boldsymbol{\gamma}})$, where

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

The full conditional posterior distribution of $\boldsymbol{\beta}$ is:

$$\begin{aligned}
&\boldsymbol{\beta}|\text{rest} \\
&\propto \pi(\beta|\sigma^2, \tau, \eta_2)\pi(\boldsymbol{Y}|\cdot) \\
&\propto \exp\{-\frac{1}{2\sigma^2}||\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{C}\boldsymbol{\gamma}||_2^2\}\prod_{j=1}^{p}\exp\{-\frac{1}{2}\frac{\eta_2\tau_j}{\sigma^2(\tau_j - 1)}\beta_j^2\} \\
&\propto \exp\{-\frac{1}{2\sigma^2}[\boldsymbol{\beta}^\top(\boldsymbol{X}^\top\boldsymbol{X} + \eta_2 D(\frac{\tau_1}{\tau_1 - 1}, \ldots, \frac{\tau_p}{\tau_p - 1}))\boldsymbol{\beta} - 2\boldsymbol{\beta}^\top\boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{C}\boldsymbol{\gamma})]\}.
\end{aligned}$$

Where $D(\cdot)$ means the diagonal meatrix. Denote $\boldsymbol{A} = \boldsymbol{X}^\top\boldsymbol{X} + \eta_2 D(\frac{\tau_1}{\tau_1 - 1}, \ldots, \frac{\tau_p}{\tau_p - 1})$, then

$\beta|\text{rest} \sim \text{N}(\mu_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}})$, where

$$\mu_{\boldsymbol{\beta}} = \boldsymbol{A}^{-1}\boldsymbol{X}^{\top}(\boldsymbol{Y} - \boldsymbol{C}\boldsymbol{\gamma}),$$

$$\Sigma_{\boldsymbol{\beta}} = \sigma^2\boldsymbol{A}^{-1}.$$

The full conditional posterior distribution of $\tau_j - 1$ is:

$$\begin{aligned}
&\tau_j - 1|\text{rest} \\
&\propto \pi(\beta_j|\tau_j, \sigma^2)\pi(\tau_j|\eta_1, \sigma^2) \\
&\propto \big(2\pi\sigma^2(\tau_j - 1)/(\eta_2\tau_j)\big)^{-1/2}\exp\{-\frac{1}{2}(\frac{\sigma^2(\tau_j - 1)}{\eta_2\tau_j})^{-1}\beta_j^2\} \\
&\times \Gamma^{-1}(\frac{1}{2}, \frac{\eta_1}{2\sigma^2})\tau_j^{-\frac{1}{2}}(\frac{\eta_1}{2\sigma^2})^{\frac{1}{2}}\exp\{-\frac{\eta_1\tau_j}{2\sigma^2}\}\text{I}(\tau_j > 1) \\
&\propto \big(\frac{\sigma^2(\tau_j - 1)}{\eta_2\tau_j}\big)^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{\eta_2\tau_j}{\sigma^2(\tau_j - 1)})\beta_j^2\}\tau_j^{-\frac{1}{2}}\exp(-\frac{\eta_1\tau_j}{2\sigma^2})\text{I}(\tau_j > 1) \\
&\propto \frac{1}{\sqrt{\tau_j - 1}}\exp\Big\{ -\frac{1}{2}[\frac{\eta_1}{\sigma^2}(\tau_j - 1) + \frac{\eta_2}{\sigma^2}\beta_j^2\frac{1}{\tau_j - 1}]\}\text{I}(\tau_j > 1).
\end{aligned}$$

Therefore, the posterior distribution for $(\tau_j - 1)^{-1}$ is Inverse-Gaussian$(\sqrt{\frac{\eta_1}{\eta_2\beta_j^2}}, \frac{\eta_1}{\sigma^2})$.

The full conditional posterior distribution of $\eta_1$ is:

$$\begin{aligned}
&\eta_1|\text{rest} \\
&\quad\propto \pi(\tau_j|\eta_1)\pi(\eta_1) \\
&\quad\propto \eta_1^{c_1-1}\exp(-d_1\eta_1)\prod_{j=1}^{p}\Gamma^{-1}(\frac{1}{2}, \frac{\eta_1}{2\sigma^2})\frac{\eta_1}{2\sigma^2}^{\frac{1}{2}}\exp\{-\frac{\eta_1\tau_j}{2\sigma^2}\} \\
&\quad\propto \Gamma^{-p}(\frac{1}{2}, \frac{\eta_1}{2\sigma^2})\eta_1^{p/2+c_1-1}\exp\{-\frac{\eta_1}{2\sigma^2}[2\sigma^2 d_1 + \sum_{j=1}^{p}\tau_j]\}.
\end{aligned}$$

Let $\tilde{\eta_1} = \frac{\eta_1}{2\sigma^2}$ and $\tilde{d_1} = 2\sigma^2 d_1$, we use a Metropolis-Hastings within Gibbs sampling. The proposal distribution for the Metropolis-Hastings step is $\tilde{\eta_1}^{p+c_1-1}\exp\{-\tilde{\eta_1}[\tilde{d_1} + \sum_{j=1}^{p}(\tau_j - 1)]\}$. Once we get the distribution of $\tilde{\eta_1}$, the distribution of $\eta_1$ can be conducted from $\eta_1 = 2\sigma^2\tilde{\eta_1}$.

The full conditional posterior distribution of $\eta_2$ is:

$$
\begin{aligned}
&\eta_2|\text{rest} \\
&\quad\propto \pi(\beta_j|\eta_2)\pi(\eta_2) \\
&\quad\propto \prod_{j=1}^{p}\left(\frac{2\pi\sigma^2(\tau_j-1)}{\eta_2\tau_j}\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\frac{\eta_2\tau_j}{\sigma^2(\tau_j-1)}\beta_j^2\right\} \\
&\quad\times \eta_2^{c_2-1}\exp(-d_2\eta_2) \\
&\quad\propto \eta_2^{\frac{p}{2}}\exp\left\{-\frac{1}{2\sigma^2}\eta_2\sum_{j=1}^{p}\frac{\tau_j}{\tau_j-1}\beta_j^2\right\}\eta_2^{c_2-1}\exp(-d_2\eta_2) \\
&\quad\propto \eta_2^{\frac{p}{2}+c_2-1}\exp\left\{-\left(\frac{1}{2\sigma^2}\sum_{j=1}^{p}\frac{\tau_j}{\tau_j-1}\beta_j^2+d_2\right)\eta_2\right\}
\end{aligned}
$$

So, the posterior distribution for $\eta_2$ is Gamma$(c_2+\frac{p}{2},\, d_2+\frac{1}{2\sigma^2}\sum_{j=1}^{p}\frac{\tau_j}{\tau_j-1}\beta_j^2)$.

The full conditional posterior distribution of $\sigma^2$ is:

$$
\begin{aligned}
&\sigma^2|\text{rest} \\
&\quad\propto \pi(\boldsymbol{\beta}|\sigma^2)\pi(\tau|\sigma^2)\pi(\sigma^2)\pi(\boldsymbol{Y}|\cdot) \\
&\quad\propto \Gamma^{-p}\left(\frac{1}{2},\frac{\eta_1}{2\sigma^2}\right)\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+p+1} \\
&\quad\times \exp\Big[-\frac{1}{2\sigma^2}\big(||\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{C}\boldsymbol{\gamma}||_2^2+\eta_1\sum_{j=1}^{p}\tau_j+\eta_2\sum_{j=1}^{p}\frac{\tau_j}{\tau_j-1}\beta_j^2\big)\Big]
\end{aligned}
$$

It can be done by the acceptance-rejection algorithm. By the definition of incomplete gamma

distributions, let

$$
\begin{aligned}
& f(\sigma^2) = \\
& \Gamma^{-p}(\frac{1}{2}, \frac{\eta_1}{2\sigma^2})(\frac{1}{\sigma^2})^{\frac{n}{2}+p+1} \\
& \times \exp\big[ - \frac{1}{2\sigma^2}\big(||\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{C}\boldsymbol{\gamma}||_2^2 + \eta_1 \sum_{j=1}^{p} \tau_j + \eta_2 \sum_{j=1}^{p} \frac{\tau_j}{\tau_j - 1}\beta_j^2\big)\big] \\
& \leq \Gamma^{-p}(\frac{1}{2})(\frac{1}{\sigma^2})^{a+1}\exp\{-\frac{1}{\sigma^2}b\} \\
& = \frac{\Gamma(a)\Gamma^{-p}(\frac{1}{2})}{b^a} g(\sigma^2)
\end{aligned}
$$

where $g(\cdot)$ is the pdf for Inverse-Gamma (a,b) and

$$
\begin{aligned}
a &= \frac{n}{2} + p \\
b &= \frac{1}{2}\big[||\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{C}\boldsymbol{\gamma}||_2^2 + \eta_1 \sum_{j=1}^{p} \tau_j + \eta_2 \sum_{j=1}^{p} \frac{\tau_j}{\tau_j - 1}\beta_j^2\big]
\end{aligned}
$$

We generate $z \sim g(\sigma^2)$ and $u \sim Uniform(0,1)$, and accept z if $u \leq \Gamma^p(\frac{1}{2})b^a f(z)/\Gamma(a)g(z)$.

### B.4.3 Bayesian quantile LASSO

#### Hierarchical model specification

$$Y_i = \boldsymbol{C}_i^\top \boldsymbol{\gamma} + \boldsymbol{X}_i^\top \boldsymbol{\beta} + \xi_1 v_i + \tau^{-1/2}\xi_2\sqrt{v_i}z_i \quad i = 1, \ldots, n$$

$$v_i|\tau \overset{iid}{\sim} \tau \exp(-\tau v_i) \quad i = 1, \ldots, n$$

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

$$\beta_j|s_{1j} \overset{iid}{\sim} \mathrm{N}(0, s_{1j}) \quad j = 1, \ldots, p$$

$$s_{1j}|\varphi_1^2 \overset{iid}{\sim} \frac{\varphi_1^2}{2}\exp(-\frac{\varphi_1^2}{2}s_{1j}) \quad j = 1, \ldots, p$$

$$\gamma_k \overset{iid}{\sim} \frac{1}{\sqrt{(2\pi\gamma_0)}}\exp(-\frac{\gamma_k^2}{2\gamma_0}) \quad k = 1, \ldots, q$$

$$\tau \sim \mathrm{Gamma}(a, b)$$

$$\varphi_1^2 \sim \mathrm{Gamma}(c_1, d_1)$$

#### Gibbs Sampler

Let $\mu_{(-\gamma_k)} = \boldsymbol{C}_i^\top \boldsymbol{\gamma} + \boldsymbol{X}_i^\top \boldsymbol{\beta} - C_{ik}\gamma_k$, $(i = 1, \ldots, n), (k = 1, \ldots, q)$, then

$$\begin{aligned}
&\pi(\gamma_k|\text{rest}) \\
&\quad \propto \pi(\gamma_k)\pi(\boldsymbol{Y}|\cdot) \\
&\quad \propto \exp\Big\{ -\sum_{i=1}^{n} \frac{(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta} - \xi_1 v_1)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \times \exp(-\frac{\gamma_k^2}{2\gamma_0}) \\
&\quad \propto \exp\Big\{ -\frac{1}{2}\big[(\sum_{i=1}^{n} \frac{\tau C_{ik}^2}{\xi_2^2 v_i} + \frac{1}{\gamma_0})\gamma_k^2 - 2\sum_{i=1}^{n} \frac{\tau(Y_i - \mu_{(-\gamma_k)} - \xi_1 v_1)C_{ik}}{\xi_2^2 v_i}\gamma_k\big] \Big\}.
\end{aligned}$$

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

$$\mu_{\gamma_k} = \big(\sum_{i=1}^{n} \frac{\tau(Y_i - \mu_{(-\gamma_k)} - \xi_1 v_1)C_{ik}}{\xi_2^2 v_i}\big)\sigma^2_{\gamma_k},$$
$$\sigma^2_{\gamma_k} = \big(\sum_{i=1}^{n} \frac{\tau C_{ik}^2}{\xi_2^2 v_i} + \frac{1}{\gamma_0}\big)^{-1}.$$

Let $\mu_{(-\boldsymbol{\beta}_j)} = \boldsymbol{C}_i^\top \boldsymbol{\gamma} + \boldsymbol{X}_i^\top \boldsymbol{\beta} - X_{ij}\beta_j$ $(j = 1, \ldots, p)$, then, $\beta_j|\text{rest} \sim N(\mu_{\boldsymbol{\beta}_j}, \sigma^2_{\boldsymbol{\beta}_j})$, where

$$\mu_{\boldsymbol{\beta}_j} = \big(\sum_{i=1}^{n} \frac{\tau(Y_i - \mu_{(-\boldsymbol{\beta}_j)} - \xi_1 v_1)X_{ij}}{\xi_2^2 v_i}\big)\sigma^2_{\boldsymbol{\beta}_j},$$
$$\sigma^2_{\beta_j} = \big(\sum_{i=1}^{n} \frac{\tau X_{ij}^2}{\xi_2^2 v_i} + \frac{1}{s_{1j}}\big)^{-1}.$$

The full conditional posterior distribution of $s_{1j}$ is:

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

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

The full conditional posterior distribution of $\varphi_1^2$:

$$
\begin{aligned}
&\varphi_1^2|\text{rest} \\
&\quad\propto \pi(s_1|\varphi_1^2)\pi(\varphi_1^2) \\
&\quad\propto \prod_{j=1}^{p} \frac{\varphi_1^2}{2}\exp(-\frac{\varphi_1^2 s_{1j}}{2})(\varphi_1^2)^{c_1-1}\exp(-d_1\varphi_1^2) \\
&\quad\propto (\varphi_1^2)^{p+c_1-1}\exp\Big(-\varphi_1^2(\sum_{j=1}^{p}\frac{s_{1j}}{2}+d_1)\Big).
\end{aligned}
$$

Therefore, the posterior distribution for $\varphi_1^2$ is Gamma$(p+c_1, \sum_{j=1}^{p}\frac{s_{1j}}{2}+d_1)$.

The full conditional posterior distribution of $\tau$:

$$
\begin{aligned}
&\tau|\text{rest} \\
&\quad\propto \pi(v|\tau)\pi(\tau)\pi(\boldsymbol{Y}|\cdot) \\
&\quad\propto \tau^{n/2}\exp\Big\{-\sum_{i=1}^{n}\frac{(Y_i-\boldsymbol{C}_i^\top\boldsymbol{\gamma}-\boldsymbol{X}_i^\top\boldsymbol{\beta}-\xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\} \\
&\quad\times \tau^n\exp(-\tau\sum_{i=1}^{n}v_i)\tau^{a-1}\exp(-b\tau) \\
&\quad\propto \tau^{a+\frac{3}{2}n-1}\exp\Big\{-\tau\big[\sum_{i=1}^{n}(\frac{(Y_i-\boldsymbol{C}_i^\top\boldsymbol{\gamma}-\boldsymbol{X}_i^\top\boldsymbol{\beta}-\xi_1 v_i}{2\xi_2^2 v_i}+v_i)+b\big]\Big\}.
\end{aligned}
$$

Therefore, the posterior distribution for $\tau$ is Gamma$(a+\frac{3}{2}n,\ \big[\sum_{i=1}^{n}(\frac{(Y_i-\boldsymbol{C}_i^\top\boldsymbol{\gamma}-\boldsymbol{X}_i^\top\boldsymbol{\beta}-\xi_1 v_i)^2}{2\xi_2^2 v_i}+v_i)+b\big]\big)$.

