High-dimensional variable selection for genomics data, from both frequentist and Bayesian perspectives

by

Jie Ren

M.S., Kansas State University, 2015

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

2020

Abstract

Variable selection is one of the most popular tools for analyzing high-dimensional genomic data. It has been developed to accommodate complex data structures and lead to structured sparse identification of important genomics features. We focus on the network and interaction structure that commonly exist in genomic data, and develop novel variable selection methods from both frequentist and Bayesian perspectives.

Network-based regularization has achieved success in variable selections for high-dimensional cancer genomic data, due to its ability to incorporate the correlations among genomic features. However, as survival time data usually follow skewed distributions, and are contaminated by outliers, network-constrained regularization that does not take the robustness into account leads to false identifications of network structure and biased estimation of patients' survival. In the first project, we develop a novel robust network-based variable selection method under the accelerated failure time (AFT) model. Extensive simulation studies show the advantage of the proposed method over the alternative methods. Promising findings are made in two case studies of lung cancer datasets with high dimensional gene expression measurements.

Gene-environment (G×E) interactions are important for the elucidation of disease etiology beyond the main genetic and environmental effects. In the second project, a novel and powerful semi-parametric Bayesian variable selection model has been proposed to investigate linear and nonlinear G×E interactions simultaneously. It can further conduct structural identification by distinguishing nonlinear interactions from main-effects-only case within the Bayesian framework. The proposed method conducts Bayesian variable selection more efficiently and accurately than alternatives. Simulation shows that the proposed model outperforms competing alternatives in terms of both identification and prediction. In the

case study, the proposed Bayesian method leads to the identification of effects with important implications in a high-throughput profiling study with high-dimensional SNP data.

In the last project, a robust Bayesian variable selection method has been developed for G×E interaction studies. The proposed robust Bayesian method can effectively accommodate heavy-tailed errors and outliers in the response variable while conducting variable selection by accounting for structural sparsity. Spike and slab priors are incorporated on both individual and group levels to identify the sparse main and interaction effects. Extensive simulation studies and analysis of both the diabetes data with SNP measurements from the Nurses' Health Study and TCGA melanoma data with gene expression measurements demonstrate the superior performance of the proposed method over multiple competing alternatives.

To facilitate reproducible research and fast computation, we have developed open source R packages for each project, which provide highly efficient C++ implementation for all the proposed and alternative approaches. The R packages *regnet* and *spinBayes*, associated with the first and second project correspondingly, are available on CRAN. For the third project, the R package *robin* is available from GitHub and will be submitted to CRAN soon.

High-dimensional variable selection for genomics data, from both frequentist and Bayesian perspectives

by

Jie Ren

M.S., Kansas State University, 2015

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

2020

Approved by:

Major Professor
Cen Wu

Copyright

Abstract

Variable selection is one of the most popular tools for analyzing high-dimensional genomic data. It has been developed to accommodate complex data structures and lead to structured sparse identification of important genomics features. We focus on the network and interaction structure that commonly exist in genomic data, and develop novel variable selection methods from both frequentist and Bayesian perspectives.

Network-based regularization has achieved success in variable selections for high-dimensional cancer genomic data, due to its ability to incorporate the correlations among genomic features. However, as survival time data usually follow skewed distributions, and are contaminated by outliers, network-constrained regularization that does not take the robustness into account leads to false identifications of network structure and biased estimation of patients' survival. In the first project, we develop a novel robust network-based variable selection method under the accelerated failure time (AFT) model. Extensive simulation studies show the advantage of the proposed method over the alternative methods. Promising findings are made in two case studies of lung cancer datasets with high dimensional gene expression measurements.

Gene-environment (G×E) interactions are important for the elucidation of disease etiology beyond the main genetic and environmental effects. In the second project, a novel and powerful semi-parametric Bayesian variable selection model has been proposed to investigate linear and nonlinear G×E interactions simultaneously. It can further conduct structural identification by distinguishing nonlinear interactions from main-effects-only case within the Bayesian framework. The proposed method conducts Bayesian variable selection more efficiently and accurately than alternatives. Simulation shows that the proposed model outperforms competing alternatives in terms of both identification and prediction. In the

case study, the proposed Bayesian method leads to the identification of effects with important implications in a high-throughput profiling study with high-dimensional SNP data.

In the last project, a robust Bayesian variable selection method has been developed for G×E interaction studies. The proposed robust Bayesian method can effectively accommodate heavy-tailed errors and outliers in the response variable while conducting variable selection by accounting for structural sparsity. Spike and slab priors are incorporated on both individual and group levels to identify the sparse main and interaction effects. Extensive simulation studies and analysis of both the diabetes data with SNP measurements from the Nurses' Health Study and TCGA melanoma data with gene expression measurements demonstrate the superior performance of the proposed method over multiple competing alternatives.

To facilitate reproducible research and fast computation, we have developed open source R packages for each project, which provide highly efficient C++ implementation for all the proposed and alternative approaches. The R packages *regnet* and *spinBayes*, associated with the first and second project correspondingly, are available on CRAN. For the third project, the R package *robin* is available from GitHub and will be submitted to CRAN soon.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to my major advisor, Dr. Cen Wu, who guided me throughout this project. I am extremely fortunate to have a caring major advisor who helped me through my graduate years. He always responded to my questions and ideas with patience. If it wasn't for his enduring support and pressure in the right direction when needed, I would not have finished this manuscript.

I want to extend my thanks to Dr. Christopher Vahl, Dr. Haiyan Wang, Dr. Weiqun Wang and Dr. Mark Ungerer, for serving on my committee and for providing support and advice through my study process. I would like to thank the families of Coyne, Fryer and Siepman for providing scholarships to graduate students in the Department of Statistics. Without this support, it would have been harder to complete my research in a timely manner. I also want to thank the coauthors of the papers that went into this dissertation. Without their collaboration and expertise, the papers would not have come together so neatly.

Finally, I want to express my deepest gratitude to my family for their unwavering support and love. To my parents, thank you for always believing in me and wanting the best for me. To my significant other, Pat, thank you for helping me going through all the ups and downs in my master and Ph.D. journey with your unending love, patience, understanding and encouragement. This accomplishment would not have been possible without them.

Chapter 1

Introduction

Due to the rapid advance of high-throughput biotechnologies, enormous amounts of omics data have been collected at various levels of biological systems. A representative example is the genotyping analysis performed by microarray technologies. The data from genotyping can be considered as a matrix with columns corresponding to variables, such as SNPs, and rows corresponding to samples. This data matrix has the "large p, small n" nature, that is, the number of genetic features measured is much larger than the sample size. This high-dimensionality of the high-throughput data has brought new challenges to the statistical modeling. Since only a small subset of the genetic features is associated with the clinic outcome of interest, an important question to answer is "How to identify the important features from a large-scale candidate pool?". With genetic features treated as variables in a statistical model, this question can be reformulated as a variable section problem. The least absolute shrinkage and selection operator (LASSO) is one of the most popular variable selection methods for analyzing high-dimensional data (Tibshirani, 1996). By shrinking the coefficients towards zero, the LASSO can effectively exclude irrelevant variables from the model and produce sparse estimation of the coefficients. The term sparsity refers to the phenomena of 0's among the estimated coefficients. Over the past decades, plenty of variants of LASSO, along with other variable selection methods have been developed for high-dimensional data. Here we provide a brief overview for popular variable selection methods

in omics studies from both frequentist and Bayesian perspectives.

1.1 Penalized Variable Selection

Penalization or regularization has become one of the most popular frameworks for selecting important features in omics studies. Let Y be a vector of the disease outcome, where Y can be a continuous disease phenotype (e.g. body weight in obesity studies), categorical disease status (e.g. cancer stages) or survival time of patients. Let X be the design matrix of the p-dimensional genomics features. X can be various omics measurements such as SNPs, DNA methylation and gene expressions, among others. The penalized regression model can be expressed as

$$L(\beta; Y, X) + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \tag{1.1}$$

where $L(\beta; Y, X)$ is the loss function that measures the lack of fit of the model, and $p_\lambda(\cdot)$ is a penalty function indexed by the regularization parameters $\lambda \geq 0$. The penalty function imposes shrinkage on the coefficient vector $\beta = (\beta_1, \ldots, \beta_p)^\top$. β_j is the coefficient corresponding to the jth feature in the high-dimensional genomics data. By minimizing the penalized loss function (1.1), variable selection can be achieved with penalized estimation simultaneously. In other words, variables whose regression coefficients are shrunk to zero ($\beta_j = 0$) are automatically excluded from the model. The well-known LASSO (Tibshirani (1996)) is a penalized least square regression with ℓ_1 penalty, which adopts the form as follows

$$\|Y - X\beta\|_2^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{1.2}$$

Fan and Li (2001) proposes three criteria for a good penalty function. In brief, a desired penalty function should result in a estimator that is continuous (continuity), while setting small estimated coefficients to zero (sparsity) and maintaining nearly unbiased estimates for large coefficients (unbiasedness). While LASSO meets the sparsity and continuity criteria, it leads to biased estimates, especially for the large coefficients, therefore fails to achieve

unbiasedness. This leads to developments of penalties for nearly unbiased variable selection, including the smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001), the minimax concave penalty (MCP) (Zhang, 2010) and adaptive LASSO (Zou, 2006). The penalties are defined as follows

$$\text{SCAD:} \quad p_{\lambda,\gamma}(\beta_j) = \begin{cases} \lambda|\beta_j| & |\beta_j| \leq \lambda \\ -\frac{\beta_j^2 - 2\gamma\lambda|\beta_j| + \lambda^2}{2(\gamma-1)} & \lambda < |\beta_j| \leq \gamma\lambda \\ \frac{1}{2}(\gamma+1)\lambda^2 & |\beta_j| > \gamma\lambda \end{cases}$$

$$\text{MCP:} \quad p_{\lambda,\gamma}(\beta_j) = \begin{cases} \lambda|\beta_j| - (2\gamma)^{-1}\beta_j^2 & |\beta_j| \leq \gamma\lambda \\ \frac{1}{2}\gamma\lambda^2 & |\beta_j| > \gamma\lambda \end{cases}$$

$$\text{Adaptive LASSO:} \quad p_{\lambda,\gamma}(\beta_j) = \lambda(|\beta_j^{(0)}|^{-\gamma})|\beta_j|$$

where $\gamma > 2$ for SCAD, $\gamma > 1$ for MCP and $\gamma > 0$ for adaptive LASSO. $\beta_j^{(0)}$ is the initial estimate of β. These penalties have been demonstrated to have attractive properties theoretically and practically. Together with LASSO, they are considered as the family of baseline penalization methods.

In order to take complex data structures into consideration, more advanced penalty functions have been developed. For example, grouping structures exist in many statistical modeling problems, such as a group of dummy variables for a categorical factor or a set of basis functions in nonparametric modeling. For selecting grouped variables, Yuan and Lin (2006) proposes the group LASSO method, with the penalty defined as

$$p_\lambda(\beta) = \lambda \sum_{g=1}^{G} \sqrt{L_g} \|\beta_g\|_2 \tag{1.3}$$

where β_g is a coefficient vector of length L_g ($\beta = (\beta_1^\top, ..., \beta_G^\top)^\top$). The term $\sqrt{L_g}$ adjusts the penalty for the varying group sizes, and $\|\cdot\|_2$ is the euclidean norm. Following

Yuan and Lin (2006), other baseline penalization methods have also been extended to the group setting (Huang et al., 2012). Another example of complex data structure is the high correlation among genetic features in omics data, which have been widely observed and reported. The elastic net (Zou and Hastie (2005)) and the fused-lasso (Tibshirani et al. (2005)) are two popular choices for analyzing correlated genomic features. The elastic net uses a combination of the ℓ_1 and ℓ_2 penalties ($\lambda_1\|\beta\|_2^2 + \lambda_2|\beta|$), which encourages a grouping effect and tends to drop or select highly correlated predictors together. The fused-lasso ($\lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=2}^{p} |\beta_j - \beta_{j-1}|$) induces smoothness among the coefficients of neighboring features. An application example of the fused-lasso is the DNA copy number variations (CNVs) data. The CNVs form block structures along their gemonic location, and the fused-lasso can promote this type of block structure on the penalized estimates. In order to utilize the correlation information in a more efficient and flexible way, the network-constrained regularization approaches have been developed, such as Li and Li (2008) and Huang et al. (2011), among many others. In particular, Huang et al. (2011) develops the sparse Laplacian shrinkage (SLS) penalty, which is built upon the combination of MCP (Zhang, 2010) and Laplacian quadratic that is associated with a graph. The SLS penalty takes the form

$$\sum_{j=1}^{p} p_{\lambda_1,\gamma}(\beta_j) + \lambda_2 \sum_{1\le j<k\le p} |a_{jk}|(\beta_j - \text{sgn}(a_{jk})\beta_k)^2 \tag{1.4}$$

where $p_{\lambda_1,\gamma}(\beta_j)$ is the MCP penalty with tuning parameter λ_1, and regularization parameter γ (Zhang, 2010), and $|a_{jk}|$ is the measure of connection intensity between variables x_j and x_k. The MCP penalty in (1.4) promotes sparsity in the model, and the second term encourages smoothness among the coefficient profiles of the related covariates. As shown in Huang et al. (2011), the penalty in (1.4) is capable of taking correlation structure into account without introducing extra bias. They also show that in high dimension settings with $p \gg n$ under certain assumptions, SLS is selection consistent and equivalent to the oracle Laplacian shrinkage estimator with high probability. Recently, as multi-layer high-dimensional omics data has become available, these popular penalization methods have also been adopted to integrate multi-omics data (Du et al. (2020); Jiang et al. (2019); Li et al. (2020)).

In the analysis of omics data, it is not uncommon to encounter model mis-specification and heterogeneity problems, like heavy-tailed errors and outliers in response variables and the contamination in predictors. Such challenges demand the development of robust methods that can accommodate data contamination and can be insensitive to model specification. For penalized regression, a practical way to obtain robustness is to adopt the "robust loss function + penalty" form. Wu and Ma (2015) surveys some broadly adopted robust loss functions, including check loss function, the least absolute deviation (LAD) loss function, rank-based loss function and their variants. The high computational burden resulting from the nonsmoothness of the loss functions is a major limitation of robust methods, especially when it comes to the analysis of high-dimensional data. The LAD loss function, as a special case of quantile-based loss functions, is especially appealing for omics studies due to its computational convenience (Wu and Ma (2015); Huang et al. (2007); Wu et al. (2018a)). The LAD-based penalization can be expressed as

$$\sum_{i=1}^{n} |y_i - x_i^\top \beta| + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \tag{1.5}$$

For a contaminated observation with y_i significantly deviating from $x_i^\top \hat{\beta}$, the predicted value from model (2.1), the ℓ_1 based loss down-weighs such a deviation, while the non-robust loss, for example, least square based loss, results in a much larger deviation.

1.2 Bayesian Variable Selection

O'Hara and Sillanpää (2009) classifies different Bayesian variable selection approaches into four categories (1) adaptive shrinkage (2) indicator model selection (3) stochastic search variable selection (SSVS) and (4) model space approach. In particular, the first category, adaptive shrinkage, has a tight connection with the frequentist variable selection methods.

From a Bayesian perspective, the LASSO estimates can be interpreted as posterior mode estimates when the regression parameters have independent and identical Laplace priors $\pi(\beta) = \prod_{j=1}^{p} \frac{\lambda}{2} \exp\left\{ -\lambda |\beta_j| \right\}$ (Tibshirani (1996)). Specifically, with a independent priori

$\pi(\sigma^2)$ on σ^2, the posterior distribution can be expressed as

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

The maximum a posteriori probability (MAP) estimate for β is equivalent to the LASSO estimate in (1.2), for any fixed value of σ^2. Park and Casella (2008) proposes a fully Bayesian analysis for lasso regression, the Bayesian LASSO, by adopting a conditional Laplace prior on β

$$\pi(\beta|\sigma^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sigma} \exp\Big\{ - \frac{\lambda}{\sigma} |\beta_j| \Big\}$$

They also demonstrate that this conditional Laplace prior guarantees the unimodality of the posterior distribution. Although this fully Bayesian analysis produces similar shrinkage on individual coefficients as the LASSO, it cannot set a posterior estimate to zero exactly. Kyung et al. (2010) extends this fully Bayesian framework to a more general form that can accommodate the group LASSO (Yuan and Lin, 2006), the fused LASSO (Tibshirani et al., 2005) and the elastic net (Zou and Hastie, 2005). In particular, the group lasso can be represented by adopting a multivariate Laplace prior

$$\pi(\beta_g|\sigma^2) \propto \exp\Big\{ - \frac{\sqrt{L_g}\lambda}{\sigma}\|\beta_g\|_2\Big\} \tag{1.6}$$

where β_g and L_g are defined at the same way as in (1.3) and $\frac{\sqrt{L_g}\lambda_v}{\sigma}$ is the scale parameter of the multivariate Laplace. These methods share the same drawback as the Bayesian LASSO in the lack of the model selection property. This difficulty can be overcome by borrowing strength from the spike-and-slab priors (Mitchell and Beauchamp, 1988) that have been widely used in other Bayesian variable selection methods, such as indicator model selection and SSVS. The spike-and-slab priors on $\beta_j (j = 1, \ldots, p)$ can be expressed in a general form

$$\beta_j|\phi_j \overset{ind}{\sim} \phi_j \pi_1(\beta_j) + (1 - \phi_j)\pi_0(\beta_j)$$

where $\pi_1(\cdot)$ is a flat "slab distribution" for modeling large effects, $\pi_0(\cdot)$ is a "spike distribution" either exactly at or concentrated around zero for modeling negligibly small effects, and $\phi_j \in \{0, 1\}$ is a auxiliary indicator variable. If $\phi_j = 1$, $\beta_j \sim \pi_1(\beta_j)$ indicates the presence of the jth genetic effect in the model. The opposite occurs when $\phi_j = 0$. Conventionally, $\pi_1(\cdot)$ is set to a normal distribution with large variance. The choices of $\pi_0(\cdot)$ leads to the different variable selection methods. The SSVS method proposed by George and McCulloch (1993) adopts a mixture prior of two normal distributions $(\phi_j \mathrm{N}(0, c_j\tau_j^2) + (1 - \phi_j)\mathrm{N}(0, \tau_j^2))$, where the spike part (the second density) has a small variance τ_j^2 and centers around zero. Kuo and Mallick (1998) adopts a point mass mixture prior where the π_0 is defined as $\delta_0(\beta_j)$, so that coefficients of unimportant effects can be set to zero in the spike part.

Many methods have been developed to combine the point mass mixture prior and Laplace shrinkage for variable selection. Yuan and Lin (2005) proposes an empirical Bayesian variable selection method with prior $\beta_j|\phi_j \overset{ind}{\sim} \phi_j \frac{\lambda}{2}\exp(-\lambda|\beta_j|) + (1 - \phi_j)\delta_0(\beta_j)$. Zhang et al. (2014a) generalizes this strategy to group variable selection. Xu and Ghosh (2015) proposes a sparse group selection method to select variables both at the group level and (within group) individual level. In Zhang et al. (2014a) and Xu and Ghosh (2015), a multivariate Laplace distribution is used as the slab part for imposing shrinkage at the group level:

$$\beta_g|\phi_g, \sigma^2 \sim \phi_g \,\text{M-Laplace}(0, \frac{\sigma}{\sqrt{L_g}\lambda}) + (1 - \phi_g)\delta_0(\beta_g)$$

where the density function of a L-dimensional multivariate Laplace distribution is

$$\text{M-Laplace}(x|0, C^{-1}) \propto C^L \exp(-C\|x\|_2)$$

Ročková and George (2018) adopts a mixture prior of two Laplace distributions in SSVS method. Specifically, they defines $\pi_0(\beta_j) = \frac{\lambda_1}{2}\exp(-\lambda_1|\beta_j|)$ with λ_1 small and $\pi_0(\beta_j) = \frac{\lambda_0}{2}\exp(-\lambda_0|\beta_j|)$ with λ_0 large. It is worth noting that there are other priors that can be incorporated within the spike and slab form, such as the Zellner's g-prior (Zhang et al. (2016)). They can also be used for variable selection purposes.

To tackle the challenges of heterogeneity, robust Bayesian variable selection methods have also been developed. Sha et al. (2006) proposes a Bayesian variable selection approach for censored survival data under the accelerated failure time (AFT) models. The robustness of the model is obtained by adopting a log-t distribution for the failure time (survival time). Another example of leveraging the heavy-tailed t distribution for robustness is the study in Yi and Xu (2008). They develop a variant of Bayesian LASSO for quantitative trait loci (QTL) mapping by assigning t distribution priors on β. Ren et al. (2020b) adopts a Bayesian formulation of the least absolute deviation (LAD) regression to accommodate data contamination and long-tailed distributions in the phenotype in Gene-environment studies.

1.3 Other Variable Selection Methods

We focus on penalization and Bayesian variable selection here since the two are the foci of this dissertation. It is worth noting that there exists a diversity of variable selection methods that are also applicable in analyzing omics data. For example, boosting and random forest are popular machine learning techniques for feature selection. In random forest, the importance measures of variables can guide the decision of whether a variable should be included in the model or not (Breiman, 2001). Boosting is an ensemble procedure that combines the outputs of many weak learners to produce a powerful strong learner. In omics study, weak learners can be individual SNP, gene expression and other genetic features. In each iteration, boosting selects the variables having the largest correlation with residuals corresponding to the current active set of selected predictors (the weak learners), and then fits the new model to recompute the residuals. Variable selection in boosting is achieved because the algorithm largely ignores non-informative predictors when fitting the models. The prediction power has improved significantly in boosting through aggregating multiple weak learners (Friedman (2001)).

1.4 Works in this dissertation

In Chapter 2, we develop a novel robust network-based variable selection method under the accelerated failure time (AFT) model for survival time in cancer genomic studies. We consider a least absolute deviation (LAD) loss function, where the ℓ_1 form introduces robustness by down weighing the deviation of the outlier. The computational cost of the proposed methods is much lower compared to other high dimensional variable selection methods for survival outcomes. To identify important genomic signatures while accommodating correlations under survival outcomes, we develop robust network-based penalization under the AFT model with Kaplan-Meier weights. The proposed penalty function is of an "MCP + L_1" form, where MCP encourages sparsity and the L_1 term incorporate network structures by promoting the smoothness among pairwise coefficient profiles of correlated gene expressions. We develop an effective algorithm that borrows strength from majorization-minimization (MM) within the coordinate descent (CD) framework. Extensive simulation studies show the advantage of the proposed method over the alternative methods. Two case studies of lung cancer datasets with high dimensional gene expression measurements demonstrate that the proposed approach has identified markers with important implications.

In Chapter 3, we explore Bayesian variable selection for Gene-Environment (G×E) interactions. We propose a novel and powerful semi-parametric Bayesian variable selection model that can investigate linear and nonlinear G×E interactions simultaneously. To accommodate possible nonlinear effects of environment factor, we first consider a partially linear varying coefficient model where the varying coefficient functions capture the possible non-linear G×E interaction, and the linear part models the G×E interactions with linear assumptions. The changing of basis with B splines is adopted to separate the coefficient functions with varying, non-zero constant and zero forms, corresponding to cases of nonlinear interaction, main effect only (no interaction) and no genetic interaction at all. Instead of using conventional Bayesian Lasso and Bayesian Group Lasso to impose penalty on individual and group levels, we adopt spike-and-slab priors with the slab parts coming from Laplace distributions, to shrink coefficients of irrelevant covariates to zero. Highly efficient Gibbs sampler has been

developed to carry out the computation. Simulation shows that the proposed model outperforms competing alternatives in terms of both identification and prediction. In the case study, the proposed Bayesian method leads to the identification of effects with important implications in a high-throughput profiling study with high-dimensional genetic variants.

Built upon our existing work in both robust network-constrained variable selection (Chapter 2) and robust penalization for G×E interactions (Wu et al., 2018a), we develop a novel robust Bayesian variable selection method to dissect G×E interactions in Chapter 4. Outliers and data contamination in disease phenotypes of G×E studies have been commonly encountered, leading to the development of a broad spectrum of robust penalization methods. Nevertheless, within the Bayesian framework, the issue has not been taken care of in existing studies. In this study, we propose a Bayesian method that can effectively accommodate heavy–tailed errors and outliers in the response variable while conducting variable selection by accounting for structural sparsity. In particular, the spike–and–slab priors have been imposed on both individual and group levels to identify important main and interaction effects. An efficient Gibbs sampler has been developed to facilitate fast computation. Extensive simulation studies and analysis of both the diabetes data with SNP measurements from the Nurses Health Study and TCGA melanoma data with gene expression measurements demonstrate the superior performance of the proposed method over multiple competing alternatives.

Chapter 2

Robust Network-Based Regularization and Variable Selection for High-Dimensional Genomic Data in Cancer Prognosis

2.1 Introduction

In cancer research, profiling studies have been extensively conducted to identify prognostic markers that may contribute to the development and progression of cancer. Important prognostic markers have the potential to shed deep insight in elucidating the genetic basis of cancer, and provide assistance in cancer prevention, diagnosis and treatment selection. The generation of unprecedented amount of high dimensional genomics data from the high-throughput profiling studies has led to the development of extensive regularized variable selection methods (Fan and Lv (2010)). The genomics features, such as gene expressions and single nucleotide polymorphism (SNPs), are treated as variables within the regularization (or penalization) framework. As the correlations among genomics features have been widely recognized, multiple studies have developed network based regularization methods to

accommodate interconnections among these features, including gene expressions (Li and Li (2008)), SNPs (Ren et al. (2017)), copy number variations (Peng et al. (2012)) and DNA methylations (Sun and Wang (2012)). Recently, network based penalization has also been adopted to integrate multi-omics data (Jiang et al. (2019)).

The network based methods have unique strength to effectively capture correlations by incorporating prior biological information via the network (or graph) structure, where the vertices of networks are the genomic features and the edges of the networks denote certain regulatory relationship among the features. Incorporation of the network structure in regularized variable selection has led to significant improvement in both identification accuracy and predictive performance, as demonstrated in aforementioned studies, as well as many other studies. Nevertheless, these methods have limitations. First, network–constrained regularization methods under survival outcomes have not received much attention. As markers identified under patients' survival have important implications in cancer prognosis, the network–based regularized variable selection will improve accuracy in both identifying prognostic markers and predicting patients' survival. However, the disease outcome investigated from published studies are mainly continuous (Li and Li (2008); Peng et al. (2012)), binary (Huang et al. (2018); Min et al. (2018); Ren et al. (2017); Sun and Wang (2012, 2013)) and multi-nomial (Tian et al. (2014)). Markers identified from these studies, though important, cannot be treated as potential prognostic markers directly. Second, existing network (or graph) based methods lack robust properties, which are critical to accommodate data contamination and long-tailed distributions. In studies that investigate the regulations of between CNVs and gene expressions (Peng et al. (2012)), as gene expressions may have heavy tailed distributions (especially at high expression levels) or be contaminated, inference of gene regulatory relationship based on non-robust methods might be biased.

We use the lung squamous cell carcinoma (LUSC) data collected by The Cancer Genome Atlas (TCGA) as a motivating example. For the 461 subjects analyzed in this study, five subjects have survival time 150.13, 151.15, 154.20, 156.54 and 173.69 months, respectively, while the rest 456 subjects have survival times ranging from 0.03 to 139.98 months. Figure 2.1 shows the plots of both empirical density function of the log survival time as well as the

corresponding best-fitted normal density. The deviation from normal is observed. Moreover, the Kolmogorov-Smirnov test yields a pvalue less than 0.01, which suggests a significant difference from normal distribution. Such a pattern may happen for multiple reasons. For example, when multiple cancer subtypes exist, the largest subtype can be viewed as being "contaminated" by small subtypes. Contamination of survival can also be caused by misclassification of causes of death (Rampatige et al. (2013)) and unreliable extraction of survival times from medical records (Fall et al. (2008)). Without taking robustness into consideration, non-robust network based methods will lead to biased estimation and thus false identification of network structure, even in the presence of only one contaminated observation. As shown in Wu and Ma (2015), for high-dimensional genomic data, the robust variable selection methods are still not well developed, which is particularly true for the network–constrained approaches, possibly due to the extra complexity from incorporating network structure to accommodate interconnections among genomic features.

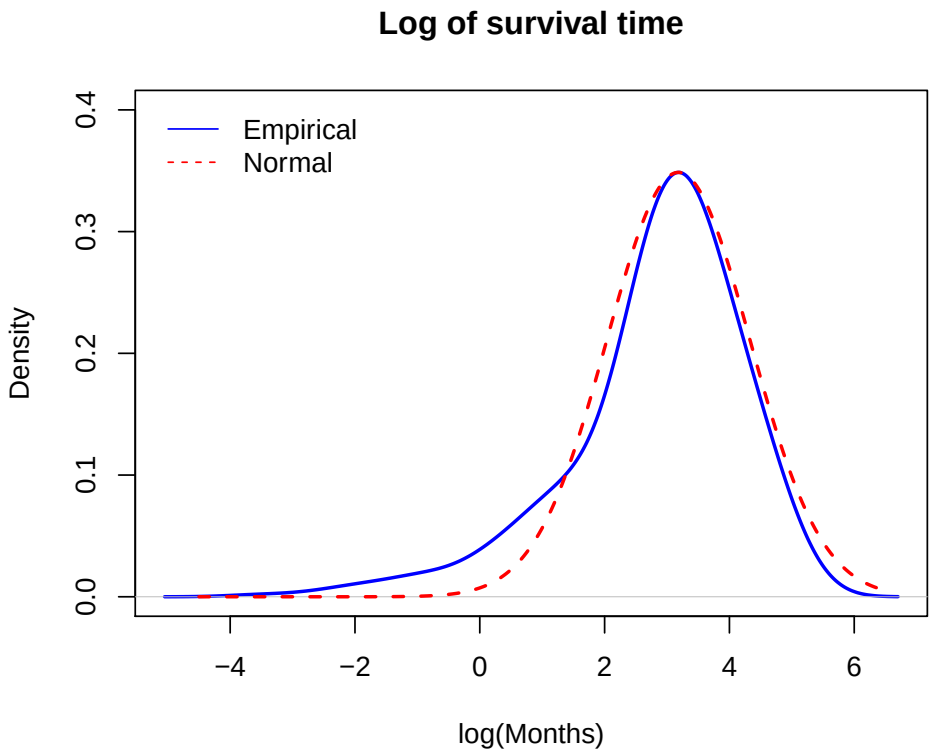

Figure 2.1: *Distribution of log(survival time) in the TCGA LUSC dataset.*

In this study, we propose a robust network–based regularization and variable selection method for high-dimensional genomics data in cancer prognosis. Our method has the following novel features to distinguish itself from existing ones. First, we adopt the least absolute deviation (LAD) loss function to accommodate heavy-tailed distribution and data contamination. Although no robust loss function universally outperforms the rest, the LAD loss function, as a special case of quantile-based loss functions, is especially appealing for high-dimensional data due to its L_1 form (Wu and Ma (2015); Huang et al. (2007); Wu et al. (2018a)). Other robust loss functions, including exponential square loss (Wang et al. (2013b)) and rank based loss (Wu et al. (2015)), do not enjoy such a computational convenience for data with high-dimensionality. Second, as our goal is to robustly identify important genomic signatures while accommodating correlations under survival outcomes, we develop robust network based penalization under the accelerated failure time (AFT) model with Kaplan-Meier weights. The proposed penalty function is of an "MCP + L_1" form, where MCP, the Minimax Concave Penalty, encourages sparsity (Zhang (2010)) and the L_1 term promotes network structure. Besides, although the weighted LAD estimator has been investigated in Huang et al. (2007), the strength of its regularized counterpart has not been fully explored, especially for network structure estimation and identification. Third, we develop an effective algorithm within the coordinate descent framework. On the contrary, the computational cost for many robust variable selection methods are prohibitively high under complicated data and model settings (Wu and Ma (2015)). The advantage of our method over alternatives has been convincingly demonstrated in both simulation studies and two case studies. To the best of our knowledge, identifying important genomic features in cancer prognostic studies through robust penalization by incorporating network structures has not been reported before. It is also noting that our method is not restricted to cancer survival only. Instead, it can be readily extended to other types of response, such as the continuous disease phenotypes.

2.2 Statistical Methods

We consider the AFT model for cancer prognosis. For high-dimensional genomics data, the AFT model is adopted over the Cox model and other alternatives due to its lower computational cost. From now on, we use gene expression as a representative example of genomics features.

2.2.1 The LAD Regression for Censored Data

Denote the ith subject by using the subscript i. Let (T_i, X_i, Z_i) $(i = 1, \ldots, n)$ be n independent and identically distributed random vectors, where T_i is the logarithm of survival time, $X_i = (x_{i1}, x_{i2} \ldots, x_{ip})^T$ is the p–dimensional vector of gene expressions, and $Z_i = (z_{i0}, z_{i1} \ldots, z_{iq})^T$ is the $(q+1)$–dimensional vector of which the first component $z_{i0} = 1$ and the last q components are clinical/environmental covariates. Usually, q and p are of low and high dimensionality, respectively. The AFT model postulates that

$$T_i = Z_i\alpha + X_i\beta + \varepsilon_i$$

$\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_q)^T$ where α_0 is the intercept and the last q components are the regression coefficients for the clinical covariates. $\beta = (\beta_1, \ldots, \beta_p)^T$ is the regression coefficient vector for the gene expressions, and ε_i is the error term with an unspecified distribution. Denote C_i as the logarithm of the censoring time. Under right censoring, we observe $(Y_i, \delta_i, Z_i, X_i)$, where $Y_i = \min(T_i, C_i)$, and $\delta_i = I(T_i \leq C_i)$ is the indicator of event. Without loss of generality, we assume that $\{(Y_i, \delta_i, Z_i, X_i), i = 1, \ldots, n\}$ have been sorted with respect to Y_i in an ascending order.