Last, we have the full conditional posterior distribution of $v_i$:

$$
\begin{aligned}
&v_i|\text{rest} \\
&\quad\propto \pi(v|\tau)\pi(\boldsymbol{Y}|\cdot) \\
&\quad\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{(Y_i-\boldsymbol{C}_i^\top\boldsymbol{\gamma}-\boldsymbol{X}_i^\top\boldsymbol{\beta}-\xi_1 v_1)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\}\times\exp(-\tau v_i) \\
&\quad\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{1}{2}\big[(\frac{\tau\xi_1^2}{\xi_2^2}+2\tau)v_i+\frac{\tau(Y_i-\boldsymbol{C}_i^\top\boldsymbol{\gamma}-\boldsymbol{X}_i^\top\boldsymbol{\beta}-\xi_1 v_1)^2}{\xi_2^2 v_i}\big]\Big\}.
\end{aligned}
$$

It can be found that

$$\frac{1}{v_i}|\text{rest} \sim \text{Inverse-Gaussian}(\sqrt{\frac{\xi_1^2+2\xi_2^2}{(Y_i-\boldsymbol{C}_i^\top\boldsymbol{\gamma}-\boldsymbol{X}_i^\top\boldsymbol{\beta}-\xi_1 v_1)^2}},\ \frac{\tau\xi_1^2}{\xi_2^2}+2\tau).$$

### B.4.4 Bayesian LASSO with spike-and-slab priors

**Hierarchical model specification**

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\boldsymbol{C}_i^\top\boldsymbol{\gamma}-\boldsymbol{X}_i^\top\boldsymbol{\beta})^2\right\}$$

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

$$\beta_j|\pi_1,\tau_j^2,\sigma^2 \overset{iid}{\sim} (1-\pi_1)\,\mathrm{N}\left(0,\ \sigma^2\tau_j^2\right)+\pi_1\,\delta_0(\beta_j) \quad j=1,\ldots,p$$

$$\tau_j^2|\lambda_1^2 \overset{iid}{\sim} \text{Gamma}(1,\frac{\lambda_1^2}{2}) \quad j=1,\ldots,p$$

$$\pi_1 \sim \text{Beta}\,(r_1,\ u_1)$$

$$\lambda_1^2 \sim \text{Gamma}\,(a_1,\ b_1)$$

$$\sigma^2 \sim \text{Inverse-Gamma}\,(s,\ h)$$

**Gibbs Sampler**

Denote $\mu_{(-\gamma)} = \boldsymbol{C}\boldsymbol{\gamma}+\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{C}\boldsymbol{\gamma}$, then $\boldsymbol{\gamma}|\text{rest} \sim \mathrm{N}(\mu_\gamma,\ \Sigma_\gamma)$, where

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

Denote $\mu_{(-\beta_j)} = \boldsymbol{C}\boldsymbol{\gamma} + \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}_j^\top \beta_j$, then $\beta_j|\text{rest} \sim (1-l_1)\text{N}(\mu_{\beta_j},\, \sigma^2\Sigma_{\beta_j}) + l_1\delta_0(\beta_j)$, where

$$\begin{aligned}
\mu_{\beta_j} &= \Sigma_{\beta_j}\boldsymbol{X}_j^\top(Y - \mu_{(-\beta_j)}),\\
\Sigma_{\beta_j} &= \left(\boldsymbol{X}_j^\top\boldsymbol{X}_j + \frac{1}{\tau_j^2}\right)^{-1},\\
l_1 &= \frac{\pi_1}{\pi_1 + (1-\pi_1)(\tau_j^2)^{-1/2}|\Sigma_{\beta_j}|^{1/2}\exp\left\{\frac{1}{2\sigma^2}\Sigma_{\beta_j}||X_j^\top(Y-\mu_{(-\beta_j)})||_2^2\right\}}.
\end{aligned}$$

The posterior of $\tau_j^2$ is:

$$\frac{1}{\tau_j^2}|\text{rest} \sim \begin{cases} \text{Inverse-Gamma}(1, \frac{\lambda_1^2}{2}) & \text{if } \beta_j = 0 \\ \text{Inverse-Gaussian}(\sqrt{\frac{\sigma^2}{\beta_j^2}\lambda_1^2},\, \lambda_1^2) & \text{if } \beta_j \neq 0 \end{cases}.$$

$\lambda_1^2$ has Gamma posterior distributions:

$$\lambda_1^2|\text{rest} \sim \text{Gamma}(a_1 + p,\, \sum_{j=1}^{p}\frac{\tau_j^2}{2} + b_1).$$

$\pi_1$ has Gamma posterior distributions:

$$\pi_1|\text{rest} \sim \text{Beta}(r_1 - \sum_{j=1}^{p}\mathbf{I}_{\{\beta_j\neq 0\}} + p,\; u_1 + \sum_{j=1}^{p}\mathbf{I}_{\{\beta_j\neq 0\}}).$$

Denote $\boldsymbol{\mu} = \boldsymbol{C}\boldsymbol{\gamma} + \boldsymbol{X}\boldsymbol{\beta}$, $\sigma^2 \sim \text{Inverse-Gamma}(\mu_{\sigma^2},\; \Sigma_{\sigma^2})$, where

$$\begin{aligned}
\mu_{\sigma^2} &= s + \frac{n + \sum_{j=1}^{p}\mathbf{I}_{\{\beta_j\neq 0\}}}{2},\\
\Sigma_{\sigma^2} &= h + \frac{(Y-\boldsymbol{\mu})^\top(Y-\boldsymbol{\mu}) + \sum_{j=1}^{p}(\tau_j^2)^{-1}\beta_j^\top\beta_j}{2}.
\end{aligned}$$

### B.4.5 Bayesian LASSO

**Hierarchical model specification**

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \boldsymbol{C}_i^\top \boldsymbol{\gamma} - \boldsymbol{X}_i^\top \boldsymbol{\beta})^2\right\}$$

$$\boldsymbol{\gamma} \sim \mathrm{N}_m(0,\, \Sigma_{\gamma 0})$$

$$\beta_j|\tau_j^2, \sigma^2 \overset{iid}{\sim} \mathrm{N}\left(0,\, \sigma^2\tau_j^2\right) \quad j = 1, \ldots, p$$

$$\tau_j^2|\lambda_1^2 \overset{iid}{\sim} \exp(\frac{\lambda_1^2}{2}) \quad k = 1, \ldots, q$$

$$\lambda_1^2 \sim \mathrm{Gamma}\,(a_c,\, b_c)$$

$$\sigma^2 \propto \frac{1}{\sigma^2}$$

**Gibbs Sampler**

Denote $\mu_{(-\boldsymbol{\gamma})} = \boldsymbol{C}\boldsymbol{\gamma} + \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{C}\boldsymbol{\gamma}$, then $\boldsymbol{\gamma}|\text{rest} \sim \mathrm{N}(\mu_{\boldsymbol{\gamma}},\, \Sigma_{\boldsymbol{\gamma}})$, where

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

Denote $\mu_{(-\boldsymbol{\beta}_j)} = \boldsymbol{C}\boldsymbol{\gamma} + \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}_j^\top\beta_j$, then $\beta_j|\text{rest} \sim \mathrm{N}(\mu_{\beta_j},\, \sigma^2\Sigma_{\beta_j})$, where

$$\mu_{\beta_j} = \Sigma_{\beta_j}\boldsymbol{X}_j^\top(\boldsymbol{Y} - \mu_{(-\boldsymbol{\beta}_j)}),$$
$$\Sigma_{\beta_j} = \left(\boldsymbol{X}_j^\top\boldsymbol{X}_j + \frac{1}{\tau_j^2}\right)^{-1}.$$

The posterior of $\tau_j^2$ is:

$$\frac{1}{\tau_j^2}|\text{rest} \sim \text{Inverse-Gaussian}(\sqrt{\frac{\sigma^2}{\beta_j^2}\lambda_1^2}, \lambda_1^2).$$

$\lambda_1^2$ has Gamma posterior distributions:

$$\lambda_1^2|\text{rest} \sim \text{Gamma}(a_1 + p,\ \sum_{j=1}^{p} \frac{\tau_j^2}{2} + b_1).$$

Denote $\boldsymbol{\mu} = \boldsymbol{C\gamma} + \boldsymbol{X\beta}$, $\sigma^2 \sim$ Inverse-Gamma$(\mu_{\sigma^2},\ \Sigma_{\sigma^2})$, where

$$\begin{aligned}
\mu_{\sigma^2} &= \frac{n+p}{2}, \\
\Sigma_{\sigma^2} &= \frac{(\boldsymbol{Y}-\boldsymbol{\mu})^\top(\boldsymbol{Y}-\boldsymbol{\mu}) + \sum_{j=1}^{p}(\tau_j^2)^{-1}\beta_j^\top\beta_j}{2}.
\end{aligned}$$

## B.5 Proof of density function(3.6)

Let $\varphi_1 = \tau\lambda_1$ and $\varphi_2 = \tau\lambda_2$, then we can set the elastic net prior on $\beta_j$ as:

$$\pi(\beta_j|\varphi_1, \varphi_2) = C(\varphi_1, \varphi_2)\frac{\varphi_1}{2}\exp\{-\varphi_1|\beta_j| - \varphi_2\beta_j^2\},$$

where $C(\varphi_1, \varphi_2)$ is a constant depending on $\varphi_1$ and $\varphi_2$. By (Andrews and Mallows (1974)), for $a > 0$, we have

$$\frac{a}{2}\exp^{-a|z|} = \int_0^\infty \frac{1}{\sqrt{2\pi s}}\exp(-\frac{z^2}{2s})\frac{a^2}{2}\exp(-\frac{a^2}{2}s)\,\mathrm{d}s.$$

Let $a = \varphi_1$, $z = \beta_j$ and $s = s_j$, we have

$$\exp(-\varphi_1|\beta_j|) = (\frac{\varphi_1}{2})^{-1}\int_0^\infty \frac{1}{\sqrt{2\pi s_j}}\exp(-\frac{\beta_j^2}{2s_j})\frac{\varphi_1^2}{2}\exp(-\frac{\varphi_1^2}{2}s_j)\,\mathrm{d}s_j.$$

Then,

$$\pi(\beta_j|\varphi_1, \varphi_2) = C(\varphi_1, \varphi_2)\int_0^\infty \frac{1}{\sqrt{2\pi s_j}}\exp(-\frac{1+2\varphi_2 s_j}{2s_j}\beta_j^2)\frac{\varphi_1^2}{2}\exp(-\frac{\varphi_1^2}{2}s_j)\,\mathrm{d}s_j.$$

Let $t_j = 1 + 2\varphi_2 s_j$, then we have

$$\pi(\beta_j|\varphi_1,\varphi_2) = C(\varphi_1,\varphi_2)\int_1^\infty \frac{t_j^{-1/2}}{\sqrt{2\pi(t_j-1)/(2\varphi_2 t_j)}}\exp(-\frac{1}{2}(\frac{t_j-1}{2\varphi_2 t_j})^{-1}\beta_j^2)\frac{\varphi_1^2}{4\varphi_2}\exp(-\frac{\varphi_1^2}{4\varphi_2}(t_j-1))\,\mathrm{d}t_j.$$

Note $C(\varphi_1,\varphi_2) = \Gamma^{-1}(\frac{1}{2},\,\frac{\varphi_1^2}{4\varphi_2})(\frac{\varphi_1^2}{4\varphi_2})^{-1/2}\exp(-\frac{\varphi_1^2}{4\varphi_2})$. Let $\eta_1 = \frac{\varphi_1^2}{4\varphi_2}$ and $\eta_2 = \varphi_2$, we have

$$\begin{aligned}\pi(\beta_j|\eta_1,\eta_2) &= C(\eta_1,\eta_2)\int_1^\infty \frac{t_j^{-1/2}}{\sqrt{2\pi(t_j-1)/(2\eta_2 t_j)}}\exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1}\beta_j^2\}\eta_1\exp(-\eta_1(t_j-1))\,\mathrm{d}t_j\\ &= \Gamma^{-1}(\frac{1}{2},\,\eta_1)\eta_1^{-1/2}\exp(-\eta_1)\int_1^\infty \frac{t_j^{-1/2}}{\sqrt{2\pi(t_j-1)/(2\eta_2 t_j)}}\exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1}\beta_j^2\}\times\\ &\quad \eta_1\exp(-\eta_1(t_j-1))\,\mathrm{d}t_j\\ &= \int_1^\infty \frac{1}{\sqrt{2\pi(t_j-1)/(2\eta_2 t_j)}}\exp\{-\frac{1}{2}(\frac{t_j-1}{2\eta_2 t_j})^{-1}\beta_j^2\}\Gamma^{-1}(\frac{1}{2},\,\eta_1)\eta_1^{\frac{1}{2}}\exp(-\eta_1 t_j)t_j^{-1/2}\,\mathrm{d}t_j.\end{aligned}$$

## B.6 More numerical results

We have conducted additional simulation studies to demonstrate that Qenetss can consistently outperform alternatives in different settings. Here we 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 $\delta$ 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 $\delta = r\sqrt{q_1(1-q_1)q_2(1-q_2)}$.