We adopt the Kaplan-Meier weights for censoring. Let $\hat{F}_n$ be the Kaplan–Meier estimator of the distribution function F of T. Then by following Stute and Wang (1993), we have

$\hat{F}_n(y) = \sum_{i=1}^n v_i 1\{Y_i \le y\}$, where the Kaplan–Meier weights v_i $(i = 1, \ldots, n.)$ are defined as

$$v_1 = \frac{\delta_1}{n}, \quad v_i = \frac{\delta_i}{n-i+1} \prod_{j=1}^{i-1} (\frac{n-j}{n-j+1})^{\delta_j}, i = 2, \ldots, n.$$

To accommodate data contamination, consider the weighted LAD loss function

$$L(\alpha, \beta) = \frac{1}{n} \sum_{i=1}^{n} v_i |Y_i - Z_i\alpha - X_i\beta\,| \tag{2.1}$$

The robustness comes from the L_1 form of the loss function. For a contaminated observation with Y_i significantly deviating from $Z_i\hat{\alpha} + X_i\hat{\beta}$, the predicted value from model (2.1), the L_1 based loss down-weighs such a deviation, while the non-robust loss, for example, least square based loss, results in a much larger deviation.

2.2.2 Robust Network-based Penalized Identification

As only a small subset of gene expressions is associated with cancer prognosis, and the total number of gene expressions is much larger than the sample size, identification of important prognostic markers is of a "large p, small n" nature, and can be achieved through regularized variable selection. Consider the regularized loss function:

$$Q(\alpha, \beta) = L(\alpha, \beta) + P(\beta; \lambda, \gamma), \tag{2.2}$$

where λ and γ are tuning parameters. A nonzero component of regularized estimate $\hat{\beta}$ indicates that the corresponding gene expression is associated with cancer prognosis. One possible choice for the penalty function is

$$P(\beta; \lambda, \gamma) = \sum_{m=1}^{p} \rho_{\lambda_1,\gamma}(|\beta_m|),$$

where $\rho_{\lambda_1,\gamma}(t) = \lambda_1 \int_0^{|t|} (1 - \frac{x}{\gamma\lambda_1})_+ dx$ is the Minimax Concave Penalty (MCP) with tuning parameter λ_1 and regularization parameter γ (Zhang (2010)).

The effects of gene expressions are represented by β, the vector of regression coefficients. We impose MCP on β, and components of non-zero regularized estimate suggests that the corresponding gene expressions are associated with cancer prognosis. A major disadvantage of this penalty is that correlations among gene expressions are not considered. Multiple studies, including aforementioned ones, have shown that failure to accommodate correlations results in biased estimation and false identification of important effects. To overcome this issue, we use a network structure to describe the interconnections among gene expressions. In the gene expression network, a node corresponds to a gene expression, and two nodes are connected by the edge if corresponding gene expressions are associated statistically or biologically. We propose the following penalty function to incorporate network information:

$$P(\beta;\lambda,\gamma)=\sum_{m=1}^{p}\rho_{\lambda_1,\gamma}(|\beta_m|)+\lambda_2\sum_{1\le m<k\le p}|a_{mk}||\beta_m-\mathrm{sgn}(a_{mk})\beta_k|, \tag{2.3}$$

where $\rho_{\lambda_1,\gamma}(\cdot)$ is the MCP defined above, β_m is the coefficient corresponding to the m-th gene expression and a_{mk} measures the strength of connection between the m-th and k-th gene expression. The first term of (2.3) imposes MCP on all the p components of β, thus it encourages sparsity in the regularized estimate. The second term promotes the smoothness among pairwise coefficient profiles of correlated gene expressions. It encourages their regression coefficients to be of similar magnitude. The extent of "encouragement" is adjusted by a_{mk}. The penalty shares certain similarity with the sparse Laplacian penalty (Huang et al. (2011)). However, it also has remarkable difference due to the L_1 form, which is adopted for the "consistency" purposes with the weighted LAD loss function.

In (2.3), $|a_{mk}|$ is the network adjacency which plays a critical role in quantifying the strength of connection between two nodes. We consider the approach from Zhang and Horvath (2005) to specify adjacency. Denote r_{mk} as the Pearson correlation coefficient between the mth and kth gene expression. Let $A=(a_{mk},1\le m,k\le p)$ be the adjacency matrix, where $a_{mk}=r_{mk}^{\alpha}\cdot I\{|r_{mk}|>r\}$ with $\alpha=5$. A properly defined network adjacency measure can keep the sign of r_{mk}, retain strong correlations and tune down weak ones (that are possibly noises). We choose the power transformation and the value of α following existing

studies (Huang et al. (2011); Zhang and Horvath (2005)). We calculate the cutoff r based on Fisher transformation $z_{mk} = 0.5\log((1 + r_{mk})/(1 - r_{mk}))$. If the correlation between mth and kth gene expressions is 0, $\sqrt{n-3}z_{mk}$ approximately follows $N(0, 1)$, which can then be adopted to calculate a threshold δ for $\sqrt{n-3}z_{mk}$. Then the threshold for r_{mk} is $r = \exp(2\delta/\sqrt{n-3}) - 1)/(\exp(2\delta/\sqrt{n-3}) + 1)$. The network is weighted and sparse. Please refer to Huang et al. (2011) and Zhang and Horvath (2005) for more details. There are alternative ways of constructing network adjacency. For instance, biological information (like pathway) is used to define adjacency in some studies. We conjecture that they are equally applicable. As our objective is not to compare different network constructions, we focus on this specific network structure.

2.2.3 Computation

Consider the following iterative algorithm:(a) initialize $\hat{\alpha}$ and $\hat{\beta}$; (b) update $\hat{\alpha}$ as the minimizer of (2.1) with β fixed at $\hat{\beta}$; (c) update $\hat{\beta}$ as the minimizer of (2.2) with α fixed at $\hat{\alpha}$; (d) iterate step (b) and (c) until convergence. The non-convexity of MCP in the penalty function (2.3) makes that computation of step (c) particularly challenging. Here, we develop an effective algorithm that borrows strength from MM (majorization-minimization) within the coordinate descent (CD) framework. More specifically, the nonconvex MCP in (2.3) is replaced by its majorization function to create a surrogate regularized loss function first, then optimization is conducted over the surrogate loss function with respect to one predictor at a time, and cycled through all predictors untill convergence.

We define a majorization function for the MCP function $\rho_{\lambda_1,\gamma}(|\beta|)$ at the d-th iteration $(d = 1, 2, ...)$ as

$$\phi_{\beta_m^{(d-1)}}(|\beta_m|) = \rho_{\lambda_1,\gamma}(|\beta_m^{(d-1)}|) + \rho'_{\lambda_1,\gamma}(|\beta_m^{(d-1)}|+)(|\beta_m| - |\beta_m^{(d-1)}|), \quad m = 1, \ldots, p$$

where $\beta_m^{(d-1)}$ is the value of β_m at the end of the (d-1)-th iteration, and $\rho'_{\lambda_1,\gamma}(|\beta_m|+)$ is the limit of $\rho'_{\lambda_1,\gamma}(t)$ as $t \to |\beta_m|$ from the above. $\rho'_{\lambda_1,\gamma}(|\beta_m|+)$ exists for all β_m due to the piecewise

differentiability of MCP. We can see that

$$\phi_{\beta_m^{(d-1)}}(|\beta_m|) \geq \rho_{\lambda_1,\gamma}(|\beta_m|) \quad \text{for all } m$$

where the equality holds when $\beta_m = \beta_m^{(d-1)}$. Hence, $\phi_{\beta_m^{(d-1)}}$, $m = 1, ..., p$ majorizes the MCP function $\rho_{\lambda_1,\gamma}(|\beta|)$. Subsequently, the regularized loss function in (2.2) is majorized at the d-th iteration by

$$Q_{\beta^{(d-1)}}(\alpha, \beta) = L(\alpha, \beta) + \sum_{m=1}^{p} \phi_{\beta_m^{(d-1)}}(|\beta_m|) + \lambda_2 \sum_{1 \leq m < k \leq p} |a_{mk}||\beta_m - \text{sgn}(a_{mk})\beta_k|$$

Next, we update the value of β at the d-th iteration by minimizing the surrogate regularized loss function:

$$\beta^{(d)} = \underset{\beta}{\text{argmin}}\, Q_{\beta^{(d-1)}}(\alpha, \beta) \tag{2.4}$$

This minimization can be conducted within the coordinate descent framework. With α and β_{-m} held fixed at the current estimate, we have

$$\begin{aligned}
\beta_m^{(d)} = \underset{\beta_m}{\text{argmin}} \Big\{ & \frac{1}{n} \sum_{i=1}^{n} v_i |Y_i - Z_i\alpha - \sum_{j \neq m} X_{ij}\beta_j - X_{im}\beta_m| \\
& + \sum_{j \neq m} \phi_{\beta_j^{(d-1)}}(|\beta_j|) + \phi_{\beta_m^{(d-1)}}(|\beta_m|) \\
& + \lambda_2 \sum_{j \neq m} \sum_{j < k \leq p} |a_{jk}||\beta_j - \text{sgn}(a_{jk})\beta_k| + \lambda_2 \sum_{m < k \leq p} |a_{mk}||\beta_m - \text{sgn}(a_{mk})\beta_k| \Big\} \\
= \underset{\beta_m}{\text{argmin}} \Big\{ & \frac{1}{n} \sum_{i=1}^{n} v_i |Y_i - Z_i\alpha - \sum_{j \neq m} X_{ij}\beta_j - X_{im}\beta_m| + \rho'_{\lambda_1,\gamma}(|\beta_m^{(d-1)}|+)|\beta_m| \\
& + \lambda_2 \sum_{m < k \leq p} |a_{mk}||\beta_m - \text{sgn}(a_{mk})\beta_k| \Big\}
\end{aligned}$$

Therefore, (2.4) can be equivalently expressed as a minimization problem for weighted median regression. We re-write (2.4) as

$$\beta_m^{(d)} = \underset{\beta_m}{\text{argmin}} \Big\{ \tfrac{1}{n+1+p-m} \textstyle\sum_{i=1}^{n+1+p-m} w_{im}|u_{im}| \Big\} \tag{2.5}$$

where

$$u_{im}^{(d)} = \begin{cases} \frac{Y_i - Z_i\alpha - \sum_{j\neq m}^{p} X_{ij}\beta_j^{(d-1)}}{X_{im}} - \beta_m & i=1,2,\ldots,n \\ 0 - \beta_m & i=n+1 \\ \text{sgn}(a_{mk})\beta_k - \beta_m & i=n+2,\ldots,n+1+p-m \end{cases} \tag{2.6}$$

and

$$w_{im}^{(d)} = \begin{cases} \frac{1}{n} v_i |X_{im}| & i=1,2,\ldots,n \\ \rho'_{\lambda_1,\gamma}(|\beta_m^{(d-1)}|+) & i=n+1 \\ \lambda_2 |a_{mk}| & i=n+2,\ldots,n+1+p-m \end{cases} \tag{2.7}$$

where m and k follow the same definition as in (2.3). The minimizer of (2.5) is the weighted median of $(n+1+p-m)$ pseudo observations. Similarly, we can update the $(q+1)$–dimensional vector $\alpha^{(d)}$ component-wisely by minimizing the loss function (2.1) using weighted median regression. Specifically, for each $l = 0, \ldots, q$, update $\alpha_l^{(d)}$ using the weighted median in (2.1) with β and α_{-l} held fixed. With fixed tuning parameters, the coordinate descent algorithm is described in Table 2.1

Table 2.1: *Coordinate descent algorithm.*

Algorithm Coordinate descent for the robust penalized network–based regularization

Initialize $d = 0$, $\alpha^{(0)}$ and $\beta^{(0)}$
Repeat
  update $\alpha^{(d+1)}$ component-wisely using weighted median regression
  for $m = 1, 2, \ldots, p$
    compute u_m and w_m via (2.6) and (2.7)
    update $\beta_m^{(d+1)}$ using the weighted median in (2.5)
    $m \leftarrow m + 1$
  end for
  $d \leftarrow d + 1$
until convergence

Selection of proper tuning parameters is crucial to the proposed method. Here, we have tuning parameters λ_1 and λ_2, as well as a regularization parameter γ. In MCP, γ balances between the concavity and unbiasedness. As suggested by Zhang (2010) and other studies, in our numerical study, we experiment γ with a sequence of values, including 1.5, 3, 5, 7 and

10, and find that the results not sensitive to the choice of γ. Therefore, we set γ=5. We choose the optimal pair of tuning parameters (λ_1, λ_2) via a two-dimensional grid search on independent testing data sets. That is, we first obtain regularized estimates from training data, then evaluate prediction performance over independently generated testing data. In simulation, the tunings determined from V-fold cross validation are very close to those based on independent testing data, but computationally more intensive. In real data analysis, we use cross validation to choose optimal tuning parameters since independent testing data sets are not available. In both simulation study and case study, convergence has been achieved in a small to moderate number of iterations. We compute the CPU time of running 100 replicates of simulated 300 $\times$ 500 gene expression data with AR structure and fixed tuning parameters on a regular laptop. The CPU time in seconds are 53.0 (LAD_Network), 36.1 (LAD_MCP), 34.9 (LAD_LASSO), 39.1 (Network), 24.3 (MCP) and 24.7 (LASSO), respectively.

To facilitate computation, we implement the proposed method, as well as the alternatives in C++ and provide the R package regnet with detailed documentation and examples (Ren et al. (2019b)).

2.3 Simulation

To demonstrate the utility of the proposed approach, we evaluate the performance through simulation study. In particular, we consider right censored survival data under the accelerated failure time (AFT) model. We generate datasets for different correlation structures and correlation levels, each with 300 subjects. For each subject, we simulate 5 clinical covariates and the expression of 500 genes, from multivariate normal distributions with marginal means equal to zero and variances equal to one. Among the 500 genes, there are 100 clusters with 5 genes per cluster. For the gene expression, we consider three correlation structures. (1) the auto-regression (AR) structure, in which gene i and j within the same cluster have correlation coefficients $\rho^{|i-j|}$, and they are independent cluster–wisely. We consider $\rho = 0.2, 0.5$ and 0.8, representing weak, moderate and strong correlation respectively. (2) banded correlation

structure where the ith and jth genes have $\rho = 0.5$ if $|i - j| = 1$ and $\rho = 0$ otherwise. Gene expressions in different clusters are independent. (3) banded correlation structure where the ith and jth genes have correlation coefficient 0.5 if $|i - j| = 1$, 0.25 if $|i - j| = 2$ and 0 otherwise. 10% of clusters are randomly selected to have nonzero regression coefficients generated from Unif[0.2, 0.8]. For the clinical covariates, we simply use a multivariate normal distribution with $\rho = 0.7$ in all scenarios. All clinical covariates have non-zero coefficients generated independently from Unif[0.2, 0.8]. The log event times are generated under the AFT model with random errors from N(0, 1) (Error1), T(1) (Error2), 85%N(0,1) + 15%Cauchy(0, 1) (Error3) or 75%N(0,1) + 25%Cauchy(0, 1) (Error4). The log censoring times are generated from uniform distribution. The average censoring rate is about 30%. We choose the tuning parameters based upon the prediction performance of the corresponding model in an independently simulated validation dataset.

For comparison, besides the developed robust network-constrained approach (LAD_Network), we also consider two robust approaches, robust MCP (LAD_MCP) and robust LASSO (LAD_LASSO), as well as three non-robust approaches, Network (Huang et al. (2011)), MCP and LASSO. All the robust methods adopt the weighted LAD loss function, while non-robust methods adopt the weighted least square loss. In particular, robust MCP is equivalent to the proposed approach when $\lambda_2 = 0$ in (2.3). Similarly, Network reduces to MCP when the tuning parameter corresponding to the Laplacian term is 0. Comparison between robust and non-robust methods has fully demonstrated the advantage of not only robustness in accommodating data contamination in survival response, but also the network based penalty from LAD_Network in accommodating interconnections among genetic measurements. Simulation results for the gene expression data under AR structure are tabulated in Table 2.2. We can observe that from the upper panel of Table 2.2 where $\rho = 0.5$, LAD_Network has better performance than LAD_MCP and LAD_LASSO for all four error types. For example, under Error2, LAD_Network identifies 31.63(sd 13.55) out of the 50 true positives, with a relatively small number of false positives 14.93(sd 9.85). LAD_MCP identifies a lower number of true positives 23.1(sd 9.64) with a higher number of false positives 56.17(sd 81.31). LAD_LASSO has a true positives 30.33(sd 6.57), but a much higher

Table 2.2: *Simulation for gene expression data $(n, p) = (300, 505)$. 50 genes have nonzero regression coefficients. 5 clinical covariates are not subject to selection. The gene expressions have AR structure with $\rho = 0.5$ (upper panel) and $\rho = 0.8$ (lower panel). mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		LAD_Network	LAD_MCP	LAD_LASSO	Network	MCP	LASSO
				AR $\rho = 0.5$			
Error1	TP	44.90(5.65)	40.67(4.59)	38.90(5.19)	40.07(4.70)	28.53(3.92)	48.27(1.20)
	FP	9.77(7.59)	8.63(8.16)	121.93(36.26)	8.13(5.37)	7.67(3.96)	75.57(10.64)
Error2	TP	31.63(13.55)	23.10(9.64)	30.33(6.57)	1.57(2.84)	1.50(2.76)	4.07(6.35)
	FP	14.93(9.85)	56.17(81.31)	103.17(49.89)	3.37(7.73)	3.27(7.94)	11.40(19.09)
Error3	TP	43.68(7.64)	36.28(5.79)	34.88(7.74)	20.03(12.84)	15.42(9.45)	31.83(16.71)
	FP	16.05(29.77)	12.64(20.27)	114.35(57.49)	9.81(6.57)	7.97(5.47)	60.73(33.39)
Error4	TP	39.03(10.15)	31.57(4.70)	34.10(6.26)	11.83(10.91)	9.73(8.51)	20.57(16.26)
	FP	14.33(11.20)	13.50(19.47)	109.87(40.55)	8.93(8.02)	7.67(7.77)	38.83(30.69)
				AR $\rho = 0.8$			
Error1	TP	46.93(5.77)	41.00(6.36)	43.70(4.94)	49.60(0.62)	23.93(2.97)	48.27(1.14)
	FP	5.27(6.35)	2.43(2.58)	94.20(38.45)	12.00(8.39)	7.70(5.38)	61.67(15.77)
Error2	TP	43.80(12.34)	23.93(5.46)	38.57(5.9)	10.97(15.16)	4.77(7.68)	9.47(10.68)
	FP	15.07(13.55)	14.20(22.23)	101.42(41.99)	18.50(33.77)	16.07(64.74)	21.82(25.00)
Error3	TP	47.23(7.11)	37.07(5.93)	43.90(4.37)	33.33(20.10)	15.47(10.68)	30.60(18.05)
	FP	4.53(5.06)	11.87(35.83)	91.37(24.94)	27.93(40.69)	19.07(66.48)	49.33(27.36)
Error4	TP	44.37(10.23)	32.30(5.03)	44.30(3.28)	32.57(19.21)	13.63(8.72)	28.27(14.97)
	FP	10.17(9.43)	6.03(8.22)	105.00(32.13)	26.90(20.39)	10.73(6.20)	47.60(26.05)

false positives 103.17(sd 49.89). Compared with non-robust methods, the proposed method has significant advantage when heterogeneity exists in the data (Error2, Error3 and Error4). When there is no heterogeneity (Error1), performance of the proposed method is comparable to that of the non-robust Network method and outperforms MCP and LASSO.

As correlation increases, the proposed one outperform other alternatives more significantly. As what we can observe from the lower panel of Table 2.2 where $\rho = 0.8$ under AR structure, LAD_Network achieves ideal true positives and satisfactory false positives. For example, LAD_Network has a TP 43.8(sd 12.34) and a FP 15.07(sd 13.55) for Error 2 and a

TP 47.23(sd 7.11) and a FP 4.53(sd 5.05) for Error 3, outperforming all other alternatives. To further examine the performance of the proposed approach, we also conduct simulation under banded structures. Results are summarized in Table A.1 in the Appendix A. The proposed LAD_Network delivers a consistent performance under different covariance structures: it outperforms robust alternatives when moderate to strong correlation exists among genetic variants, and it has significant advantage over non-robust methods when heterogeneity exists in the data.

In the second set of simulation, we consider more realistic correlation structures. Specifically, We generate gene expression datasets based on correlation structure extracted from real data in cancer studies. 500 genes are selected from Non-small cell lung cancer (NSCLC) data and Lung squamous cell carcinoma (LUSC) data, respectively. Two gene expression datasets, each with 300 subjects, are simulated with a multivariate normal distribution with marginal means zero and correlation matrix computed from genes selected from NSCLC data and LUSC data, respectively. 10% of genes are assigned to have nonzero regression coefficients generated from Unif[0.2, 0.8]. The 500 genes from real data are selected in a way that they form group-wise correlation structure. Unlike the first set of simulation where there are 5 genes per cluster, the clusters in this setting form more closely to real data based upon the calculated correlation coefficient. Results are shown in Table A.2 and A.3 in the Appendix A. In Table A.2, under Error 3, LAD_Network has the highest TP, 43.00(6.79), and the lowest FP, 3.14(3.91), among all the six approaches. The superior performance has also been observed under other heavy-tailed distributions. With standard normal error (Error 1), LAD_Network is comparable with the non-robust Network method, and outperforms the other two non-robust methods. Similar patterns have also been observed from Table A.3. The conclusions from the simulations based on real gene expression data are consistent with the ones we have from the first setting.

In the third set of simulation, we examine whether the proposed one demonstrates superior performance over the alternatives on simulated single-nucleotide polymorphism (SNP) data. We consider two schemes to simulate SNP data. With the first SNP generating scheme, the SNP data are simulated by dichotomizing expression values of each gene at

the 1st and 3rd quartiles, with the 3–level (2,1,0) for genotypes (AA,Aa,aa) respectively, where the gene expresison values are generated under the first set of simulation. Results are given in Table A.4 and A.5 under AR structure and banded structure respectively in the Appendix A. Under the second approach, the SNP genotype data are simulated based on a pairwise linkage disequilibrium (LD) structure. Let q_1 and q_2 be the minor allele frequencies (MAFs) of two alleles A and B for two adjacent SNPs. We denote LD as δ, and the frequencies of four haplotypes are calculated as $p_{AB} = q_1q_2 + \delta$, $p_{Ab} = q_1(1-q_2) - \delta$, $p_{aB} = (1-q_1)q_2 - \delta$, and $p_{ab} = (1-q_1)(1-q_2) + \delta$. Under Hardy-Weinberg equilibrium, SNP genotype (AA, Aa, aa) at locus 1 can be generated 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 (Cui et al. (2008)), we can simulate the genotypes for locus 2. With MAFs 0.3 and pairwise correlation $r = 0.6$, we have $\delta = r\sqrt{q_1(1-q_1)q_2(1-q_2)}$. The simulation results based on LD structure are given in Table A.6 in the Appendix A. Under both SNP generating schemes, the patterns are similar as those observed from the gene expression data.

2.4 Real Data Analysis

We analyze lung cancer data with gene expression measurements from two studies, separately. The first dataset is from the study of Xie et al. (2011), and the second one is the Lung squamous cell carcinoma (LUSC) data from TCGA (`https://cancergenome.nih.gov/`).

2.4.1 Non-small cell lung cancer (NSCLC) data

In the USA, lung cancer is the most common cause of cancer death. About 80% to 85% of lung cancers are non-small cell lung cancer (NSCLC). To identify genetic markers associated with the prognosis of NSCLC, gene profiling studies have been extensively conducted. As individual studies usually have small sample sizes, we follow the study of Xie et al. (2011) and collect data from four independent studies with gene expression measurements. After matching clinical variables and gene expression data, we have total 348 subjects and 22,283

gene expressions. Among the 348 subjects, 180 died during follow up, with survival times ranging from 0.03 to 204 months (median 26.19 months). To reduce the computational cost, we rank the probes by their variations and select the top 700 for downstream analysis. We include five clinical covariates, age, gender, smoking history, tumor stage and chemotherapy. Age is a normalized continuous variable, and dummy variables are created for categorical variables: smoking history, tumor stage and chemotherapy.

We apply all the methods to the lung cancer dataset. First, we conduct the logrank test to evaluate the prediction performance after splitting the patient group into training and testing sets. By dichotomizing the patients according to the median risk scores from the testing set, two risk groups can be created. Larger log rank test statistic indicates more significant survival difference between the low-risk and high-risk groups, thus better prediction performance. The log-rank statistics are 206.5 (LAD_Network), 130.7 (LAD_MCP), 132.7 (LAD_LASSO), 77.0 (Network), 11.1 (MCP) and 133.0 (LASSO), respectively. The proposed method has the best predictive performance, as indicated by the log-rank test statistic.

As a representative example, we examine the sub-network of gene PCLAF, PCNA Clamp Associated Factor. PCLAF is identified by five methods (all methods except MCP) as one of the most important genes. PCLAF encodes a PCNA-binding protein and is a regulator of DNA repair during DNA replication. It has been found to be overexpressed in various tumors, including lung tumor tissues (Hosokawa et al. (2007); Kato et al. (2012b); Yu et al. (2001)). Figure 2.2 shows the sub-network of PCLAF, where the red nodes indicate the probe of PCLAF. Thickness of the edges denotes the strength of correlation between genes. Comparing different methods, it can be clearly observed that the proposed approach has identified much more highly correlated prognostic genes, since the interconnections among genes have been accommodated by the approach that incorporates the network structure information. Eight genes are directly connected to PCLAF in the sub-network identified by the proposed approach. They are TOP2A, ASPM, SELENBP1, MAD2L1, CDC20, PRC1, TYMS and DLGAP5. All of them are positively correlated to PCLAF, except SELENBP1. PRC1 (Protein regulator of Cytokinesis 1) has the highest correlation with PCLAF (r=0.83). It is interesting that PCLAF and PRC1 are located closely on Chromosome 15. Similar as

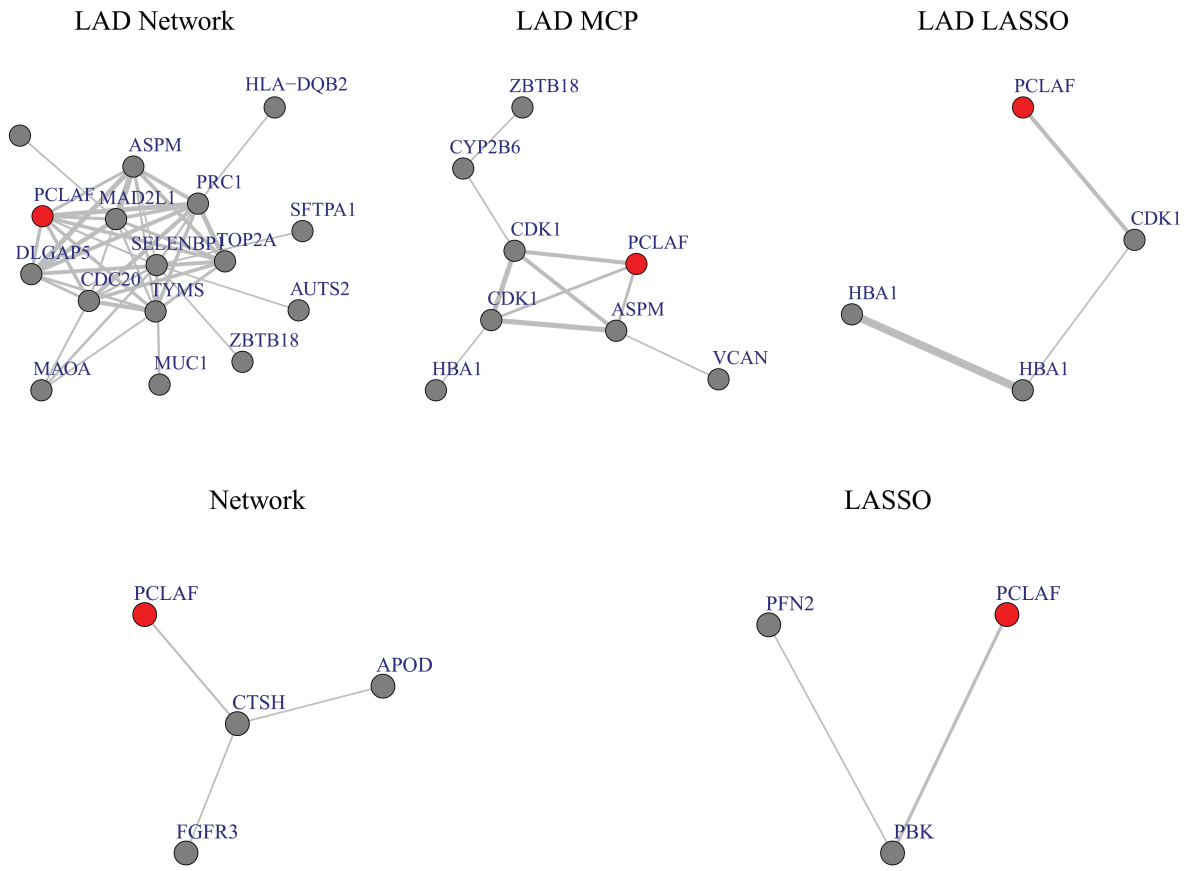

Figure 2.2: *Sub-network for PCLAF.*

PCLAF, PRC1 is overexpressed in lung cancer cells. Higher level of PRC1 is found to be associated with poor survival of lung cancer patients (Hanselmann et al. (2017); Zhan et al. (2017)).However, none of the alternative methods capture this important prognostic marker in the PCLAF network. In addition, TOP2A (DNA Topoisomerase II Alpha) (Hou et al. (2017); Huang et al. (2015)), CDC20 (Cell Division Cycle 20) (Kato et al. (2012a); Wang et al. (2013c)), DLGAP5 (DLG Associated Protein 5) (Schneider et al. (2017); Shi et al. (2017)) and MAD2L1 (MAD2 mitotic arrest deficient-like 1) (Shi et al. (2016)) have been identified as negative prognostic markers in NSCLC by recent studies. Studies report that the over-expression of TYMS (thymidylate synthase) (Chamizo et al. (2015); Wang et al. (2013a)) and ASPM (Kuo et al. (2015)) are related to drug-resistance in advanced NSCLC. Among the genes, SELENBP1 (selenium-binding protein 1) is negatively correlated

with PCLAF and other genes in the network. Selenium-binding proteins are known to play important roles in cancer prevention effects of selenium. Down-regulation of SELENBP1 is associated with poor prognosis in NSCLC patients (Tan et al. (2016); Zeng et al. (2013)). Overall, the proposed approach identifies more informative prognostic markers.

We also examine the biological similarity of the identified genes by the Gene Ontology (GO) analysis. An obvious difference between the proposed method and the alternatives is observed. The GO analysis results are provided in Figure A.1 in the Appendix.

2.4.2 Lung squamous cell carcinoma (LUSC) data

Lung squamous cell carcinoma (LUSC) is one of the most common types of NSCLC. It comprises 25–30% of all lung cancer cases (Zappa and Mousa (2016)). LUSC is more strongly correlated with cigarette smoking history than most other subtypes of NSCLC (Kenfield et al. (2008)). We analyze TCGA (The Cancer Genome Atlas) data on the prognosis of LUSC (The Cancer Genome Atlas Research Network (2012)). We consider four clinical covariates: age at diagnosis, gender, smoking history and tumor stage. The total number of genes is 20,499 and the sample size is 461. 203 died during follow-up among all the subjects. The survival times range from 0.03 to 173.69 months, with a median of 17.84 months. Similar as the NSCLC study, we select the top 700 genes for further analysis.

We applied the six methods to the working dataset. The log-rank statistics are 155.0 (LAD_Network), 116.9 (LAD_MCP), 102.8 (LAD_LASSO), 76.4 (Network), 40.6 (MCP) and 96.1 (LASSO), respectively. The proposed method has the largest log-rank statistic and thus superior prediction performance.

We use the sub-network of gene IRS4 (Insulin receptor substrate 4) as a representative example. IRS4 is identified by five methods (all expect Network) as a prognostic gene. IRS4 plays a tumor-promoting role in NSCLC (Hoxhaj et al. (2013); Weischenfeldt et al. (2017)). The proposed method identifies 13 genes in the sub-network of IRS4 (Figure 2.3). Ten genes are uniquely identified by the proposed method, and the rest three (PSMD10, CMTX5 and LOC158602) are also identified by other methods. Both PSMD10 (Proteasome 26S Subunit,

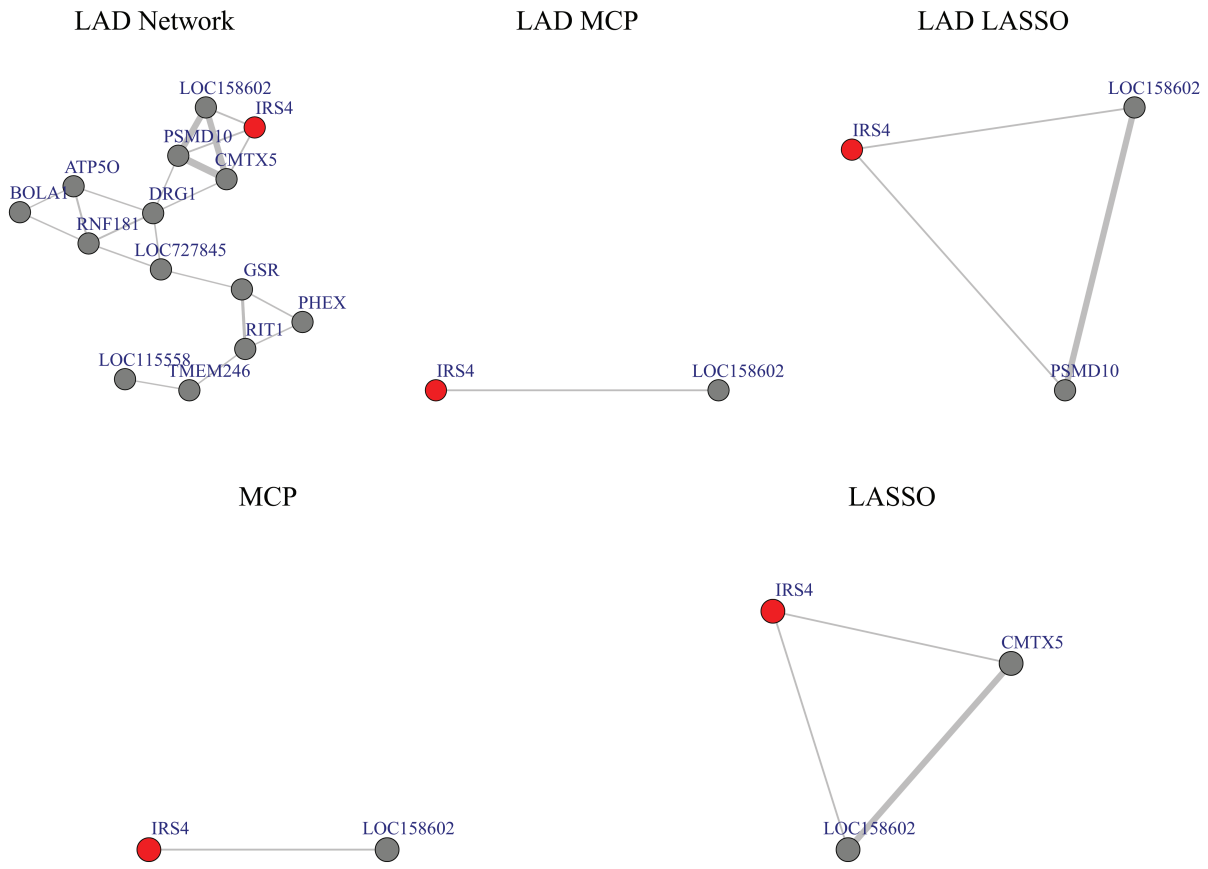

Figure 2.3: *Sub-network for IRS4.*

Non-ATPase 10) and CMTX5 (also known as PRPS1, phosphoribosyl pyrophosphate synthetase 1) are positively correlated with IRS4 and have been reported as oncogenes (He et al. (2017); Luo et al. (2016)). Among the 13 genes in IRS sub-network, three of them (PSMD10, CMTX5, PHEX) are located on chromosome X, the same as IRS4. LOC158602 is a gene with unknown function, but highly correlated with both PSMD10 and CMTX5. DRG1 (Developmentally regulated GTP binding protein 1) is only identified by the proposed method. DRG1 plays important roles in regulating cell growth. Overexpression of DRG1 leads to chromosome missegregation and promotes tumor progression in NSCLC (Lu et al. (2016)). GSR (glutathione reductase) is one of enzymes in the glutathione (GSH) metabolism system, which is a major redox regulatory systems in mammals that support increased tumor growth (Tobe et al. (2015)). It has been reported that GSH levels in cells, regulated by

GSH-synthesising enzymes such as GSR, is associated with resistance to epidermal growth factor receptor (EGFR) inhibitors in NSCLC (Li et al. (2016)). In this network, GSR has a strong correlation with RIT1 (Ras Like Without CAAX 1) (r=0.69). RIT1 encodes a RAS-family small GTPase. It has been reported as an oncogene. Mutations in RIT1 may also induce resistance to EGFR inhibition, but in a MEK-dependent manner (Berger et al. (2014)).

To further investigate the biological similarity of the identified genes, the GO analysis is conducted for the LUSC data. The results are provided in Figure A.1, which suggests moderate similarity.

2.5 Discussion

In cancer genomics studies, much effort has been devoted to developing variable selection methods to identify important genomics features associated with survival outcomes (Tibshirani (1997); Huang and Ma (2010); Sha et al. (2006)). In recent decades, it has been recognized that network (or graph) based regularization methods are particularly effective in accommodating the correlation among genomic variants in a number of studies, nevertheless, their development and application in cancer survival studies are quite limited. Besides, although the lack of robustness might lead to biased estimation and false identification of sparse network structures, robust network–based variable selection has not received much attention in cancer prognosis studies. Motivated by the limitations of existing studies and analysis of the cancer genetic data, we have proposed a robust network constrained regularization and variable selection method to accommodate correlations among gene expressions in the search of important prognostic markers. The proposed method outperforms alternatives, both robust and non-robust, under a diversity of simulation setups. In the analysis of cancer prognosis data with high-dimensional gene expression measurements, it leads to biologically sensible findings and improved prediction.

Our method significantly distinguishes from and complements existing ones in the following aspects. We adopt a weighted LAD objective function to accommodate data con-

tamination, with Kaplan-Meier weights for censoring. To incorporate the interconnections among gene expressions, we propose a network-constrained penalty of the "MCP+L_1" form, and develop an efficient algorithm within the coordinate descent framework. The MM step is critical for the formulation of the convex surrogate objective function, which naturally leads to a weighted median regression problem. The effectiveness of smoothing the non-convex penalty function has been demonstrated in Peng and Wang (2015) and studies alike. To achieve high computational efficiency, we adopt AFT model for cancer prognosis in this study. Cox proportional hazard model can also be used in cancer genomic studies (Huang and Liang (2018, 2019)), but has higher computational cost than AFT model in general.

Here we describe the correlation among genomic variants through network structures. We acknowledge that, first, different network structures can be constructed (Huang et al. (2011)) and, second, there exists a variety of ways to incorporate correlations in penalized estimation and identification, not necessarily through network based penalty functions. For example, the spatial correlation among CNVs can be taken into account by using the adaptive fused LASSO penalty (Gao and Huang (2010a)). Also, in addition to the data-driven approach, network structures can be constructed based on prior biological knowledge (Jiang et al. (2017); Li and Li (2008); Min et al. (2018); Sun and Wang (2012)). Comparisons to other network structures and structures other than networks are not the focus of this project, thus not pursued. We also acknowledge that Bayesian methods can be robust depending on the prior distribution assumptions. For example, Sha et al. (2006) consider AFT models with the t prior distribution. Note that the robustness of our proposed method is not only restricted to certain type of heavy–tailed distribution or data contamination, and Sha et al. (2006) will not lead to networks among genomic variants. Moreover, comparisons between frequentist and Bayesian methods is beyond of the scope of this project, and will be postponed to the future.

The proposed algorithm for LAD_Network under survival response is essentially a first order method. The first order method, such as gradient descent and proximal gradient descent, can enjoy a linear convergence rate when the objective function has strong convexity (Boyd and Vandenberghe (2004)). The LAD_Network loss function is, however, not differentiable

and not strongly convex, which poses challenge on establishing the rate of convergence. We conjecture that the rate of convergence of LAD_Network can be shown by following that of the subgradient method (Bertsekas (2010)). It is also worth noting that Wu and Lange (2008) has given a detailed discussion of LASSO in LAD regression, although the rate of convergence has not been provided. The iteration cost of our algorithm is not cheap, due to the MM step and the sorting step for solving weighted median regression. From a practical point of view, the fast convergence of our algorithm is guaranteed by the C++ core module of R package regnet. In addition, Gao and Huang (2010b) has investigated estimation and selection consistency of LAD_LASSO, which is important for developing consistency properties of LAD_Network case. In this project, we focused on the development of statistical methodology. Investigations on the theoretical properties will be conducted in future studies.

Regularized objective function of robust penalization methods share a common structure of "robust objective function + penalty function" (Wu and Ma (2015)). The computational advantage of the proposed method roots in the L_1 form of the objective function. It is conjectured that the robustness can be achieved by coupling the penalty function with other robust loss functions, such as the exponential squared loss (Wang et al. (2013b)) and rank based loss (Wu et al. (2015)). However, since additional tunings and smoothing are demanded for these loss functions, the computational expenses are even high under low dimensional settings.

In this study, we focus on prognostic outcomes. Extension of our method to continuous disease phenotypes can be made readily by changing Kaplan-Meier weights to equal weights. In addition, the proposed method is not limited to the analysis of gene expression measurement. The network structure has been widely adopted to describe correlations among other genomics features, such as SNPs (Ren et al. (2017)), CNVs (Peng et al. (2012)) and DNA methylations (Sun and Wang (2012)), where robust network based penalization is also of great interest.

Chapter 3

Semi-parametric Bayesian variable selection for gene-environment interactions

3.1 Introduction

It has been widely recognized that the genetic and environmental main effects alone are not sufficient to decipher an overall picture of the genetic basis of complex diseases. The Gene-Environment (G×E) interactions also play vital roles in dissecting and understanding complex diseases beyond the main effects (Hunter (2005); Hutter et al. (2013)). Significant amount of efforts have been made to conducting analysis for the investigation of the associations between disease phenotypes and interaction effects marginally, especially in GWAS (Mukherjee et al. (2011)). As the disease etiology and prognosis are generally attributable to the coordinated effects of multiple genetic and environment factors, as well as the G×E interactions, joint analysis has provided a powerful alternative to dissect G×E interactions.

From the statistical modeling perspective, the interactions can be described as the product of variables corresponding to genetic and environmental factors. With the main G and E effects, as well as their interactions, the contribution of genetic variants to disease phenotype

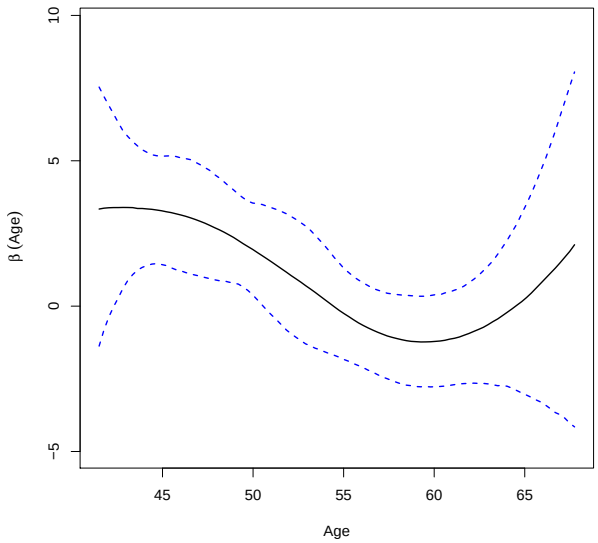

Figure 3.1: *Non-linear G×E effect of SNP rs1106380 from the Nurses' Health Study (NHS) data. The blue dashed lines represent the 95% credible region.*

can be expressed as a linear function of the environmental factor. Such a linear interaction assumption does not necessarily hold true in practice. Taking the Nurses' Health Study (NHS) data analyzed in this article as an example, we are interested in examining how the SNP effects on weight are mediated by age as the environmental factor. The range of subjects' age in the NHS data is from 41 to 68. As reported, for type 2 diabetes, the average age for the onset is 45 years (Centers for Disease Control and Prevention (2020)). Therefore, the presence of rs1106380×age interaction is roughly within such a range. We fit a Bayesian marginal model to SNP rs1106380 by using a non-parametric method to model the G×E interaction while accounting for effects from clinical covariates. A 95% credible region has also been provided. Figure 3.1 clearly suggests that the linear interaction assumption is violated. Mis–specifying the form of interactions will lead to biased identification of important effects and inferior prediction performance.

The non-linear G×E interactions have been first conducted in marginal analysis, including Ma et al. (2011) and Wu and Cui (2013a). Motivated by the set based association analysis, the modeling strategy has been adopted to investigate how genetic variants in a set, such as the gene set, pathways or networks, are mediated by one or multiple types of environmental

exposures to influence disease risk. The set–based modeling incorporating the nonlinear G×E interactions is essentially a joint analysis with high-dimensional covariates. Recently, penalized variable selection methods have emerged as a promising tool to capture G×E interactions that might be only weak or moderate individually, but that are strong collectively (Du et al. (2020); Wu et al. (2013, 2014, 2015, 2018c); Zhou et al. (2020b)).

Penalization methods have been first coined in Tibshirani (1996), which has also pointed out the connection between penalization and the corresponding Bayesian variable selection methods. In particular, the LASSO estimate can be interpreted as the posterior mode estimate when identical and independent Laplace prior has been imposed on each component of the coefficient vector under penalized least square loss. Park and Casella (2008) has further refined the prior as a conditional Laplace prior within the fully Bayesian framework to guarantee the unimodality of the posterior distribution. As LASSO belongs to the family of penalized estimate induced by the ℓ_q norm penalty with q=1, the Bayesian counterpart of penalization methods have been generalized to accommodate more complex data structure with other penalty functions, such as elastic net, fused LASSO and group LASSO. These extensions can also be formulated within the Bayesian framework with a similar rationale of specifying priors (Kyung et al. (2010)).

As penalization is tightly connected to Bayesian methods, the development of novel Bayesian variable selection will significantly broaden the scope of variable selection methods for G×E interaction studies, which will provide us fresh perspectives and promising results not offered by the existing studies. However, our limited literature review indicates that Bayesian variable selection has not been thoroughly conducted in existing G×E studies, especially for nonlinear interactions. For example, Liu et al. (2015) has developed a Bayesian mixture model to identify important G×E and G×G interaction effects through indicator model selection. Variable selection has been achieved by examining the posterior inclusion probability. However, this study cannot handle nonlinear interactions. More pertinent to the penalization, Li et al. (2015) has developed a Bayesian group LASSO for non-parametric varying coefficient models, where the non-linear interaction is expressed as a linear combinations of Legendre polynomials, and the identification of G×E interactions amounts to the

shrinkage selection of polynomials on the group level using multivariate Laplace priors. Li et al. (2015) has been built upon the Laplace prior adopted in Bayesian LASSO, therefore the coefficients cannot be shrunken to zero exactly in order to achieve the "real" sparsity,

Accounting for nonlinear effects in G×E studies has deeply rooted in structured variable selection for high dimensional data (Zhang et al. (2011)). An efficient selection procedure is expected to not only accurately pinpoint the form of nonlinear interactions, but also avoid modeling the main-effect-only case (corresponding to the non-zero constant effects) as nonparametric ones, since this type of misspecification may over–fit the data and result in loss of efficiency. To the best of our knowledge, automatic structure identification involving nonlinear effects has not been conducted in Bayesian G×E studies. To overcome the aforementioned limitations, we develop a novel semi-parametric Bayesian variable selection method for G×E interactions. We consider both linear and nonlinear interactions simultaneously. The interactions between a genetic factor and a discrete environmental factor are modeled parametrically, while the nonlinear interactions are modeled using varying coefficient functions. In particular, we conduct automatic structure identification via Bayesian regularization to separate the cases of G×E interactions, main-effect-only and no genetic effects at all, which more flexibly captures the main and interaction effects. Besides, to shrink the coefficients of unimportant linear and nonlinear effects to zero exactly, we adopt the spike-and-slab priors in our model. The spike-and-slab priors have recently been shown as effective when being incorporated in Bayesian hierarchical framework for penalization methods, including the spike–and–slab LASSO (Ročková and George (2018); Tang et al. (2017)), Bayesian fused LASSO (Zhang et al. (2014a)) and Bayesian sparse group LASSO (Xu and Ghosh (2015)). It leads to sparsity in the sense of exact 0 posterior estimates which are not available in Bayesian LASSO type of Bayesian shrinkage methods including Li et al. (2015).

Motivated by the pressing need to conduct efficient Bayesian G×E interaction studies accounting for the nonlinear interaction effects, the proposed semi-parametric model significantly advances from existing Bayesian variable selection methods for G×E interactions in the following aspects. First, compared to studies that solely focus on linear (Liu et al. (2015)) or non-linear effects (Li et al. (2015)), the proposed one can accommodate both

types of effects concurrently, thus more comprehensively describe the overall genetic architecture of complex diseases. Second, to the best of our knowledge, for G×E interactions, automatic structure discovery has been considered in the Bayesian framework for the first time. Compared to Li et al. (2015), one of the very few (or perhaps the only) literature in Bayesian variable selection for non-linear effects, our method is more fine tuned for the structured sparsity by distinguishing whether the genetic variants have nonlinear interaction, main effects only and no genetic effects at all, with the forms of coefficient functions being varying, non-zero constant and zero respectively. Third, borrowing strength from the spike–and–slab priors, we efficiently perform Bayesian shrinkage on the individual and group level simultaneously. In particular, with B–spline basis expansion, the identification of nonlinear interaction is equivalent to the selection of a group of basis functions. We develop an efficient MCMC algorithm for semi–parametric Bayesian hierarchical model. We show in both simulations and a case study that the exact sparsity significantly improves accuracy in identification of relevant main and interaction effects, as well as prediction. For fast computation and reproducible research, we implement the proposed and alternative methods in C++ and encapsulate them in a publicly available R package spinBayes (Ren et al. (2019c)).

The rest of the article is organized as follows. In Section 3.2, we formulate the semiparametric Bayesian variable selection model and derive a Gibbs sampler to compute the posterior estimates of the coefficients. We carry out the simulation studies to demonstrate the utility of our method in Section 3.3. A case study of Nurses' Health Study (NHS) data is conducted in Section 3.4.

3.2 Data and Model Settings

3.2.1 Partially linear varying coefficient model

We denote the ith subject using subscript i. Let $(X_i, Y_i, Z_i, E_i, W_i)$, $i = 1, \ldots, n$ be independent and identically distributed random vectors. Y_i is the response variable. X_i is the p-dimensional design vector of genetic factors, and Z_i and E_i are the continuous and dis-

crete environment factors, respectively. The clinical covariates are denoted by q-dimensional vector W_i. In the NHS data, the response variable is weight, and X_i represents SNPs. We consider age and the indicator of history of hypertension for Z_i and E_i, correspondingly. Height and total physical activity are used as clinical covariates, so q is 2. Now consider the following partially linear varying coefficient model

$$Y_i = \beta_0(Z_i) + \sum_{j=1}^{p} \beta_j(Z_i) X_{ij} + \sum_{t=1}^{q} \alpha_t W_{it} + \zeta_0 E_i + \sum_{j=1}^{p} \zeta_j E_i X_{ij} + \epsilon_i \tag{3.1}$$

where $\beta_j(\cdot)$ is a smoothing varying coefficient function, α_t is the coefficient of the tth clinical covariates, ζ_0 is the coefficient of the discrete E factor, and ζ_j is the coefficient of the interaction between the jth G factor $X_j = (X_{1j}, \ldots, X_{nj})^\top$ and E_i. The random error $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$.

Here only two environmental factors, Z_i and E_i, are considered for the simplicity of notation. Their interactions with the G factor are modeled as non–linear and linear forms, respectively. The model can be readily extended to accommodate multiple E factors.

3.2.2 Basis expansion for structure identification

As we discussed, distinguishing the case of main-effect-only from nonlinear G×E interaction is necessary since mis-specification of the effects cause over-fitting. The following basis expansion is necessary for the separation of different types of effects. We approximate the varying coefficient function $\beta_j(Z_i)$ via basis expansion. Let q_n be the number of basis functions

$$\beta_j(Z_i) \approx \sum_{k=1}^{q_n} \tilde{B}_{jk}(Z_i)\tilde{\gamma}_{jk} = \tilde{B}_j(Z_i)^\top \tilde{\gamma}_j$$

where $\tilde{B}_j(Z_i) = (\tilde{B}_{j1}(Z_i), \ldots, \tilde{B}_{jq_n}(Z_i))^\top$ is a set of normalized B spline basis, and $\tilde{\gamma}_j = (\tilde{\gamma}_{j1}, \ldots, \tilde{\gamma}_{jq_n})^\top$ is the coefficient vector. By changing of basis, the aforementioned basis

expansion is equivalent to

$$\beta_j(\cdot) \approx \sum_{k=1}^{q_n} \tilde{B}_{jk}(\cdot)\tilde{\gamma}_{jk} \doteq \gamma_{j1} + \tilde{B}_{j*}(\cdot)^\top \gamma_{j*}$$

where $\tilde{B}_{j*}(Z_i) = (\tilde{B}_{j2}(Z_i), \ldots, \tilde{B}_{jq_n}(Z_i))^\top$. γ_{j1} and $\gamma_{j*} = (\gamma_{j2}, \ldots, \gamma_{jq_n})^\top$ correspond to the constant and varying components of $\beta_j(\cdot)$, respectively. The intercept function can be approximated similarly as $\beta_0(\cdot) \approx \sum_{k=1}^{q_n} \tilde{B}_{0k}(\cdot)\tilde{\eta}_k \doteq \eta_1 + \tilde{B}_{0*}(\cdot)^\top \eta_*$. Define $\gamma_j = (\gamma_{j1}, (\gamma_{j*})^\top)^\top$, $\eta = (\eta_1, (\eta_*)^\top)^\top$, $B_j(Z_i) = (1, (\tilde{B}_{j*}(Z_i))^\top)^\top \doteq (B_{j1}(Z_i), \ldots, B_{jq_n}(Z_i))^\top$ and $B_0(Z_i) = (1, (\tilde{B}_{0*}(Z_i))^\top)^\top$. Collectively, model (3.1) can be rewritten as

$$\begin{aligned} Y_i &= B_0(Z_i)^\top \eta + \sum_{j=1}^{p} B_j(Z_i)^\top \gamma_j X_{ij} + \sum_{t=1}^{q} \alpha_t W_{it} + \zeta_0 E_i + \sum_{j=1}^{p} \zeta_j E_i X_{ij} + \epsilon_i \\ &= B_0(Z_i)^\top \eta + \sum_{j=1}^{p} (X_{ij}\gamma_{j1} + U_{ij}^\top \gamma_{j*}) + W_i^\top \alpha + E_i^\top \zeta_0 + T_i^\top \zeta + \epsilon_i \end{aligned}$$

where $U_{ij} = (B_{j2}(Z_i)X_{ij}, \ldots, B_{jq_n}(Z_i)X_{ij})^\top$, $\alpha = (\alpha_1, \ldots, \alpha_q)^\top$, $T_i = (X_{i1}E_i, \ldots, X_{ip}E_i)^\top$, and $\zeta = (\zeta_1, \ldots, \zeta_p)^\top$. Note that basis functions have been widely adopted for modeling the functional type of coefficient in general semi-parametric models, as well as functional regression analysis (Huang et al. (2002, 2004); Zhu et al. (2010)). For a comprehensive review of literature in this area, please refer to Morris (Morris et al. (2007)).

3.2.3 Semi-parametric Bayesian variable selection

The proposed semi-parametric model is of "large p, small n" nature. First, not all the main and interaction effects are associated with the phenotype. Second, we need to further determine for the genetic variants, whether they have nonlinear interactions, or main effect merely, or no genetic contribution to the phenotype at all. Therefore, variable selection is demanded.

From the Bayesian perspective, variable selection falls into the following four categories: (1) indicator model selection, (2) stochastic search variable selection, (3) adaptive shrinkage and (4) model space method (O'Hara and Sillanpää (2009)). Among them, adaptive shrink-

age methods solicit priors based on penalized loss function, which leads to sparsity in the Bayesian shrinkage estimates. For example, within the Bayesian framework, LASSO and group LASSO estimates can be understood as the posterior mode estimates when univariate and multivariate independent and identical Laplace priors are placed on the individual and group level of regression coefficients, respectively (Li et al. (2015); Park and Casella (2008)).

The proposed one belongs to the family of adaptive shrinkage Bayesian variable selection. For convenience of notation, we first define the approximated least square loss function as follows:

$$\tilde{L}(\eta, \gamma, \alpha, \zeta_0, \zeta) = \|Y - B_0\eta - \sum_{j=1}^{p} X_j\gamma_{j1} - \sum_{j=1}^{p} U_j\gamma_{j*} - W\alpha - E\zeta_0 - T\zeta\|^2$$

where $Y = (Y_1, \ldots, Y_n)^\top$, $B_0 = (B_0(Z_1), \ldots, B_0(Z_n))^\top$, $U_j = (U_{1j}, \ldots, U_{nj})^\top$, $W = (W_1, \ldots, W_n)^\top$ and $T = (T_1, \ldots, T_n)^\top$. Let $\theta = (\eta^\top, \gamma^\top, \alpha^\top, \zeta_0, \zeta^\top)^\top$ be the vector of all the parameters. Then the corresponding penalized loss function is

$$\tilde{L}(\eta, \gamma, \alpha, \zeta_0, \zeta) + \lambda_e \sum_{j=1}^{p} |\zeta_j| + \lambda_c \sum_{j=1}^{p} |\gamma_{j1}| + \lambda_v \sum_{j=1}^{p} \|\gamma_{j*}\|_2 \tag{3.2}$$

The formulation of (3.2) has been primarily driven by the need to accommodate linear and nonlinear G×E interaction while avoiding mis-specification of the main-effect-only as nonlinear interactions. Here γ_{j1} is the coefficient for the main effect of the jth genetic factor X_j, and the ℓ_2 norm of the spline coefficients $\|\gamma_{j*}\|_2$ is corresponding to the varying parts of $\beta_j(\cdot)$. If $\|\gamma_{j*}\|_2 = 0$, then there is no nonlinear interaction between X_j and continuous environment factor Z. Furthermore, if $\gamma_{j1} = 0$, then X_j has no main effect and is not associated with the phenotype. Similarly, the linear interaction between X_j and the discrete environment factor E is determined by ζ_j. ζ_j=0 indicates that there is no linear interaction. Overall, the penalty functions in (3.2) provide us the flexibility to achieve identification of structured sparsity through variable selection. Note that the main effects of environmental exposures Z and E are of low dimensionality, thus they are not subject to selection. Therefore, for the current G×E interaction study, we are particular interested in conducting Bayesian variable

selection on both the **individual level** of γ_{j1} and ζ_j $(j = 1, \ldots, p)$, and the **group level** of γ_{j*} $(j = 1, \ldots, p)$.

Laplacian shrinkage on individual level effects. Following the fully Bayesian analysis for LASSO proposed in Park and Casella (2008), we impose the individual-level shrinkage on genetic main effects and linear G×E interactions by adopting i.i.d. conditional Laplace prior on γ_{j1} and ζ_j $(j = 1, \ldots, p)$

$$\begin{aligned} \pi(\gamma_{11}, \ldots, \gamma_{p1}|\sigma^2) &= \prod_{j=1}^{p} \frac{\lambda_c}{2\sigma} \exp\Big\{ -\frac{\lambda_c}{\sigma} |\gamma_{j1}| \Big\} \\ \pi(\zeta_1, \ldots, \zeta_p|\sigma^2) &= \prod_{j=1}^{p} \frac{\lambda_e}{2\sigma} \exp\Big\{ -\frac{\lambda_e}{\sigma} |\zeta_j| \Big\} \end{aligned} \tag{3.3}$$

The above Laplace priors can be expressed as scale mixture of normals (Andrews and Mallows (1974))

$$\begin{aligned} \pi(\gamma_{j1}|\tau_{cj}^2, \sigma^2) &\overset{ind}{\sim} \mathrm{N}(0,\, \sigma^2\tau_{cj}^2) \\ \tau_{cj}^2 &\overset{ind}{\sim} \frac{\lambda_c^2}{2} \exp\Big\{ -\frac{\lambda_c^2}{2}\tau_{cj}^2 \Big\} \\ \pi(\zeta_j|\tau_{ej}^2, \sigma^2) &\overset{ind}{\sim} \mathrm{N}(0,\, \sigma^2\tau_{ej}^2) \\ \tau_{ej}^2 &\overset{ind}{\sim} \frac{\lambda_e^2}{2} \exp\Big\{ -\frac{\lambda_e^2}{2}\tau_{ej}^2 \Big\} \end{aligned} \tag{3.4}$$

It is easy to show that, after integrating out τ_{cj}^2 and τ_{ej}^2, (3.4) leads to the same priors in (3.3).

Laplacian shrinkage on group level effects. Kyung et al. (2010) extended the Bayesian LASSO to a more general form that can represent the group LASSO by adopting a multivariate Laplace prior. We follow the strategy and let the prior for γ_{j*} $(j = 1, \ldots, p)$ be

$$\pi(\gamma_{j*}|\sigma^2) \propto \exp\Big\{ -\frac{\sqrt{L}\lambda_v}{\sigma} \|\gamma_{j*}\|_2 \Big\} \tag{3.5}$$

where $L = q_n - 1$ is the size of the group, $(\frac{\sqrt{L}\lambda_v}{\sigma})^{-1}$ is the scale parameter of the multivariate Laplace and $\sqrt{L}$ terms adjusts the penalty for the group size. $\sqrt{L}$ can be dropped from the formula when all the groups have the same size. In this study, we use the same number of

basis functions for all parameters, and thus L is the same for all groups. For completeness, we still include $\sqrt{L}$ in (3.5) for possible extension to varying group sizes in the future. Similar to the (3.4), this prior can be expressed as a gamma mixture of normals

$$\begin{aligned} \pi(\gamma_{j*}|\tau_{vj}^2, \sigma^2) &\overset{ind}{\sim} \mathrm{N}_L(0,\, \sigma^2\tau_{vj}^2\mathbf{I}_L) \\ \tau_{vj}^2 &\overset{ind}{\sim} \mathrm{Gamma}\Big(\frac{L+1}{2},\, \frac{L\lambda_v^2}{2}\Big) \end{aligned} \tag{3.6}$$

where $\frac{L+1}{2}$ is the shape parameter and $\frac{L\lambda^2}{2}$ is the rate parameter of the Gamma distribution. After integrating out τ_{vj}^2 in (3.6), the conditional prior on γ_{j*} has the desired form in (3.5). Priors in (3.4) and (3.6) can lead to a similar performance as the general LASSO model in (3.2), by imposing individual shrinkage on γ_{j1} and ζ_j and group level shrinkage on γ_{j*}, respectively.

Spike-and-slab priors on both individual and group level effects. Compared with (3.2), priors in (3.4) and (3.6) cannot shrink the posterior estimates to exact 0. Li et al. (2015) has such a limitation since multivariate Laplace priors have been imposed on the group level effects. One of the significant advancements of our study over existing Bayesian G×E interaction studies, including Li et al. (2015), is the incorporation of spike–and–slab priors to achieve sparsity. For γ_{j*}, we have

$$\begin{aligned} \gamma_{j*}|\phi_{vj}, \tau_{vj}^2, \sigma^2 &\overset{ind}{\sim} \phi_{vj}\mathrm{N}_L(0,\, \mathrm{diag}(\sigma^2\tau_{vj}^2, \ldots, \sigma^2\tau_{vj}^2)) + (1-\phi_{vj})\delta_0(\gamma_{j*}) \\ \phi_{vj}|\pi_v &\overset{ind}{\sim} \mathrm{Bernoulli}(\pi_v) \\ \tau_{vj}^2|\lambda_v &\overset{ind}{\sim} \mathrm{Gamma}(\frac{L+1}{2},\, \frac{L\lambda_v^2}{2}) \end{aligned} \tag{3.7}$$

where $\delta_0(\gamma_{j*})$ denotes a point mass at $0_{L\times 1}$ and $\pi_v \in [0,1]$. We introduce a latent binary indicator variable ϕ_{vj} for each group $j, (j = 1, \ldots, p)$. ϕ_{vj} facilitates the variable selection by indicating whether or not the jth group is included in the final model. Specifically, when $\phi_{vj} = 0$, the coefficient vector γ_{j*} has a point mass density at zero which implies all predictors in the jth group are excluded from the final model. This is equivalent to concluding that the jth G factor X_j does not have an interaction effect with the environment factor Z. On

the other hand, when $\phi_{vj} = 1$, the prior in (3.7) reduces to the prior in (3.6) and induces the same behavior as Bayesian group LASSO. Thus, the coefficients in vector γ_{j*} have non-zero values and the jth group is included in the final model. Note that, after integrating out ϕ_{vj} and τ^2_{vj} in (3.7), the marginal prior on γ_{j*} is a mixture of a multivariate Laplace and a point mass at $0_{L\times 1}$ as follows

$$\pi(\gamma_{j*}|\sigma^2) \sim \pi_v \,\text{M-Laplace}(0, \frac{\sigma}{\sqrt{L}\lambda_v}) + (1-\pi_v)\delta_0(\gamma_{j*}) \tag{3.8}$$

When $\pi_v = 1$, (3.8) is equivalent to (3.5). Fixing $\pi_v = 0.5$ makes the prior essentially non-informative since it gives the equal prior probabilities to all sub-models. Instead of fixing π_v, we assign it a conjugate beta prior $\pi_v \sim \text{Beta}(r_v,\, w_v)$ with fixed parameters r_v and w_v. The value of λ_v controls the shape of the slab part of (3.8) and determines the amount of shrinkage on the γ_{j*}. For computation convenience, we assign a conjugate Gamma hyperprior $\lambda_v^2 \sim \text{Gamma}(a_v,\, b_v)$ which can automatically accounts for the uncertainty in choosing λ_v and ensure it is positive. We set a_v and b_v to small values so that the priors are essentially non-informative.

Remark: The form in (3.8) shows that our prior combines the strength of the Laplacian shrinkage and the spike–and–slab prior. The Laplacian shrinkage is used as the slab part of the prior, which captures the signal in the data and provides the estimation for large effects. Compared with (3.5), the additional spike part (point mass at zero) in (3.8) shrinks the negligibly small effects to zeros and achieve the variable selection.