**Table B.9**: *Simulation results of the homogeneous errors. n=300, p=400. Mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

| | | bqenetss | bqenet | benet | qbl | blss | bl | enet | lasso |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\theta = 0.3$ | | | | |
| **Error1** | TP | 25.00(0.01) | 23.36(1.57) | 22.64(1.63) | 22.73(1.27) | 24.91(0.30) | 24.09(1.14) | 22.18(2.09) | 22.00(1.95) |
| | FP | 3.00(1.84) | 7.00(4.07) | 47.18(8.82) | 5.64(2.77) | 0.64(0.81) | 15.36(4.08) | 57.82(10.06) | 49.55(12.92) |
| | ESM | 0.21(0.07) | 2.68(0.44) | 7.82(1.82) | 2.19(0.31) | 0.13(0.04) | 2.46(0.37) | 3.51(0.84) | 3.50(0.980 |
| **Error2** | TP | 22.83(1.53) | 13.97(3.85) | 16.67(2.92) | 14.00(3.55) | 15.83(5.88) | 15.43(4.12) | 18.17(3.77) | 17.67(4.03) |
| | FP | 1.87(1.63) | 0.63(0.81) | 80.07(7.03) | 0.70(1.02) | 1.87(1.63) | 23.30(5.28) | 44.77(16.59) | 38.80(15.73) |
| | ESM | 0.77(0.37) | 4.42(0.70) | 36.50(24.21) | 4.10(0.80) | 3.55(2.63) | 14.44(7.57) | 5.43(1.36) | 5.40(1.41) |
| **Error3** | TP | 25.00(0.01) | 17.00(2.73) | 16.57(2.93) | 16.90(2.75) | 17.63(4.92) | 16.87(3.37) | 18.67(3.65) | 18.50(2.80) |
| | FP | 0.10(0.31) | 0.43(0.68) | 77.50(10.89) | 0.40(0.67) | 1.63(1.40) | 22.67(5.92) | 48.17(18.94) | 37.57(14.98) |
| | ESM | 0.06(0.02) | 3.32(0.52) | 25.62(9.95) | 2.57(0.39) | 2.30(1.62) | 10.35(3.72) | 4.81(1.05) | 4.76(0.90) |
| **Error4** | TP | 19.57(2.51) | 14.14(2.27) | 17.43(2.57) | 14.71(3.64) | 18.29(2.43) | 16.71(2.21) | 20.86(1.77) | 20.29(1.89) |
| | FP | 3.14(2.41) | 0.71(1.11) | 77.71(3.90) | 1.43(1.51) | 1.43(2.07) | 23.57(5.16) | 52.86(15.72) | 46.00(16.12) |
| | ESM | 1.59(0.76) | 4.33(0.78) | 26.68(2.36) | 4.65(0.64) | 1.99(0.58) | 11.52(2.48) | 4.44(0.68) | 4.39(0.67) |
| **Error5** | TP | 18.36(3.11) | 8.73(3.41) | 14.82(1.99) | 9.36(2.11) | 12.09(2.47) | 13.73(1.90) | 17.91(4.16) | 17.27(3.44) |
| | FP | 13.36(10.15) | 0.73(1.01) | 83.64(3.38) | 2.00(1.10) | 1.09(1.38) | 25.45(7.90) | 40.73(14.01) | 39.45(18.94) |
| | ESM | 3.23(0.71) | 5.32(0.55) | 36.55(9.32) | 5.94(0.64) | 4.30(1.01) | 15.75(1.98) | 5.30(0.71) | 5.38(0.87) |
| | | | | | $\theta = 0.5$ | | | | |
| **Error1** | TP | 24.97(0.18) | 23.37(1.25) | 22.77(1.61) | 23.20(1.27) | 24.93(0.37) | 24.10(0.84) | 21.47(1.93) | 20.97(0.18) |
| | FP | 2.23(2.36) | 4.57(3.72) | 51.07(8.01) | 4.57(3.01) | 0.97(1.16) | 13.83(4.36) | 55.10(16.61) | 49.00(2.05) |
| | ESM | 0.19(0.09) | 2.56(0.47) | 7.99(1.61) | 2.18(0.32) | 0.14(0.07) | 2.38(0.36) | 4.01(0.87) | 3.91(0.43) |
| **Error2** | TP | 22.77(2.73) | 13.10(5.08) | 15.17(4.96) | 13.40(4.58) | 13.07(7.17) | 14.10(5.64) | 16.60(7.00) | 16.03(7.02) |
| | FP | 0.60(0.72) | 0.50(0.73) | 81.53(7.83) | 0.23(0.57) | 1.20(1.21) | 23.57(8.47) | 43.07(27.15) | 36.67(23.09) |
| | ESM | 0.64(0.67) | 4.15(0.87) | 182.58(50.88) | 3.47(0.59) | 4.40(2.98) | 57.51(13.99) | 5.59(1.81) | 5.54(1.80) |
| **Error3** | TP | 24.73(0.52) | 17.07(2.92) | 17.97(3.39) | 17.23(2.90) | 19.80(5.20) | 18.20(3.51) | 19.57(3.79) | 19.13(4.11) |
| | FP | 0.40(0.62) | 0.60(1.07) | 74.87(9.50) | 0.83(1.32) | 1.53(1.89) | 21.10(5.89) | 46.60(17.79) | 41.67(14.51) |
| | ESM | 0.18(0.12) | 3.49(0.67) | 23.06(11.86) | 2.96(0.63) | 1.76(1.80) | 9.12(4.67) | 4.73(1.24) | 4.7(1.32) |
| **Error4** | TP | 22.30(1.66) | 12.20(2.87) | 17.27(2.02) | 12.93(2.29) | 17.33(2.89) | 16.93(2.15) | 19.73(2.73) | 19.27(2.75) |
| | FP | 1.13(1.14) | 0.60(0.89) | 78.33(5.97) | 0.90(1.37) | 2.23(1.72) | 22.10(5.68) | 47.73(18.09) | 42.33(18.66) |
| | ESM | 0.84(0.46) | 4.42(0.72) | 26.94(4.36) | 4.31(0.81) | 2.46(0.89) | 10.90(1.76) | 4.62(0.86) | 4.58(0.86) |
| **Error5** | TP | 18.90(3.01) | 8.27(2.61) | 15.20(2.38) | 9.40(1.94) | 12.23(4.47) | 13.37(2.17) | 17.57(3.35) | 16.67(3.54) |
| | FP | 3.73(4.36) | 0.43(0.63) | 85.53(4.39) | 0.80(1.10) | 1.33(1.35) | 22.70(7.54) | 41.73(17.88) | 38.60(19.77) |
| | ESM | 2.21(1.06) | 5.08(0.62) | 40.21(9.28) | 5.08(0.76) | 4.55(1.77) | 15.56(3.23) | 5.77(1.18) | 5.85(1.50) |
| | | | | | $\theta = 0.7$ | | | | |
| **Error1** | TP | 24.87(0.43) | 22.87(1.31) | 22.40(1.43) | 22.60(1.33) | 24.93(0.25) | 23.60(1.22) | 21.50(1.76) | 21.43(1.83) |
| | FP | 3.97(2.93) | 6.80(7.32) | 49.00(4.83) | 5.6(3.77) | 0.93(1.01) | 15.20(4.25) | 54.87(17.24) | 53.53(19.10) |
| | ESM | 0.28(0.12) | 2.63(0.44) | 7.63(0.84) | 2.27(0.34) | 0.14(0.06) | 2.53(0.31) | 3.87(0.79) | 3.86(0.85) |
| **Error2** | TP | 22.04(2.23) | 13.64(5.15) | 15.20(3.69) | 13.32(4.67) | 14.00(8.06) | 14.76(4.94) | 17.32(5.16) | 17.40(5.00) |
| | FP | 1.92(1.71) | 0.72(1.17) | 81.04(9.94) | 0.88(1.45) | 1.64(1.38) | 23.00(5.87) | 37.36(14.81) | 35.28(11.61) |
| | ESM | 0.99(0.58) | 4.36(0.75) | 44.21(36.19) | 3.93(0.68) | 4.18(3.19) | 16.74(12.96) | 5.42(1.77) | 5.31(1.80) |
| **Error3** | TP | 22.63(1.69) | 15.93(3.50) | 18.30(3.08) | 15.87(3.83) | 19.90(4.12) | 18.23(3.53) | 20.17(2.80) | 19.63(3.06) |
| | FP | 3.43(2.47) | 1.57(1.33) | 74.47(7.89) | 2.77(1.92) | 1.60(1.35) | 22.97(7.78) | 51.27(15.67) | 43.63(13.52) |
| | ESM | 0.98(0.53) | 4.51(0.68) | 21.17(6.45) | 4.83(1.05) | 1.71(1.63) | 9.17(3.61) | 4.75(1.30) | 4.72(1.37) |
| **Error4** | TP | 21.03(1.75) | 12.77(2.78) | 16.70(1.99) | 13.27(2.63) | 16.67(2.25) | 15.90(2.44) | 19.67(2.11) | 19.17(2.07) |
| | FP | 3.83(3.82) | 0.93(1.20) | 77.67(6.88) | 1.87(1.36) | 1.93(2.03) | 24.03(7.92) | 53.4(17.26) | 43.47(13.11) |
| | ESM | 1.43(0.60) | 4.72(0.66) | 26.96(5.52) | 5.00(0.75) | 2.67(0.76) | 11.46(1.84) | 4.90(0.95) | 4.85(0.90) |
| **Error5** | TP | 18.20(3.37) | 9.50(3.24) | 14.90(2.75) | 10.13(2.92) | 12.07(5.19) | 13.40(3.11) | 18.03(4.97) | 18.33(3.44) |
| | FP | 9.27(9.79) | 1.03(1.30) | 83.23(4.75) | 1.27(1.64) | 1.47(7.66) | 24.00(7.66) | 43.83(21.97) | 40.30(14.21) |
| | ESM | 3.14(1.16) | 5.73(0.74) | 42.05(14.11) | 5.93(1.13) | 4.80(3.87) | 17.14(3.87) | 5.86(1.07) | 5.65(1.07) |

**Table B.10**: *Simulation results of the heterogeneous errors. n=300, p=400. Mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

| | | bqenetss | bqenet | benet | qbl | blss | bl | enet | lasso |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\theta = 0.3$ | | | | |
| **Error1** | TP | 24.97(0.18) | 20.20(2.17) | 21.40(1.96) | 20.60(2.33) | 24.23(1.04) | 22.33(1.99) | 21.67(2.51) | 21.57(2.27) |
| | FP | 0.01(0.01) | 2.50(2.35) | 59.73(7.24) | 2.47(2.42) | 1.80(1.40) | 16.37(5.73) | 57.87(18.65) | 54.5(14.59) |
| | ESM | 0.04(0.02) | 3.21(0.61) | 12.45(2.22) | 2.97(0.73) | 0.68(0.22) | 4.62(1.02) | 4.16(1.04) | 4.08(0.98) |
| **Error2** | TP | 22.80(3.23) | 8.53(4.85) | 13.40(3.37) | 10.13(4.37) | 7.93(6.60) | 11.07(4.83) | 14.93(4.99) | 14.07(5.39) |
| | FP | 0.20(0.66) | 0.17(0.38) | 86.60(9.53) | 0.40(0.72) | 0.97(1.35) | 24.63(8.26) | 36.07(19.76) | 29.93(17.64) |
| | ESM | 0.59(0.81) | 4.87(0.70) | 72.81(49.85) | 4.27(0.67) | 6.88(2.79) | 27.37(17.31) | 6.44(1.25) | 6.49(1.37) |
| **Error3** | TP | 25.00(0.01) | 15.13(3.31) | 15.77(3.71) | 14.80(3.46) | 14.10(6.18) | 14.63(3.98) | 17.87(4.65) | 17.17(4.68) |
| | FP | 0.02(0.01) | 0.20(0.41) | 79.07(10.57) | 0.23(0.50) | 1.17(1.51) | 24.43(7.86) | 45.20(15.43) | 35.10(13.49) |
| | ESM | 0.03(0.02) | 3.57(0.78) | 72.03(19.50) | 2.89(0.56) | 4.90(2.76) | 31.38(88.65) | 5.55(1.83) | 5.59(1.85) |
| **Error4** | TP | 22.64(2.11) | 7.82(2.14) | 13.55(1.69) | 8.27(1.74) | 9.36(3.32) | 11.45(2.42) | 18.27(1.56) | 17.64(2.25) |
| | FP | 0.36(0.67) | 0.09(0.30) | 86.82(6.45) | 0.45(0.69) | 1.00(0.77) | 24.82(5.53) | 47.27(15.17) | 40.64(18.34) |
| | ESM | 0.81(0.69) | 5.25(0.73) | 53.10(7.28) | 5.12(0.76) | 6.63(1.49) | 21.60(3.89) | 5.87(0.79) | 6.00(0.91) |
| **Error5** | TP | 15.00(2.62) | 3.88(1.64) | 11.88(1.96) | 5.38(2.72) | 5.38(3.70) | 9.00(2.27) | 12.75(3.77) | 11.88(5.41) |
| | FP | 1.38(0.92) | 0.12(0.35) | 90.88(3.52) | 0.38(0.74) | 0.38(0.52) | 27.62(4.93) | 29.38(15.31) | 26.12(23.75) |
| | ESM | 3.16(0.71) | 6.14(0.96) | 72.42(15.71) | 6.10(1.59) | 8.55(1.52) | 31.19(5.03) | 7.24(1.10) | 7.47(1.20) |
| | | | | | $\theta = 0.5$ | | | | |
| **Error1** | TP | 25.00(0.01) | 20.50(2.00) | 21.50(2.39) | 20.25(2.12) | 24.12(0.64) | 22.25(1.58) | 23.25(1.04) | 22.50(1.20) |
| | FP | 0.01(0.02) | 2.00(1.41) | 60.88(6.49) | 3.00(2.20) | 2.12(1.96) | 21.38(5.37) | 69.12(19.79) | 57.00(26.37) |
| | ESM | 0.03(0.02) | 3.22(0.50) | 11.41(1.49) | 3.05(0.58) | 0.47(0.18) | 4.66(0.72) | 3.82(0.65) | 3.80(0.59) |
| **Error2** | TP | 24.03(1.50) | 8.93(3.59) | 14.23(4.02) | 9.97(3.37) | 8.53(6.38) | 12.00(4.62) | 15.13(5.20) | 14.23(5.62) |
| | FP | 0.03(0.18) | 0.10(0.31) | 86.07(7.70) | 0.23(0.50) | 1.07(1.36) | 25.33(5.86) | 36.83(19.67) | 32.07(19.53) |
| | ESM | 0.25(0.31) | 4.80(0.79) | 63.99(64.57) | 3.94(0.63) | 6.31(3.00) | 24.80(20.17) | 6.42(1.63) | 6.50(1.76) |
| **Error3** | TP | 25.00(0.01) | 13.73(3.85) | 15.43(3.78) | 14.47(3.51) | 13.43(6.73) | 14.27(4.63) | 17.50(4.97) | 16.83(4.81) |
| | FP | 0.01(0.02) | 0.17(0.38) | 82.93(9.54) | 0.27(0.52) | 1.43(1.63) | 22.90(6.58) | 42.87(19.42) | 35.47(15.44) |
| | ESM | 0.04(0.02) | 3.78(0.63) | 46.59(37.97) | 2.96(0.55) | 4.53(3.04) | 17.62(12.21) | 5.50(1.39) | 5.48(1.42) |
| **Error4** | TP | 23.43(1.91) | 7.87(2.96) | 13.67(2.78) | 8.83(2.59) | 8.93(4.57) | 11.77(3.57) | 16.13(2.97) | 15.13(3.51) |
| | FP | 0.07(0.37) | 0.10(0.31) | 87.9(6.67) | 0.33(0.66) | 0.93(1.28) | 23.47(6.45) | 40.00(18.85) | 30.37(16.39) |
| | ESM | 0.35(0.33) | 4.93(0.69) | 50.72(12.29) | 4.33(0.70) | 5.88(1.92) | 19.38(3.55) | 5.92(1.13) | 5.92(1.19) |
| **Error5** | TP | 18.43(4.07) | 4.30(2.68) | 12.33(2.54) | 6.27(2.27) | 4.93(2.89) | 9.73(2.13) | 14.43(3.35) | 13.83(3.73) |
| | FP | 0.43(0.73) | 0.02(0.01) | 87.53(4.90) | 0.07(0.25) | 0.83(0.99) | 26.07(6.26) | 35.80(16.49) | 31.53(16.08) |
| | ESM | 1.86(1.32) | 5.53(0.67) | 70.45(20.24) | 5.04(0.73) | 8.30(1.52) | 28.90(7.05) | 6.62(1.01) | 6.66(1.06) |
| | | | | | $\theta = 0.7$ | | | | |
| **Error1** | TP | 24.92(0.29) | 20.83(2.69) | 20.75(2.86) | 19.92(2.91) | 23.75(1.06) | 21.92(2.07) | 20.67(1.67) | 20.50(1.73) |
| | FP | 0.25(0.45) | 2.83(1.80) | 62.33(6.65) | 4.08(2.61) | 1.92(1.56) | 19.08(4.70) | 54.25(16.53) | 47.58(13.69) |
| | ESM | 0.09(0.07) | 3.56(0.32) | 11.56(1.51) | 3.37(0.40) | 0.86(0.28) | 4.74(0.59) | 4.49(0.95) | 4.39(0.98) |
| **Error2** | TP | 23.60(1.71) | 10.80(5.12) | 12.90(3.57) | 12.10(4.98) | 8.20(8.35) | 11.90(5.02) | 15.20(4.10) | 14.70(4.37) |
| | FP | 0.01(0.01) | 0.20(0.42) | 86.50(6.77) | 0.40(0.70) | 0.50(0.85) | 23.60(4.70) | 38.00(16.48) | 29.40(13.28) |
| | ESM | 0.42(0.44) | 4.90(1.00) | 74.84(6.90) | 4.32(0.58) | 6.65(3.64) | 26.62(18.59) | 6.53(1.55) | 6.57(1.63) |
| **Error3** | TP | 21.90(2.60) | 8.20(4.42) | 14.30(4.11) | 10.00(3.92) | 10.50(8.25) | 12.00(5.23) | 13.70(6.93) | 13.50(6.96) |
| | FP | 0.50(0.53) | 0.6(0.70) | 86.30(8.38) | 0.80(0.92) | 1.10(1.29) | 24.70(6.38) | 36.2(21.57) | 34.70(23.44) |
| | ESM | 0.87(0.74) | 5.08(0.72) | 60.94(42.18) | 4.77(0.84) | 5.96(3.50) | 26.03(18.30) | 6.20(1.96) | 6.27(2.06) |
| **Error4** | TP | 22.57(2.57) | 8.57(2.93) | 13.10(2.98) | 8.97(2.58) | 9.33(4.03) | 11.80(2.93) | 17.00(3.25) | 16.20(4.15) |
| | FP | 0.50(0.82) | 0.33(0.55) | 86.00(6.13) | 0.37(0.56) | 0.80(1.24) | 23.50(7.82) | 45.80(18.17) | 40.40(19.78) |
| | ESM | 0.81(0.77) | 5.28(0.67) | 50.04(11.30) | 5.22(0.80) | 6.00(1.75) | 19.88(4.56) | 6.33(1.00) | 6.41(1.16) |
| **Error5** | TP | 16.23(2.90) | 4.63(2.66) | 12.77(2.49) | 5.33(2.50) | 4.13(3.36) | 9.23(3.26) | 13.10(4.87) | 12.43(4.83) |
| | FP | 3.33(4.39) | 0.17(0.46) | 89.97(5.73) | 0.30(0.53) | 0.30(0.70) | 22.83(6.20) | 30.87(17.64) | 29.5(23.23) |
| | ESM | 3.18(1.26) | 5.86(0.73) | 77.27(41.37) | 6.04(0.76) | 8.09(1.70) | 28.21(11.09) | 7.35(1.45) | 7.53(1.36) |