Likewise, for γ_{j1} and ζ_j $(j = 1, \ldots, p)$ corresponding to the individual level effects, the spike-and-slab priors can be written as

$$\begin{aligned}
\gamma_{j1}|\phi_{cj}, \tau^2_{cj}, \sigma^2 &\overset{ind}{\sim} \phi_{cj}\text{N}(0,\, \sigma^2\tau^2_{cj}) + (1-\phi_{cj})\delta_0(\gamma_{j1}) \\
\phi_{cj}|\pi_c &\overset{ind}{\sim} \text{Bernoulli}(\pi_c) \\
\tau^2_{cj}|\lambda_c &\overset{ind}{\sim} \text{Gamma}(1,\, \frac{\lambda_c^2}{2})
\end{aligned} \tag{3.9}$$

and

$$\begin{aligned}
\zeta_j|\phi_{ej}, \tau_{ej}^2, \sigma^2 &\overset{ind}{\sim} \phi_{ej}\mathrm{N}(0,\, \sigma^2\tau_{ej}^2) + (1-\phi_{ej})\delta_0(\zeta_j) \\
\phi_{ej}|\pi_e &\overset{ind}{\sim} \text{Bernoulli}(\pi_e) \\
\tau_{ej}^2|\lambda_e &\overset{ind}{\sim} \text{Gamma}(1,\, \frac{\lambda_e^2}{2})
\end{aligned} \tag{3.10}$$

We assign conjugate beta prior $\pi_c \sim \text{Beta}(r_c,\, w_c)$ and $\pi_e \sim \text{Beta}(r_e,\, w_e)$, and Gamma priors $\lambda_c^2 \sim \text{Gamma}(a_c,\, b_c)$ and $\lambda_e^2 \sim \text{Gamma}(a_e,\, b_e)$. An inverted gamma prior for σ^2 can maintain conjugacy. The limiting improper prior $\pi(\sigma^2) = 1/\sigma^2$ is another popular choice. Parameters η, α and ζ_0 may be given independent flat priors.

3.2.4 Gibbs sampler

The binary indicator variables can cause an absorbing state in the MCMC algorithm which violates the convergence condition (Carlin and Chib (1995)). To avoid this problem, we integrate out the indicator variables ϕ_c, ϕ_v and ϕ_e in (3.7), (3.9) and (3.10). We will show that, even though ϕ_c, ϕ_v and ϕ_e are not part of the MCMC chain, their values still can be easily computed at every iterations. Let $\mu = E(Y)$, the joint posterior distribution of all the

unknown parameters conditional on data can be expressed as

$$
\begin{aligned}
&\pi(\eta, \alpha, \zeta_0, \gamma_{j1}, \tau_c^2, \pi_c, \lambda_c, \gamma_{j*}, \tau_v^2, \pi_v, \lambda_v, \zeta_j, \pi_e, \lambda_e, \tau_e^2, \sigma^2 | Y) \\
&\qquad \propto (\sigma^2)^{-\frac{n}{2}} \exp\Big\{ -\frac{1}{2\sigma^2}(Y-\mu)^\top (Y-\mu) \Big\} \\
&\qquad\quad \times \exp\Big(-\frac{1}{2}\eta^\top \Sigma_{\eta 0}^{-1} \eta \Big) \exp\Big(-\frac{1}{2}\alpha^\top \Sigma_{\alpha 0}^{-1} \alpha \Big) \exp\Big(-\frac{1}{2\sigma_{\zeta_0}^2}\zeta_0^2 \Big) \\
&\qquad\quad \times \prod_{j=1}^{p} \left(\pi_v (2\pi\sigma^2\tau_{vj}^2)^{-\frac{L}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{vj}^2}\gamma_{j*}^\top \gamma_{j*} \Big) \mathbf{I}_{\{\gamma_{j*} \neq 0\}} + (1-\pi_v)\delta_0(\gamma_{j*}) \right) \\
&\qquad\quad \times (\lambda_v^2)^{a_v - 1} \exp(-b_v \lambda_v^2) \prod_{j=1}^{p} \left(\frac{L\lambda_v^2}{2} \right)^{\frac{L+1}{2}} (\tau_{vj}^2)^{\frac{L+1}{2}-1} \exp\left(-\frac{L\lambda_v^2}{2}\tau_{vj}^2 \right) \\
&\qquad\quad \times \pi_v^{\gamma_v - 1}(1-\pi_v)^{w_v - 1} \\
&\qquad\quad \times \prod_{j=1}^{p} \left(\pi_c (2\pi\sigma^2\tau_{cj}^2)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{cj}^2}\gamma_{j1}^2 \Big) \mathbf{I}_{\{\gamma_{j1} \neq 0\}} + (1-\pi_c)\delta_0(\gamma_{j1}) \right) \\
&\qquad\quad \times (\lambda_c^2)^{a_c - 1} \exp(-b_c \lambda_c^2) \prod_{j=1}^{p} \frac{\lambda_c^2}{2} \exp\left(-\frac{\lambda_c^2}{2}\tau_{cj}^2 \right) \\
&\qquad\quad \times \pi_c^{\gamma_c - 1}(1-\pi_c)^{w_c - 1} \\
&\qquad\quad \times \prod_{j=1}^{p} \left(\pi_e (2\pi\sigma^2\tau_{ej}^2)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{ej}^2}\zeta_j^2 \Big) \mathbf{I}_{\{\zeta_j \neq 0\}} + (1-\pi_e)\delta_0(\zeta_j) \right) \\
&\qquad\quad \times (\lambda_e^2)^{a_e - 1} \exp(-b_e \lambda_e^2) \prod_{j=1}^{p} \frac{\lambda_e^2}{2} \exp\left(-\frac{\lambda_e^2}{2}\tau_{ej}^2 \right) \\
&\qquad\quad \times \pi_e^{\gamma_e - 1}(1-\pi_e)^{w_e - 1} \\
&\qquad\quad \times (\sigma^2)^{-s-1} \exp(-\frac{h}{\sigma^2})
\end{aligned}
$$

Let $\mu_{(-\eta)} = E(Y) - B_0\eta$, representing the mean effect without the contribution of $\beta_0(Z_i)$.

The posterior distribution of η conditional on all other parameters can be expressed as

$$\begin{aligned}
&\pi(\eta|\text{rest}) \\
&\quad \propto \pi(\eta)\pi(y|\cdot) \\
&\quad \propto \exp\Big(-\frac{1}{2}\eta^\top\Sigma_{\eta 0}^{-1}\eta\Big)\exp\Big(-\frac{1}{2\sigma^2}(Y-\mu)^\top(Y-\mu)\Big) \\
&\quad \propto \exp\Big(-\frac{1}{2}\eta^\top\Sigma_{\eta 0}^{-1}\eta-\frac{1}{2\sigma^2}(Y-B_0\eta-\mu_{(-\eta)})^\top(Y-B_0\eta-\mu_{(-\eta)})\Big) \\
&\quad \propto \exp\Big(\eta^\top(\Sigma_{\eta 0}^{-1}+\frac{1}{\sigma^2}B_0^\top B_0)\eta-\frac{2}{\sigma^2}(Y-\mu_{(-\eta)})^\top B_0\eta\Big)
\end{aligned}$$

Hence, the full conditional distribution of m is multivariate normal $N(\mu_\eta, \Sigma_\eta)$ with mean

$$\mu_\eta = \Big(\Sigma_{\eta 0}^{-1}+\frac{1}{\sigma^2}B_0^\top B_0\Big)^{-1}\Big(\frac{1}{\sigma^2}(Y-\mu_{(-\eta)})^\top B_0\Big)^\top$$

and variance

$$\Sigma_\eta = \Big(\Sigma_{\eta 0}^{-1}+\frac{1}{\sigma^2}B_0^\top B_0\Big)^{-1}$$

The full conditional distribution of α and ζ_0 can be obtained in similar way.

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

where $\mu_\alpha = \Sigma_\alpha(\frac{1}{\sigma^2}(Y-\mu_{(-\alpha)})^\top W)^\top$ and $\Sigma_\alpha = (\Sigma_{\alpha 0}^{-1}+\frac{1}{\sigma^2}W^\top W)^{-1}$

$$\zeta_0|\text{rest} \sim N(\mu_{\zeta_0}, \Sigma_{\zeta_0})$$

where $\mu_{\zeta_0} = \Sigma_{\zeta_0}(\frac{1}{\sigma^2}(Y-\mu_{(-\zeta_0)})^\top E)$ and $\Sigma_{\zeta_0} = (1/\sigma_{\zeta_0}^2+\frac{\sum_{i=1}^n E_i^2}{\sigma^2})^{-1}$.

Denote $\mu_{(-\gamma_{j*})} = E(Y) - U_j\gamma_{j*}$ and $l_{vj} = \pi(\gamma_{j*} \neq 0|\text{rest})$, the conditional posterior distribution of γ_{j*} is a multivariate spike-and-slab distribution:

$$\gamma_{j*}|\text{rest} \sim l_{vj}\mathrm{N}(\mu_{\gamma_{j*}},\, \sigma^2\Sigma_{\gamma_{j*}}) + (1-l_{vj})\delta_0(\gamma_{j*}) \tag{3.11}$$

where $\mu_{\gamma_{j*}} = \Sigma_{\gamma_{j*}} U_j^\top (Y - \mu_{(-\gamma_{j*})})$ and $\Sigma_{\gamma_{j*}} = (U_j^\top U_j + \frac{1}{\tau_{vj}^2}\mathbf{I}_L)^{-1}$. It is easy to compute that l_{vj} is equal to

$$l_{vj} = \frac{\pi_v}{\pi_v + (1-\pi_v)(\tau_{vj}^2)^{\frac{L}{2}} |\Sigma_{\gamma_{j*}}|^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2} \|\Sigma_{\gamma_{j*}}^{\frac{1}{2}} U_j^\top (Y - \mu_{(-\gamma_{j*})})\|_2^2 \Big)}$$

The posterior distribution (4.10) is a mixture of a multivariate normal and a point mass at 0. Specifically, at the gth iteration of MCMC, $\gamma_{j*}^{(g)}$ is drawn from $N(\mu_{\gamma_{j*}}, \Sigma_{\gamma_{j*}})$ with probability l_{vj} and is set to 0 with probability $1 - l_{vj}$. If $\gamma_{j*}^{(g)}$ is set to 0, we have $\phi_{vj}^{(g)} = 0$. Otherwise $\phi_{vj}^{(g)} = 1$.

Likewise, the conditional posterior distributions of γ_{j1} and ζ_j are also spike-and-slab distributions. Let $\mu_{\gamma_{j1}} = \Sigma_{\gamma_{j1}} X_j^\top (Y - \mu_{(-\gamma_{j1})})$ and $\Sigma_{\gamma_{j1}} = (X_j^\top X_j + \frac{1}{\tau_{cj}^2})^{-1}$, the full conditional distribution of γ_{j1} is

$$\gamma_{j1}|\text{rest} \sim l_{cj} \mathrm{N}(\mu_{\gamma_{j1}}, \sigma^2 \Sigma_{\gamma_{j1}}) + (1 - l_{cj})\delta_0(\gamma_{j1})$$

where

$$\begin{aligned} l_{cj} &= \pi(\gamma_{j1} \neq 0|\text{rest}) \\ &= \frac{\pi_c}{\pi_c + (1-\pi_c)(\tau_{cj}^2)^{\frac{1}{2}} (\Sigma_{\gamma_{j1}})^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2} \Sigma_{\gamma_{j1}} \|(Y - \mu_{(-\gamma_{j1})})^\top X_j\|_2^2 \Big)} \end{aligned}$$

Let $\mu_{\zeta_j} = \Sigma_{\zeta_j} (Y - \mu_{(-\zeta_j)})^\top T_j$ and $\Sigma_{\zeta_j} = (T_j^\top T_j + \frac{1}{\tau_{ej}^2})^{-1}$, the full conditional distribution of ζ_j is

$$\zeta_j|\text{rest} \sim l_{ej} \mathrm{N}(\mu_{\zeta_j}, \sigma^2 \Sigma_{\zeta_j}) + (1 - l_{ej})\delta_0(\zeta_j)$$

where

$$\begin{aligned} l_{ej} &= \pi(\zeta_j \neq 0|\text{rest}) \\ &= \frac{\pi_e}{\pi_e + (1-\pi_e)(\tau_{ej}^2)^{\frac{1}{2}} (\Sigma_{\zeta_j})^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2} \Sigma_{\zeta_j} \|(Y - \mu_{(-\zeta_j)})^\top T_j\|_2^2 \Big)} \end{aligned}$$

At the gth iteration, the values of $\phi_{cj}^{(g)}$ and $\phi_{ej}^{(g)}$ can be determined by whether the $\gamma_{j1}^{(g)}$ and $\zeta_j^{(g)}$ are set to 0 or not, respectively. We list the conditional posterior distributions of other

unknown parameters here. The details can be found in the Supplementary Materials.

$$(\tau_{vj}^2)^{-1}|\text{rest} \sim \begin{cases} \text{Inverse-Gamma}(\frac{L+1}{2},\ \frac{L\lambda_v^2}{2}) & \text{if } \gamma_{j*} = 0 \\ \text{Inverse-Gaussian}(L\lambda_v^2,\ \sqrt{\frac{L\lambda_v^2\sigma^2}{\|\gamma_{j*}\|_2^2}}) & \text{if } \gamma_{j*} \neq 0 \end{cases}$$

$$(\tau_{cj}^2)^{-1}|\text{rest} \sim \begin{cases} \text{Inverse-Gamma}(1,\ \frac{\lambda_c^2}{2}) & \text{if } \gamma_{j1} = 0 \\ \text{Inverse-Gaussian}(\lambda_c^2,\ \sqrt{\frac{\lambda_c^2\sigma^2}{\gamma_{j1}^2}}) & \text{if } \gamma_{j1} \neq 0 \end{cases}$$

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

λ_v^2, λ_c^2 and λ_e^2 all have Gamma posterior distributions

$$\begin{aligned} \lambda_v^2|\text{rest} &\sim \text{Gamma}(a_v + \frac{p(L+1)}{2},\ b_v + \frac{L\sum_{j=1}^p \tau_{vj}^2}{2}) \\ \lambda_c^2|\text{rest} &\sim \text{Gamma}(a_c + p,\ b_c + \frac{\sum_{j=1}^p \tau_{cj}^2}{2}) \\ \lambda_e^2|\text{rest} &\sim \text{Gamma}(a_e + p,\ b_e + \frac{\sum_{j=1}^p \tau_{ej}^2}{2}) \end{aligned}$$

π_v, π_c and π_e have beta posterior distributions

$$\begin{aligned} \pi_v|\text{rest} &\sim \text{Beta}(r_v + \sum_{j=1}^p \mathbf{I}_{\{\gamma_{j*}\neq 0\}},\ w_v + \sum_{j=1}^p \mathbf{I}_{\{\gamma_{j*}=0\}}) \\ \pi_c|\text{rest} &\sim \text{Beta}(r_c + \sum_{j=1}^p \mathbf{I}_{\{\gamma_{j1}\neq 0\}},\ w_c + \sum_{j=1}^p \mathbf{I}_{\{\gamma_{j1}=0\}}) \\ \pi_e|\text{rest} &\sim \text{Beta}(r_e + \sum_{j=1}^p \mathbf{I}_{\{\zeta_j\neq 0\}},\ w_e + \sum_{j=1}^p \mathbf{I}_{\{\zeta_j=0\}}) \end{aligned}$$

Last, the full conditional distribution for σ^2 the posterior distribution for σ^2 is Inverse-

Gamma(μ_{σ^2}, Σ_{σ^2}) where

$$\mu_{\sigma^2} = s + \frac{n + \sum \mathbf{I}_{\{\gamma_{j1} \neq 0\}} + L \sum \mathbf{I}_{\{\gamma_{j*} \neq 0\}} + \sum \mathbf{I}_{\{\zeta_j \neq 0\}}}{2}$$

$$\Sigma_{\sigma^2} = h + \frac{(Y-\mu)^\top (Y-\mu) + \sum_{j=1}^{p} \left((\tau_{cj}^2)^{-1} \gamma_{j1}^2 + (\tau_{vj}^2)^{-1} \gamma_{j*}^\top \gamma_{j*} + (\tau_{ej}^2)^{-1} \zeta_j^2 \right)}{2}$$

Under our priors setting, conditional posterior distributions of all unknown parameters have closed forms by conjugacy. Therefore, efficient Gibbs sampler can be used to simulate from the posterior distribution.

To facilitate fast computation and reproducible research, we have implemented the proposed and all the alternative methods in C++ from the R package spinBayes (Ren et al. (2019c)) on CRAN.

3.3 Simulation

We compare the performance of the proposed method, Bayesian spike and slab variable selection with structural identification, termed as BSSVC-SI, to four alternatives termed as BSSVC, BVC-SI, BVC and BL, respectively. BSSVC is the proposed method but without implementing structural identification. It does not distinguish the nonzero constant effect from the nonlinear effect. Specifically, in BSSVC, coefficients of q_n basis functions of β_j are treated as one group and are subject to selection at the group level. Comparison of BSSVC-SI with BSSVC demonstrate the importance of structural identification in the detection of interaction effects. BVC-SI is similar to the proposed method, except that it does not adopt the spike-and-slab prior. BVC does not use the spike-and-slab prior and does not distinguish the constant and varying effects. All these three alternative methods, BSSVC, BVC-SI and BVC, are different variations of the proposed BSSVC-SI, aiming to evaluate the strength of using the spike-and-slab prior and demonstrate the necessity of including structural identification. The last alternative BL is the well-known Bayesian LASSO (Park and Casella (2008)). BL assumes all interactions are linear. Details of the alternatives,

including the prior and posterior distributions, are available in the Supplementary Materials.

We consider four examples in our simulations. Under all four settings, the responses are generated from model (3.1) with $n = 500, p = 100$ and $q = 2$. Note that, the dimension of regression coefficients to be estimated after basis expansion is larger than the sample size ($n = 500$). For example, when the number of basis function $q_n = 5$, the effective dimension of regression coefficient is 604. In each example, we assess the performance in terms of identification, estimation, and prediction accuracy. We use the integrated mean squared error (IMSE) to evaluate estimation accuracy on the nonlinear effects. Let $\hat{\beta}_j(z)$ be the estimate of a nonparametric function $\beta_j(z)$, and $\{z_m\}_{m=1}^{n_{grid}}$ be the grid points where β_j is assessed. The IMSE of $\hat{\beta}_j(z)$ is defined as IMSE $(\hat{\beta}_j(z)) = \frac{1}{n_{grid}} \sum_{m=1}^{n_{grid}} \left\{\hat{\beta}_j(z_m) - \beta_j(z_m)\right\}^2$. Note that IMSE$(\hat{\beta}_j)$ reduces to MSE$(\hat{\beta}_j)$ when β_j is a constant. Identification accuracy is assessed by the number of true/false positives. Prediction performance is evaluated using the mean prediction errors on an independently generated testing dataset under the same settings.

Example 1

We first generate a $n \times p$ matrix of gene expressions, where $n = 500$ and $p = 100$, from a multivariate normal distribution with zero mean vector. We consider an auto-regression (AR) correlation structure for gene expression data, in which gene j and k have correlation coefficient $\rho^{|j-k|}$, with $\rho = 0.5$. For each observation, we simulate two clinical covariates from a multivariate normal distribution with $\rho = 0.5$. The continuous and discrete environment factors Z_i and E_i are simulated from a Unif[0, 1] distribution and a binomial distribution, respectively. The random error $\epsilon \sim N(0, 1)$.

The coefficients are set as $\mu(z) = 2\sin(2\pi z)$, $\beta_1(z) = 2\exp(2z - 1)$, $\beta_2(z) = -6z(1 - z)$, $\beta_3(z) = -4z^3$, $\beta_4(z) = 0.5$, $\beta_5(z) = 0.8$, $\beta_6(z) = -1.2$, $\beta_7(z) = 0.7$, $\beta_8(z) = -1.1$, $\alpha_1 = -0.5$, $\alpha_2 = 1$, $\zeta_0 = 1.5$, $\zeta_1 = 0.6$, $\zeta_2 = 1.5$, $\zeta_3 = -1.3$, $\zeta_4 = 1$, $\zeta_5 = -0.8$. We set all the rest of the coefficients to 0.

Example 2

We examine whether the proposed method demonstrates superior performance over the alternatives on simulated single-nucleotide polymorphism (SNP) data. The SNP genotype

data X_i are simulated by dichotomizing expression values of each gene at the 1st and 3rd quartiles, with the 3–level (2,1,0) for genotypes (AA,Aa,aa) respectively, where the gene expression values are generated from Example 1.

Example 3

In the third example, we consider a different scheme to simulate SNP data. The SNP genotype data are simulated based on 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 calculated as $p_{AB} = q_1q_2 + \delta$, $p_{ab} = (1-q_1)(1-q_2) + \delta$, $p_{Ab} = q_1(1-q_2) - \delta$, and $p_{aB} = (1-q_1)q_2 - \delta$, where δ denotes the LD. Under Hardy-Weinberg equilibrium, SNP genotype (AA, Aa, aa) at locus 1 can be generated 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 (Cui et al. (2008)), we can simulate the genotypes for locus 2. With MAFs 0.3 and pairwise correlation $r = 0.6$, we have $\delta = r\sqrt{q_1(1-q_1)q_2(1-q_2)}$.

Example 4

In the last example, we consider more realistic correlation structures. Specifically, we use the real data analyzed in the next section. To reduce the computational cost, we use the first 100 SNPs from the case study. For each simulation replicate, we randomly sample 500 subjects from the dataset. The same coefficients and error distribution are adopted.

Posterior samples are collected from a Gibbs Sampler running 10,000 iterations in which the first 5,000 are burn-ins. The Bayesian estimates are the posterior medians. To estimate the prediction errors, we compute the mean squared error in 100 simulations. For both BSSVC-SI and BSSVC, we consider the median probability model (MPM) (Barbieri and Berger (2004); Xu and Ghosh (2015)) to identify predictors that are significantly associated with the response variable. Suppose we collect G posterior samples from MCMC after burn-ins. The jth predictor is included in the regression model at the gth MCMC iterations if the indicator $\phi_j^{(g)} = 1$. Thus, the posterior probability of including the jth predictor in the final

model is defined as

$$p_j = \hat{\pi}(\phi_j = 1|y) = \frac{1}{G}\sum_{g=1}^{G} \phi_j^{(g)}, \quad j = 1, \ldots, p \tag{3.12}$$

A higher posterior inclusion probability p_j can be interpreted as a stronger empirical evidence that the jth predictor has a non-zero coefficient and therefore is associated with the response variable. The MPM model is defined as the model consisting of predictors that have posterior inclusion probability at least $\frac{1}{2}$. When the goal is to select a single model, Barbieri and Berger (2004) recommends using MPM due to its optimal prediction performance.

Table 3.1: *Simulation results.* (n, p, q) *= (500, 100, 2). mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		BSSVC-SI			BSSVC		
		Varying	Constant	Nonzero	Varying	Constant	Nonzero
Example 1	TP	3.00(0.00)	4.93(0.25)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.20(0.41)	0.00(0.00)	0.00(0.00)	5.00(0.26)	0.00(0.00)	0.10(0.31)
Example 2	TP	3.00(0.00)	5.00(0.00)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.20(0.41)	0.00(0.00)	0.03(0.18)	5.00(0.26)	0.00(0.00)	0.03(0.18)
Example 3	TP	3.00(0.00)	4.97(0.18)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.03(0.18)	0.07(0.37)	0.00(0.00)	5.03(0.18)	0.00(0.00)	0.10(0.31)
Example 4	TP	3.00(0.00)	4.97(0.18)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.17(0.38)	0.03(0.18)	0.00(0.00)	5.10(0.31)	0.00(0.00)	0.13(0.35)

Table 3.1 summarized the results on model selection accuracy. The identification performance for the varying and nonzero constant effects corresponding to the continuous environment factor, and nonzero effect (linear interaction) corresponding to the discrete environment factor are evaluated separately. We can observe that the proposed model has superior performance over BSSVC. BSSVC fails to identify any nonzero constant effect and has high false positive for identifying varying effect since it lacks structural identification to separate main-effect-only case from the varying effects. On the other hand, BSSVC-SI identifies most of the

true effects with very lower false positives. For example, considering the MPM in Example 1, BSSVC-SI identifies all 3 true varying effects in every iteration, with a small number of false positives 0.20(sd 0.41). It also identifies 4.93(sd 0.25) out of the 5 true constant effects without false positives. Besides, all the 5 true nonzero effects are identified without any false positives. We demonstrate the sensitivity of BSSVC-SI for variable selection to the choice of the hyper-parameters for π_v, π_c and π_e and the the choice of the hyper-parameters for λ_v, λ_c and λ_e in the Appendix. The results are tabulated in Table C.2 and Table B.2, respectively. Both tables show that the MPM model is insensitive to different specification of the hyper-parameters. An alternative way for selecting variables with posterior inclusion probability is to use a FDR-based method (Morris et al. (2007); Storey (2003)), which control the overall average Bayesian FDR rate by selecting variables with marginal inclusion probability larger than certain threshold. The results of FDR-based variable selection is summarized in (Table B.3). Overall the MPM and FDR models have very similar results in all four examples. The alternatives BVC-SI and BVC are not included here due to the lack of variable selection property. Li et al. (2015) adopts a method that is based on 95% credible interval (95%CI) for selecting important varying effects. In the Appendix, we show that, even adopting the 95%CI-based selection method, the identification performance of BVC-SI and BVC are unsatisfied, especially in terms of selecting a large number of false positives (Table B.4).

We also examine the estimation performance. We show the results from Example 1 (Table 3.2) here. The IMSE for all true varying effects, MSE for constant and nonzero effects, as well as the total squared errors for all coefficient estimates and prediction errors are provided in the Table. We observe that, across all the settings, the proposed method has the smallest prediction errors and total squared errors of coefficients estimates than all alternatives. For example, in Table 3.2, the BSSVC-SI has the smallest total squared errors 0.268(sd 0.080) and prediction error 1.159(0.066) among all the approaches. The key of the superior performance lies in (1) accurate modeling of different types of main and interaction effects, and (2) the spike and slab priors for achieving sparsity. Compared with BVC-SI which has (1) but does not spike and slab prior, BSSVC-SI performs better when estimating

both varying and constant coefficients. For example, the IMSE and MSE on $\beta_0(Z)$ and α_1 are 0.049 (sd 0.017) and 0.004 (sd 0.004), respectively. While BVC-SI yields 0.067(sd 0.030) and 0.008(0.010), correspondingly. Besides, compared with BSSVC which adopts the spike and slab priors without considering structured Bayesian variable selection, BSSVC-SI has comparable estimation performance on coefficients even though BSSVC overfits the data. In addition, similar patterns have been observed in Table B.6, Table B.7 and Table B.8 for Examples 2, 3 and 4 respectively, in the Appendix.

As a demonstrating example, Figure B.1 shows the estimated varying coefficients of the proposed model for the gene expression data in Example 1. Results from the proposed method fit the underlying trend of varying effects reasonably well. Following Li et al. Li et al. (2015), we assess the convergence of the MCMC chains by the potential scale reduction factor (PSRF) (Brooks and Gelman (1998); Gelman and Rubin (1992)). PSRF values close to 1 indicate that chains converge to the stationary distribution. Gelman et al. (2004) recommend using PSRF$\leq$ 1.1 as the cutoff for convergence, which has been adopted in our study. We compute the PSRF for each parameter and find all chains converge after the burn-ins. For the purpose of demonstration, Figure C.1 shows the pattern of PSRF after burn-ins for each parameter in Figure B.1. The figure clearly shows the convergence of the proposed Gibbs sampler.

We conduct sensitivity analysis on how the smoothness specification of the parameters in the B spline affects variable selection. The results summarized in Table B.5 in the Appendix shows that the proposed model is insensitive to the smoothness specification as long as the choices on number of spline basis are sensible. In simulation, we set the degree of B spline basis $O = 2$ and the number of interior knots $K = 2$, which makes $q_n = 5$.

Computation feasibility is an important practical consideration for high-dimensional Bayesian variable selection methods. We examine the computational cost of the proposed method for finishing 10,000 MCMC iterations under different combinations of sample sizes and SNP numbers. We focus on SNP numbers since the increase is computationally more challenging than that of the covariate numbers due to basis expansion. The results summarized in Table B.9 show that the proposed method is highly computationally efficient. For

Table 3.2: *Simulation results in Example 1. Gene expression data* (n, p, q) *= (500, 100, 2). mean(sd) of the integrated mean squared error (IMSE), mean squared error (MSE), total squared errors for all estimates and prediction errors based on 100 replicates.*

	BSSVC-SI	BSSVC	BVC-SI	BVC	BL
IMSE					
$\beta_0(Z)$	0.049(0.017)	0.050(0.017)	0.067(0.030)	0.066(0.028)	0.806(0.039)
$\beta_1(Z)$	0.052(0.028)	0.027(0.019)	0.090(0.051)	0.107(0.051)	0.139(0.060)
$\beta_2(Z)$	0.035(0.020)	0.026(0.014)	0.045(0.023)	0.050(0.021)	0.252(0.049)
$\beta_3(Z)$	0.033(0.025)	0.024(0.019)	0.081(0.057)	0.106(0.062)	0.256(0.062)
MSE					
α_1	0.004(0.004)	0.004(0.005)	0.008(0.010)	0.008(0.011)	0.012(0.015)
α_2	0.004(0.005)	0.004(269)	0.009(0.013)	0.009(0.013)	0.011(0.012)
ζ_0	0.033(0.025)	0.024(0.019)	0.081(0.057)	0.106(0.062)	0.032(0.045)
ζ_1	0.004(0.005)	0.003(0.004)	0.007(0.008)	0.006(0.007)	0.026(0.043)
ζ_2	0.011(0.014)	0.009(0.011)	0.017(0.016)	0.017(0.016)	0.055(0.067)
ζ_3	0.008(0.011)	0.008(0.010)	0.017(0.024)	0.017(0.022)	0.055(0.052)
ζ_4	0.014(0.017)	0.019(0.028)	0.020(0.025)	0.020(0.023)	0.042(0.052)
ζ_5	0.009(0.013)	0.010(0.016)	0.020(0.030)	0.024(0.032)	0.048(0.052)
Total	0.268(0.080)	0.304(0.132)	2.181(0.373)	2.119(0.363)	4.916(0.564)
Pred. Error	1.159(0.066)	1.167(0.067)	2.112(0.175)	2.075(0.170)	9.417(0.914)

example, when sample size $n = 1500$ and the number of gene $p = 300$, the CPU time for 10,000 iterations is approximately 121 seconds. Please note that the number of regression coefficients to be estimated after basis expansion is on the order of $q_n p + p$, where q_n is the number of basis functions. The term $q_n p$ gives the number of spline coefficients of nonlinear G×E interactions and p is the number of linear G×E interactions. In this example, the number of regression coefficients to be estimated is approximately 1800, higher than the sample size $n = 1500$. The efficient C++ implementation of the Gibbs sampler is an important guarantee for the computational scalability. The proposed method can be potentially applied to larger datasets with a reasonable computation time.

3.4 Real Data Analysis

We analyze the data from Nurses' Health Study (NHS). We use weight as the response and focus on SNPs on chromosome 10. We consider two environment factors. The first is age which is continuous and is known to be related to the variations in the obesity level. The second is the binary indicator of whether an individual has a history of hypertension (hbp), which is a sensible candidate for a discrete environment factor. In addition, we consider two clinical covariates: height and total physical activity. In NHS study, about half of the subjects are diagnosed of type 2 diabetes (T2D) and the other half are controls without the disease. We only use health subjects in this study. After cleaning the data through matching phenotypes and genotypes, removing SNPs with minor allele frequency (MAF) less than 0.05 or deviation from Hardy–Weinberg equilibrium, the working dataset contains 1716 subjects with 35099 SNPs.

For computational convenience prescreening can be conducted to reduce the feature space to a more attainable size for variable selection. For example, Li et al. (2015) use the single SNP analysis to filter SNPs in a GWA study before downstream analysis. Machine learning has also being used for screening important genetic variants in T2D studies (Jung et al. (2020)). In this study, we follow the procedure described in Ma et al. (2011) and Wu and Cui (2013a) to screen SNPs. Specifically, we use three likelihood ratio tests with weight as the response variable to evaluate the penetrance effect of a variant under the environmental exposure. The three likelihood ratio tests have been developed to test whether the interaction effects are nonlinear, linear, constant or zero, respectively. The SNPs with p-values less than a certain cutoff (0.005) from any of the tests are kept. 269 SNPs pass the screening.

We analyze the data by using the proposed method as well as BSSVC, the alternative without structural identification. As methods BVC-SI, BVC and BL show inferior performance in simulation, they are not considered in real data analysis. The proposed method identifies three SNPs with constant effects only, eleven SNPs with varying effects and sixteen SNPs with interactions with the hbp indicator. The BSSVC identifies twelve SNPs with varying effects and 10 SNPs with interactions with the hbp indicator. The Identifi-

cation results for varying and constant effects are summarized in Table 3.3. In this table, we can see that the three SNPs (rs11014290, rs2368945 and rs10787374) that are identified as constant effects only by BSSVC-SI are also selected by BSSVC. However, due to lack of structural identification, BSSVC identified them as SNPs with varying effects. The proposed method identifies rs1816002, a SNP located within gene ADAMTS14 as an important SNP with varying effect. ADAMTS14 is a member of ADAMTS metalloprotease family. Studies have shown that two members in the family, ADAMTS1 and ADAMTS13 are related to the development of obesity (Liu et al. (2012); Porter et al. (2005)), which suggests that ADAMTS14 may also have implications in obesity. The alternative method BSSCV fails to identify this important gene. The varying effect of the DIP2C gene SNP rs4880704 is identified by both BSSVC-SI and BSSVC. DIP2C (disco interacting protein 2 homolog C) has been found a potential epigenetic mark associated with obesity in children (Fradin et al. (2017)) and plays an important role in the association between obesity and hyperuricemia (Li et al. (2013)). The identification results for nonzero effects (representing the interactions with the binary indicator of a history of hypertension (hbp)) are summarized in Table 3.4. The interaction between rs593572 in gene KCNMA1 and hbp is identified by the proposed method. KCNMA1 (potassium calcium-activated channel subfamily M alpha 1) has been reported as an obesity gene that contributes to excessive accumulation of adipose tissue in obesity (Jiao et al. (2011)). Interestingly, the main effect of KCNMA1 is not identified, which suggests that KCNMA1 only has effect in the hypertension patients group. This result could be partially explained by the observation of significant association between the genetic variation in the KCNMA1 and hypertension (Tomás et al. (2008)).