## B.7 Barplots of the simulation results

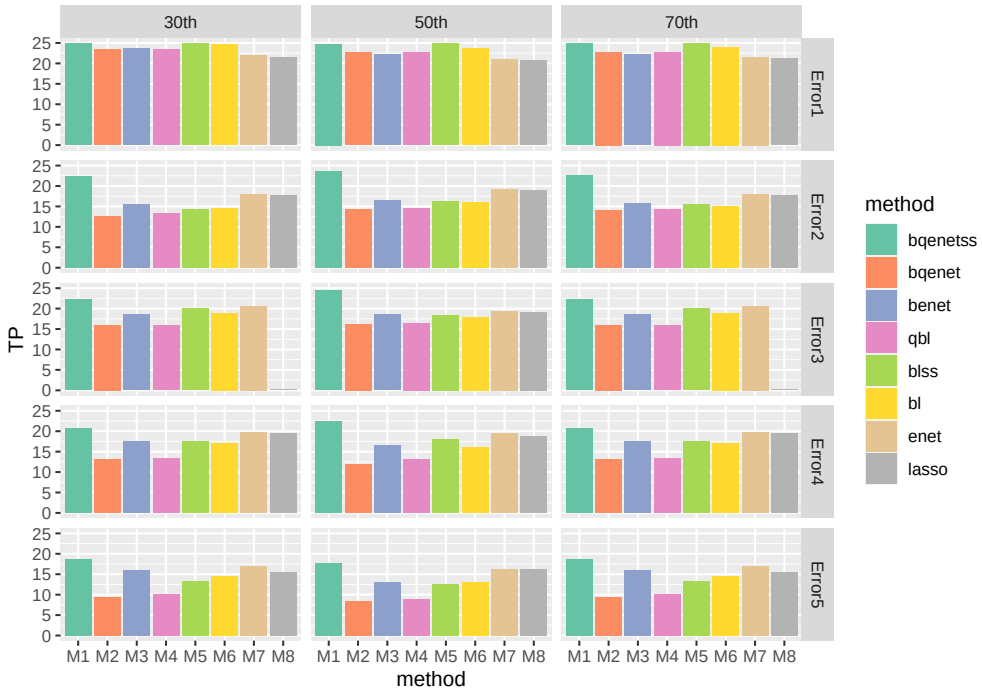


**Figure B.2**: *Ture positive of the first setting with homogeneous error and* $(n, p) = (300, 400)$. *Methods M1-M8 correspond to bqenetss, bqenet, benet, qbl, blss, bl, enet, lasso.*

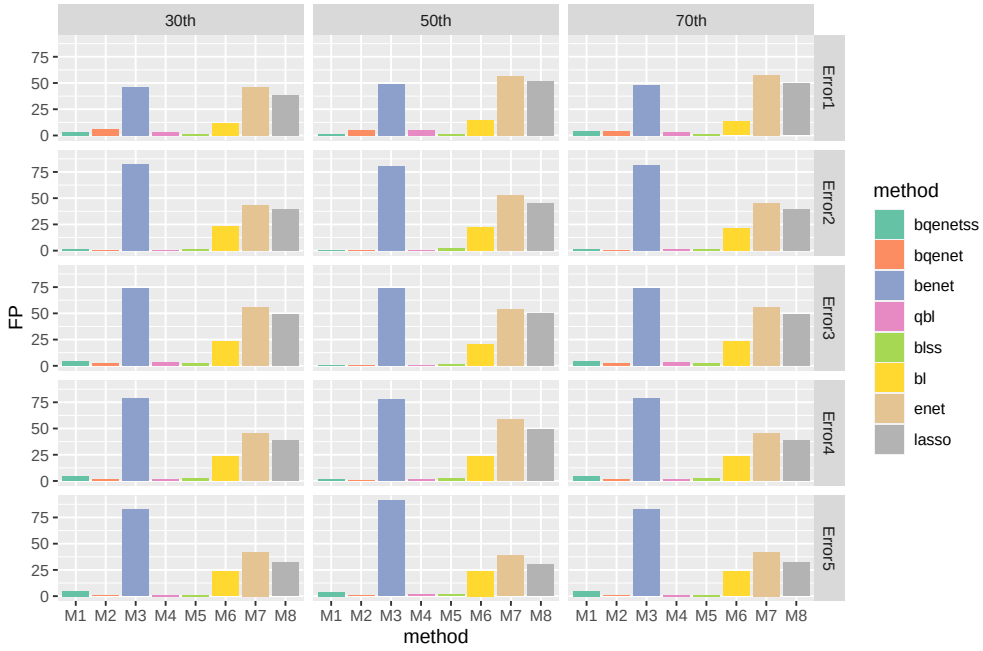


**Figure B.3**: *False positive of the first setting with homogeneous error and* $(n, p) = (300, 400)$. *Methods M1-M8 correspond to bqenetss, bqenet, benet, qbl, blss, bl, enet, lasso.*

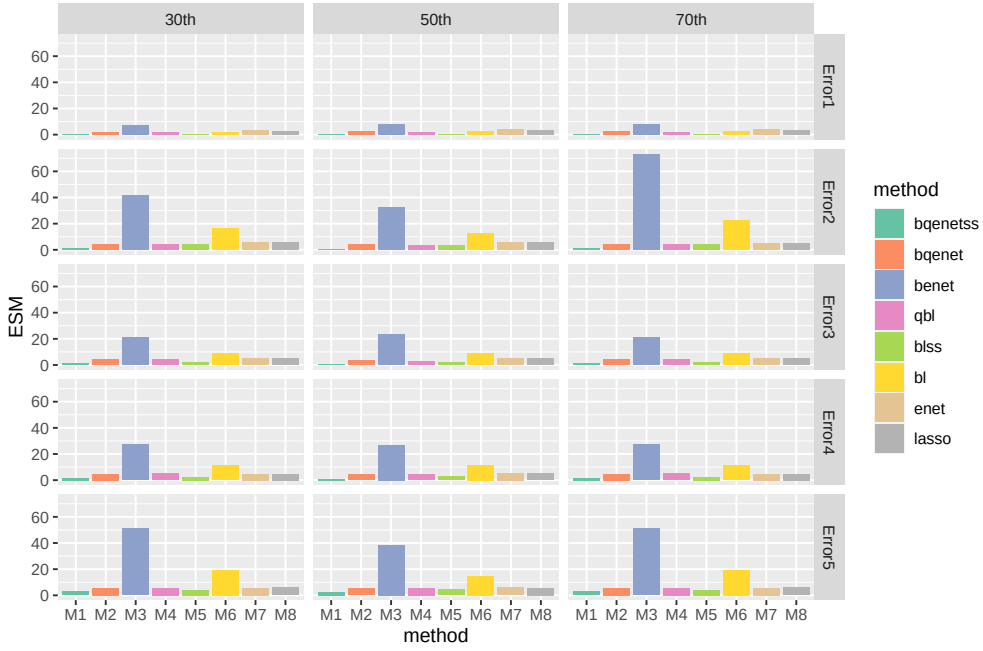


**Figure B.4**: *Estimation of the first setting with homogeneous error and* $(n, p) = (300, 400)$. *Methods M1-M8 correspond to bqenetss, bqenet, benet, qbl, blss, bl, enet, lasso.*


**Figure B.5**: *Ture positive of the first setting with heterogeneous error and* $(n, p) = (300, 400)$. *Methods M1-M8 correspond to bqenetss, bqenet, benet, qbl, blss, bl, enet, lasso.*

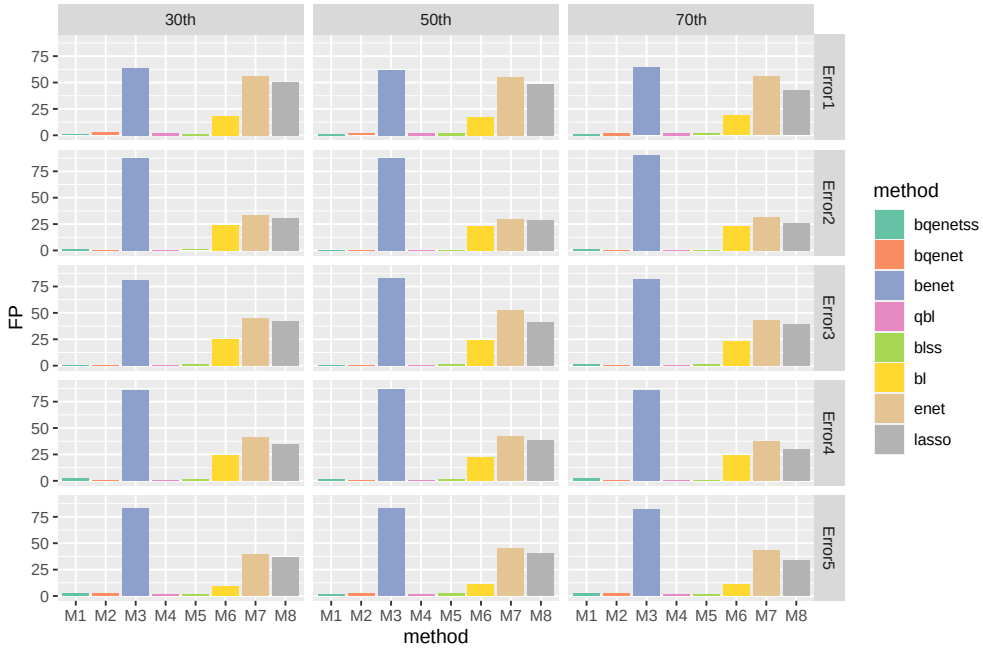


**Figure B.6**: *False positive of the first setting with heterogeneous error and* $(n, p) = (300, 400)$. *Methods M1-M8 correspond to bqenetss, bqenet, benet, qbl, blss, bl, enet, lasso.*

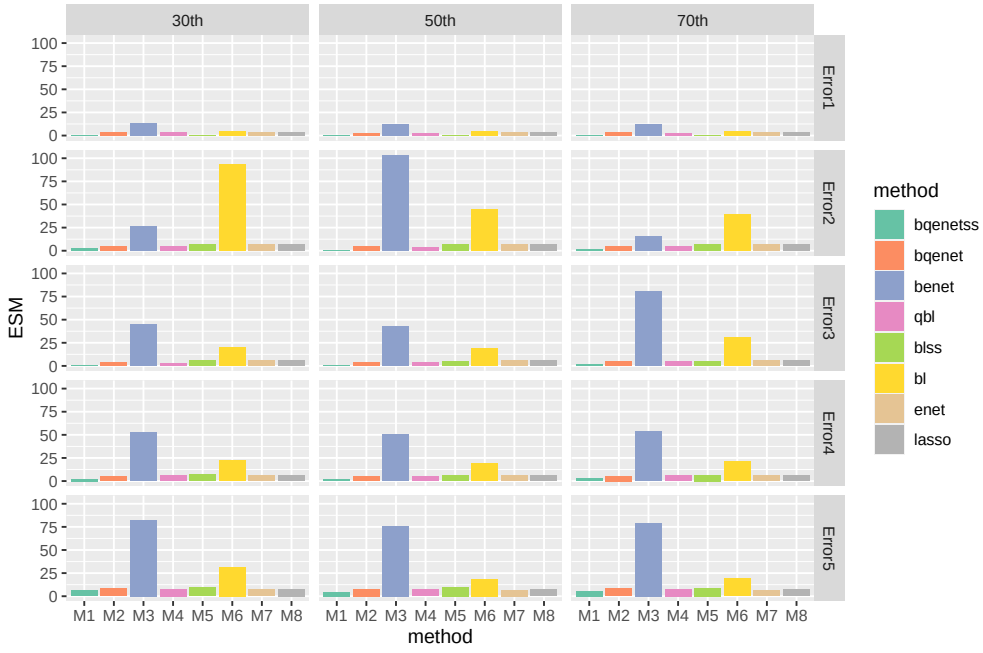


**Figure B.7**: *Estimation of the first setting with heterogeneous error and* $(n, p) = (300, 400)$. *Methods M1-M8 correspond to bqenetss, bqenet, benet, qbl, blss, bl, enet, lasso.*

## B.8 Comparison of different criterions

Besides credible interval criterion and inclusion probability criterion, the scaled neighborhood criterion is also a popular method. Considering the posterior probability in $[-\sqrt{var(\beta_j|y)}, \sqrt{var(\beta_j|y)}]$, a predictor is excluded if the posterior probability exceeds a certain probability threshold (Li and Lin (2010)). Here, we conduct the simulation with the data of the first setting with homogeneous error $t(2)$ and $\theta = 0.5$ with Bayesian elastic net methods bqenetss, bqenet and benet. We consider the simulation results using credible interval criterion for method without spike-and-slab priors and inclusion probability criterion for method with spike-and-slab priors and compare the results with those of using scaled neighborhood criterion. Denote the credible interval level with $\alpha$ and the probability threshold with $p$ in the scaled neighborhood criterion. Different pairs of $(\alpha, p)$ have been demonstrated to show the performance of these three Bayesian methods. Recommended by Li and Lin (2010), $\alpha = 0.5$ and $p = 0.5$ would lead to higher power by Bayesian method. This setting been tested in our dataset. The advantages of the proposed method over the others can be found through the simulation results. However, as $\alpha = 0.5$ is seldomly used in credible interval criterion and 95% credible interval is most commonly used statistical analysis, the setting of $(\alpha, p) = (0.05, 0.05)$has been conducted. The advantages of credible interval criterion and inclusion probability criterion over the scaled neighborhood criterion can be clearly observed from the results below. Additionally, other pairs of parameters $(\alpha, p) = (0.1, 0.1)$ and $(\alpha, p) = (0.25\ 0.25)$ have also been included to support the conclusion.

## B.9 Coverage probability

To compare the performance of all Bayesian methods, we compare the coverage probability of bqenetss, bqenet, benet, qbl, blss and bl. The data used for comparison is generated from the first setting with homogeneous error $t(2)$ and $(n, p) = (300, 400)$. Three non-zero coefficients are chosen at the position 1, 5 and 12. The corresponing value of the coefficients are 1.1, 1.2 and 1.5, seperately.

**Table B.11**: *Simulation results with the data of the first setting with homogeneous error* $t(2)$ *and* $\theta = 0.5$. *n=300, p=400. Mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

| **Method** | bqenetss | bqenetss | bqenet | bqenet | benet | benet |
|---|---|---|---|---|---|---|
| **Criteria** | Inclusion Probability | Scaled Neighborhood | Credible Interval | Scaled Neighborhood | Credible Interval | Scaled Neighborhood |
| | | | $\alpha = 0.05, p = 0.05$ | | | |
| TP | 23.033(1.564) | 18.900(3.315) | 13.467(4.265) | 6.567(4.125) | 15.467(3.767) | 12.067(3.778) |
| FP | 0.933 (1.172) | 0.001(0.001) | 0.433(0.679) | 0.033(0.183) | 80.400(10.820) | 35.067(6.405) |
| | | | $\alpha = 0.1, p = 0.1$ | | | |
| TP | 23.167(1.533) | 20.967(2.125) | 17.300(2.548) | 9.233(3.803) | 17.967(2.965) | 14.600(3.480) |
| FP | 0.867(1.106) | 0.167(0.461) | 1.867(1.697) | 0.100(0.305) | 111.933(11.12) | 57.067(8.682) |
| | | | $\alpha = 0.25, p = 0.25$ | | | |
| TP | 23.300(1.705) | 21.433(2.046) | 21.067(2.180) | 13.567(3.549) | 19.267(2.586) | 16.467(3.256) |
| FP | 0.700(0.952) | 0.033(0.183) | 12.900(5.241) | 0.800(0.925) | 176.200(9.342) | 111.233(9.641) |
| | | | $\alpha = 0.5, p = 0.5$ | | | |
| TP | 22.517(2.064) | 22.379(2.060) | 23.552(1.242) | 21.379(1.821) | 21.966(1.802) | 20.862(2.232) |
| FP | 0.931(1.132) | 0.586(0.867) | 73.103(15.089) | 16.655(6.720) | 250.414(9.579) | 207.138(9.516) |

**Table B.12**: *Comprehensive comparison of bqenetss, bqenet, benet, qbl, blss and bl and for the datasets from the first setting with homogeneous error $t(2)$ and $(n, p) = (300, 400)$ over 200 replication.*

| | | Methods | | | | | |
|---|---|---|---|---|---|---|---|
| | | bqenetss | bqenet | benet | qbl | blss | bl |
| $\theta = 0.3$ | Coverage probability | | | | | | |
| | $\beta_1$ | 0.9306 | 0.0100 | 0.5800 | 0.8200 | 0.8750 | 0.7050 |
| | $\beta_5$ | 0.9444 | 0.0050 | 0.4900 | 0.8100 | 0.8850 | 0.7250 |
| | $\beta_{12}$ | 0.9444 | 0.0000 | 0.3500 | 0.8550 | 0.9200 | 0.7400 |
| | average | 0.9398(0.0080) | 0.0050(0.0050) | 0.4733(0.1159) | 0.8283(0.0236) | 0.8933(0.0236) | 0.7233(0.0176) |
| | Credible interval length | | | | | | |
| | $\beta_1$ | 0.3537(0.0445) | 0.4832(0.0379) | 0.7492(0.9844) | 0.6521(0.0777) | 0.6558(0.3683) | 0.7630(0.3662) |
| | $\beta_5$ | 0.3574(0.0426) | 0.5007(0.0372) | 0.7911(0.8954) | 0.7008(0.0835) | 0.6510(0.3668) | 0.8312(0.4303) |
| | $\beta_{12}$ | 0.3549(0.0478) | 0.5224(0.0532) | 0.7948(0.9070) | 0.7012(0.0869) | 0.6768(0.3698) | 0.8401(0.4290) |
| $\theta = 0.5$ | Coverage probability | | | | | | |
| | $\beta_1$ | 0.8889 | 0.0052 | 0.5900 | 0.8650 | 0.9050 | 0.7550 |
| | $\beta_5$ | 0.9630 | 0.0052 | 0.5050 | 0.7550 | 0.9350 | 0.7250 |
| | $\beta_{12}$ | 0.9630 | 0.0000 | 0.4550 | 0.8150 | 0.9300 | 0.7450 |
| | average | 0.9383(0.0428) | 0.0035(0.0030) | 0.5167(0.0683) | 0.8117(0.0551) | 0.9233(0.0161) | 0.7417(0.0153) |
| | Credible interval length | | | | | | |
| | $\beta_1$ | 0.3278(0.0324) | 0.4734(0.0334) | 0.6712(0.2944) | 0.6396(0.0753) | 0.6191(0.2651) | 0.7556(0.3877) |
| | $\beta_5$ | 0.3253(0.0381) | 0.4870(0.0363) | 0.7254(0.3356) | 0.6927(0.0779) | 0.6275(0.2743) | 0.8224(0.4329) |
| | $\beta_{12}$ | 0.3313(0.0262) | 0.5112(0.0375) | 0.7251(0.3285) | 0.6908(0.0790) | 0.6140(0.2343) | 0.8338(0.4791) |
| $\theta = 0.7$ | Coverage probability | | | | | | |
| | $\beta_1$ | 0.9586 | 0.0000 | 0.5400 | 0.8100 | 0.9200 | 0.7900 |
| | $\beta_5$ | 0.9379 | 0.0050 | 0.4900 | 0.8150 | 0.9050 | 0.7550 |
| | $\beta_{12}$ | 0.9586 | 0.0000 | 0.4000 | 0.8500 | 0.9000 | 0.7750 |
| | average | 0.9517(0.0119) | 0.0017(0.0029) | 0.4767(0.0709) | 0.8250(0.0218) | 0.9083(0.0104) | 0.7733(0.0176) |
| | Credible interval length | | | | | | |
| | $\beta_1$ | 0.3635(0.0482) | 0.4801(0.0383) | 0.6944(0.3928) | 0.6413(0.0733) | 0.6096(0.2984) | 0.7500(0.3841) |
| | $\beta_5$ | 0.3598(0.0509) | 0.4993(0.0445) | 0.7443(0.4250) | 0.6873(0.0738) | 0.6269(0.3079) | 0.8021(0.3554) |
| | $\beta_{12}$ | 0.3657(0.0556) | 0.5193(0.0498) | 0.7431(0.4224) | 0.6943(0.0756) | 0.6084(0.2413) | 0.8123(0.3952) |

# Appendix C

# Appendices for chapter 4

## C.1 Posterior inference

### C.1.1 Bayesian quantile subgroup (qsubgroup)

#### Hierarchical model specification

$$Y_i = \mu_i + \boldsymbol{X}_i\boldsymbol{\beta} + \xi_1 v_i + \tau^{-1/2}\xi_2\sqrt{v_i}z_i \quad i = 1, \ldots, n$$

$$v_i|\tau \overset{iid}{\sim} \tau\exp(-\tau v_i) \quad i = 1, \ldots, n$$

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

$$\beta_j|s_j \overset{iid}{\sim} \mathrm{N}(0, s_j) \quad j = 1, \ldots, p$$

$$s_j|\eta_2^2 \overset{iid}{\sim} \frac{\eta_2^2}{2}\exp(-\frac{\eta_2^2}{2}s_j) \quad j = 1, \ldots, p$$

$$\tau \sim \mathrm{Gamma}(a, b)$$

$$\eta_2^2 \sim \mathrm{Gamma}(c_2, d_2)$$

$$\boldsymbol{\mu}|\varphi_{ij}^2 \sim \mathrm{N}_n(0, \Sigma_\mu)$$

$$\varphi_{ij}^2 \sim \frac{\eta_1^2}{2}\exp(-\frac{\eta_1^2}{2}\varphi_{ij}^2)$$

$$\eta_1^2 \sim \mathrm{Gamma}(c_1, d_1)$$

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

$$
\begin{aligned}
&\pi(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{v}, \tau, \eta_1^2, \eta_2^2, \varphi_{ij}^2, \pi_1 | Y) \\
&\quad \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau^{-1}\xi_2^2 v_i}} \exp\Big\{ - \frac{(y_i - \mu_i - \boldsymbol{X}_i\boldsymbol{\beta} - \xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \\
&\quad \times \prod_{i=1}^{n} \tau \exp(-\tau v_i) \tau^{a-1} \exp(-b\tau) \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} z_i^2) \\
&\quad \times \prod_{j=1}^{p} (2\pi s_j)^{-1/2} \exp\{-\frac{\beta_j^2}{2s_j}\} \times (\eta_1^2)^{c_1 - 1} \exp(-d_1\eta_1) \\
&\quad \times (\eta_2^2)^{c_2 - 1} \exp(-d_2\eta_2) \times |\Sigma_\mu|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\boldsymbol{\mu}^\top (\Sigma_\mu)^{-1}\boldsymbol{\mu}\} \\
&\quad \times \prod_{1 \le i < j \le n} \frac{\eta_1^2}{2} \exp(-\frac{\eta_1^2}{2}\varphi_{ij}^2)
\end{aligned}
$$

**Gibbs Sampler**

Let $\mu_{(-\boldsymbol{\beta}_j)} = \mu_i + \boldsymbol{X}_i\boldsymbol{\beta} + \xi_1 v_i - X_j\beta_j$ $(j = 1, \ldots, p)$, then, $\beta_j|\text{rest} \sim N(\mu_{\beta_j}, \sigma_{\beta_j}^2)$, where

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

The full conditional posterior distribution of $s_j$ is:

$$
\begin{aligned}
&s_j|\text{rest} \\
&\quad \propto \pi(\beta_j|s_j)\pi(s_{1j}|\eta_2^2) \\
&\quad \propto \frac{1}{\sqrt{s_j}} \exp(-\frac{\eta_2^2}{2}s_j) \exp(-\frac{\beta_j^2}{2s_j}) \\
&\quad \propto \frac{1}{\sqrt{s_j}} \exp\Big\{ - \frac{1}{2}[\eta_2^2 s_j + \frac{\beta_j^2}{s_j}] \Big\}.
\end{aligned}
$$

Therefore, the posterior distribution for $s_j^{-1}$ is Inverse-Gaussian$(\sqrt{\frac{\eta_2^2}{\beta_j^2}}, \eta_2^2)$.

The full conditional posterior distribution of $\eta_2^2$:

$$
\begin{aligned}
&\eta_2^2|\text{rest} \\
&\quad\propto \pi(s_1|\eta_2^2)\pi(\eta_2^2) \\
&\quad\propto \prod_{j=1}^{p} \frac{\eta_2^2}{2}\exp(-\frac{\eta_2^2 s_j}{2})(\eta_2^2)^{c_2-1}\exp(-d_2\eta_2^2) \\
&\quad\propto (\eta_2^2)^{p+c_2-1}\exp\Big(-\eta_2^2(\sum_{j=1}^{p}\frac{s_j}{2}+d_2)\Big).
\end{aligned}
$$

Therefore, the posterior distribution for $\eta_2^2$ is Gamma$(p+c_2, \sum_{j=1}^{p}\frac{s_j}{2}+d_2)$.

The full conditional posterior distribution of $\tau$:

$$
\begin{aligned}
&\tau|\text{rest} \\
&\quad\propto \pi(\boldsymbol{v}|\tau)\pi(\tau)\pi(\boldsymbol{Y}|\cdot) \\
&\quad\propto \tau^{n/2}\exp\Big\{-\sum_{i=1}^{n}\frac{(y_i-\mu_i-\boldsymbol{X}_{ij}\beta_j-\xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\} \\
&\quad\times \tau^n\exp(-\tau\sum_{i=1}^{n}v_i)\tau^{a-1}\exp(-b\tau) \\
&\quad\propto \tau^{a+\frac{3}{2}n-1}\exp\Big\{-\tau\big[\sum_{i=1}^{n}(\frac{(y_i-\mu_i-X_{ij}\beta_j-\xi_1 v_i}{2\xi_2^2 v_i}+v_i)+b\big]\Big\}.
\end{aligned}
$$

Therefore, the posterior distribution for $\tau$ is Gamma$(a+\frac{3}{2}n,\ \big[\sum_{i=1}^{n}(\frac{(y_i-\mu_i-X_{ij}\beta_j-\xi_1 v_i)^2}{2\xi_2^2 v_i}+v_i)+b\big])$.