The eleven varying coefficients of age that are identified by BSSVC-SI and the intercept are shown in Figure B.3 in the Appendix. All estimates have clear curvature and cannot be appropriately approximated by a model assuming linear effects. It is difficult to objectively evaluate the selection performance with real data. The prediction performance may provide partial information on the relative performance of different methods. Following Yan and Huang (2012) and Li et al. (2015), we refit the models selected by BSSVC-SI and BSSVC by Bayesian LASSO. The prediction mean squared errors (PMSE) based on the posterior

Table 3.3: *Identification results for varying and constant effects.*

		BSSVC-SI		BSSVC
SNP	Gene	V(Age)	C	V(Age)
rs11014290	PRTFDC1		-1.864	Varying
rs2368945	RPL21P93		1.494	Varying
rs4880704	DIP2C	Varying		Varying
rs1106380	CACNB2	Varying		Varying
rs2245456	MALRD1	Varying		
rs17775990	OGDHL	Varying		Varying
rs7922576	ZNF365	Varying		Varying
rs1816002	ADAMTS14	Varying		
rs2784761	RPL22P18	Varying		Varying
rs181652	AC005871.1	Varying		
rs10765108	DOCK1	Varying		
rs2764375	LINC00959	Varying		Varying
rs10787374	RPS6P15		2.020	Varying
rs11006525	MRPL50P4	Varying		
rs1698417	AC026884.1			Varying
rs7084791	PPP1R3C			Varying
rs12354542	BTF3P15			Varying

median estimates are computed. The PMSEs are 90.66 and 95.21 for BSSVC-SI and BSSVC, respectively. We also compute the prediction performance of BVC-SI, BVC and BL, based on the models selected by the 95% CI-based method. The PMSE is 106.26 for BVC-SI, 110.19 for BVC and 107.82 for BL. The proposed method outperforms all the competitors.

3.5 Discussion

The importance of G×E interactions in deciphering the genetic architecture of complex diseases have been increasingly recognized. A considerable amount of effort has been developed to dissect the G×E interactions. In marginal analysis, statistical testing of G×E interactions prevails, which spans from the classical linear model with interactions in a wide range of stud-

Table 3.4: *Identification results for nonzero effect corresponds to the discrete environment effect.*

		BSSVC-SI	BSSVC
rs10740217	CTNNA3	-1.06	-1.18
rs10787374	RPS6P15	-1.56	-1.42
rs10795690	AC044784.1		1.23
rs10829152	ANKRD26	1.29	1.73
rs10999234	PRKG1		1.97
rs11187761	PIPSL	1.04	
rs11245023	C10orf90	-0.92	
rs11250578	ADARB2	-1.62	
rs12267702	LYZL1	1.30	0.96
rs17767748	BTRC	1.18	1.15
rs2495763	PAX2	-1.33	-1.12
rs4565799	MCM10	-0.84	-0.98
rs593572	KCNMA1	1.70	
rs685578	AL353149.1		-1.13
rs7075347	AL357037.1	1.00	
rs7911264	HHEX	-1.30	
rs796945	RNLS	1.89	
rs9419280	LINC01168	1.57	
rs997064	PCDH15	1.31	

ies, such as case-control study, case only study and the two-stage screening study, to more sophisticated models, such as empirical Bayesian models, non- and semi-parametric models (Cornelis et al. (2011)). On the other hand, the joint methods, especially the penalized variable selection methods, for G×E interactions, have been motivated by the success of gene set based association analysis over marginal analysis, as demonstrated in Wu and Cui (2013b), Wu et al. (2012) and Schaid et al. (2012). Recently, multiple penalization methods have been proposed to identify important G×E interactions under parametric, semi-parametric and non-parametric models recently (Wu et al. (2014, 2015, 2018a,c)).

Within the Bayesian framework, non-linear interaction has not been sufficiently consid-

ered for G×E interactions. Furthermore, incorporation of the structured identification to determine whether the genetic variants have non-linear interaction, or main-effect-only, or no genetic influences at all is particularly challenging. In this study, we have proposed a novel semi-parametric Bayesian variable selection method to simultaneously pinpoint important G×E interactions in both linear and nonlinear forms while conducting automatic structure discovery. We approximate the nonlinear interaction effects using B splines, and develop a Bayesian hierarchical model to accommodate the selection of linear and nonlinear G×E interactions. For the nonlinear effects, we achieve the separation of varying, non-zero constant and zero coefficient functions through changing of spline basis, corresponding to cases of G×E interactions, main effects only (no G×E interactions) and no genetic effects. This automatic separation of different effects, together with the identification of linear interaction, lead to selection of important coefficients on both individual and group levels. Within our Bayesian hierarchical model, the group and individual level shrinkage are induced through assigning spike-and-slab priors with the slab parts coming from a multivariate Laplace distribution on the group of spline coefficients and univariate Laplace distribution on the individual coefficient, correspondingly. We have developed an efficient Gibbs sampler and implemented in R with core modules developed in C++, which guarantees fast computation in MCMC estimation. The superior performance of the proposed method over multiple alternatives has been demonstrated through extensive simulation studies and a case study.

The cumulative evidence has indicated the effectiveness of penalized variable selection methods to pinpoint important G×E interactions. Bayesian variable selection methods, however, have not been widely adopted in existing G×E studies. The proposed semi-parametric Bayesian variable selection method has the potential to be extended to accommodate a diversity forms of complex interaction structures under the varying index coefficient models and models alike, as summarized in Ma and Song (2015). Other possible extensions include Bayesian semi-parametric interaction analysis for integrating multiple genetic datasets (Li et al. (2019)). Investigations of all the aforementioned extensions are postponed to the future.

Chapter 4

Robust Bayesian variable selection for gene-environment interactions

4.1 Introduction

Deciphering the genetic architecture of complex diseases is a challenging task, as it demands the elucidation of the coordinated function of multiple genetic factors, their interactions, as well as gene-environment interactions. How the genetic contributions to influence the variations in the disease phenotypes are mediated by the environmental factors reveals a unique perspective of the disease etiology beyond the main genetic effects and their interactions (or epistasis) (Hunter (2005); Simonds et al. (2016)). Till now, G×E interaction analyses have been extensively conducted, especially within the framework of genetic association studies, to search for the important main and interaction effects that are associated with the disease trait (Mukherjee et al. (2011)).

With the availability of a large amount of genetic factors, such as SNPs or gene expressions, G×E interactions are of high dimensionality even though the preselected environmental factors are usually low dimensional. Therefore, the genetic association studies essentially aim to "find a needle from a haystack". Such a "large dimensionality, small sample size" problem can also be effectively addressed using penalization and Bayesian variable selection

methods (Fan and Lv (2010); Wu and Ma (2015)), and a surging amount of G×E studies have recently been conducted along this line (Zhou et al. (2020a)).

A prominent trend among these studies is to incorporate robustness in penalized identification of main and interaction in order to accommodate data contamination and heavy-tailed distributions in the disease phenotypes. Take the datasets analyzed in this article for example. The disease outcomes of interest are weight from the Nurses Health Study (NHS) and (log-transformed) Breslows depth from The Cancer Genome Atlas (TCGA) Skin Cutaneous Melanoma (SKCM) data. We plot the two in Figure 4.1, where the long tails can be clearly observed. In practice, such a heavy-tailed distribution is frequently encountered and arise due to multiple reasons. For instance, some phenotypes have skewness in nature. For the subjects recruited for the NHS, their ages are in the range from 41 to 68 as the average age for the onset of type 2 diabetes is 45 (Centers for Disease Control and Prevention (2020)). The subjects weight among this age group does have a right-skewed tendency. In addition, in the study of complex diseases such as cancer, even patients of similar profiles may have different subtypes as rigorous accrual of patients is usually not affordable. The data from the major disease subtype can be viewed as being contaminated by other subtypes or outliers. As nonrobust approaches cannot efficiently accommodate data contamination and long tailed distributions, which inevitably leads to biased estimates and false identifications, the robust penalization methods have thus been extensively developed for G×E studies (Wu and Ma (2015); Zhou et al. (2020a)).

Note that the interconnections between penalization and Bayesian methods has already been established in literature. For instance, it has been pointed out in Park and Casella (2008) that within the Bayesian framework, regularized estimate under LASSO is equivalent to the posterior mode of regression coefficients when the same conditional Laplace prior is placed on each regression coefficient independently. Such an interconnection can be further generalized to the Bayesian counterparts of group LASSO, fused LASSO and elastic net (Kyung et al. (2010)). Despite such a deep connection and remarkable successes of robust penalization-based analysis, robust Bayesian methods have not been investigated for gene-environment interactions by far. In fact, our literature search indicates that only limited

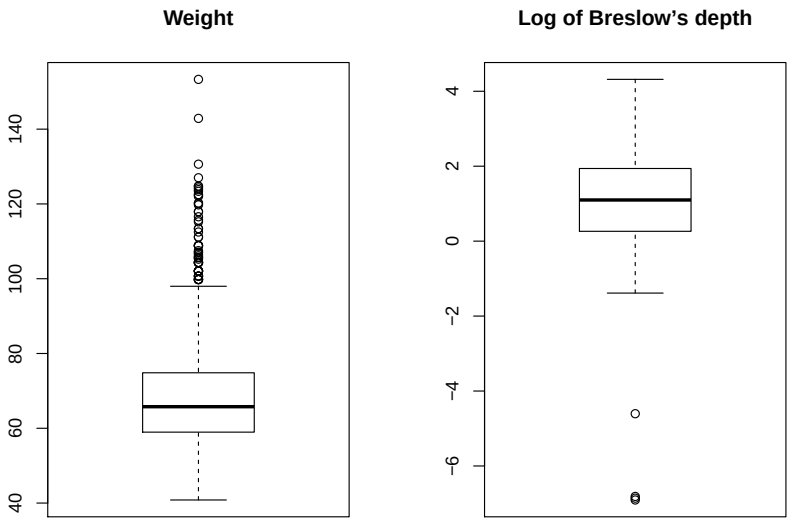

Figure 4.1: *Distribution of the outcome variables for the NHS (left) and SKCM (right) data.*

number of Bayesian variable selection methods have been developed for G×E studies, and none of them is robust (Zhou et al. (2020a)).

Driven by the urgent need to conduct robust Bayesian analysis for gene-environment interactions, we propose robust Bayesian variable selection methods tailored for interaction analysis. To illustrate, before we discuss the proposed joint study, let us first consider a marginal conceptual model: Outcome $\sim$ Es + G + G×(Es), where G and Es denote one G factor and several E factors, respectively. The term G×(Es) represents the interaction between the G factor and all the environmental variables. With a slight abuse of notation, the last two terms in the conceptual model can be rewritten as G×(1, Es), which is a group of main genetic and G×E interaction effects with respect to the G factor. Therefore, to determine whether the genetic factor is associated with the phenotype, and if so, to further examine which effects from the group are associated with the phenotype, essentially amount to a sparse group (or bi-level) variable selection problem. The sparse group selection remains and becomes even more challenging for our G×E study when a large number of G factors are jointly analyzed. Therefore, within the Bayesian framework, the proposed method should

aim to incorporate robustness and bi-level selection simultaneously.

We adopt a Bayesian formulation of the least absolute deviation (LAD) regression to accommodate data contamination and long-tailed distributions in the phenotype. Such a formulation is a special case of the Bayesian quantile regression (Yu and Moyeed (2001)). The LAD loss has been a very popular choice for developing robust penalization methods due to computational simplicity when being coupled with complex penalty functions, including the network-constrained penalty (Ren et al. (2019a); Wu et al. (2018b)) and sparse group penalty (Wu et al. (2018a)). Interestingly, its computational convenience has also been revealed within the Bayesian framework as efficient Gibbs sampler can be constructed when the loss is combined with LASSO, group LASSO and elastic net penalties (Li et al. (2010)). Furthermore, following the strategy of eliciting the prior for bilevel selection from a nonrobust Bayesian method (Xu and Ghosh (2015)), we have developed the Bayesian LAD sparse group LASSO for robust G×E interaction studies. The spike-and-slab priors have been imposed on both the individual and group level to ensure the shrinkage of posterior estimates corresponding to unimportant main and interaction effects to zero exactly. Such a prior lead to the real sparsity and is superior over Laplacian shrinkage in terms of identification and prediction results (George and McCulloch (1993); Ročková and George (2018); Tang et al. (2017)).

In this study, our objective is to conduct robust Bayesian variable selection for G×E interactions, which has been well motivated from the success of penalization methods (especially those robust ones) in G×E studies and a lack of robust interaction analysis within the Bayesian framework. The significance of the proposed study lies in the following aspects. First, it advances from existing Bayesian G×E studies by incorporating robustness to accommodate data contamination and heavy-tailed distributions in the disease phenotype. Second, on a broader scope, although robust Bayesian quantile regression based variable selection has been proposed under LASSO, group LASSO and elastic net, the more complicated sparse group (or bi-level) structure, which is of particular importance in high dimensional data analysis in general (Breheny and Huang (2009)), has not been examined by far. We are among the first to develop robust Bayesian sparse group LASSO for bi-level variable selec-

tion. Third, unlike existing Bayesian regularized quantile regression methods which build upon the priors under Laplacian type of shrinkage, we conduct efficient Bayesian regularization on both the individual and group levels by borrowing strength from the spike-and-slab priors, thus leading to better identification and prediction performance over the competing alternatives, as demonstrated in extensive simulation studies and case studies of NHS data with SNP measurements and TCGA melanoma data with gene expression measurements. To facilitate reproducible research and fast computation using our MCMC algorithm, we implement the proposed and alternative approaches in C++, which are available from an open source R package robin that will be available at CRAN soon.

4.2 Data and Model Settings

Use subscript i to denote the ith subject. Let (X_i, Y_i, E_i, W_i), $(i = 1, \ldots, n)$ be independent and identically distributed random vectors. Y_i is a continuous response variable representing the disease phenotype. X_i is the p–dimensional vector of G factors. The environmental factors and clinical covariates are denoted as the k-dimensional vector E_i and the q-dimensional vector W_i, respectively. Considering the following model:

$$\begin{aligned} Y_i &= \sum_{t=1}^{q} \alpha_t W_{it} + \sum_{m=1}^{k} \theta_m E_{im} + \sum_{j=1}^{p} \gamma_j X_{ij} + \sum_{j=1}^{p}\sum_{m=1}^{k} \zeta_{jm} E_{im} X_{ij} + \epsilon_i \\ &= \sum_{t=1}^{q} \alpha_t W_{it} + \sum_{m=1}^{k} \theta_m E_{im} + \sum_{j=1}^{p} \left(\gamma_j X_{ij} + \sum_{m=1}^{k} \zeta_{jm} E_{im} X_{ij}\right) + \epsilon_i \\ &= \sum_{t=1}^{q} \alpha_t W_{it} + \sum_{m=1}^{k} \theta_m E_{im} + \sum_{j=1}^{p} \left(U_{ij}^{\top} \beta_j\right) + \epsilon_i, \end{aligned} \tag{4.1}$$

where α_t's, θ_m's, γ_j's and ζ_{jm}'s are the regression coefficients for the clinical covariates, environmental factors, genetic factors and G×E interactions, respectively. We define $\beta_j = (\gamma_j, \zeta_{j1}, \ldots, \zeta_{jk})^{\top} \equiv (\beta_{j1}, \ldots, \beta_{jL})^{\top}$ and $U_{ij} = (X_{ij}, X_{ij}E_{i1} \ldots, X_{ij}E_{ik})^{\top} \equiv (U_{ij1}, \ldots, U_{ijL})^{\top}$, where $L = k + 1$. The coefficient vector β_j represents all the main and interaction effects corresponding to the jth genetic measurement. The ϵ_i's are random errors. Without loss

of generality, we assume that the data have been properly normalized so that the intercept can be omitted. Denote $U_i = (U_{i1}^\top, \ldots, U_{ip}^\top)^\top$, $\alpha = (\alpha_1, \ldots, \alpha_q)^\top$, $\theta = (\theta_1, \ldots, \theta_k)^\top$ and $\beta = (\beta_1^\top, \ldots, \beta_p^\top)^\top$. The vector β is of length $p \times L$. Then model (4.1) can be written in a more concise form as

$$Y_i = W_i^\top \alpha + E_i^\top \theta + U_i^\top \beta + \epsilon_i \tag{4.2}$$

4.2.1 Bayesian LAD Regression

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

$$f(\epsilon_i|\nu) = \frac{\nu}{2} \exp\{-\nu|\epsilon_i|\} \quad i = 1, \ldots, n, \tag{4.3}$$

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

$$f(Y|W, E, U, \alpha, \theta, \beta, \nu) = \prod_{i=1}^{n} \frac{\nu}{2} \exp\{-\nu\,|Y_i - \mu_i|\}, \tag{4.4}$$

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

Based on Kozumi and Kobayashi (2011), the Laplace distribution is equivalent to the mixture of an exponential and a scaled normal distribution. Specifically, let z and $\tilde{u}$ be the standard normal and exponential random variables, respectively. If a random variable ϵ follows the Laplace distribution with parameter ν, then it can be represented as follows

$$\epsilon = \nu^{-1}\kappa\sqrt{\tilde{u}}z, \tag{4.5}$$

where $\kappa = \sqrt{8}$ is a constant. Therefore, the response Y_i can be rewritten as $Y_i = \mu_i + \nu^{-1}\kappa\sqrt{\tilde{u}_i}z_i$, where $z_i \sim N(0,1)$ and $\tilde{u}_i \sim \text{Exp}(1)$. Let $u = \nu^{-1}\tilde{u}$. Then u follows the

exponential distribution $\text{Exp}(\nu^{-1})$. We thus have the following hierarchical representation of the Laplace likelihood:

$$Y_i = \mu_i + \nu^{-\frac{1}{2}}\kappa\sqrt{u_i}z_i,$$

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

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

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

Remark: The Laplace distribution in Bayesian LAD regression can be treated as a special case of the asymmetric Laplace distribution (ALD) in Bayesian quantile regression (Yu and Moyeed (2001); Yu and Zhang (2005)). In Bayesian quantile regression, we assume that ϵ_i follows the asymmetric Laplace distribution with density

$$f(\epsilon_i|\tau,\nu) = \tau(1-\tau)\nu \exp\left\{-\nu\rho_\tau(\epsilon_i)\right\} \quad i = 1,\ldots,n, \tag{4.6}$$

where the check loss function is $\rho_\tau(\epsilon_i) = \epsilon_i\left\{\tau - I(\epsilon_i < 0)\right\}$ for the τth quantile $(0 < \tau < 1)$. Note that, when $\tau = 0.5$, the ALD in (4.6) reduces to a symmetric Laplace distribution defined in (4.3). Yu and Moyeed (Yu and Moyeed (2001)) have shown that maximizing a likelihood function under the asymmetric Laplace error distribution (4.6) is equivalent to minimizing the check loss function in quantile regression. Kozumi and Kobayashi (2011) have proposed a Gibbs sampling algorithm for Bayesian quantile regression based on a location-scale mixture representation of the ALD. Specifically, with $\tilde{u}$ and z defined as above, the asymmetric Laplace error in (4.6) can be represented as

$$\epsilon = \nu^{-1}\psi z + \nu^{-1}\kappa\sqrt{\tilde{u}}z, \tag{4.7}$$

where

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

When $\tau = 0.5$, we have $\psi = 0$ and $\kappa = \sqrt{8}$, and equation (4.7) reduces to the Laplace error in (4.5).

4.2.2 Bayesian sparse group variable selection for G×E interactions

The proposed Bayesian sparse group variable selection method is motivated by the following considerations. In model (4.1), the coefficient vector β_j corresponds to the main and interaction effects with respect to the jth genetic variant. Whether the genetic variant is associated with the phenotype or not can be determined by whether $\beta_j = 0$. A zero coefficient vector suggests that the variant does not have any effect on the disease outcome. If $\beta_j \neq 0$, then a further investigation on the presence of the main effect, or the interaction or both is of interest, which can be facilitated by examining the nonzero component in β_j. Therefore, a tailored robust Bayesian variable selection method for G×E studies should accommodate the selection on both group (the entire vector of β_j) and individual (each component of β_j) levels at the same time.

In order to impose sparsity on both group and individual level to identify important main and interaction effects, we conduct the decomposition of β_j by following the reparmetrization from Xu and Ghosh (2015). Specifically, β_j is defined as

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

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

$$\begin{aligned} b_j|\phi_j^b &\overset{ind}{\sim} \phi_j^b \, \text{N}_L\left(0, \mathbf{I}_L\right) + (1 - \phi_j^b)\delta_0(b_j), \\ \phi_j^b|\pi_0 &\overset{ind}{\sim} \text{Bernoulli}(\pi_0), \end{aligned} \tag{4.8}$$

where $\mathbf{I}_L$ is an identity matrix, $\delta_0(b_j)$ denotes a point mass at $0_{L\times 1}$ and $\pi_0 \in [0, 1]$. We

introduce a latent binary indicator variable ϕ_j^b for each group $j(j = 1, \ldots, p)$ to tackle the group–level selection. In particular, when $\phi_j^b = 0$, the coefficient vector b_j has a point mass density at zero and all predictors representing the main and interaction effects in the jth group are excluded from the model, indicating that the jth genetic variant is not associated with the phenotype. On the other hand, when $\phi_j^b = 1$, the components in coefficient vector b_j have non-zero values.

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

$$\begin{aligned} \omega_{jl}|\phi_{jl}^w &\overset{ind}{\sim} \phi_{jl}^w \, \mathrm{N}^+\left(0,\, s^2\right) + (1 - \phi_{jl}^w)\delta_0(\omega_{jl}), \\ \phi_{jl}^w|\pi_1 &\overset{ind}{\sim} \text{Bernoulli}(\pi_1), \end{aligned} \tag{4.9}$$

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

In (4.8) and (4.9), π_0 and π_1 control the sparsity on the group and individual level, respectively. Their values should be carefully tuned. Fixing their values at 0.5 makes the prior essentially non-informative since equal prior probabilities are given to all the sub-models. Instead of fixing π_0 and π_1, we assign conjugate beta priors $\pi_0 \sim \text{Beta}(a_0,\, b_0)$ and $\pi_1 \sim \text{Beta}(a_1,\, b_1)$, which can automatically account for the uncertainty in choosing π_0 and π_1. We fixed parameters $a_0 = b_0 = a_1 = b_1 = 1$, so that the priors are essentially non-informative. For computation convenience, we assign a conjugate Inverse–Gamma hyperprior on s^2

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

η is estimated with the Monte Carlo EM algorithm (Park and Casella (2008); Xu and Ghosh (2015)). For the gth EM update,

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

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

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

We set c and d to small values.

4.2.3 Gibbs sampler

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

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

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

$$
b_j | \text{rest} \sim l_j^b \, \mathrm{N}_L(\mu_{b_j}, \, \Sigma_{b_j}) + (1 - l_j^b) \, \delta_0(b_j), \tag{4.10}
$$

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

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

The posterior distribution (4.10) is a mixture of a multivariate normal and a point mass at 0. Specifically, at the gth iteration of MCMC, $b_j^{(g)}$ is drawn from $N(\mu_{b_j}, \Sigma_{b_j})$ with probability l_j^b and is set to 0 with probability $1-l_j^b$. If $b_j^{(g)}$ is set to 0, we have $\phi_j^{b(g)} = 0$, which suggests that the jth genetic variant is not associated with the phenotype at the gth iteration. Otherwise, $\phi_j^{b(g)} = 1$.

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

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

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

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

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

The full conditional distribution for u_i is Inverse-Gaussian:

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

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

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

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

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

$$\pi_0|\text{rest} \sim \text{Beta}\left(a_0 + \sum_{j=1}^{p} \mathbf{I}_{\{b_j\neq 0\}},\ b_0 + \sum_{j=1}^{p} \mathbf{I}_{\{b_j = 0\}}\right),$$
$$\pi_1|\text{rest} \sim \text{Beta}\left(a_1 + \sum_{j,l} \mathbf{I}_{\{\omega_{jl}\neq 0\}},\ b_1 + \sum_{j,l} \mathbf{I}_{\{\omega_{jl} = 0\}}\right).$$

Last, the full conditional distribution for ν is Gamma distribution

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

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

To facilitate fast computation and reproducible research, we have implemented the proposed and all the alternative methods in C++

4.2.4 A summary of proposed and alternative methods

All the methods under comparison can be grouped according to three criteria: with or without robustness, with or without spike-and-slab priors, and the types of structured sparsity (individual-, group- and bi-level) accommodated through variable selection. We first describe the robust Bayesian methods with spike-and-slab priors: RBSG–SS, RBG–SS and RBL–SS, which have all been proposed for the first time. Among them, RBSG–SS is the

golden method developed for conducting robust sparse group variable selection for G×E interactions with spike-and-slab priors on both the group and individual levels. Besides, RBG–SS and RBL–SS are robust Bayesian group level and individual level selection with spike–and–slab priors, respectively. The spike–and–slab prior has only been imposed on the group level in RBG–SS. Compared to RBSG–SS, it does not induce within group sparsity. On the other hand, RBL–SS conducts individual-level selection without accounting for group structure. An immediate family of robust methods related to the three are RBSG, RBG and RBL, which do not adopt spike–and–slab priors and cannot shrink coefficients corresponding to the main and interaction effects to zero exactly. While RBG and RBL can be directly derived based on Li et al. (2010), RBSG, robust Bayesian sparse group selection, has not been investigated in existing studies so far.

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

For clarification, we list all the methods under comparison in Table C.1 in the Appendix. Our contribution includes developing the 4 robust Bayesian variable selection approaches, RBSG–SS, RBG–SS, RBL–SS and RBSG among the first time. For all the rest of the approaches, a modification to the methods from the references provided in Table C.1 by including clinical covariates is necessary. Otherwise, these methods cannot be adopted for a direct comparison with the four newly developed ones.

4.3 Simulation

We comprehensively evaluate the proposed and alternative methods through simulation studies. Under all the settings, the responses are generated from model (4.1) with $n = 500, q = 3, p = 100$ and $k = 5$, which leads to a total of 105 main effects, 500 interactions and 3 additional clinical covariates. Thus, the actual dimension of coefficient vector is 608, higher than the sample size ($n = 500$). The genetic main effects and G×E interactions form 100 groups with group size $L = 6$. Within each one of the following examples, we consider six different error distributions for ϵ_i's: N(0, 1)(Error 1), Laplace(μ,b) with the mean $\mu = 0$ and the scale parameter $b = 2$ (Error 2), 10%Laplace(0,1) + 90%Laplace(0,$\sqrt{5}$) (Error 3), 90%N(0,1) + 10%Cauchy(0, 1) (Error 4), t-distribution with 2 degrees of freedom ($t(2)$) (Error 4), LogNormal(0,1) (Error 5). All of them are heavy-tailed distributions except the first one.

We assess the performance in terms of identification and prediction accuracy. For methods incorporating spike–and–slab priors, we consider the median probability model (MPM) (Barbieri and Berger (2004); Xu and Ghosh (2015)) to identify important effects. In particular, for the proposed RBSG–SS, we define $\phi_{jl} = \phi_j^b \phi_{jl}^w$ for the lth predictor in the jth group. At the gth MCMC iterations, this predictor is included in the model if the indicator $\phi_{jl}^{(g)}$ is 1. Suppose we have collected G posterior samples from the MCMC after burn-ins, then the posterior probability of including the lth predictor from the jth group in the final model is

$$p_{jl} = \hat{\pi}(\phi_{jl} = 1|y) = \frac{1}{G}\sum_{g=1}^{G} \phi_{jl}^{(g)}, \quad j = 1, \ldots, p \text{ and } l = 1, \ldots, L. \tag{4.11}$$

A higher posterior inclusion probability p_{jl} can be interpreted as a stronger empirical evidence that the corresponding predictor has a non-zero coefficient and is associated with the phenotype. The MPM model is defined as the model consisting of predictors with at least $\frac{1}{2}$ posterior inclusion probability. When the goal is to select a single model, Barbieri and Berger (Barbieri and Berger (2004)) recommends using MPM because of its optimal prediction performance. Meanwhile, the 95% credible interval (95%CI) (Li et al. (2015)) is

adopted for methods without spike–and–slab priors.

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

Example 1

We generate a $n \times p$ matrix of gene expressions with $n = 500$ and $p = 100$, from a multivariate normal distribution with marginal mean 0 and marginal variance 1. We consider an auto-regression (AR) correlation structure for gene expression data, in which gene j and h have correlation coefficient $\rho^{|j-h|}$, with $\rho = 0.3$ $(1 \leqslant j, h \leqslant p)$. For E factors, five continuous variables are generated from a multivariate normal distribution with marginal mean 0, marginal variance 1 and AR correlation structure with $\rho = 0.5$. We then dichotomize one of them at 0 to create a binary variable. Thus, there are four continuous and one binary E factor. At last, we simulate three clinical covariates from a multivariate normal distribution and AR correlation structure with $\rho = 0.5$, and dichotomize one of them at 0 to create a binary clinical covariate.

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

Example 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 Example 1.

Example 3

Consider simulating the SNP data under a pairwise linkage disequilibrium (LD) structure. Let the minor allele frequencies (MAFs) of two neighboring SNPs with risk alleles A and

B be r_1 and r_2, respectively. The frequencies of four haplotypes are as $p_{AB} = r_1 r_2 + \delta$, $p_{ab} = (1 - r_1)(1 - r_2) + \delta$, $p_{Ab} = r_1(1 - r_2) - \delta$, and $p_{aB} = (1 - r_1)r_2 - \delta$, where δ denotes the LD. Assuming Hardy-Weinberg equilibrium and given the allele frequency for A at locus 1, we can generate the SNP genotype (AA, Aa, aa) from a multinomial distribution with frequencies $(r_1^2, 2r_1(1 - r_1), (1 - r_1)^2)$. The genotypes at locus 2 can be simulated according to the conditional genotype probability matrix in Cui et al. (2008). We have $\delta = r_p\sqrt{r_1(1 - r_1)r_2(1 - r_2)}$ with MAFs 0.3 and pairwise correlation $r_p = 0.6$.

Example 4

A more practical correlation structure is adopted in this example. We extract the first 100 SNPs from the NHS data analyzed in the case study, so the correlation is based on the real data. For each simulation replicate, we randomly sample 500 subjects from the dataset. The same coefficients and error distributions from the first 3 examples are adopted.

Table 4.1: *Simulation results in Example 1.* (n, q, k, p) *= (500, 2, 5, 100). mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.*

		RBSG–SS	RBG–SS	RBL–SS	BSG–SS	BG–SS	BL–SS
Error 1	TP	24.97(0.18)	25.00(0.00)	24.93(0.25)	24.97(0.18)	25.00(0.00)	24.93(0.25)
N	FP	1.30(1.24)	29.60(2.42)	1.30(1.44)	0.47(0.68)	29.00(0.00)	0.43(0.73)
	Pred	0.83(0.03)	0.86(0.03)	0.84(0.04)	1.07(0.07)	1.13(0.07)	1.08(0.08)
Error 2	TP	21.66(1.72)	24.84(0.55)	18.58(2.14)	19.98(1.95)	24.58(0.86)	15.54(2.04)
L	FP	1.32(1.33)	30.96(4.27)	1.62(1.64)	1.82(1.53)	30.98(4.83)	0.92(0.94)
	Pred	2.15(0.10)	2.17(0.09)	2.24(0.12)	9.32(0.97)	8.98(0.79)	10.09(1.08)
Error 3	TP	21.28(2.24)	24.80(0.73)	18.14(2.68)	19.00(2.61)	24.40(1.09)	14.24(2.39)
Mix.L	FP	1.48(1.34)	30.64(4.23)	1.42(1.63)	2.04(1.73)	30.20(4.73)	1.18(1.16)
	Pred	2.29(0.12)	2.32(0.11)	2.41(0.12)	11.11(1.12)	10.59(0.95)	12.02(1.12)
Error 4	TP	23.80(1.30)	24.93(0.37)	21.80(1.94)	16.20(6.45)	21.83(5.24)	12.53(5.79)
t2	FP	0.53(0.86)	29.47(2.56)	0.20(0.41)	3.73(4.61)	35.77(23.92)	1.93(2.49)
	Pred	1.50(0.14)	1.52(0.13)	1.53(0.14)	12.48(6.56)	12.34(7.27)	13.35(6.72)
Error 5	TP	24.33(0.76)	25.00(0.00)	22.93(1.20)	22.93(1.26)	25.00(0.00)	18.00(2.17)
logNor	FP	0.26(0.45)	29.00(0.00)	0.13(0.35)	4.30(3.40)	34.80(8.11)	1.23(1.55)
	Pred	1.16(0.10)	1.18(0.10)	1.18(0.10)	4.75(1.24)	4.78(1.23)	5.18(1.34)

Table 4.2: *Simulation results in Example 1.* (n, q, k, p) = *(500, 2, 5, 100). mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.*

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

We have collected the posterior samples from the Gibbs sampler running 15,000 iterations while discarding the first 7,500 samples as burn-ins. The Bayesian estimates are calculated using the posterior medians. Simulation results for the gene expression data in Example 1 are tabulated in Table (4.1) and (4.2). We can observe that the performance of methods that adopt spike–and–slab priors in Table (4.1) is consistently better than methods without spike–and–slab priors in Table (4.2). Although, methods without spike–and–slab priors have slightly lower FPs than their counterparts with spike–and–slab priors under some error distributions, they tend to have much lower TPs and higher prediction errors under all the error distributions. For example, under Error2, RBSG identifies 14.48(SD 2.04) out of the 25 true positives, much lower than the true positives of 21.66(SD 1.72) from RBSG–SS. Meanwhile, its false positives 0.64(SD 0.85) is only slightly lower than the FP of RBSG–SS (1.32(SD 1.33)). The prediction error of RBSG, 2.57 with a SD of 0.11, is also inferior than

that of the RBSG–SS (2.17(SD 0.10)). Such an advantage can also be observed by comparing other methods in Table (4.1) with their counterparts (without spike–and–slab priors) from Table (4.2).

Among all the methods with spike–and–slab priors, as shown in Table (4.1), the proposed RBSG–SS has the best performance in both identification and prediction in the presence of data contamination and heavy–tailed errors. Under the mixture Laplace error (Error 3), RSGB–SS identifies 21.28(SD 2.24) true positives, with a small number of false positives, 1.48(SD 1.34). RBG–SS has a true positive of 24.80(SD 0.73), however, the number of false positives, 30.64(SD 4.23), is much higher. This is due to the fact that RBG–SS only conducts group level selection and does not impose the within-group sparsity. Compared to RBSG–SS, RBL–SS ignores the group structure, leading to fewer true positives of 18.14(SD2.68). In terms of prediction, RBSG–SS has the smallest L1 error, 2.29(0.12), among all the 3 robust methods with spike–and–slab priors. Although the difference in prediction error between RBSG–SS and RBG–SS is not distinct, considering the much smaller number of false positive main and interaction effects, we can fully observe the advantage of RSGB–SS over RBG–SS in prediction.

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

Similar patterns have been observed in Table C.3, C.4, C.5, C.6, C.7 and C.8 for Examples 2, 3 and 4, respectively, in the Appendix. Overall, based on the investigations over all the methods through comprehensive simulation studies, we can establish the advantage of

conducting robust Bayesian bi–level selection incorporating spike–and–slab priors.

We demonstrate the sensitivity of RBSG–SS for variable selection to the choice of the hyper–parameters for π_0, and π_1 in the Appendix. The results are tabulated in Table C.2, showing that the MPM model is insensitive to different specification of the hyper-parameters. Following Li et al. Li et al. (2015), we assess the convergence of the MCMC chains by the potential scale reduction factor (PSRF).Brooks and Gelman (1998); Gelman and Rubin (1992) PSRF values close to 1 indicate that chains converge to the stationary distribution. Gelman et al. Gelman et al. (2004) recommend using PSRF$\leq$ 1.1 as the cutoff for convergence, which has been adopted in our study. We compute the PSRF for each parameter and find all chains converge after the burn-ins. For the purpose of demonstration, Figure C.1 shows the pattern of PSRF the first five groups of coefficients in Example 1 under Error 2. The figure clearly shows the convergence of the proposed Gibbs sampler.

4.4 Real Data Analysis

4.4.1 Nurses' Health Study (NHS) data

Nurses' Health Study (NHS) is one of the largest investigations into the risk factors for major chronic diseases in women. As part of the the Gene Environment Association Studies initiative (GENEVA), the NHS provides SNP genotypes data as well as detailed information on dietary and lifestyle variables. Obesity level is one of the most important risk factors for Type 2 diabetes mellitus (T2D), a chronic disease determined by both genetic and environmental factors. In this study, we analyze the NHS type 2 diabetes data to identify genetic factors that are associated with obesity via genetic main effect or gene-environment interactions. We use weight as the response and focus on SNPs on chromosome 10. We consider five environment factors, including the total physical activity (act), glycemic load (gl), cereal fiber intake (ceraf), alcohol intake (alcohol) and a binary indicator of whether an individual has a history of high cholesterol (chol). All these environmental exposures have been suggested to be associated with obesity and diabetes (Hu et al. (2001)). In addition,

we consider three clinical covariates: height, age and a binary indicator of whether an individual has a history of hypertension (hbp). In NHS study, about half of the subjects are diagnosed of type 2 diabetes and the other half are controls without the disease. We only use health subjects in this study. After cleaning the data through matching phenotypes and genotypes, removing SNPs with minor allele frequency (MAF) less than 0.05 or deviation from Hardy–Weinberg equilibrium, the working dataset contains 1732 subjects with 35099 SNPs.

For computational convenience prescreening can be conducted to reduce the feature space to a more attainable size for variable selection. For example, Li et al. Li et al. (2015) use the single SNP analysis to filter SNPs in a GWA study before downstream analysis. In this study, we use a marginal linear model with weight as the response variable to evaluate the penetrance effect of a variant under the environmental exposure. The marginal linear model use a group of genetic main effect and G×E interactions of a SNP as the predictors, and test whether this SNP has any effect, main or G×E interaction, at all. The SNPs with p-values less than a certain cutoff (0.001) for any effect, main or interaction, from the test are kept. 253 SNPs pass the screening.

The proposed approach RBSG-SS identifies 22 main SNP effects and 45 G×E interactions. The detailed estimation results are provided in Table C.9 in the Appendix. We observe that the proposed method identifies main and interaction effects of SNPs with important implications in obesity. For example, two important SNPs, rs6482836 and rs10741150, that located within gene DOCK1 are identified. DOCK1 (Dedicator Of Cytokinesis 1) has been reported as a putative candidate for obesity related to adiponectin and triceps skinfold by previous studies (Kim et al. (2019); Vaughan et al. (2015)). RBSG-SS identifies the main effect of rs6482836 and its interaction with the E factor act. Physical activity plays an important role in the prevention of overweight and obese (Wareham et al. (2005)). This result suggests that the expression level of DOCK1 in an individual may influence the effect of physical activity in obesity prevention. RBSG-SS also identifies the interaction between rs10741150 and the E factor chol, suggesting that the effect of cholesterol level can be mediated by DOCK1. Interestingly, a previous study has shown that the expression level of DOCK5, an important

paralog of DOCK1, is increased in individuals exposed to a diet high in saturated fatty acids (El-Sayed Moustafa et al. (2012)). Our results provide more evidence of the importance of DOCK1 in diet-induced obesity. Another example is the SNP rs11196539, located within gene NRG3. NRG3(Neuregulin 3) has been found to be associated with both the basal metabolic rate (BMR) and body mass index (BMI) (Lee et al. (2016)). RBSG-SS identifies its interaction with the E factors, gl and alcohol. Both glycemic load and alcohol intake are important dietary variables associated with obesity. The continued intake of high-glycemic load meals leads to an increased risk of obesity (Brand-Miller et al. (2002)). The increasing alcohol consumption is associated with a decline in body mass index in women (Nanchahal et al. (2000)), however, heavy drinking can increase risk of the metabolic syndrome (Baik and Shin (2008)). Our results suggest that further investigation of NRG3 may help explain the mechanism of the effects of glycemic load and alcohol intake on obesity. For the environment main effects, two E factors, chol and gl, have positive coefficients, and the other three, act, ceraf and alcohol, have negative coefficients, which are consistent with findings in the previous literature.

In addition to the proposed approach, we also conduct analysis using the alternatives RBL-SS, BSG-SS and BL-SS. As other alternative methods show inferior performance in simulation, they are not considered in real data analysis. Detailed estimation results are provided in Table C.10, C.11 and C.12 in the Appendix. In Table 4.3, we provide the numbers of main G effects and interactions identified by different approaches and their overlaps. We can observe that the proposed method identifies different main G effects and more significantly different interactions from those identified by the alternatives. To further investigate the biological similarity of the identified genes, we conduct the the Gene Ontology (GO) analysis. We observe that there is an obvious difference between the proposed method and the three alternatives. The GO analysis results are provided in Figure C.2.

With real data, it is difficult to objectively evaluate the selection performance. The prediction performance may provide additional information on the relative performance of different methods. Following Yan and Huang (2012) and Li et al. (2015), we refit the models selected by RBSG-SS and RBL-SS by the Robust Bayesian Lasso, and refit the models

selected by BSG-SS and BL-SS by the Bayesian Lasso. For the comparison between the robust methods, the prediction mean absolute deviations (PMAD) are computed based on the posterior median estimates. The PMADs are 8.64 and 8.88 for RBSG-SS and RBL-SS, respectively. The proposed method outperforms than the competitor. For the comparison between the non-robust methods, the prediction mean squared errors (PMSE) are computed. The PMSEs are 128.39 and 137.77 for BSG-SS and BL-SS, respectively. The sparse group method outperforms than the individual selection method.

4.4.2 TCGA skin cutaneous melanoma data

In this case study, we analyze the Cancer Genome Atlas (TCGA) skin cutaneous melanoma (SKCM) data. TCGA ia a cancer genomics program organized by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI). It publishes high quality clinical, environmental, and genetic data. For this study, we use the level-3 gene expression data of SKCM downloaded from the cBio Cancer Genomics Portal (Cerami et al. (2012); Krauthammer et al. (2015)). Our goal is to identify genes that have genetic main effect or G×E interaction effects on the Breslow' thickness, an important prognostic variable for SKCM (Marghoob et al. (2000)). The log-transformed Breslows depth is used as the response variable and four E factors are considered, age, AJCC pathologic tumor stage, gender and Clark level. Data are available on 294 subjects and 20,531 gene expressions. We adopt the same screening method used in the first case study to select 109 genes for further analysis.

The proposed approach RBSG-SS identifies 16 main SNP effects and 32 G×E interactions. The detailed estimation results are provided in Table C.13 in the Appendix. One important gene identified is CXCL6 (C-X-C Motif Chemokine Ligand 6), a chemokine with neutrophil chemotactic and angiogenic activites. It has been reported that CXCL6 plays an important role in melanoma growth and metastasis (Verbeke et al. (2011)). RBSG-SS identifies its main effect and its interactions with E factors, stage and Clark level. This suggests that CXCL6 can have different effects at different stages of melanoma. Another important finding is the

gene MAGED4, one of member in MAGE(Melanoma-associated antigen) family. MAGE family contains genes that are highly attractive targets for cancer immunotherapy (Zhang et al. (2014b)). MAGED4 has been found to be an potential target for glioma immunotherapy (Sang et al. (2011)). RBSG-SS identifies the main effect of MAGED4 and its interaction with the E factor tumor stage, suggesting that MAGED4 may also play an important role in SKCM and its effect may change over different tumor stages. For the main effects of the E factors, Clark level and tumor stage have positive coefficients, and age and gender have negative coefficients, which match observations in the literature.

Analysis is further conducted using the three alternatives, and the comparison results are summarized in Table 4.3. Detailed estimation results are provided in Table C.14, C.15 and C.16 in the Appendix. As for the previous case study, the proposed approach identifies different sets of main and G×E interaction effects from those identified by the alternatives. We also investigate the biological similarity of the identified genes by GO analysis (Figure C.2). The results show that there is an obvious difference between the proposed method and the three alternatives. Prediction performance is also evaluated. The PMADs are 0.69 and 0.83 for RBSG-SS and RBL-SS, respectively. The proposed approach again has better prediction performance than RBL-SS. The PMSEs are 0.93 and 1.05 for BSG-SS and BL-SS, respectively. The sparse group method still outperforms than the individual selection method.

4.5 Discussion

In this study, we have developed robust Bayesian variable selection methods for gene-environment interaction studies. The robustness of our methods comes from Bayesian formulation of LAD regression. In G×E studies, the demand for robustness arises in heavy-tailed distribution/ data contamination in both the response and predictors, as well as model misspecification. We have focused on the first case, which is frequently encountered in practice. Investigations of the robust Bayesian methods accommodating the other two cases are interesting and will be pursued in the future.

Table 4.3: *The numbers of main G effects and interactions identified by different approaches and their overlaps.*

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

Recently, penalization has emerged as a power tool for dissecting G×E interactions Zhou et al. (2020a). Our literature review suggests that Bayesian variable selection methods, although tightly related to penalization, has not been fully explored for interaction analyses, let alone the robust ones. We are among the first to conduct robust G×E analysis within the Bayesian framework. The proposed Bayesian LAD sparse group LASSO are not only specifically tailored for G×E studies, and but also generally applicable for problems incorporating the bi-level structure in a broader context, such as simultaneously selection of prognostic genes and pathways Liu et al. (2019). The spike-and-slab priors have been incorporated to further improve identification and prediction performances. As a byproduct, the Bayesian LAD LASSO and group LASSO, both with spike-and-slab priors, have also been investigated for the first time. The computational feasibility of the Gibbs samplers are guaranteed by the R package robin, with the core modules of the MCMC algorithms developed in C++.

In G×E studies, the form of interaction effects can be linear, nonlinear, and both linear and nonlinear, resulting in parametric Wu et al. (2018a); Zhou et al. (2019a,b), nonparametric Li et al. (2015); Wu et al. (2018c) and semiparametric variable selection methods Ren

et al. (2020a); Wu et al. (2014, 2015) to dissect G×E interactions, respectively. The proposed study can be potentially generalized to these studies within robust Bayesian framework. For example, variable selection for multiple semiparametric G×E studies can be formulated as a combination of individual and group level selection problem, where the robust Bayesian methods based on sparse group, group and individual level selection are directly applicable. The proposed robust Bayesian framework has paved the way for the future investigations.

Summary

This dissertation focuses on developing penalized variable selection methods, from both frequentist and Bayesian perspectives, to conduct efficient variable selection of high dimensional omics data with complicated network and interaction structures.

Incorporation of the correlation structure as networks within the framework of penalized variable selection, can lead to more accurate identification of important omics features and improved prediction (Li and Li (2008) and Huang et al. (2011)). However, the existing network–based methods lack robust properties, which are critical to accommodate data contamination and longtailed distributions in survival time data. In Chapter 2, we develop a novel robust network-based variable selection method under the AFT model for survival time in cancer genomic studies. Our method significantly distinguishes from existing ones in (1) we adopt a weighted LAD objective function to accommodate data contamination, with Kaplan-Meier weights for censoring; (2) to incorporate the interconnections among gene expressions, we propose a network-constrained penalty of the "MCP+L_1" form, and develop an efficient algorithm within the coordinate descent framework. The paper associated with this study is published at the Genetic Epidemiology (Ren et al. (2019a)).

Gene–environment (G×E) interactions plays an important in elucidating the disease etiology for complex diseases, such as cancer, Type 2 Diabetes and asthma. Many studies have demonstrated the advantages of penalization methods in detecting G×E interactions from frequentist point of view (Wu et al. (2019)). Bayesian variable selection, however, has not been widely developed for interaction studies. In Chapter 3, we have proposed a novel semi-parametric Bayesian variable selection method to simultaneously pinpoint important G×E interactions in both linear and nonlinear forms while conducting automatic structure discovery. The superior performance of the proposed method over multiple alternatives has been demonstrated through extensive simulation studies and a case study. This work is

published at Statistics in Medicine (Ren et al. (2020a)).

In the last part of the dissertation, we have developed a novel robust Bayesian variable selection method to dissect G×E interactions in genomic studies. The proposed Bayesian LAD sparse group method can effectively accommodate heavy-tailed errors and outliers in the response variable while conducting variable selection by accounting for structural sparsity. To the best of our knowledge, we are among the first to conduct robust G×E analysis within the Bayesian framework. The proposed method are not only specifically tailored for G×E studies, but also generally applicable for problems incorporating the bi-level structure in a broader context.

To facilitate reproducible research and fast computation, we have developed open source R packages for each project, which provide highly efficient C++ implementation for all the proposed and alternative approaches. The R packages regnet (Ren et al. (2019b)) and spinBayes (Ren et al. (2019c)), associated with the first and second project correspondingly, are available on CRAN. For the third project, the R package robin is available from GitHub and will be submitted to CRAN soon.

Bibliography

D. F. Andrews and C. L. Mallows. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(1):99–102, 1974. ISSN 00359246. URL http://www.jstor.org/stable/2984774.

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

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

A. H. Berger, M. Imielinski, F. Duke, J. Wala, N. Kaplan, G.X. Shi, and M. Meyerson. Oncogenic rit1 mutations in lung adenocarcinoma. *Oncogene*, 33(35):4418–44231, 2014. doi: 10.1038/onc.2013.581.

Dimitri P. Bertsekas. Incremental gradient, subgradient, and proximal methods for convex optimization: A survey. *Optimization for Machine Learning*, 3:1–38, 2010.

S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

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

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

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001. doi: 10.1023/A:1010933404324.

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

Bradley P. Carlin and Siddhartha Chib. Bayesian model choice via markov chain monte carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(3): 473–484, 1995. ISSN 00359246. URL http://www.jstor.org/stable/2346151.

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

Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, Benjamin E. Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, Caitlin J. Byrne, Michael L. Heuer, Erik Larsson, Yevgeniy Antipin, Boris Reva, Arthur P. Goldberg, Chris Sander, and Nikolaus Schultz. The cbio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, 2(5):401–404, 2012. doi: 10.1158/2159-8290.CD-12-0095.

Cristina Chamizo, Sandra Zazo, Manuel Dmine, Ion Cristbal, Jess Garca-Foncillas, Federico Rojo, and Juan Madoz-Grpide. Thymidylate synthase expression as a predictive biomarker of pemetrexed sensitivity in advanced non–small–cell lung cancer. *BMC Pulmonary Medicine*, 15(1):132, 2015. ISSN 1471-2466. doi: 10.1186/s12890-015-0132-x.

Marilyn C. Cornelis, Eric J. Tchetgen Tchetgen, Liming Liang, Lu Qi, Nilanjan Chatterjee, Frank B. Hu, and Peter Kraft. Gene–Environment Interactions in Genome–Wide Association Studies: A Comparative Study of Tests Applied to Empirical Studies of Type 2 Diabetes. *American Journal of Epidemiology*, 175(3):191–202, 12 2011. ISSN 0002-9262. doi: 10.1093/aje/kwr368.

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

Yinhao Du, Jie Ren, Fei Zhou, Yu Jiang, Shuangge Ma, and Cen Wu. Integrating muti–omics data for gene–environment interactions. 2020.

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

Katja Fall, Fredrik Stromberg, Johan Rosell, Ove Andren, Eberhard Varenhorst, and The South–East Region Prostate Cancer Group. Reliability of death certificates in prostate cancer patients. *Scandinavian Journal of Urology and Nephrology*, 42(4):352–357, 2008. doi: 10.1080/00365590802078583.

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001. doi: 10.1198/016214501753382273.

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

Delphine Fradin, Pierre-Yves Bolle, Marie-Pierre Belot, Fanny Lachaux, Jorg Tost, Cline Besse, Jean-Franois Deleuze, Gianpaolo De Filippo, and Pierre Bougnres. Genome–wide methylation analysis identifies specific epigenetic marks in severely obese children. *Scientific Reports*, 7, 2017.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *Ann. Statist.*, 29(5):1189–1232, 10 2001. doi: 10.1214/aos/1013203451. URL https://doi.org/10.1214/aos/1013203451.

Xiaoli Gao and Jian Huang. A robust penalized method for the analysis of noisy DNA copy number data. *BMC Genomics*, 11(1):517, 2010a. doi: 10.1186/1471-2164-11-517.

Xiaoli Gao and Jian Huang. Asymptotic analysis of high–dimensional LAD regression with lasso. *Statistica Sinica*, 20(4):1485–1506, 2010b.

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

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

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

S. Hanselmann, P. Wolter, J. Malkmus, and S. Gaubatz. The microtubule–associated protein PRC1 is a potential therapeutic target for lung cancer. *Oncotarget*, 9(4):4985–4997, 2017. doi: 10.18632/oncotarget.23577.

Min He, Lin Chao, and YiPing You. PRPS1 silencing reverses cisplatin resistance in human breast cancer cells. *Biochemistry and Cell Biology*, 95(3):385–393, 2017. doi: 10.1139/bcb-2016-0106.

Masayo Hosokawa, Akio Takehara, Koichi Matsuda, Hidetoshi Eguchi, Hiroaki Ohigashi, Osamu Ishikawa, Yasuhisa Shinomura, Kohzoh Imai, Yusuke Nakamura, and Hidewaki Nakagawa. Oncogenic role of KIAA0101 interacting with proliferating cell nuclear antigen in pancreatic cancer. *Cancer Research*, 67(6):2568–2576, 2007. doi: 10.1158/0008-5472.CAN-06-4356.

Guoxin Hou, Panpan Liu, Jing Yang, and Shijun Wen. Mining expression and prognosis of topoisomerase isoform in non–small–cell lung cancer lung cancer by using oncomine and kaplan–meier plotter. *PLOS ONE*, 12(3):1–16, 03 2017. doi: 10.1371/journal.pone.0174515.

Gerta Hoxhaj, Kumara Dissanayake, and Carol MacKintosh. Effect of IRS4 levels on PI3–kinase signalling. *PLOS ONE*, 8(9):1–9, 09 2013. doi: 10.1371/journal.pone.0073327.

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

H. Huang, J. Liu, Q. Meng, and G. Niu. Multidrug resistance protein and topoisomerase 2 alpha expression in non–small cell lung cancer are related with brain metastasis postoperatively. *International Journal of Clinical and Experimental Pathology*, 8(9):11537–11542, 2015.

Haihui Huang and Yong Liang. Hybrid l 1/2+2 method for gene selection in the cox proportional hazards model. *Computer Methods and Programs in Biomedicine*, 164:65–73, 2018. ISSN 0169-2607. doi: https://doi.org/10.1016/j.cmpb.2018.06.004.

Haihui Huang and Yong Liang. A novel cox proportional hazards model for high - dimensional genomic data in cancer prognosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 1–1, 2019. ISSN 1557-9964. doi: 10.1109/TCBB.2019.2961667.

Haihui Huang, Jingguo Dai, and Yong Liang. Clinical drug response prediction by using a lq penalized network-constrained logistic regression method. *Cell Physiol Biochem*, 51: 2073–2084, 2018. doi: 10.1159/000495826.

Jian Huang and Shuangge Ma. Variable selection in the accelerated failure time model via the bridge method. *Lifetime data analysis*, 16(2):176–195, 2010. doi: 10.1007/s10985-009-9144-2.

Jian Huang, Shuangge Ma, and Huiliang Xie. Least absolute deviations estimation for the accelerated failure time model. *Statistica Sinica*, 17(4):1533–1548, 2007. ISSN 10170405, 19968507.

Jian Huang, Shuangge Ma, Hongzhe Li, and Cun-Hui Zhang. The sparse laplacian shrinkage estimator for high–dimensional regression. *Ann. Statist.*, 39(4):2021–2046, 08 2011. doi: 10.1214/11-AOS897.

Jian Huang, Patrick Breheny, and Shuangge Ma. A selective review of group selection in high-dimensional models. *Statist. Sci.*, 27(4):481–499, 11 2012. doi: 10.1214/12-STS392.

Jianhua Z. Huang, Colin O. Wu, and Lan Zhou. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128, 2002.

Jianhua Z. Huang, Colin O. Wu, and Lan Zhou. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, 14(3):763–788, 2004.

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

Carolyn M. Hutter, Leah E. Mechanic, Nilanjan Chatterjee, Peter Kraft, and Elizabeth M. Gillanders. Gene-environment interactions in cancer epidemiology: A national cancer institute think tank report. *Genetic Epidemiology*, 37(7):643–657, 2013. doi: 10.1002/gepi.21756.

Duo Jiang, Courtney R. Armour, Chenxiao Hu, Meng Mei, Chuan Tian, Thomas J. Sharpton, and Yuan Jiang. Microbiome multi-omics network analysis: Statistical considerations, limitations, and opportunities. *Frontiers in Genetics*, 10:995, 2019. ISSN 1664-8021. doi: 10.3389/fgene.2019.00995.

Yu Jiang, Yuan Huang, Yinhao Du, Yinjun Zhao, Jie Ren, Shuangge Ma, and Cen Wu. Identification of prognostic genes and pathways in lung adenocarcinoma using a bayesian approach. *Cancer informatics*, 15, 11 2017. doi: 10.1177/1176935116684825.

Hong Jiao, Peter Arner, Johan Hoffstedt, David Brodin, Beatrice Dubern, Sébastien Czernichow, Ferdinand van't Hooft, Tomas Axelsson, Oluf Pedersen, Torben Hansen, Thork-

ild IA Sørensen, Johannes Hebebrand, Juha Kere, Karin Dahlman-Wright, Anders Hamsten, Karine Clement, and Ingrid Dahlman. Genome wide association study identifies kcnma1contributing to human obesity. *BMC Medical Genomics*, 4(1):51, 2011. doi: 10.1186/1755-8794-4-51.

Luann C. Jung, Haiyan Wang, Xukun Li, and Cen Wu. A machine learning method for selection of genetic variants to increase prediction accuracy of type 2 diabetes mellitus using sequencing data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 2020. doi: 10.1002/sam.11456.

Tatsuya Kato, Yataro Daigo, Masato Aragaki, Keidai Ishikawa, Masaaki Sato, and Mitsuhito Kaji. Overexpression of CDC20 predicts poor prognosis in primary non–small–cell lung cancer patients. *Journal of Surgical Oncology*, 106(4):423–430, 2012a. doi: 10.1002/jso.23109.

Tatsuya Kato, Yataro Daigo, Masato Aragaki, Keidai Ishikawa, Masaaki Sato, and Mitsuhito Kaji. Overexpression of kiaa0101 predicts poor prognosis in primary lung cancer patients. *Lung Cancer*, 75(1):110–118, 2012b. doi: 10.1016/j.lungcan.2011.05.024.

S A Kenfield, E K Wei, M J Stampfer, B A Rosner, and G A Colditz. Comparison of aspects of smoking among the four histological types of lung cancer. *Tobacco Control*, 17 (3):198–204, 2008. doi: 10.1136/tc.2007.022582.

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

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

Michael Krauthammer, Yong Kong, Antonella Bacchiocchi, Perry Evans, Natapol Pornputtapong, Cen Wu, James P McCusker, Shuangge Ma, Elaine Cheng, Robert Straub, Merdan Serin, Marcus Bosenberg, Stephan Ariyan, Deepak Narayan, Mario Sznol, Harriet M Kluger, Shrikant Mane, Joseph Schlessinger, Richard P Lifton, and Ruth Halaban. Exome sequencing identifies recurrent mutations in nf1 and rasopathy genes in sun-exposed melanomas. *Nature genetics*, 47(9):996–1002, sep 2015. ISSN 1546-1718 (Electronic). doi: 10.1038/ng.3361.

Lynn Kuo and Bani Mallick. Variable selection for regression models. *Sankhy: The Indian Journal of Statistics, Series B (1960-2002)*, 60(1):65–81, 1998.

W.Y. Kuo, C.Y. Wu, L. Hwu, J.S. Lee, C.H. Tsai, K.P. Lin, and R.S. Liu. Enhancement of tumor initiation and expression of kcnma1, morf4l2 and aspm genes in the adenocarcinoma of lung xenograft after vorinostat treatment. *Oncotarget*, 16(11):8663–8675, 2015. ISSN 1471-2466. doi: 10.18632/oncotarget.3536.

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

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

C. Li and H. Li. Network–constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, 24(9):1175–1182, 2008. doi: 10.1093/bioinformatics/btn081.

H. Li, W. Stokes, E. Chater, R. Roy, E. de Bruin, Y. Hu, and O. E. Pardo. Decreased glutathione biosynthesis contributes to egfr t790m–driven erlotinib resistance in non–small–cell lung cancer. *Cell Discovery*, 2:16031, 2016. doi: 10.1038/celldisc.2016.31.

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

Jun Li, Qing Lu, and Yalu Wen. Multi-kernel linear mixed model with adaptive lasso for prediction analysis on high-dimensional multi-omics data. *Bioinformatics*, 36(6):1785–1794, 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz822.

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

Wei Dong Li, Hongxiao Jiao, Kai Wang, Clarence K. Zhang, Joseph T. Glessner, Struan F.A. Grant, Hongyu Zhao, Hakon Hakonarson, and R. Arlen Price. A genome wide association study of plasma uric acid levels in obese cases and never-overweight controls. *Obesity*, 21 (9):E490–E494, 2013. doi: 10.1002/oby.20303.

Yang Li, Rong Li, Cunjie Lin, Yichen Qin, and Shuangge Ma. Li, yang, et al. "penalized integrative semiparametric interaction analysis for multiple genetic datasets. *Statistics in Medicine*, pages 1–22, 2019. doi: 10.1002/sim.8172.

Changlu Liu, Jianzhong Ma, and Christopher I. Amos. Bayesian variable selection for hierarchical gene–environment and gene–gene interactions. *Human Genetics*, 134(1):23–36, 1 2015. ISSN 1432-1203. doi: 10.1007/s00439-014-1478-5.

Mine Y. Liu, Zhou Zhou, Ruidong Ma, Zhenyin Tao, Huiwan Choi, Angela L. Bergeron, Huaizhu Wu, and Jing fei Dong. Gender-dependent up-regulation of adamts-13 in mice with obesity and hypercholesterolemia. *Thrombosis Research*, 129(4):536 – 539, 2012. ISSN 0049-3848. doi: https://doi.org/10.1016/j.thromres.2011.11.039.

Wei Liu, Songyun Ouyang, Zhigang Zhou, Meng Wang, Tingting Wang, Yu Qi, Chunling Zhao, Kuisheng Chen, and Liping Dai. Identification of genes associated with cancer progression and prognosis in lung adenocarcinoma: Analyses based on microarray from on-

comine and the cancer genome atlas databases. *Molecular Genetics & Genomic Medicine*, 7(2):e00528, 2019. doi: 10.1002/mgg3.528.

L. Lu, Y. Lv, J. Dong, S. Hu, and R. Peng. DRG1 is a potential oncogene in lung adenocarcinoma and promotes tumor progression via spindle checkpoint signaling regulation. *Oncotarget*, 7(45):72795–72806, 2016. doi: 10.18632/oncotarget.11973.

Tao Luo, Jing Fu, An Xu, Bo Su, Yibing Ren, Ning Li, Junjie Zhu, Xiaofang Zhao, Rongyang Dai, Jie Cao, Bibo Wang, Wenhao Qin, Jinhua Jiang, Juan Li, Mengchao Wu, Gensheng Feng, Yao Chen, and Hongyang Wang. PSMD10/gankyrin induces autophagy to promote tumor progression through cytoplasmic interaction with atg7 and nuclear transactivation of atg7 expression. *Autophagy*, 12(8):1355–1371, 2016. doi: 10.1080/15548627.2015.1034405.

Shujie Ma and Peter X.-K. Song. Varying index coefficient models. *Journal of the American Statistical Association*, 110(509):341–356, 2015. doi: 10.1080/01621459.2014.903185.

Shujie Ma, Lijian Yang, Roberto Romero, and Yuehua Cui. Varying coefficient model for gene–environment interaction: a non-linear look. *Bioinformatics*, 27(15):2119–2126, 06 2011. ISSN 1367-4803. doi: 10.1093/bioinformatics/btr318.

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

W. Min, J. Liu, and S. Zhang. Network–regularized Sparse Logistic Regression Models for Clinical Risk Prediction and Biomarker Discovery. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(3):944–953, 2018. doi: 10.1109/TCBB.2016.2640303.

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

Jeffrey S Morris, Philip J Brown, Richard C Herrick, Keith A Baggerly, and Kevin R Coombes. Bayesian analysis of mass spectrometry proteomic data using wavelet-based functional mixed models. *Biometrics*, 64(2):479–89, 2007. doi: 10.1111/j.1541-0420.2007.00895.x.

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

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

R. B. O'Hara and M. J. Sillanpää. A review of bayesian variable selection methods: what, how and which. *Bayesian Anal.*, 4(1):85–117, 03 2009. doi: 10.1214/09-BA403. URL https://doi.org/10.1214/09-BA403.

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

Bo Peng and Lan Wang. An iterative coordinate descent algorithm for high–dimensional nonconvex penalized quantile regression. *Journal of Computational and Graphical Statistics*, 24(3):676–694, 2015. doi: 10.1080/10618600.2014.913516.

Jie Peng, Ji Zhu, Anna Bergamaschi, Wonshik Han, Dong Young Noh, Jonathan R. Pollack, and Pei Wang. Regularized multivariate regression for identifying master predictors with application to integrative genomics study of breast cancer. *The Annals of Applied Statistics*, 4(1):53–77, 3 2012. ISSN 1932-6157. doi: 10.1214/09-AOAS271.

Sarah Porter, Ian M. Clark, Lara Kevorkian, and Dylan R. Edwards. The adamts metalloproteinases. *Biochemical Journal*, 386(1):15–27, 2005. ISSN 0264-6021. doi: 10.1042/BJ20040424.

Rasika Rampatige, Saman Gamage, Sharika Peiris, and Alan D Lopez. Assessing the reliability of causes of death reported by the vital registration system in Sri Lanka: Medical records review in Colombo. *Health Information Management Journal*, 42(3):20–28, 2013. doi: 10.1177/183335831304200302.

Jie Ren, Tao He, Ye Li, Sai Liu, Yinhao Du, Yu Jiang, and Cen Wu. Network–based regularization for high dimensional snp data in the case–control study of type 2 diabetes. *BMC Genetics*, 18(1):44, 2017. ISSN 1471-2156. doi: 10.1186/s12863-017-0495-5. URL https://doi.org/10.1186/s12863-017-0495-5.