We have the full conditional posterior distribution of $v_i$:

$$
\begin{aligned}
&v_i|\text{rest} \\
&\quad\propto \pi(v|\tau)\pi(Y|\cdot) \\
&\quad\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{(y_i-\mu_i-\boldsymbol{X}_i\boldsymbol{\beta})^2}{2\tau^{-1}\xi_2^2 v_i}\Big\}\times\exp(-\tau v_i) \\
&\quad\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{1}{2}\big[(\frac{\tau\xi_1^2}{\xi_2^2}+2\tau)v_i+\frac{\tau(y_i-\mu_i-\boldsymbol{X}_i\boldsymbol{\beta})^2}{\xi_2^2 v_i}\big]\Big\}.
\end{aligned}
$$

It can be found that

$$\frac{1}{v_i}|\text{rest}\sim \text{ Inverse-Gaussian}(\sqrt{\frac{\xi_1^2+2\xi_2^2}{(y_i-\mu_i-\boldsymbol{X}_i\boldsymbol{\beta})^2}},\ \frac{\tau\xi_1^2}{\xi_2^2}+2\tau).$$

The full conditional posterior distribution of $\boldsymbol{\mu}$:

$$\begin{aligned}
&\boldsymbol{\mu}|\text{rest}\\
&\quad\propto \exp\{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v)^\top\Sigma_2^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v)\}\\
&\quad\times|\Sigma_\mu|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\mu}^\top\Sigma_\mu^{-1}\boldsymbol{\mu}\}\\
&\quad\propto \exp\{-\frac{1}{2}[\boldsymbol{\mu}^\top({\Sigma_2}^{-1}+{\Sigma_\mu}^{-1})\boldsymbol{\mu}-2\boldsymbol{\mu}^\top{\Sigma_2}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v)]\}
\end{aligned}$$

Where $\mu_2^{-1}=diag(\frac{1}{\tau^{-1}\xi_2^2 v})$. It can be found that the posterior distribution of $\boldsymbol{\mu}$ is $N(\mu_\mu,\sigma_\mu^2)$, where

$$\begin{aligned}
\mu_\mu&=\sigma_\mu^2\Sigma_2^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v),\\
\sigma_\mu^2&=(\Sigma_2^{-1}+{\Sigma_\mu}^{-1})^{-1}.
\end{aligned}$$

The full conditional posterior distribution of $\varphi_{ij}$:

$$\begin{aligned}
&\varphi_{ij}^2|\text{rest}\\
&\quad\propto\frac{1}{\sqrt{2\pi\varphi_{ij}^2}}\exp\{-\frac{(\mu_i-\mu_j)^2}{2\varphi_{ij}^2}\}\exp(-\frac{\lambda^2\varphi_{ij}^2}{2})\\
&\quad=\frac{1}{\sqrt{2\pi\varphi_{ij}^2}}\exp\{-\frac{1}{2}[\frac{1}{\varphi_{ij}^2}(\mu_i-\mu_j)^2+\eta_1^2\varphi_{ij}^2]\}
\end{aligned}$$

Then,

$$\frac{1}{\varphi_{ij}^2}|\text{rest}\sim \text{ Inverse-Gaussian}(\sqrt{\frac{\eta_1^2}{(\mu_i-\mu_j)^2}},\ \eta_1^2), 1\le i<j\le n.$$

The full conditional posterior distribution of $\eta_1^2$:

$$\begin{aligned}
&\eta_1^2|\text{rest} \\
&\propto \prod_{1\le i<j\le n} \frac{\eta_1^2}{2}\exp(-\frac{\eta_1^2}{2}\varphi_{ij}^2) \times (\eta_1^2)^{c_1-1}\exp(-d_1\eta_1) \\
&= (\eta_1^2)^{\frac{n(n-1)}{2}+c_1-1}\exp\{-\eta_1^2(\sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+d_1)\}.
\end{aligned}$$

Therefore, the posterior distribution for $\eta_1^2$ is Gamma($\frac{n(n-1)}{2}+c_1$, $\sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+d_1$).

### C.1.2 Bayesian quantile subgroup with spike-and-slab priors (Normal) (qsgss)

**Hierarchical model specification**

$$Y_i = \mu_i + \boldsymbol{X}_i\boldsymbol{\beta} + \xi_1 v_i + \tau^{-1/2}\xi_2\sqrt{v_i}z_i \quad i=1,\dots,n$$

$$v_i|\tau \overset{iid}{\sim} \tau\exp(-\tau v_i) \quad i=1,\dots,n$$

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

$$\beta_j \overset{iid}{\sim} (1-\pi_1)\mathrm{N}(0,\sigma_0)+\pi_1\delta_0(\beta_j) \quad j=1,\dots,p$$

$$\tau\sim\mathrm{Gamma}(a,b)$$

$$\pi_1\sim\mathrm{Beta}(r_1,u_1)$$

$$\boldsymbol{\mu}|\varphi_{ij}^2\sim\mathrm{N}_n(0,\Sigma_\mu)$$

$$\varphi_{ij}^2\sim\frac{\lambda^2}{2}\exp(-\frac{\lambda^2}{2}\varphi_{ij}^2)$$

$$\lambda^2\sim\mathrm{Gamma}(c,d)$$

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

be expressed as

$$
\begin{aligned}
&\pi(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{v}, \tau, \eta_1^2, \eta_2^2, \varphi_{ij}^2, \pi_1 | Y) \\
&\quad \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau^{-1}\xi_2^2 v_i}} \exp\Big\{ - \frac{(y_i - \mu_i - \boldsymbol{X}_i\boldsymbol{\beta} - \xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \\
&\quad \times \prod_{i=1}^{n} \tau \exp(-\tau v_i) \tau^{a-1} \exp(-b\tau) \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} z_i^2) \\
&\quad \times \prod_{j=1}^{p} \Big( (1-\pi_1)(2\pi\sigma_0)^{-1/2} \exp\{-\frac{\beta_j^2}{2\sigma_0}\} \mathbf{I}_{\{\boldsymbol{\beta}_j \neq 0\}} + \pi_1 \delta_0(\boldsymbol{\beta}_j) \Big) \\
&\quad \times (\lambda^2)^{c-1} \exp(-d\lambda^2) \times \pi_1^{r_1 - 1} (1-\pi_1)^{u_1 - 1} \\
&\quad \times |\Sigma_\mu|^{-\frac{1}{2}} \exp\{-\frac{1}{2} \boldsymbol{\mu}^\top (\Sigma_\mu)^{-1} \boldsymbol{\mu}\} \times \prod_{1 \leq i < j \leq n} \frac{\lambda^2}{2} \exp(-\frac{\lambda^2}{2} \varphi_{ij}^2)
\end{aligned}
$$

**Gibbs Sampler**

Let $\mu_{(-\boldsymbol{\beta}_j)} = \mu_i + \boldsymbol{X}_i\boldsymbol{\beta} + \xi_1 v_i - X_{ij}\beta_j$ and $l_j = \pi(\beta_j = 0 | \text{rest})$, the conditional posterior distribution of the coefficient of genetic factor $\beta_j$ is a spike-and-slab distribution:

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

where

$$
\begin{aligned}
\mu_{\beta_j} &= \big( \sum_{i=1}^{n} \frac{\tau (y_i - \mu_{(-\beta_j)}) X_{ij}}{\xi_2^2 v_i} \big) \sigma_{\beta_j}^2, \\
\sigma_{\beta_j}^2 &= \big( \sum_{i=1}^{n} \frac{\tau X_{ij}^2}{\xi_2^2 v_i} + \frac{1}{\sigma_0} \big)^{-1}.
\end{aligned}
$$

We can show that

$$l_j = \frac{\pi_1}{\pi_1 + (1-\pi_1)\sigma_0^{-1/2} (\sigma_{\beta_j}^2)^{1/2} \exp\{\frac{1}{2} (\sum_{i=1}^{n} \frac{\tau (y_i - \mu_{(-\beta_j)}) X_{ij}}{\xi_2^2 v_i})^2 \sigma_{\beta_j}^2\}}.$$

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

The full conditional posterior distribution of $\tau$:

$$
\begin{aligned}
&\tau|\text{rest} \\
&\propto \pi(\boldsymbol{v}|\tau)\pi(\tau)\pi(\boldsymbol{Y}|\cdot) \\
&\propto \tau^{n/2}\exp\Big\{-\sum_{i=1}^{n}\frac{(y_i-\mu_i-X_{ij}\beta_j-\xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\} \\
&\times \tau^n \exp(-\tau\sum_{i=1}^{n}v_i)\tau^{a-1}\exp(-b\tau) \\
&\propto \tau^{a+\frac{3}{2}n-1}\exp\Big\{-\tau\big[\sum_{i=1}^{n}(\frac{(y_i-\mu_i-X_{ij}\beta_j-\xi_1 v_i}{2\xi_2^2 v_i}+v_i)+b\big]\Big\}.
\end{aligned}
$$

Therefore, the posterior distribution for $\tau$ is Gamma$(a+\frac{3}{2}n,\ \big[\sum_{i=1}^{n}(\frac{(y_i-\mu_i-X_{ij}\boldsymbol{\beta}_j-\xi_1 v_i)^2}{2\xi_2^2 v_i}+v_i)+b\big])$.

The full conditional posterior distribution of $\pi_1$:

$$
\begin{aligned}
&\pi_1|\text{rest} \\
&\propto \pi(\pi_1|\beta)\pi(\pi_1) \\
&\propto \pi_1^{r_1-1}(1-\pi_1)^{u_1-1}\times\prod_{j=1}^{p}\Big((1-\pi_1)(2\pi s_j)^{-1/2}\exp(-\frac{\beta_j^2}{2s_j})\mathbf{I}_{\{\boldsymbol{\beta}_j\neq 0\}}+\pi_1\delta_0(\beta_j)\Big).
\end{aligned}
$$

So, the posterior distribution for $\pi_1$ is Beta $(1+r_1-\sum_{j=1}^{p}\mathbf{I}(\beta_j\neq 0), u_1+\sum_{j=1}^{p}\mathbf{I}(\beta_j\neq 0))$.

We have the full conditional posterior distribution of $v_i$:

$$
\begin{aligned}
&v_i|\text{rest} \\
&\propto \pi(v|\tau)\pi(Y|\cdot) \\
&\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{(y_i-\mu_i-\boldsymbol{X}_i\boldsymbol{\beta})^2}{2\tau^{-1}\xi_2^2 v_i}\Big\}\times\exp(-\tau v_i) \\
&\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{1}{2}\big[(\frac{\tau\xi_1^2}{\xi_2^2}+2\tau)v_i+\frac{\tau(y_i-\mu_i-\boldsymbol{X}_i\boldsymbol{\beta})^2}{\xi_2^2 v_i}\big]\Big\}.
\end{aligned}
$$

It can be found that

$$\frac{1}{v_i}|\text{rest} \sim \text{Inverse-Gaussian}(\sqrt{\frac{\xi_1^2+2\xi_2^2}{(y_i-\mu_i-\boldsymbol{X}_i\boldsymbol{\beta})^2}},\ \frac{\tau\xi_1^2}{\xi_2^2}+2\tau).$$

The full conditional posterior distribution of $\boldsymbol{\mu}$:

$$\begin{aligned}
&\boldsymbol{\mu}|\text{rest} \\
&\propto \exp\{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v)^\top \Sigma_2^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v)\} \\
&\times |\Sigma_\mu|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\mu}^\top\Sigma_\mu^{-1}\boldsymbol{\mu}\} \\
&\propto \exp\{-\frac{1}{2}[\boldsymbol{\mu}^\top({\Sigma_2}^{-1}+{\Sigma_\mu}^{-1})\boldsymbol{\mu}-2\boldsymbol{\mu}^\top{\Sigma_2}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v)]\}
\end{aligned}$$

Where $\mu_2^{-1} = diag(\frac{1}{\tau^{-1}\xi_2^2 v})$. It can be found that the posterior distribution of $\boldsymbol{\mu}$ is $N(\mu_\mu, \sigma_\mu^2)$, where

$$\begin{aligned}
\mu_\mu &= \sigma_\mu^2\Sigma_2^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}-\xi_1 v), \\
\sigma_\mu^2 &= (\Sigma_2^{-1}+{\Sigma_\mu}^{-1})^{-1}.
\end{aligned}$$

The full conditional posterior distribution of $\varphi_{ij}$:

$$\begin{aligned}
&\varphi_{ij}^2|\text{rest} \\
&\propto \frac{1}{\sqrt{2\pi\varphi_{ij}^2}}\exp\{-\frac{(\mu_i-\mu_j)^2}{2\varphi_{ij}^2}\}\exp(-\frac{\lambda^2\varphi_{ij}^2}{2}) \\
&= \frac{1}{\sqrt{2\pi\varphi_{ij}^2}}\exp\{-\frac{1}{2}[\frac{1}{\varphi_{ij}^2}(\mu_i-\mu_j)^2+\lambda^2\varphi_{ij}^2]\}
\end{aligned}$$

Then,

$$\frac{1}{\varphi_{ij}^2}|\text{rest} \sim \text{Inverse-Gaussian}(\sqrt{\frac{\lambda^2}{(\mu_i-\mu_j)^2}},\ \lambda^2), 1 \le i < j \le n.$$

The full conditional posterior distribution of $\lambda^2$:

$$\begin{aligned}
&\lambda^2|\text{rest} \\
&\propto \prod_{1\le i<j\le n} \frac{\lambda^2}{2}\exp(-\frac{\lambda^2}{2}\varphi_{ij}^2) \times (\lambda^2)^{c-1}\exp(-d\lambda^2) \\
&= (\lambda^2)^{\frac{n(n-1)}{2}+c-1}\exp\{-\lambda^2(\sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+d)\}.
\end{aligned}$$

Therefore, the posterior distribution for $\lambda^2$ is Gamma($\frac{n(n-1)}{2}+c$, $\sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+d$).