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

Jie Ren, Luann C. Jung, Yinhao Du, Cen Wu, Yu Jiang, and Junhao Liu. *regnet: Network-Based Regularization for Generalized Linear Models*, 2019b. URL https://CRAN.R-project.org/package=regnet. R package version 0.4.0.

Jie Ren, Fei Zhou, Xiaoxi Li, Cen Wu, and Yu Jiang. *spinBayes: Semi-Parametric Gene-Environment Interaction via Bayesian Variable Selection*, 2019c. URL https://CRAN.R-project.org/package=spinBayes. R package version 0.1.0.

Jie Ren, Fei Zhou, Xiaoxi Li, Qi Chen, Hongmei Zhang, Shuangge Ma, Yu Jiang, and Cen Wu. Semiparametric bayesian variable selection for gene-environment interactions. *Statistics in Medicine*, 39(5):617–638, 2020a. doi: 10.1002/sim.8434.

Jie Ren, Fei Zhou, Xi Lu, Shuangge Ma, Yu Jiang, and Cen Wu. Robust bayesian variable selection for gene–environment interactions. 2020b.

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

Meixiang Sang, Lifang Wang, Chunyan Ding, Xinliang Zhou, Bin Wang, Ling Wang, Yishui

Lian, and Baoen Shan. Melanoma-associated antigen genes an update. *Cancer Letters*, 302(2):85 – 90, 2011. ISSN 0304-3835. doi: https://doi.org/10.1016/j.canlet.2010.10.021.

Daniel J. Schaid, Jason P. Sinnwell, Gregory D. Jenkins, Shannon K. McDonnell, James N. Ingle, Michiaki Kubo, Paul E. Goss, Joseph P. Costantino, D. Lawrence Wickerham, and Richard M. Weinshilboum. Using the gene ontology to scan multilevel gene sets for associations in genome wide association studies. *Genetic Epidemiology*, 36(1):3–16, 2012. doi: 10.1002/gepi.20632.

M. A. Schneider, P. Christopoulos, T. Muley, A. Warth, U. Klingmueller, M. Thomas, F. J. Herth, H. Dienemann, N. S. Mueller, F. Theis, and M. Meister. Aurka, dlgap5, tpx2, kif11 and ckap5: Five specific mitosis–associated genes correlate with poor prognosis for non–small–cell lung cancer patients. *International Journal of Oncology*, 50(2):365–372, 2017. doi: 10.3892/ijo.2017.3834.

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. doi: 10.1093/bioinformatics/btl362.

Y.X. Shi, T. Zhu, T. Zou, W. Zhuo, Y.X. Chen, M.S. Huang, and Z.Q. Liu. Prognostic and predictive values of cdk1 and mad2l1 in lung adenocarcinoma. *Oncotarget*, 7(51): 85235–85243, 2016. doi: 10.18632/oncotarget.13252.

Y.X. Shi, J.Y. Yin, Y. Shen, W. Zhang, H.H. Zhou, and Z.Q. Liu. Genome–scale analysis identifies nek2, dlgap5 and ect2 as promising diagnostic and prognostic biomarkers in human lung cancer. *Scientific Reports*, 7(1):8072, 2017. doi: 10.1038/s41598-017-08615-5.

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

John D. Storey. The positive false discovery rate: a bayesian interpretation and the q -value. *Ann. Statist.*, 31(6):2013–2035, 12 2003. doi: 10.1214/aos/1074290335. URL https://doi.org/10.1214/aos/1074290335.

W. Stute and J.L. Wang. The strong law under random censorship. *The Annals of Statistics*, 21(3):1591–1607, 1993. ISSN 00905364.

Hokeun Sun and Shuang Wang. Penalized logistic regression for high–dimensional dna methylation data with case–control studies. *Bioinformatics*, 28(10):1368–1375, 2012. doi: 10.1093/bioinformatics/bts145.

Hokeun Sun and Shuang Wang. Network-based regularization for matched case-control analysis of high-dimensional dna methylation data. *Statistics in Medicine*, 32(12):2127–2139, 2013. doi: 10.1002/sim.5694.

Xing Tan, Li Liao, YanPing Wan, MeiXiang Li, SiHan Chen, WenJuan Mo, QiongLan Zhao, LiFang Huang, and GuQing Zeng. Downregulation of selenium–binding protein 1 is associated with poor prognosis in lung squamous cell carcinoma. *World Journal of Surgical Oncology*, 14(1):70, 2016. ISSN 1477-7819. doi: 10.1186/s12957-016-0832-6.

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

The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, 489(7417):519–525, 2012. doi: 10.1038/nature11404.

Xinyu Tian, Xuefeng Wang, and Jun Chen. Network–constrained group lasso for high–dimensional multinomial classification with application to cancer subtype prediction. *Cancer Informatics*, 13s6:CIN.S17686, 2014. doi: 10.4137/CIN.S17686.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996. ISSN 00359246.

Robert Tibshirani. The lasso method for variable selection in the cox model. *Statistics in Medicine*, 16(4):385–395, 1997. doi: 10.1002/(SICI)1097-0258(19970228)16:4⟨385::AID-SIM380⟩3.0.CO;2-3.

Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005. doi: 10.1111/j.1467-9868.2005.00490.x.

Ryuta Tobe, Bradley A. Carlson, Petra A. Tsuji, Byeong Jae Lee, Vadim N. Gladyshev, and Dolph L. Hatfield. Differences in redox regulatory systems in human lung and liver tumors suggest different avenues for therapy. *Cancers*, 7(4):2262–2276, 2015. doi: 10.3390/cancers7040889.

Marta Tomás, Esther Vázquez, José M Fernández-Fernández, Isaac Subirana, Cristina Plata, Magda Heras, Joan Vila, Jaume Marrugat, Miguel A Valverde, and Mariano Sentí. Genetic variation in the kcnma1 potassium channel alpha subunit as risk factor for severe essential hypertension and myocardial infarction. *Journal of Hypertension*, 26(11):0263–6352, 2008.

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

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

Ting Wang, Chang Chuan Pan, Jing Rui Yu, Yu Long, Xiao Hong Cai, Xu De Yin, Li Qiong

Hao, and Li Li Luo. Association between tyms expression and efficacy of pemetrexed–based in advanced non–small–cell lung cancer: A meta analysis. *PLOS ONE*, 8(9), 09 2013a. doi: 10.1371/journal.pone.0074284.

Xueqin Wang, Yunlu Jiang, Mian Huang, and Heping Zhang. Robust variable selection with exponential squared loss. *Journal of the American Statistical Association*, 108(502): 632–643, 2013b. doi: 10.1080/01621459.2013.766613.

Z. Wang, L. Wan, J. Zhong, H. Inuzuka, P. Liu, F. H. Sarkar, and W. Wei. Cdc20: a potential novel therapeutic target for cancer treatment. *Current Pharmaceutical Design*, 19(18):3210–3214, 2013c.

Nicholas J. Wareham, Esther M. F. van Sluijs, and Ulf Ekelund. Physical activity and obesity prevention: a review of the current evidence. *Proceedings of the Nutrition Society*, 64(2): 229247, 2005. doi: 10.1079/PNS2005423.

J. Weischenfeldt, T. Dubash, A. P. Drainas, B. R. Mardin, Y. Chen, A. M. Stutz, and J. O. Korbel. Pan–cancer analysis of somatic copy number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nature Genetics*, 49(1):65–74, 2017. doi: 10.1038/ng.3722.

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

Cen Wu and Yuehua Cui. Boosting signals in gene–based association studies via efficient SNP selection. *Briefings in Bioinformatics*, 15(2):279–291, 01 2013b. ISSN 1467-5463. doi: 10.1093/bib/bbs087.

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.

Cen Wu, Shaoyu Li, and Yuehua Cui. Genetic association studies: An information content

perspective. *Current Genomics*, 13(7):566–573, 2012. ISSN 1389-2029/1875-5488. doi: 10.2174/138920212803251382.

Cen Wu, Ping-Shou Zhong, and Yuehua Cui. High dimensional variable selection for gene-environment interactions. *Technical Report. Michigan State University*, 2013.

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

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

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

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

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

Cen Wu, Fei Zhou, Jie Ren, Xiaoxi Li, Yu Jiang, and Shuangge Ma. A selective review of multi-level omics data integration using variable selection. *High-Throughput*, 8(1), 2019. ISSN 2571-5135. doi: 10.3390/ht8010004. URL https://www.mdpi.com/2571-5135/8/1/4.

Tong Tong Wu and Kenneth Lange. Coordinate descent algorithms for lasso penalized regression. *The Annals of Applied Statistics*, 2(1):224–244, 2008. ISSN 19326157.

Yang Xie, Guanghua Xiao, Kevin R. Coombes, Carmen Behrens, Luisa M. Solis, Gabriela Raso, Luc Girard, Heidi S. Erickson, Jack Roth, John V. Heymach, Cesar Moran, Kathy Danenberg, John D. Minna, and Ignacio I. Wistuba. Robust gene expression signature from formalin–fixed paraffin–embedded samples predicts prognosis of non–small–cell lung cancer patients. *Clinical Cancer Research*, 17(17):5705–5714, 2011. ISSN 1078-0432. doi: 10.1158/1078-0432.CCR-11-0196.

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

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

Nengjun Yi and Shizhong Xu. Bayesian lasso for quantitative trait loci mapping. *Genetics*, 179(2):1045–1055, 2008. doi: 10.1534/genetics.107.085589.

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

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

Peiwen Yu, Betty Huang, Mary Shen, Clorinda Lau, Eva Chan, Jennifer Michel, Yue Xiong, Donald G Payan, and Ying Luo. p15PAF, a novel PCNA associated factor with increased expression in tumor tissues. *Oncogene*, 20:484–489, 2001. doi: 10.1038/sj.onc.12041138.

Ming Yuan and Yi Lin. Efficient empirical bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, 100(472):1215–1225, 2005. doi: 10.1198/016214505000000367.

Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables.

Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1):49–67, 2006. doi: 10.1111/j.1467-9868.2005.00532.x.

Cecilia Zappa and Shaker A. Mousa. Non–small–cell lung cancer: current treatment and future advances. *Translational Lung Cancer Research*, 5(3):288–300, 2016. doi: 10.21037/tlcr.2016.06.07.

Guqing Zeng, Hong Yi, Pengfei Zhang, Xinhui Li, Rong Hu, Maoyu Li, Cui Li, Jiaquan Qu, Xingming Deng, and Zhiqiang Xiao. The function and significance of selenbp1 downregulation in human bronchial epithelial carcinogenic process. *PLOS ONE*, 8(8):1–9, 08 2013. doi: 10.1371/journal.pone.0071865.

Ping Zhan, Bin Zhang, GuangMin Xi, Ying Wu, HongBing Liu, YaFang Liu, WuJian Xu, QingQing Zhu, Feng Cai, ZeJun Zhou, YingYing Miu, XiaoXia Wang, JiaJia Jin, Qian Li, LiPing Qian, TangFeng Lv, and Yong Song. PRC1 contributes to tumorigenesis of lung adenocarcinoma in association with the wnt/β-catenin signaling pathway. *Molecular Cancer*, 16(1):108, 2017. doi: 10.1186/s12943-017-0682-z.

Bin Zhang and Steve Horvath. A general framework for weighted gene co–expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 2005. doi: 10.2202/1544-6115.1128.

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

Hao Helen Zhang, Guang Cheng, and Yufeng Liu. Linear or nonlinear? automatic structure discovery for partially linear models. *Journal of the American Statistical Association*, 106 (495):1099–1112, 2011. doi: 10.1198/jasa.2011.tm10281. PMID: 22121305.

Hongmei Zhang, Xianzheng Huang, Jianjun Gan, Wilfried Karmaus, and Tara Sabo-Attwood. A two-component g -prior for variable selection. *Bayesian Anal.*, 11(2):353–380, 06 2016. doi: 10.1214/15-BA953. URL https://doi.org/10.1214/15-BA953.

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

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

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

Fei Zhou, Jie Ren, Xiaoxi Li, Cen Wu, and Yu Jiang. *interep: Interaction Analysis of Repeated Measure Data*, 2019b. URL https://CRAN.R-project.org/package=interep. R package version 0.3.0.

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

Fei Zhou, Jie Ren, Shuangge Ma, Yu Jiang, and Cen Wu. Sparse group variable selection for gene–environment interaction in longitudinal studies. 2020b.

Hongxiao Zhu, Marina Vannucci, and Dennis D. Cox. A bayesian hierarchical model for classification with selection of functional predictors. *Biometrics*, 66(2):463–473, 2010.

Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006. doi: 10.1198/016214506000000735.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(2):301–320, 2005.

Appendix A

Appendices for Chapter 2

A.1 Additional simulation results

Table A.1: *Simulation for gene expression data* $(n, p) = (300, 505)$*. 50 genes have nonzero regression coefficients. 5 clinical covariates are not subject to selection. The gene expressions have Banded.1 (upper panel) or Banded.2 structure (lower panel) with* $\rho = 0.5$*. mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		LAD_Network	LAD_MCP	LAD_LASSO	Network	MCP	LASSO
			Banded.1	$\rho = 0.5$			
Error1	TP	41.30(6.26)	36.87(3.16)	32.80(6.73)	35.37(4.41)	27.93(2.35)	45.30(2.38)
	FP	13.43(10.74)	8.57(8.39)	105.53(45.21)	8.37(3.65)	8.63(6.67)	85.40(10.19)
Error2	TP	24.93(9.84)	19.77(15.93)	26.93(12.59)	2.00(5.55)	2.47(8.90)	3.17(5.34)
	FP	23.00(24.92)	79.87(124.43)	105.27(67.09)	11.93(44.98)	17.97(79.54)	14.10(33.67)
Error3	TP	39.63(9.88)	32.00(4.88)	33.70(6.93)	18.77(10.76)	16.87(9.36)	30.17(14.13)
	FP	11.80(9.28)	17.97(31.91)	111.17(46.28)	9.40(6.98)	7.03(4.06)	60.30(32.13)
Error4	TP	36.03(11.87)	30.43(5.37)	30.33(7.91)	12.97(11.78)	11.40(10.67)	22.80(16.40)
	FP	11.90(10.46)	21.57(29.67)	109.40(48.10)	8.23(8.18)	6.73(7.25)	52.83(36.92)
			Banded.2	$\rho = 0.5$			
Error1	TP	43.63(7.89)	42.43(3.22)	33.53(9.72)	34.20(4.34)	25.20(4.13)	47.40(1.69)
	FP	8.07(7.72)	10.83(14.28)	101.17(48.16)	9.17(4.89)	10.07(4.64)	80.93(12.3)
Error2	TP	31.77(11.31)	26.60(10.75)	33.33(5.57)	3.43(8.64)	3.27(7.98)	6.27(8.48)
	FP	24.73(43.87)	100.50(119.98)	129.90(45.73)	17.00(69.35)	15.03(61.68)	19.40(24.64)
Error3	TP	41.40(8.58)	35.50(5.47)	33.50(7.66)	15.17(11.58)	12.37(8.46)	29.93(15.21)
	FP	12.97(27.18)	10.80(17.94)	114.40(54.39)	8.80(8.26)	7.33(5.57)	61.27(30.46)
Error4	TP	37.57(11.66)	32.87(5.69)	34.03(7.3)	14.87(13.23)	11.83(10.39)	26.33(16.11)
	FP	9.60(8.62)	19.00(31.61)	113.30(50.56)	9.30(9.34)	7.97(7.45)	55.17(33.04)

Table A.2: *Simulation for gene expression data using correlations calculated from LUSC data.* $(n, p) = (300, 505)$*. 50 genes have nonzero regression coefficients. 5 clinical covariates are not subject to selection. mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		LAD_Network	LAD_MCP	LAD_LASSO	Network	MCP	LASSO
Error1	TP	46.88(3.61)	46.29(2.89)	40.47(6.29)	45.66(2.21)	43.42(2.38)	48.09(1.18)
	FP	1.28(2.00)	3.83(4.83)	1.97(2.69)	1.44(1.92)	1.58(2.20)	13.27(3.66)
Error2	TP	33.65(6.54)	32.59(8.21)	33.69(5.54)	11.02(7.35)	18.25(7.12)	18.14(8.35)
	FP	9.92(23.64)	28.03(45.27)	13.27(14.75)	23.94(52.35)	73.50(71.82)	17.14(21.58)
Error3	TP	43.00(6.79)	41.63(5.95)	41.15(6.03)	30.24(11.93)	28.81(8.41)	39.35(9.46)
	FP	3.14(3.91)	5.58(6.60)	3.85(6.80)	5.52(12.42)	12.43(29.37)	21.83(11.31)
Error4	TP	40.91(6.53)	40.02(5.67)	39.72(5.82)	22.99(11.13)	23.39(8.16)	35.02(9.72)
	FP	3.07(3.41)	8.60(10.37)	3.12(6.07)	7.07(22.44)	18.22(37.61)	26.29(25.11)

Table A.3: *Simulation for gene expression data using correlations calculated from NSCLC data.* $(n, p) = (300, 505)$*. 50 genes have nonzero regression coefficients. 5 clinical covariates are not subject to selection. mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		LAD_Network	LAD_MCP	LAD_LASSO	Network	MCP	LASSO
Error1	TP	43.10(6.57)	43.32(3.41)	41.01(6.91)	47.12(3.08)	39.67(3.12)	45.91(1.65)
	FP	1.40(2.00)	3.02(3.77)	5.68(9.65)	0.83(1.44)	1.44(2.27)	23.01(12.09)
Error2	TP	36.44(9.83)	33.73(6.75)	34.46(5.28)	25.23(6.41)	11.27(3.03)	19.72(5.92)
	FP	13.80(25.85)	34.36(55.80)	23.14(42.02)	43.94(34.98)	15.86(18.95)	14.23(8.96)
Error3	TP	41.51(8.41)	39.23(6.68)	38.17(6.52)	38.20(10.43)	24.93(9.19)	35.45(8.77)
	FP	4.75(6.41)	11.10(12.24)	13.47(20.48)	13.66(37.16)	8.64(34.40)	22.71(21.76)
Error4	TP	42.13(7.50)	38.58(6.86)	39.72(6.14)	34.56(10.60)	19.90(9.03)	30.23(9.58)
	FP	7.03(9.27)	13.33(25.09)	16.63(28.77)	30.02(72.28)	13.17(42.50)	19.57(27.40)

Table A.4: *Simulation for SNP data* $(n, p) = (300, 505)$ *under AR structures. 50 genes have nonzero regression coefficients. 5 clinical covariates are not subject to selection. The SNPs have AR structure with* $\rho = 0.5$ *(upper panel) and* $\rho = 0.8$ *(lower panel). mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		LAD_Network	LAD_MCP	LAD_LASSO	Network	MCP	LASSO
				AR $\rho = 0.5$			
Error1	TP	39.97(9.05)	39.77(3.54)	34.50(6.24)	33.50(6.60)	29.83(6.88)	46.03(2.30)
	FP	8.83(5.80)	9.33(7.71)	106.03(37.94)	9.13(6.26)	8.77(5.17)	84.53(11.25)
Error2	TP	23.23(10.57)	22.43(10.93)	27.43(8.64)	5.17(9.41)	6.93(14.32)	6.10(8.79)
	FP	23.93(40.56)	57.97(95.48)	110.17(66.55)	32.03(86.43)	45.67(126.72)	23.67(39.83)
Error3	TP	37.03(10.07)	36.17(4.53)	34.30(6.39)	16.17(11.50)	15.07(10.51)	28.30(16.74)
	FP	9.00(6.44)	15.93(26.06)	107.77(43.29)	8.50(5.76)	8.30(5.45)	58.40(34.86)
Error4	TP	38.13(7.35)	34.17(6.01)	34.90(7.04)	10.63(10.55)	9.83(9.89)	19.90(17.49)
	FP	10.17(6.86)	27.17(41.85)	109.70(45.23)	6.73(6.10)	6.07(6.49)	40.63(41.16)
				AR $\rho = 0.8$			
Error1	TP	46.87(5.11)	45.30(2.87)	42.87(6.06)	48.70(1.42)	26.07(3.49)	48.73(1.01)
	FP	3.80(2.96)	3.47(2.66)	104.20(30.42)	10.00(7.89)	6.67(3.39)	60.00(12.99)
Error2	TP	38.80(10.71)	26.17(6.93)	35.93(4.16)	5.97(10.07)	6.10(12.41)	5.70(7.53)
	FP	12.20(8.02)	44.20(83.64)	105.40(32.2)	35.30(89.14)	43.57(111.1)	25.77(42.43)
Error3	TP	46.70(6.02)	39.90(5.94)	43.17(3.65)	33.23(16.9)	18.10(9.78)	33.97(15.77)
	FP	6.77(5.88)	6.10(10.6)	105.63(28.03)	18.97(16.06)	20.83(65.59)	60.93(29.47)
Error4	TP	42.87(10.92)	36.03(6.83)	42.17(4.47)	25.93(16.99)	13.43(8.39)	29.33(15.39)
	FP	5.67(4.56)	10.30(21.71)	120.90(48.61)	19.77(13.39)	7.57(6.82)	50.30(31.1)

Table A.5: *Simulation for SNP data (n, p) = (300, 505) under banded structures. 50 genes have nonzero regression coefficients. 5 clinical covariates are not subject to selection. The SNPs have Banded.1 (upper panel) or Banded.2 structure (lower panel) with $\rho = 0.5$. mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		LAD_Network	LAD_MCP	LAD_LASSO	Network	MCP	LASSO
			Banded.1	$\rho = 0.5$			
Error1	TP	39.17(9.57)	36.87(8.19)	30.13(8.40)	31.80(3.95)	26.57(4.39)	43.97(3.18)
	FP	9.10(5.92)	13.53(21.54)	91.63(43.74)	8.27(5.30)	8.53(6.22)	81.93(12.6)
Error2	TP	19.17(9.70)	18.30(13.42)	25.43(8.69)	2.83(7.39)	3.60(9.97)	3.17(5.17)
	FP	14.93(13.16)	57.30(93.44)	97.63(51.97)	17.30(60.78)	24.20(89.50)	13.10(27.14)
Error3	TP	33.33(10.57)	32.47(4.61)	29.97(8.37)	14.60(13.06)	12.83(11.25)	24.40(17.98)
	FP	8.00(4.61)	18.57(38.32)	101.23(52.47)	8.13(7.21)	6.63(6.19)	53.10(37.71)
Error4	TP	33.53(13.54)	29.43(8.05)	31.00(6.77)	11.43(8.63)	10.83(7.73)	24.40(12.76)
	FP	16.37(12.94)	18.30(36.76)	108.90(39.14)	7.07(7.91)	6.53(5.78)	56.57(30.96)
			Banded.2	$\rho = 0.5$			
Error1	TP	41.23(8.78)	40.57(3.30)	34.23(7.57)	32.23(6.74)	27.77(5.10)	45.93(1.76)
	FP	10.83(7.68)	9.17(7.45)	113.63(46.59)	9.63(6.7)	9.30(4.23)	84.00(12.91)
Error2	TP	21.10(7.69)	20.53(13.05)	27.27(8.2)	1.60(1.96)	1.63(2.11)	3.03(4.19)
	FP	17.27(10.43)	76.43(122.77)	93.57(49.76)	5.10(10.5)	5.00(10.3)	11.40(15.40)
Error3	TP	35.60(9.11)	34.90(4.48)	34.20(5.11)	15.30(10.17)	14.37(9.68)	28.67(16.14)
	FP	10.60(7.02)	15.63(38.21)	117.63(38.53)	9.80(7.23)	9.50(7.10)	66.77(37.00)
Error4	TP	38.50(8.44)	36.47(5.17)	33.20(7.07)	12.43(10.86)	11.80(10.26)	21.80(18.26)
	FP	18.77(13.23)	41.57(69.17)	109.73(42.16)	11.13(26.16)	12.90(30.37)	45.90(38.12)

Table A.6: *Simulation for SNP data based on the linkage disequilibrium (LD) structure. (n, p) = (300, 505). 50 genes have nonzero regression coefficients. 5 clinical covariates are not subject to selection. mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		LAD_Network	LAD_MCP	LAD_LASSO	Network	MCP	LASSO
Error1	TP	46.47(4.62)	42.94(5.09)	43.10(3.54)	46.25(2.17)	44.94(2.62)	45.93(2.03)
	FP	4.30(6.29)	9.43(16.91)	28.10(13.34)	2.59(3.12)	2.85(4.23)	21.52(5.25)
Error2	TP	38.22(7.42)	34.44(7.76)	27.45(4.70)	23.90(5.28)	12.15(9.04)	10.34(9.56)
	FP	18.84(18.76)	46.88(69.94)	49.31(16.43)	95.05(59.21)	36.10(82.64)	20.73(22.59)
Error3	TP	45.38(4.71)	40.16(5.59)	39.12(5.01)	26.16(15.89)	27.16(13.09)	33.03(13.71)
	FP	5.85(6.66)	11.17(23.56)	35.25(12.73)	10.22(32.97)	25.37(77.51)	36.70(18.09)
Error4	TP	42.65(6.21)	39.28(5.35)	36.40(5.03)	21.66(15.24)	25.14(12.31)	28.30(14.80)
	FP	5.99(5.63)	17.00(31.33)	39.82(14.83)	13.95(38.37)	44.58(108.55)	36.91(21.17)

A.2 Biological similarity analysis

GO enrichment analysis was conducted using the R package GOSim. GO biological processes that are associated with identified genes are divided into four categories: positive regulation (P), negative regulation (N), regulation (R, without a well-defined direction) and other (O). The proportions of genes that involve in the four categories of processes are computed for each methods. The results are provided in Figure A.1.

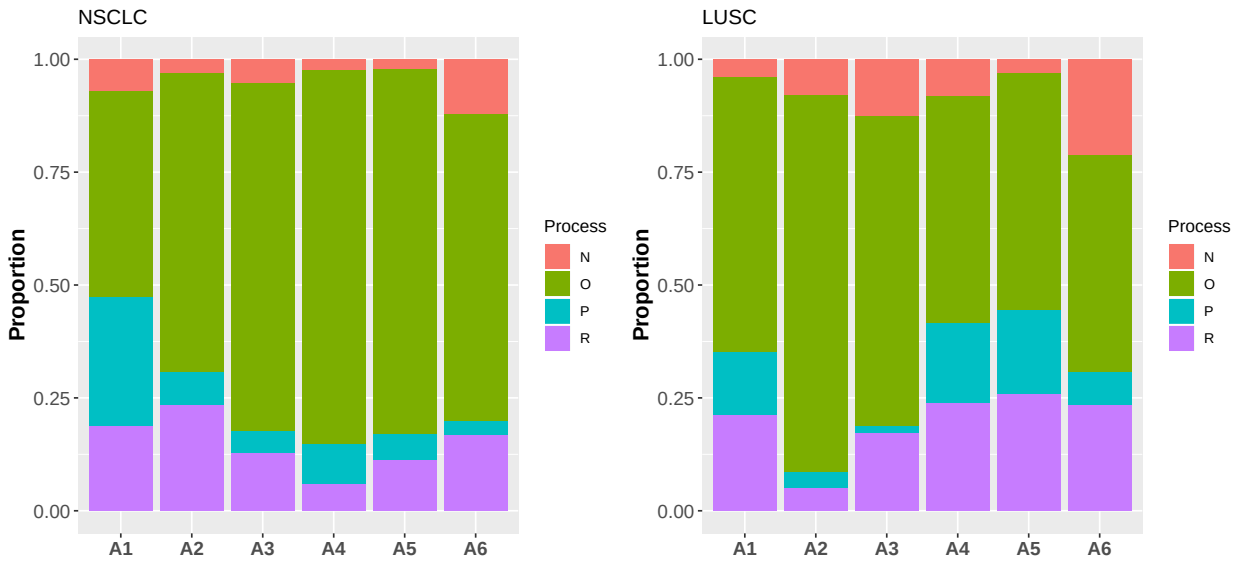

Figure A.1: *Gene Ontology (GO) analysis: proportions of genes that have the four categories of processes with different approaches. Left: NSCLC data. Right: LUSC data. A1: LAD_Network. A4: LAD_MCP. A3: LAD_LASSO. A4: Network. A5: MCP. A6: LASSO.*

Appendix B

Appendices for Chapter 3

B.1 Hyper-parameters sensitivity analysis

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

We also evaluate the sensitivity of BSSVC-SI to the choice of the Gamma hyperpriors on λ_v, λ_c and λ_e. We test the shape parameter of the Gamma prior for five different values: $\{0.1, 0.5, 1, 2, 5\}$. This ranges from highly skewed exponential shape to highly diffuse unimodal shape. We fix the rate parameter at $\{1, 2, 5\}$ and test different combinations of shape and rate parameters on a two-dimensional grid. In Table B.2, we show the simulation results of some representative cases under the scenarios of Example 2. BSSVC-SI model has

stable performance with high TP and low FP for different Gamma priors. Similar patterns are observed for all other cases. In this study, we use Gamma(1, 1) for λ_v, λ_c and λ_e under all scenarios.

Table B.1: *Sensitivity analysis.* (n, p, q) *= (500, 100, 2). mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		BSSVC-SI			BSSVC		
		Varying	Constant	Nonzero	Varying	Constant	Nonzero
Beta(0.5, 0.5)	TP	3.00(0.00)	5.00(0.00)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.07(0.25)	0.00(0.00)	0.00(0.00)	5.07(0.25)	0.00(0.00)	0.03(0.18)
Beta(1, 1)	TP	3.00(0.00)	5.00(0.00)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.07(0.25)	0.00(0.00)	0.03(0.18)	5.00(0.00)	0.00(0.00)	0.10(0.31)
Beta(2, 2)	TP	3.00(0.00)	4.93(0.25)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.20(0.48)	0.00(0.00)	0.00(0.00)	4.97(0.18)	0.00(0.00)	0.10(0.31)
Beta(1, 5)	TP	3.00(0.00)	4.97(0.18)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.17(0.46)	0.00(0.00)	0.00(0.00)	5.00(0.00)	0.00(0.00)	0.03(0.18)
Beta(5, 1)	TP	3.00(0.00)	5.00(0.00)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.27(0.52)	0.07(0.25)	0.03(0.18)	5.07(0.25)	0.00(0.00)	0.27(0.58)

Table B.2: *Sensitivity analysis.* (n, p, q) *= (500, 100, 2). mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		BSSVC-SI			BSSVC		
		Varying	Constant	Nonzero	Varying	Constant	Nonzero
Gamma(0.1, 1)	TP	3.00(0.00)	4.93(0.26)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.20(0.41)	0.07(0.26)	0.00(0.00)	5.00(0.00)	0.00(0.00)	0.07(0.26)
Gamma(0.5, 2)	TP	3.00(0.00)	5.00(0.00)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.07(0.26)	0.00(0.00)	0.00(0.00)	5.00(0.00)	0.00(0.00)	0.00(0.00)
Gamma(1, 1)	TP	3.00(0.00)	5.00(0.00)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.07(0.25)	0.00(0.00)	0.03(0.18)	5.00(0.00)	0.00(0.00)	0.10(0.31)
Gamma(1, 5)	TP	3.00(0.00)	4.93(0.26)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.07(0.26)	0.00(0.00)	0.07(0.26)	5.00(0.00)	0.00(0.00)	0.07(0.26)
Gamma(2, 5)	TP	3.00(0.00)	4.93(0.26)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.13(0.35)	0.07(0.26)	0.00(0.00)	4.93(0.26)	0.00(0.00)	0.00(0.00)
Gamma(5, 1)	TP	3.00(0.00)	5.00(0.00)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.20(0.41)	0.00(0.00)	0.00(0.00)	5.07(0.26)	0.00(0.00)	0.20(0.41)

B.2 FDR-based variable selection

The FDR-based variable selection selects variables with posterior inclusion probability larger than certain threshold that is chosen to control the overall Bayesian FDR rate. Specifically, a threshold ψ_α can be determined for flagging the set of predictors $\Psi_{\psi_\alpha} = \{j : p_j > \psi_\alpha\}$ as significant, for a given overall FDR bound $\alpha \in (0, 1)$. The threshold ψ_α is chosen in a way that we expect less than $100\alpha\%$ of the predictors in set Ψ_{ψ_α} are false positive on average. To compute threshold ψ_α, we first sort the $\{p_j, j = 1, \ldots, p\}$ in descending order to obtain $\{p_{(j)}, j = 1, \ldots, p\}$. To compute threshold ψ_α, we first sort the $\{p_j, j = 1, \ldots, p\}$ in descending order to obtain $\{p_{(j)}, j = 1, \ldots, p\}$. Then $\psi_\alpha = p_{(\xi)}$ with $\xi = \max\{j^* : \frac{1}{j^*}\sum_{j=1}^{j^*}(1 - p_{(j)}) \leqslant \alpha\}$. The $(1 - p_j)$ is interpreted as the estimate of the local FDR (Storey (2003)) that measures the probability of including the jth predictor when the jth predictor is not in the true model.