### C.1.3 Bayesian quantile subgroup (Normal) (qsg)

**Hierarchical model specification**

$$Y_i = \mu_i + \boldsymbol{X}_i\boldsymbol{\beta} + \xi_1 v_i + \tau^{-1/2}\xi_2\sqrt{v_i}z_i \quad i=1,\dots,n$$

$$v_i|\tau \overset{iid}{\sim} \tau\exp(-\tau v_i) \quad i=1,\dots,n$$

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

$$\beta_j \overset{iid}{\sim} \mathrm{N}(0,\sigma_0) \quad j=1,\dots,p$$

$$\tau\sim\mathrm{Gamma}(a,b)$$

$$\boldsymbol{\mu}|\varphi_{ij}^2\sim\mathrm{N}_n(0,\Sigma_\mu)$$

$$\varphi_{ij}^2\sim\frac{\lambda^2}{2}\exp(-\frac{\lambda^2}{2}\varphi_{ij}^2)$$

$$\lambda^2\sim\mathrm{Gamma}(c,d)$$

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

be expressed as

$$
\begin{aligned}
&\pi(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{v}, \tau, \eta_1^2, \eta_2^2, \varphi_{ij}^2, \pi_1 | Y) \\
&\quad \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau^{-1}\xi_2^2 v_i}} \exp\Big\{ - \frac{(y_i - \mu_i - \boldsymbol{X}_i\boldsymbol{\beta} - \xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i} \Big\} \\
&\quad \times \prod_{i=1}^{n} \tau \exp(-\tau v_i) \tau^{a-1} \exp(-b\tau) \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} z_i^2) \\
&\quad \times \prod_{j=1}^{p} (2\pi\sigma_0)^{-1/2} \exp\{-\frac{\beta_j^2}{2\sigma_0}\} \times (\lambda^2)^{c-1} \exp(-d\lambda^2) \\
&\quad \times |\Sigma_\mu|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\boldsymbol{\mu}^\top (\Sigma_\mu)^{-1} \boldsymbol{\mu}\} \times \prod_{1 \le i < j \le n} \frac{\lambda^2}{2} \exp(-\frac{\lambda^2}{2} \varphi_{ij}^2)
\end{aligned}
$$

**Gibbs Sampler**

Let $\mu_{(-\beta_j)} = \mu_i + \boldsymbol{X}_i\boldsymbol{\beta} + \xi_1 v_i - X_{ij}\beta_j$, the conditional posterior distribution of the coefficient of genetic factor $\beta_j$ is a Normal distribution:

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

where

$$
\begin{aligned}
\mu_{\beta_j} &= \big( \sum_{i=1}^{n} \frac{\tau(y_i - \mu_{(-\beta_j)}) X_{ij}}{\xi_2^2 v_i} \big) \sigma_{\beta_j}^2, \\
\sigma_{\beta_j}^2 &= \big( \sum_{i=1}^{n} \frac{\tau X_{ij}^2}{\xi_2^2 v_i} + \frac{1}{\sigma_0} \big)^{-1}.
\end{aligned}
$$

The full conditional posterior distribution of $\tau$:

$$
\begin{aligned}
&\tau|\text{rest} \\
&\propto \pi(\boldsymbol{v}|\tau)\pi(\tau)\pi(\boldsymbol{Y}|\cdot) \\
&\propto \tau^{n/2}\exp\Big\{-\sum_{i=1}^{n}\frac{(y_i-\mu_i-X_{ij}\beta_j-\xi_1 v_i)^2}{2\tau^{-1}\xi_2^2 v_i}\Big\} \\
&\times \tau^n \exp(-\tau\sum_{i=1}^{n}v_i)\tau^{a-1}\exp(-b\tau) \\
&\propto \tau^{a+\frac{3}{2}n-1}\exp\Big\{-\tau\big[\sum_{i=1}^{n}(\frac{(y_i-\mu_i-X_{ij}\beta_j-\xi_1 v_i}{2\xi_2^2 v_i}+v_i)+b\big]\Big\}.
\end{aligned}
$$

Therefore, the posterior distribution for $\tau$ is Gamma$(a+\frac{3}{2}n,\ \big[\sum_{i=1}^{n}(\frac{(y_i-\mu_i-X_{ij}\beta_j-\xi_1 v_i)^2}{2\xi_2^2 v_i}+v_i)+b\big])$.

We have the full conditional posterior distribution of $v_i$:

$$
\begin{aligned}
&v_i|\text{rest} \\
&\propto \pi(v|\tau)\pi(Y|\cdot) \\
&\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{(y_i-\mu_i-\boldsymbol{X}_i\boldsymbol{\beta})^2}{2\tau^{-1}\xi_2^2 v_i}\Big\}\times\exp(-\tau v_i) \\
&\propto \frac{1}{\sqrt{v_i}}\exp\Big\{-\frac{1}{2}\big[(\frac{\tau\xi_1^2}{\xi_2^2}+2\tau)v_i+\frac{\tau(y_i-\mu_i-\boldsymbol{X}_i\boldsymbol{\beta})^2}{\xi_2^2 v_i}\big]\Big\}.
\end{aligned}
$$

It can be found that

$$
\frac{1}{v_i}|\text{rest}\sim \text{ Inverse-Gaussian}(\sqrt{\frac{\xi_1^2+2\xi_2^2}{(y_i-\mu_i-\boldsymbol{X}_i\boldsymbol{\beta})^2}},\ \frac{\tau\xi_1^2}{\xi_2^2}+2\tau).
$$

The full conditional posterior distribution of $\boldsymbol{\mu}$:

$$\begin{aligned}
&\boldsymbol{\mu}|\text{rest} \\
&\quad\propto \exp\{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X\beta}-\xi_1 v)^\top \Sigma_2^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X\beta}-\xi_1 v)\} \\
&\quad\times |\Sigma_\mu|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\mu}^\top \Sigma_\mu^{-1}\boldsymbol{\mu}\} \\
&\quad\propto \exp\{-\frac{1}{2}[\boldsymbol{\mu}^\top({\Sigma_2}^{-1}+{\Sigma_\mu}^{-1})\boldsymbol{\mu}-2\boldsymbol{\mu}^\top{\Sigma_2}^{-1}(\boldsymbol{Y}-\boldsymbol{X\beta}-\xi_1 v)]\}
\end{aligned}$$

Where $\mu_2^{-1} = diag(\frac{1}{\tau^{-1}\xi_2^2 v})$. It can be found that the posterior distribution of $\boldsymbol{\mu}$ is $N(\mu_\mu, \sigma_\mu^2)$, where

$$\begin{aligned}
\mu_\mu &= \sigma_\mu^2\Sigma_2^{-1}(\boldsymbol{Y}-\boldsymbol{X\beta}-\xi_1 v), \\
\sigma_\mu^2 &= (\Sigma_2^{-1}+{\Sigma_\mu}^{-1})^{-1}.
\end{aligned}$$

The full conditional posterior distribution of $\varphi_{ij}$:

$$\begin{aligned}
&\varphi_{ij}^2|\text{rest} \\
&\quad\propto \frac{1}{\sqrt{2\pi\varphi_{ij}^2}}\exp\{-\frac{(\mu_i-\mu_j)^2}{2\varphi_{ij}^2}\}\exp(-\frac{\lambda^2\varphi_{ij}^2}{2}) \\
&\quad= \frac{1}{\sqrt{2\pi\varphi_{ij}^2}}\exp\{-\frac{1}{2}[\frac{1}{\varphi_{ij}^2}(\mu_i-\mu_j)^2+\lambda^2\varphi_{ij}^2]\}
\end{aligned}$$

Then,

$$\frac{1}{\varphi_{ij}^2}|\text{rest}\sim \text{Inverse-Gaussian}(\sqrt{\frac{\lambda^2}{(\mu_i-\mu_j)^2}},\ \lambda^2), 1\le i<j\le n.$$

The full conditional posterior distribution of $\lambda^2$:

$$\begin{aligned}
&\lambda^2|\text{rest} \\
&\quad\propto \prod_{1\le i<j\le n}\frac{\lambda^2}{2}\exp(-\frac{\lambda^2}{2}\varphi_{ij}^2)\times(\lambda^2)^{c-1}\exp(-d\lambda^2) \\
&\quad= (\lambda^2)^{\frac{n(n-1)}{2}+c-1}\exp\{-\lambda^2(\sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+d)\}.
\end{aligned}$$

Therefore, the posterior distribution for $\lambda^2$ is Gamma($\frac{n(n-1)}{2}+c$, $\sum_{1\leq i<j\leq n}\frac{\varphi_{ij}^2}{2}+d$).

### C.1.4 Bayesian subgroup with spike-and-slab priors (subgroupss)

**Hierarchical model specification**

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i-\boldsymbol{X}_i\boldsymbol{\beta})^2\right\}$$

$$\beta_j|\pi_1,\tau_j^2,\sigma^2 \overset{iid}{\sim} (1-\pi_1)\,\mathrm{N}\left(0,\,\sigma^2\tau_j^2\right)+\pi_1\,\delta_0(\beta_j) \quad j=1,\ldots,p$$

$$\tau_j^2|\lambda_2^2 \overset{iid}{\sim} \mathrm{Gamma}(1,\frac{\lambda_2^2}{2}) \quad j=1,\ldots,p$$

$$\pi_1 \sim \mathrm{Beta}\,(r_1,\,u_1)$$

$$\lambda_2^2 \sim \mathrm{Gamma}\,(a_2,\,b_2)$$

$$\sigma^2 \sim \frac{1}{\sigma^2}$$

$$\boldsymbol{\mu}|\sigma^2,\varphi_{ij}^2 \sim \mathrm{N}_n(0,\sigma^2\Sigma_\mu)$$

$$\varphi_{ij}^2 \sim \frac{\lambda_1^2}{2}\exp(-\frac{\lambda_1^2}{2}\varphi_{ij}^2)$$

$$\lambda_1^2 \sim \mathrm{Gamma}(a_1,b_1)$$

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

be expressed as

$$
\begin{aligned}
&\pi(\boldsymbol{\mu}, \boldsymbol{\beta}, \tau_j^2, \lambda_1^2, \lambda_2^2, \sigma^2, \varphi_{ij}^2, \pi_1 | Y) \\
&\quad \propto \prod_{i=1}^{n} (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu_i - \boldsymbol{X}_i \boldsymbol{\beta})^2 \right\} \\
&\quad \times \prod_{j=1}^{p} \Big( (1-\pi_1)(2\pi\sigma^2\tau_j^2)^{-1/2} \exp\{ -\frac{\beta_j^2}{2\sigma^2\tau_j^2} \} \mathbf{I}_{\{\boldsymbol{\beta}_j \neq 0\}} + \pi_1 \delta_0(\boldsymbol{\beta}_j) \Big) \\
&\quad \times \prod_{j=1}^{p} \frac{\lambda_2^2}{2} \exp(-\frac{\lambda_2^2}{2}\tau_j^2) \times (\lambda_2^2)^{a_2 - 1} \exp(-b_2 \lambda_2^2) \times \pi_1^{r_1 - 1} (1-\pi_1)^{u_1 - 1} \times \frac{1}{\sigma^2} \\
&\quad \times (2\pi\sigma^2)^{-\frac{n}{2}} |\Sigma_\mu|^{-\frac{1}{2}} \exp\{ -\frac{1}{2} \boldsymbol{\mu}^\top (\sigma^2 \Sigma_\mu)^{-1} \boldsymbol{\mu} \} \times \prod_{1 \leq i < j \leq n} \frac{\lambda_1^2}{2} \exp(-\frac{\lambda_1^2}{2} \varphi_{ij}^2) \times (\lambda_1^2)^{a_1 - 1} \exp(-b_1 \lambda_1^2)
\end{aligned}
$$

**Gibbs Sampler**

The full conditional posterior distribution of $\boldsymbol{\mu}$:

$$
\begin{aligned}
&\boldsymbol{\mu} | \text{rest} \\
&\quad \propto \exp\{ -\frac{1}{2\sigma^2} [ (\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\mu}^\top (\Sigma_\mu)^{-1} \boldsymbol{\mu} ] \} \\
&\quad \propto \exp\{ -\frac{1}{2\sigma^2} [ \boldsymbol{\mu}^\top (\mathbf{I}_n + {\Sigma_\mu}^{-1}) \boldsymbol{\mu} - 2\boldsymbol{\mu}^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) ] \}
\end{aligned}
$$

It can be found that the posterior distribution of $\boldsymbol{\mu}$ is $N(\mu_\mu, \sigma_\mu^2)$, where

$$
\begin{aligned}
\mu_\mu &= (\mathbf{I}_n + {\Sigma_\mu}^{-1})^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), \\
\sigma_\mu^2 &= \sigma_2 (\mathbf{I}_n + {\Sigma_\mu}^{-1})^{-1}.
\end{aligned}
$$

The full conditional posterior distribution of $\varphi_{ij}$:

$$
\begin{aligned}
&\varphi_{ij}^2 | \text{rest} \\
&\quad \propto \frac{1}{\sqrt{2\pi\sigma^2\varphi_{ij}^2}} \exp\{ -\frac{(\mu_i - \mu_j)^2}{2\sigma^2\varphi_{ij}^2} \} \exp(-\frac{\lambda_1^2 \varphi_{ij}^2}{2}) \\
&\quad = \frac{1}{\sqrt{2\pi\sigma^2\varphi_{ij}^2}} \exp\{ -\frac{1}{2} [ \frac{1}{\sigma^2\varphi_{ij}^2} (\mu_i - \mu_j)^2 + \lambda_1^2 \varphi_{ij}^2 ] \}
\end{aligned}
$$

Then,

$$\frac{1}{\varphi_{ij}^2}|\text{rest} \sim \text{Inverse-Gaussian}(\sqrt{\frac{\lambda_1^2\sigma^2}{(\mu_i-\mu_j)^2}},\ \lambda_1^2), 1 \le i < j \le n.$$

The full conditional posterior distribution of $\lambda_1^2$:

$$\begin{aligned}
&\lambda_1^2|\text{rest} \\
&\quad\propto \prod_{1\le i<j\le n} \frac{\lambda_1^2}{2}\exp(-\frac{\lambda_1^2}{2}\varphi_{ij}^2) \times (\lambda_1^2)^{a_1-1}\exp(-b_1\lambda_1^2) \\
&\quad= (\lambda_1^2)^{\frac{n(n-1)}{2}+a_1-1}\exp\{-\lambda_1^2(\sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+b_1)\}.
\end{aligned}$$

Therefore, the posterior distribution for $\lambda_1^2$ is Gamma($\frac{n(n-1)}{2}+a_1\ \sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+b_1$).

Denote $\mu_{(-\beta_j)} = \boldsymbol{\mu}+\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{X}_j\beta_j$, then $\beta_j|\text{rest} \sim (1-l_j)\text{N}(\mu_{\beta_j},\ \sigma^2\Sigma_{\beta_j}) + l_j\delta_0(\beta_j)$, where

$$\begin{aligned}
\mu_{\boldsymbol{\beta}_j} &= \Sigma_{\boldsymbol{\beta}_j}\boldsymbol{X}_j^\top(\boldsymbol{Y}-\mu_{(-\boldsymbol{\beta}_j)}), \\
\Sigma_{\boldsymbol{\beta}_j} &= \left(\boldsymbol{X}_j^\top\boldsymbol{X}_j+\frac{1}{\tau_j^2}\right)^{-1}, \\
l_j &= \frac{\pi_1}{\pi_1+(1-\pi_1)(\tau_j^2)^{-1/2}|\Sigma_{\boldsymbol{\beta}_j}|^{1/2}\exp\{\frac{1}{2\sigma^2}\Sigma_{\boldsymbol{\beta}_j}||\boldsymbol{X}_j^\top(\boldsymbol{Y}-\mu_{(-\beta_j)}||_2^2\}}.
\end{aligned}$$

The posterior of $\tau_j^2$ is:

$$\frac{1}{\tau_j^2}|\text{rest} \sim \begin{cases} \text{Inverse-Gamma}(1, \frac{\lambda_2^2}{2}) & \text{if } \beta_j = 0 \\ \text{Inverse-Gaussian}(\sqrt{\frac{\sigma^2}{\beta_j^2}\lambda_2^2},\ \lambda_2^2) & \text{if } \beta_j \ne 0 \end{cases}.$$

$\lambda_2^2$ has Gamma posterior distributions:

$$\lambda_2^2|\text{rest} \sim \text{Gamma}(a_2+p,\ \sum_{j=1}^{p}\frac{\tau_j^2}{2}+b_2).$$

$\pi_1$ has Gamma posterior distributions:

$$\pi_1|\text{rest} \sim \text{Beta}(r_1 - \sum_{j=1}^{p} \mathbf{I}_{\{\boldsymbol{\beta}_j \neq 0\}} + p,\ u_1 + \sum_{j=1}^{p} \mathbf{I}_{\{\boldsymbol{\beta}_j \neq 0\}}).$$

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

$$\mu_{\sigma^2} = n + \frac{\sum_{j=1}^{p} \mathbf{I}_{\{\boldsymbol{\beta}_j \neq 0\}}}{2},$$
$$\Sigma_{\sigma^2} = \frac{(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{X}\boldsymbol{\beta}) + \sum_{j=1}^{p} (\tau_j^2)^{-1} \boldsymbol{\beta}_j^\top \boldsymbol{\beta}_j + \boldsymbol{\mu}^\top (\Sigma_\mu)^{-1} \boldsymbol{\mu}}{2}.$$

### C.1.5 Bayesian subgroup (sgroup)

**Hierarchical model specification**

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu_i - \boldsymbol{X}_i \boldsymbol{\beta})^2 \right\}$$
$$\beta_j | \tau_j^2, \sigma^2 \overset{iid}{\sim} \text{N}\left(0,\ \sigma^2 \tau_j^2\right) \quad j = 1, \ldots, p$$
$$\tau_j^2 | \lambda_2^2 \overset{iid}{\sim} \text{Gamma}(1, \frac{\lambda_2^2}{2}) \quad j = 1, \ldots, p$$
$$\lambda_2^2 \sim \text{Gamma}\left(a_2,\ b_2\right)$$
$$\sigma^2 \sim \frac{1}{\sigma^2}$$
$$\boldsymbol{\mu} | \sigma^2, \varphi_{ij}^2 \sim \text{N}_n(0, \sigma^2 \Sigma_\mu)$$
$$\varphi_{ij}^2 \sim \frac{\lambda_1^2}{2} \exp(-\frac{\lambda_1^2}{2} \varphi_{ij}^2)$$
$$\lambda_1^2 \sim \text{Gamma}(a_1, b_1)$$

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

be expressed as

$$
\begin{aligned}
&\pi(\boldsymbol{\mu}, \boldsymbol{\beta}, \tau_j^2, \lambda_1^2, \lambda_2^2, \sigma^2, \varphi_{ij}^2, \pi_1 | Y) \\
&\quad \propto \prod_{i=1}^{n} (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu_i - \boldsymbol{X}_i \boldsymbol{\beta})^2 \right\} \\
&\qquad \times \prod_{j=1}^{p} (2\pi\sigma^2\tau_j^2)^{-1/2} \exp\{ -\frac{\beta_j^2}{2\sigma^2\tau_j^2} \} \times \prod_{j=1}^{p} \frac{\lambda_2^2}{2} \exp(-\frac{\lambda_2^2}{2}\tau_j^2) \times (\lambda_2^2)^{a_2 - 1} \exp(-b_2\lambda_2^2) \times \frac{1}{\sigma^2} \\
&\qquad \times (2\pi\sigma^2)^{-\frac{n}{2}} |\Sigma_\mu|^{-\frac{1}{2}} \exp\{ -\frac{1}{2} \boldsymbol{\mu}^\top (\sigma^2 \Sigma_\mu)^{-1} \boldsymbol{\mu} \} \times \prod_{1 \le i < j \le n} \frac{\lambda_1^2}{2} \exp(-\frac{\lambda_1^2}{2}\varphi_{ij}^2) \times (\lambda_1^2)^{a_1 - 1} \exp(-b_1\lambda_1^2)
\end{aligned}
$$

**Gibbs Sampler**

The full conditional posterior distribution of $\boldsymbol{\mu}$:

$$
\begin{aligned}
&\boldsymbol{\mu} | \text{rest} \\
&\quad \propto \exp\{ -\frac{1}{2\sigma^2} [ (\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\mu}^\top (\Sigma_\mu)^{-1} \boldsymbol{\mu} ] \} \\
&\quad \propto \exp\{ -\frac{1}{2\sigma^2} [ \boldsymbol{\mu}^\top (\mathbf{I}_n + {\Sigma_\mu}^{-1}) \boldsymbol{\mu} - 2\boldsymbol{\mu}^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) ] \}
\end{aligned}
$$

It can be found that the posterior distribution of $\boldsymbol{\mu}$ is $N(\mu_\mu, \sigma_\mu^2)$, where

$$
\begin{aligned}
\mu_\mu &= (\mathbf{I}_n + {\Sigma_\mu}^{-1})^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), \\
\sigma_\mu^2 &= \sigma_2 (\mathbf{I}_n + {\Sigma_\mu}^{-1})^{-1}.
\end{aligned}
$$

The full conditional posterior distribution of $\varphi_{ij}$:

$$
\begin{aligned}
&\varphi_{ij}^2 | \text{rest} \\
&\quad \propto \frac{1}{\sqrt{2\pi\sigma^2\varphi_{ij}^2}} \exp\{ -\frac{(\mu_i - \mu_j)^2}{2\sigma^2\varphi_{ij}^2} \} \exp(-\frac{\lambda_1^2\varphi_{ij}^2}{2}) \\
&\quad = \frac{1}{\sqrt{2\pi\sigma^2\varphi_{ij}^2}} \exp\{ -\frac{1}{2} [ \frac{1}{\sigma^2\varphi_{ij}^2} (\mu_i - \mu_j)^2 + \lambda_1^2\varphi_{ij}^2 ] \}
\end{aligned}
$$

Then,

$$\frac{1}{\varphi_{ij}^2}|\text{rest} \sim \text{Inverse-Gaussian}(\sqrt{\frac{\lambda_1^2\sigma^2}{(\mu_i-\mu_j)^2}},\ \lambda_1^2), 1 \le i < j \le n.$$

The full conditional posterior distribution of $\lambda_1^2$:

$$\begin{aligned}
&\lambda_1^2|\text{rest} \\
&\propto \prod_{1\le i<j\le n} \frac{\lambda_1^2}{2}\exp(-\frac{\lambda_1^2}{2}\varphi_{ij}^2) \times (\lambda_1^2)^{a_1-1}\exp(-b_1\lambda_1^2) \\
&= (\lambda_1^2)^{\frac{n(n-1)}{2}+a_1-1}\exp\{-\lambda_1^2(\sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+b_1)\}.
\end{aligned}$$

Therefore, the posterior distribution for $\lambda_1^2$ is Gamma($\frac{n(n-1)}{2}+a_1\ \sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+b_1$).

Denote $\mu_{(-\beta_j)} = \boldsymbol{\mu} + \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}_j\beta_j$, then $\beta_j|\text{rest} \sim \text{N}(\mu_{\beta_j}, \sigma^2\Sigma_{\beta_j})$, where

$$\begin{aligned}
\mu_{\boldsymbol{\beta}_j} &= \Sigma_{\boldsymbol{\beta}_j}\boldsymbol{X}_j^\top(\boldsymbol{Y}-\mu_{(-\boldsymbol{\beta}_j)}), \\
\Sigma_{\boldsymbol{\beta}_j} &= \left(\boldsymbol{X}_j^\top\boldsymbol{X}_j+\frac{1}{\tau_j^2}\right)^{-1},
\end{aligned}$$

The posterior of $\tau_j^2$ is:

$$\frac{1}{\tau_j^2}|\text{rest} \sim \text{Inverse-Gaussian}(\sqrt{\frac{\sigma^2}{\beta_j^2}\lambda_2^2}, \lambda_2^2).$$

$\lambda_2^2$ has Gamma posterior distributions:

$$\lambda_2^2|\text{rest} \sim \text{Gamma}(a_2+p,\ \sum_{j=1}^{p}\frac{\tau_j^2}{2}+b_2).$$

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

$$\begin{aligned}
\mu_{\sigma^2} &= n+\frac{p}{2}, \\
\Sigma_{\sigma^2} &= \frac{(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X}\boldsymbol{\beta})+\sum_{j=1}^{p}(\tau_j^2)^{-1}\boldsymbol{\beta}_j^\top\boldsymbol{\beta}_j+\boldsymbol{\mu}^\top(\Sigma_\mu)^{-1}\boldsymbol{\mu}}{2}.
\end{aligned}$$

### C.1.6 Bayesian subgroup (subgroup)

**Hierarchical model specification**

$$Y \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu_i - \boldsymbol{X}_i\boldsymbol{\beta})^2\right\}$$

$$\beta_j \overset{iid}{\sim} \mathrm{N}\,(0,\, \sigma_0) \quad j = 1, \ldots, p$$

$$\sigma^2 \sim \frac{1}{\sigma^2}$$

$$\boldsymbol{\mu}|\sigma^2, \varphi_{ij}^2 \sim \mathrm{N}_n(0, \sigma^2\Sigma_\mu)$$

$$\varphi_{ij}^2 \sim \frac{\lambda^2}{2}\exp(-\frac{\lambda^2}{2}\varphi_{ij}^2)$$

$$\lambda^2 \sim \mathrm{Gamma}(a, b)$$

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

$$\begin{aligned}
&\pi(\boldsymbol{\mu}, \boldsymbol{\beta}, \tau_j^2, \lambda_1^2, \lambda_2^2, \sigma^2, \varphi_{ij}^2, \pi_1|Y) \\
&\quad \propto \prod_{i=1}^{n}(\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu_i - \boldsymbol{X}_i\boldsymbol{\beta})^2\right\} \\
&\qquad \times \prod_{j=1}^{p}(2\pi\sigma_0)^{-1/2}\exp\{-\frac{\beta_j^2}{2\sigma_0}\} \times \frac{1}{\sigma^2} \times (2\pi\sigma^2)^{-\frac{n}{2}}|\Sigma_\mu|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\mu}^\top(\sigma^2\Sigma_\mu)^{-1}\boldsymbol{\mu}\} \\
&\qquad \times \prod_{1\le i<j\le n} \frac{\lambda^2}{2}\exp(-\frac{\lambda^2}{2}\varphi_{ij}^2) \times (\lambda^2)^{a-1}\exp(-b\lambda^2)
\end{aligned}$$

**Gibbs Sampler**

The full conditional posterior distribution of $\boldsymbol{\mu}$:

$$\begin{aligned}
&\boldsymbol{\mu}|\text{rest} \\
&\quad\propto \exp\{-\frac{1}{2\sigma^2}[(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X\beta})^\top(\boldsymbol{Y}-\boldsymbol{\mu}-\boldsymbol{X\beta})+\boldsymbol{\mu}^\top(\Sigma_\mu)^{-1}\boldsymbol{\mu}]\} \\
&\quad\propto \exp\{-\frac{1}{2\sigma^2}[\boldsymbol{\mu}^\top(\mathbf{I}_n+{\Sigma_\mu}^{-1})\boldsymbol{\mu}-2\boldsymbol{\mu}^\top(\boldsymbol{Y}-\boldsymbol{X\beta})]\}
\end{aligned}$$

It can be found that the posterior distribution of $\boldsymbol{\mu}$ is $N(\mu_\mu, \sigma_\mu^2)$, where

$$\begin{aligned}
\mu_\mu &= (\mathbf{I}_n+{\Sigma_\mu}^{-1})^{-1}(\boldsymbol{Y}-\boldsymbol{X\beta}), \\
\sigma_\mu^2 &= \sigma_2(\mathbf{I}_n+{\Sigma_\mu}^{-1})^{-1}.
\end{aligned}$$

The full conditional posterior distribution of $\varphi_{ij}$:

$$\begin{aligned}
&\varphi_{ij}^2|\text{rest} \\
&\quad\propto \frac{1}{\sqrt{2\pi\sigma^2\varphi_{ij}^2}}\exp\{-\frac{(\mu_i-\mu_j)^2}{2\sigma^2\varphi_{ij}^2}\}\exp(-\frac{\lambda^2\varphi_{ij}^2}{2}) \\
&\quad= \frac{1}{\sqrt{2\pi\sigma^2\varphi_{ij}^2}}\exp\{-\frac{1}{2}[\frac{1}{\sigma^2\varphi_{ij}^2}(\mu_i-\mu_j)^2+\lambda^2\varphi_{ij}^2]\}
\end{aligned}$$

Then,

$$\frac{1}{\varphi_{ij}^2}|\text{rest}\sim \text{Inverse-Gaussian}(\sqrt{\frac{\lambda^2\sigma^2}{(\mu_i-\mu_j)^2}},\ \lambda^2), 1\le i<j\le n.$$

The full conditional posterior distribution of $\lambda^2$:

$$\begin{aligned}
&\lambda^2|\text{rest} \\
&\quad\propto \prod_{1\le i<j\le n}\frac{\lambda_1^2}{2}\exp(-\frac{\lambda^2}{2}\varphi_{ij}^2)\times(\lambda_1^2)^{a-1}\exp(-b\lambda^2) \\
&\quad= (\lambda^2)^{\frac{n(n-1)}{2}+a-1}\exp\{-\lambda^2(\sum_{1\le i<j\le n}\frac{\varphi_{ij}^2}{2}+b)\}.
\end{aligned}$$

Therefore, the posterior distribution for $\lambda_1^2$ is Gamma($\frac{n(n-1)}{2} + a$ $\sum_{1 \le i < j \le n} \frac{\varphi_{ij}^2}{2} + b$).

Denote $\mu_{(-\beta_j)} = \boldsymbol{\mu} + \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}_j\beta_j$, then $\beta_j|\text{rest} \sim \text{N}(\mu_{\beta_j}, \Sigma_{\beta_j})$, where

$$\begin{aligned}
\mu_{\beta_j} &= \frac{1}{\sigma^2}\Sigma_{\beta_j}\boldsymbol{X}_j^\top(\boldsymbol{Y} - \mu_{(-\beta_j)}), \\
\Sigma_{\beta_j} &= \left(\frac{1}{\sigma^2}\boldsymbol{X}_j^\top\boldsymbol{X}_j + \sigma_0^{-1}\right)^{-1},
\end{aligned}$$

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

$$\begin{aligned}
\mu_{\sigma^2} &= n, \\
\Sigma_{\sigma^2} &= \frac{(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\mu}^\top(\Sigma_\mu)^{-1}\boldsymbol{\mu}}{2}.
\end{aligned}$$