Table B.3: *Simulation results for FDR-based variable selection.* $(n, p, q) = (500, 100, 2)$. *mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		BSSVC-SI			BSSVC		
		Varying	Constant	Nonzero	Varying	Constant	Nonzero
Example 1	TP	3.00(0.00)	4.93(0.25)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.17(0.38)	0.07(0.25)	0.00(0.00)	5.20(0.41)	0.00(0.00)	0.13(0.35)
Example 2	TP	3.00(0.00)	5.00(0.00)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.10(0.31)	0.10(0.31)	0.03(0.18)	5.23(0.50)	0.00(0.00)	0.13(0.35)
Example 3	TP	2.97(0.18)	4.97(0.18)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.03(0.18)	0.17(0.46)	0.00(0.00)	5.10(0.31)	0.00(0.00)	0.10(0.31)
Example 4	TP	3.00(0.00)	5.00(0.00)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	0.07(0.25)	0.00(0.00)	0.07(0.25)	5.13(0.35)	0.00(0.00)	0.20(0.41)

B.3 Variable selection based on 95% credible interval

Alternatives BVC-SI and BVC lack for the variable selection property. In order to create sparsity on the coefficients estimated by these three methods, we consider a 95% credible interval based method used in Li et al. (2015). Specifically, a varying effect is included in the final model if at least one of its spline coefficients has a two-sided 95% credible interval that does not cover zero. Similarly, a constant effect is included in the final model if two-sided 95% credible interval od its spline coefficient does not cover zero. The same rule applies to the linear interaction effects. The results are tabulated in Table B.4.

Table B.4: *Simulation results.* (n, p, q) *= (500, 100, 2). mean(sd) of true positives (TP) and false positives (FP) based on 100 replicates.*

		BVC-SI			BVC		
		Varying	Constant	Nonzero	Varying	Constant	Nonzero
Example 1	TP	2.98(0.15)	4.73(0.45)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	1.89(1.40)	0.42(0.69)	4.07(2.27)	6.13(1.18)	0.00(0.00)	3.16(2.02)
Example 2	TP	3.00(0.00)	4.76(0.48)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	3.27(2.38)	0.36(0.57)	5.13(2.32)	6.78(1.52)	0.00(0.00)	4.20(2.21)
Example 3	TP	3.00(0.00)	4.78(0.42)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	2.09(1.86)	0.24(0.53)	4.33(2.32)	6.04(1.30)	0.00(0.00)	3.42(2.11)
Example 4	TP	3.00(0.00)	4.78(0.52)	5.00(0.00)	3.00(0.00)	0.00(0.00)	5.00(0.00)
	FP	3.33(1.98)	0.24(0.43)	6.47(2.66)	6.51(1.36)	0.00(0.00)	5.07(2.61)

B.4 Sensitivity analysis on smoothness specification

Let O denotes the degree of B spline basis, and K denotes the number of interior knots. Huang et al. (2002, 2004) show that $n^{1/(2O+3)}$ is the optimal order of the number of spline knots K. For quadratic and cubic splines corresponding to $O = 2$ and 3 respectively, We conduct a sensitivity analysis for the proposed model under the setting of Example 2 for $K \in [1, 4]$. Table B.5 shows that $K = 1$ leads to unsatisfactory performance, especially for prediction. When $K \geq 2$, different values of K lead to similar performance under $O = 2$ and $O = 3$. This suggests the model performance is insensitive with respect to the smoothness specification. We conduct a sensitivity analysis for the proposed model under the setting of Example 2 given $K \in [1, 4]$,

Table B.5: *Sensitivity analysis on smoothness specification.* $(n, p, q) = (500, 100, 2)$. *mean(sd) of true positives (TP), false positives (FP) and prediction error based on 100 replicates.*

$O = 2$	Varying		Constant		Nonzero		
K	TP	FP	TP	FP	TP	FP	Pred.Error
1	2.97(0.18)	0.20(0.55)	4.87(0.35)	0.10(0.31)	4.97(0.18)	0.03(0.18)	1.998(0.152)
2	3.00(0.00)	0.03(0.18)	5.00(0.00)	0.00(0.00)	5.00(0.00)	0.00(0.00)	1.172(0.071)
3	3.00(0.00)	0.00(0.00)	5.00(0.00)	0.03(0.18)	5.00(0.00)	0.13(0.43)	1.140(0.093)
4	3.00(0.00)	0.07(0.25)	4.93(0.25)	0.00(0.00)	5.00(0.00)	0.00(0.00)	1.200(0.118)
$O = 3$	Varying		Constant		Nonzero		
K	TP	FP	TP	FP	TP	FP	Pred.Error
1	3.00(0.00)	0.17(0.38)	4.87(0.35)	0.03(0.18)	5.00(0.00)	0.00(0.00)	1.676(0.159)
2	3.00(0.00)	0.00(0.00)	5.00(0.00)	0.03(0.18)	5.00(0.00)	0.03(0.18)	1.089(0.054)
3	3.00(0.00)	0.00(0.00)	4.97(0.18)	0.03(0.18)	5.00(0.00)	0.00(0.00)	1.185(0.072)
4	3.00(0.00)	0.00(0.00)	5.00(0.00)	0.03(0.18)	5.00(0.00)	0.03(0.18)	1.156(0.078)

B.5 Additional simulation results

Table B.6: *Simulation results in Example 2. SNP genotype data* (n, p, q) *= (500, 100, 2). mean(sd) of the integrated median mean squared error (IMSE), median mean squared error (MSE), total squared errors for all estimates and prediction errors based on 100 replicates.*

	BSSVC-SI	BSSVC	BVC-SI	BVC	BL
IMSE					
$\mu(Z)$	0.043(0.013)	0.043(0.012)	0.055(0.024)	0.055(0.022)	200.857(3.626)
$\beta_1(Z)$	0.042(0.020)	0.021(0.012)	0.069(0.031)	0.085(0.031)	181.351(4.523)
$\beta_2(Z)$	0.027(0.018)	0.021(0.012)	0.044(0.025)	0.049(0.025)	25.653(10.366)
$\beta_3(Z)$	0.030(0.026)	0.026(0.022)	0.074(0.034)	0.094(0.038)	136.250(5.223)
MSE					
α_1	0.011(0.012)	0.012(0.013)	0.022(0.023)	0.022(0.022)	0.065(0.083)
α_2	0.003(0.003)	0.003(0.004)	0.007(0.009)	0.007(0.008)	0.023(0.033)
ζ_0	0.033(0.025)	0.024(0.019)	0.081(0.057)	0.106(0.062)	135.875(3.861)
ζ_1	0.005(0.005)	0.006(0.007)	0.009(0.013)	0.008(0.013)	0.025(0.036)
ζ_2	0.008(0.009)	0.006(0.008)	0.019(0.023)	0.019(0.022)	0.065(0.104)
ζ_3	0.009(0.015)	0.009(0.013)	0.017(0.023)	0.019(0.025)	0.065(0.070)
ζ_4	0.009(0.014)	0.011(0.019)	0.011(0.015)	0.010(0.014)	0.045(0.080)
ζ_5	0.006(0.007)	0.006(0.008)	0.020(0.028)	0.024(0.030)	0.065(0.086)
Total	0.227(0.083)	0.253(0.104)	2.020(0.260)	1.931(0.228)	548.430(13.003)
Pred. Error	1.160(0.071)	1.169(0.064)	2.196(0.180)	2.155(0.154)	7.870(0.445)

Table B.7: *Simulation results in Example 3. SNP genotype data based on the linkage disequilibrium (LD) structure (n, p, q) = (500, 100, 2). mean(sd) of the integrated median mean squared error (IMSE), median mean squared error (MSE), total squared errors for all estimates and prediction errors based on 100 replicates.*

	BSSVC-SI	BSSVC	BVC-SI	BVC	BL
IMSE					
$\mu(Z)$	0.046(0.015)	0.045(0.014)	0.060(0.023)	0.058(0.022)	200.784(3.630)
$\beta_1(Z)$	0.059(0.025)	0.025(0.013)	0.112(0.045)	0.120(0.045)	182.705(6.130)
$\beta_2(Z)$	0.035(0.018)	0.023(0.017)	0.051(0.024)	0.054(0.025)	28.617(12.114)
$\beta_3(Z)$	0.032(0.019)	0.027(0.018)	0.083(0.049)	0.105(0.051)	142.424(15.015)
MSE					
α_1	0.003(0.005)	0.003(0.004)	0.006(0.009)	0.006(0.009)	0.017(0.020)
α_2	0.005(0.006)	0.004(0.006)	0.010(0.015)	0.009(0.013)	0.024(0.026)
ζ_0	0.010(0.013)	0.008(0.011)	0.024(0.034)	0.023(0.033)	0.075(0.095)
ζ_1	0.008(0.014)	0.008(0.015)	0.012(0.016)	0.011(0.014)	0.051(0.093)
ζ_2	0.010(0.014)	0.008(0.012)	0.024(0.035)	0.025(0.035)	0.088(0.092)
ζ_3	0.009(0.008)	0.010(0.009)	0.024(0.034)	0.024(0.032)	0.102(0.114)
ζ_4	0.013(0.017)	0.022(0.022)	0.026(0.034)	0.023(0.030)	0.064(0.090)
ζ_5	0.017(0.026)	0.038(0.034)	0.032(0.039)	0.030(0.036)	0.050(0.065)
Total	0.307(0.107)	0.407(0.141)	2.176(0.219)	2.015(0.207)	559.260(24.011)
Pred. Error	1.203(0.064)	1.209(0.068)	2.164(0.137)	2.088(0.132)	7.683(0.424)

Table B.8: *Simulation results in Example 4. SNP genotype from T2D data* (n, p, q) = *(500, 100, 2). mean(sd) of the integrated median mean squared error (IMSE), median mean squared error (MSE), total squared errors for all estimates and prediction errors based on 100 replicates.*

	BSSVC-SI	BSSVC	BVC-SI	BVC	BL
IMSE					
$\mu(Z)$	0.051(0.019)	0.051(0.019)	0.066(0.021)	0.064(0.020)	202.409(4.069)
$\beta_1(Z)$	0.032(0.015)	0.018(0.011)	0.052(0.027)	0.068(0.030)	181.747(6.418)
$\beta_2(Z)$	0.015(0.010)	0.014(0.009)	0.029(0.021)	0.033(0.020)	24.012(3.920)
$\beta_3(Z)$	0.023(0.018)	0.019(0.013)	0.051(0.027)	0.066(0.030)	137.823(6.639)
MSE					
α_1	0.003(0.003)	0.003(0.004)	0.007(0.013)	0.007(0.013)	0.021(0.031)
α_2	0.003(0.004)	0.003(0.005)	0.005(0.005)	0.005(0.004)	0.013(0.016)
ζ_0	0.007(0.014)	0.007(0.016)	0.015(0.015)	0.014(0.014)	0.098(0.157)
ζ_1	0.004(0.006)	0.004(0.005)	0.008(0.013)	0.008(0.011)	0.050(0.086)
ζ_2	0.009(0.009)	0.006(0.007)	0.019(0.022)	0.018(0.020)	0.030(0.044)
ζ_3	0.007(0.010)	0.007(0.009)	0.012(0.018)	0.012(0.021)	0.040(0.057)
ζ_4	0.004(0.004)	0.004(0.004)	0.006(0.008)	0.006(0.008)	0.026(0.033)
ζ_5	0.003(0.003)	0.003(0.004)	0.013(0.014)	0.014(0.014)	0.043(0.059)
Total	0.178(0.052)	0.194(0.049)	1.751(0.194)	1.648(0.157)	550.070(12.987)
Pred. Error	1.141(0.073)	1.147(0.064)	2.164(0.134)	2.109(0.125)	8.575(0.458)

B.6 Computational cost

Table B.9: *Computational cost analysis for BSSVC-SI under the setting of Example 1. p: number of genes. time: CPU time (in seconds) for 10,000 MCMC iterations. The number of regression coefficients to be estimated after basis expansion is approximately $q_n p + p$, where q_n is the number of basis function. In this study, $q_n = 5$.*

$n = 500$		$n = 1500$		$n = 3000$	
# of genes	time	# of genes	time	# of genes	time
$p = 100$	11.707	$p = 300$	121.396	$p = 600$	552.043
$p = 200$	24.878	$p = 600$	236.571	$p = 900$	834.645
$p = 300$	36.372	$p = 900$	341.366	$p = 1200$	988.939

B.7 The estimated varying coefficient functions

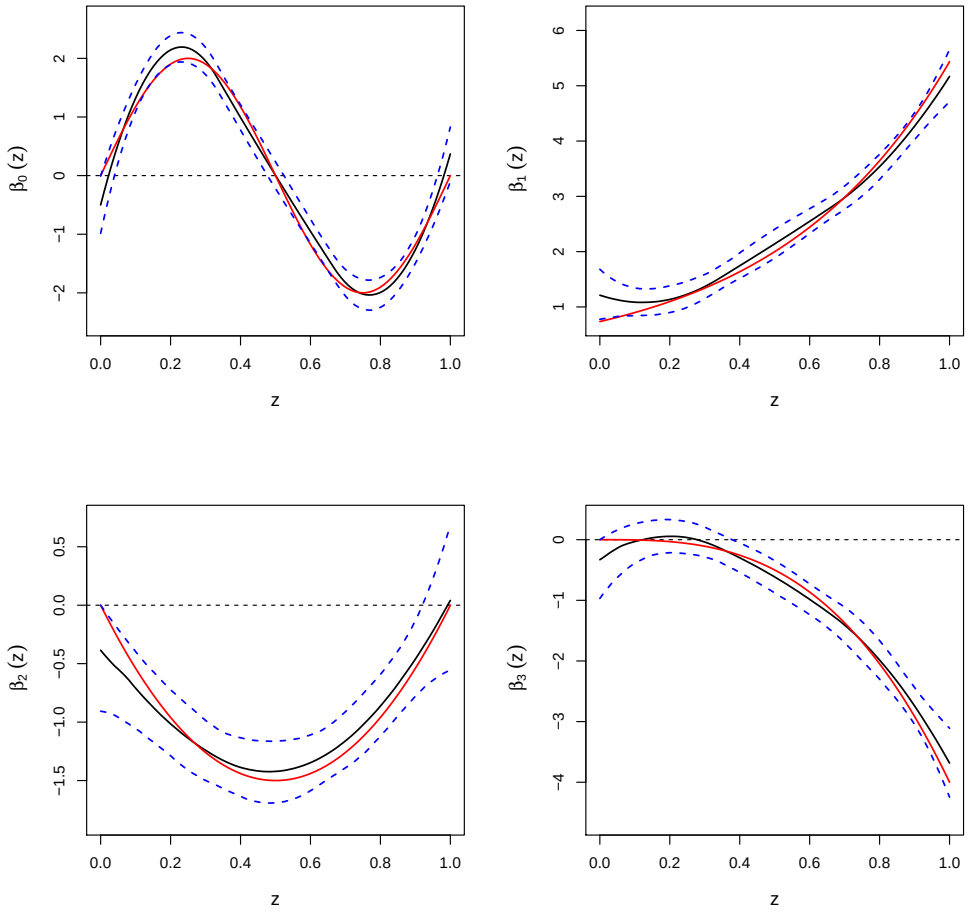

Figure B.1: *Simulation study in Example 1 for the proposed method (BSSVC-SI). Red line: true parameter values. Black line: median estimates of varying coefficients for BSSVC-SI. Blue dashed lines: 95% credible intervals for the estimated varying coefficients.*

B.8 Assessment of the convergence of MCMC chains

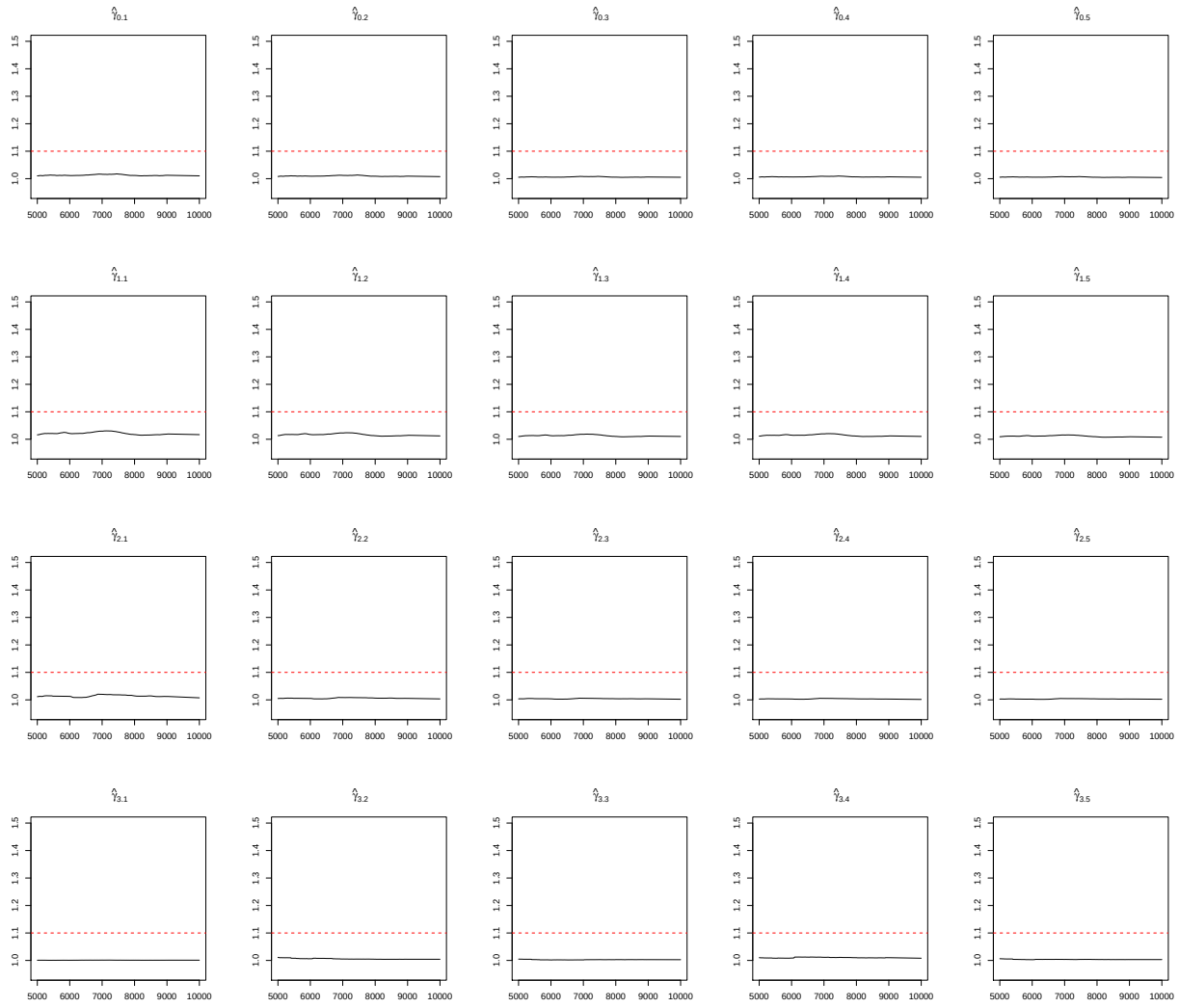

Figure B.2: *Potential scale reduction factor (PSRF) against iterations for varying coefficient functions in Figure B.1. Black line: the PSRF. Red line: the threshold of 1.1. The* $\hat{\gamma}_{j1}$ *to* $\hat{\gamma}_{j5}$, $(j = 0, \ldots, 3)$, *represent the five estimated spline coefficients for the varying coefficient function* β_j, *respectively.*

B.9 Additional results for real data analysis

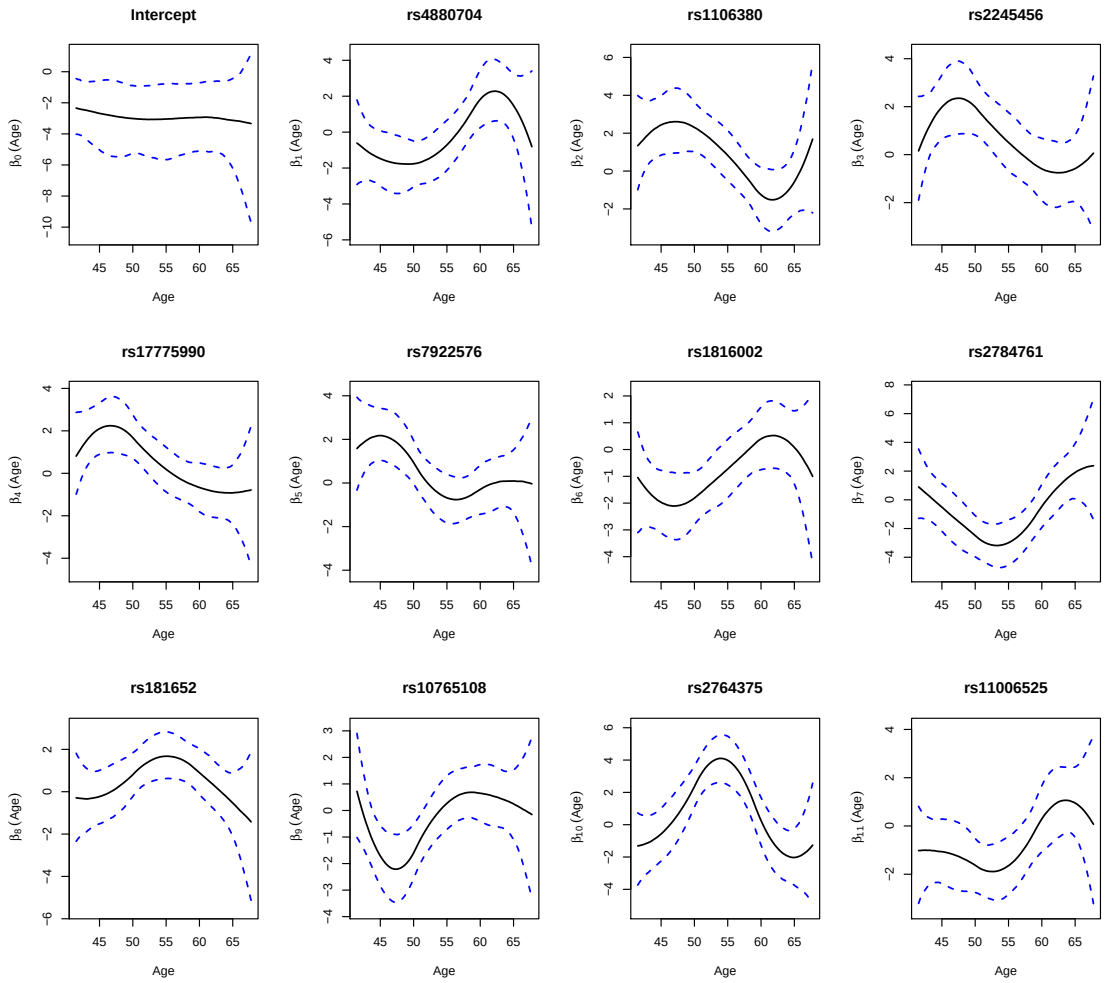

Figure B.3: *Real data analysis for the proposed method (BSSVC-SI). Black line: median estimates of varying coefficients for BSSVC-SI. Blue dashed lines: 95% credible intervals for the estimated varying coefficients.*

B.10 Posterior inference for the BSSVC-SI method

B.10.1 Priors

$$Y|\eta, \gamma_{11}, \ldots, \gamma_{p1}, \gamma_{1*}, \ldots, \gamma_{p*}, \alpha_1, \ldots, \alpha_q, \zeta_0, \zeta_1, \ldots, \zeta_p, \sigma^2$$
$$\propto (\sigma^2)^{-\frac{n}{2}} \exp\Big\{ -\frac{1}{2\sigma^2}(Y-\mu)^\top (Y-\mu)\Big\}$$

$$\eta \sim \mathrm{N}_{q_n}(0,\, \Sigma_{\eta 0})$$

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

$$\zeta_0 \sim \mathrm{N}(0,\, \sigma^2_{\zeta_0})$$

$$\gamma_{j1}|\pi_c, \tau^2_{cj}, \sigma^2 \sim \pi_c \mathrm{N}(0,\, \sigma^2\tau^2_{cj}) + (1-\pi_c)\delta_0(\gamma_{j1}), \quad j = 1, \ldots, p$$

$$\tau^2_{cj}|\lambda_c \sim \frac{\lambda_c^2}{2}\exp(-\frac{\lambda_c^2\tau^2_{cj}}{2}), \quad j = 1, \ldots, p$$

$$\gamma_{j*}|\pi_v, \tau^2_{vj}, \sigma^2 \sim \pi_v \mathrm{N}_L(0,\, \mathrm{Diag}(\sigma^2\tau^2_{vj}, \ldots, \sigma^2\tau^2_{vj})) + (1-\pi_v)\delta_0(\gamma_{j*}), \quad j = 1, \ldots, p$$

$$\tau^2_{vj}|\lambda_v \sim \mathrm{Gamma}(\frac{L+1}{2},\, \frac{L\lambda_v^2}{2}), \quad j = 1, \ldots, p$$

$$\zeta_j|\pi_e, \tau^2_{ej}, \sigma^2 \sim \pi_e \mathrm{N}(0,\, \sigma^2\tau^2_{ej}) + (1-\pi_e)\delta_0(\zeta_j), \quad j = 1, \ldots, p$$

$$\tau^2_{ej}|\lambda_e \sim \frac{\lambda_e^2}{2}\exp(-\frac{\lambda_e^2\tau^2_{ej}}{2}), \quad j = 1, \ldots, p$$

$$\sigma^2 \sim (\sigma^2)^{-s-1}\exp(-\frac{h}{\sigma^2})$$

Consider the following conjugate gamma priors for λ_c^2, λ_v^2 and λ_e^2

$$\lambda_c^2 \sim \mathrm{Gamma}(a_c,\, b_c), \quad \lambda_v^2 \sim \mathrm{Gamma}(a_v,\, b_v) \quad \text{and} \quad \lambda_e^2 \sim \mathrm{Gamma}(a_e,\, b_e)$$

and conjugate beta priors for π_c, π_v and π_e

$$\pi_c \sim \mathrm{Beta}(r_c,\, w_c), \quad \pi_v \sim \mathrm{Beta}(r_v,\, w_v) \quad \text{and} \quad \pi_e \sim \mathrm{Beta}(r_e,\, w_e)$$

B.10.2 Gibbs Sampler

$\pi(\eta|\text{rest})$

$$\propto \pi(\eta)\pi(y|\cdot)$$

$$\propto \exp\Big(-\frac{1}{2}\eta^\top\Sigma_{\eta 0}^{-1}\eta\Big)\exp\Big(-\frac{1}{2\sigma^2}(Y-\mu)^\top(Y-\mu)\Big)$$

$$\propto \exp\left(-\frac{1}{2}\eta^\top\Sigma_{\eta 0}^{-1}\eta-\frac{1}{2\sigma^2}(Y-B_0\eta-\mu_{(-\eta)})^\top(Y-B_0\eta-\mu_{(-\eta)})\right)$$

$$\propto \exp\left(\eta^\top\Sigma_{\eta 0}^{-1}\eta+\frac{1}{\sigma^2}\eta^\top B_0^\top B_0\eta-\frac{2}{\sigma^2}(Y-\mu_{(-\eta)})^\top(B_0\eta)\right)$$

$$\propto \exp\left(\eta^\top\Big(\Sigma_{\eta 0}^{-1}+\frac{1}{\sigma^2}B_0^\top B_0\Big)\eta-\frac{2}{\sigma^2}(Y-\mu_{(-\eta)})^\top B_0\eta\right)$$

where $B_0 = (B_0(Z_1),\ldots,B_0(Z_n))^\top$. Hence, the full conditional distribution of η is multivariate normal with mean

$$\mu_\eta = \Big(\Sigma_{\eta 0}^{-1}+\frac{1}{\sigma^2}B_0^\top B_0\Big)^{-1}\Big(\frac{1}{\sigma^2}(Y-\mu_{(-\eta)})^\top B_0\Big)^\top$$

and variance

$$\Sigma_\eta = \Big(\Sigma_{\eta 0}^{-1}+\frac{1}{\sigma^2}B_0^\top B_0\Big)^{-1}$$

Similarly, the full conditional distribution of α is $N(\mu_\alpha,\Sigma_\alpha)$ with

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

and variance

$$\Sigma_\alpha = \Big(\Sigma_{\alpha 0}^{-1}+\frac{1}{\sigma^2}W^\top W\Big)^{-1}$$

where $W = (W_1,\ldots,W_n)^\top$. And the full conditional distribution of ζ_0 is $N(\mu_{\zeta_0},\Sigma_{\zeta_0})$ with

$$\mu_{\zeta_0} = \Big(1/\sigma_{\zeta_0}^2+\frac{1}{\sigma^2}\sum_{i=1}^n E_i^2\Big)^{-1}\Big(\frac{1}{\sigma^2}\sum_{i=1}^n(y_i-\mu_{(-\zeta_0)})E_i\Big)$$

and variance

$$\Sigma_{\zeta_0} = \Big(1/\sigma^2_{\zeta_0} + \frac{1}{\sigma^2}\sum_{i=1}^{n} E_i^2\Big)^{-1}$$

The full conditional distribution of γ_{j*}

$$\begin{aligned}
&\pi(\gamma_{j*}|\text{rest}) \\
&\quad\propto \pi(\gamma_{j*}|\tau^2_{vj}, \sigma^2)\pi(y|\cdot) \\
&\quad\propto (\sigma^2)^{-\frac{n}{2}} \exp\bigg(-\frac{1}{2\sigma^2}(Y - U_j\gamma_{j*} - \mu_{(-\gamma_{j*})})^\top (Y - U_j\gamma_{j*} - \mu_{(-\gamma_{j*})})\bigg) \\
&\quad\times \bigg(\pi_v(2\pi\sigma^2\tau^2_{vj})^{-\frac{L}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau^2_{vj}}\gamma_{j*}^\top\gamma_{j*}\Big)\mathbf{I}_{\{\gamma_{j*}\neq 0\}} + (1-\pi_v)\delta_0(\gamma_{j*})\bigg)
\end{aligned} \tag{B.1}$$

where $U_j = (U_{1j}, \ldots, U_{nj})^\top$ is a $n \times L$ matrix.

$$\begin{aligned}
l_{vj} &= \pi(\gamma_{j*} \neq 0|\text{rest}) \\
&= \frac{\pi_v}{\pi_v + (1-\pi_v)(\tau^2_{vj})^{\frac{L}{2}}|\Sigma_{\gamma_{j*}}|^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2}\|\Sigma^{\frac{1}{2}}_{\gamma_{j*}} U_j^\top (Y - \mu_{(-\gamma_{j*})})\|_2^2\Big)}
\end{aligned}$$

Hence, the full conditional distribution of γ_{j*} is a spike and slab distribution

$$\gamma_{j*}|\text{rest} \sim l_{vj}\text{N}(\mu_{\gamma_{j*}},\, \sigma^2\Sigma_{\gamma_{j*}}) + (1 - l_{vj})\delta_0(\gamma_{j*})$$

with mean

$$\mu_{\gamma_{j*}} = \Sigma_{\gamma_{j*}} U_j^\top (Y - \mu_{(-\gamma_{j*})})$$

This posterior distribution is a mixture of a multivariate normal and a point mass at 0. To sample from this posterior distribution at the gth iteration, we follow the steps:

- Generate u from Unif[0,1]
- If $u \leq l_{vj}$
  - Generate t from $\text{N}(\mu_{\gamma_{j*}},\, \sigma^2\Sigma_{\gamma_{j*}})$

– set $\gamma_{j*}^{(g)} = t$ and $\phi_{vj}^{(g)} = 1$

- If $u > l_{vj}$

  – set $\gamma_{j*}^{(g)} = 0$ and $\phi_{vj}^{(g)} = 0$

Note that, when we sample $\gamma_{j*}^{(g)}$, we also compute the value of $\phi_{vj}^{(g)}$. The full conditional distribution of γ_{j1} can be expressed as

$$
\begin{aligned}
&\pi(\gamma_{j1}|\text{rest}) \\
&\quad \propto \pi(\gamma_{j1}|\tau_{cj}^2, \sigma^2)\pi(y|\cdot) \\
&\quad \propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(Y - X_j\gamma_{j1} - \mu_{(-\gamma_{j1})})^\top (Y - X_j\gamma_{j1} - \mu_{(-\gamma_{j1})})\right) \\
&\quad \times \left(\pi_c (2\pi\sigma^2\tau_{cj}^2)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{cj}^2}\gamma_{j1}^2\Big)\mathbf{I}_{\{\gamma_{j1}\neq 0\}} + (1-\pi_c)\delta_0(\gamma_{j1})\right)
\end{aligned}
$$

Let $\Sigma_{\gamma_{j1}} = (X_j^\top X_j + \frac{1}{\tau_{cj}^2})^{-1}$, we have

$$
\begin{aligned}
l_{cj} &= \pi(\gamma_{j1} \neq 0|\text{rest}) \\
&= \frac{\pi_c}{\pi_c + (1-\pi_c)(\tau_{cj}^2)^{\frac{1}{2}}(\Sigma_{\gamma_{j1}})^{-\frac{1}{2}} \exp\Big(-\frac{\Sigma_{\gamma_{j1}}}{2\sigma^2}\|(Y - \mu_{(-\gamma_{j1})})^\top X_j\|_2^2\Big)}
\end{aligned}
$$

Hence, the full conditional distribution of γ_{j1} is a spike and slab distribution

$$\gamma_{j1}|\text{rest} \sim l_{cj} N(\mu_{\gamma_{j1}},\, \sigma^2\Sigma_{\gamma_{j1}}) + (1 - l_{cj})\delta_0(\gamma_{j1})$$

with mean

$$\mu_{\gamma_{j1}} = \Sigma_{\gamma_{j1}} X_j^\top (Y - \mu_{(-\gamma_{j1})})$$

The full conditional distribution of $\zeta_j, j = 1, \ldots, p$

$$
\begin{aligned}
&\pi(\zeta_j|\text{rest}) \\
&\quad \propto \pi(\zeta_j|\tau_j^2, \sigma^2)\pi(y|\cdot) \\
&\quad \propto (\sigma^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma^2}(Y - T_j\zeta_j - \mu_{(-\zeta_j)})^\top (Y - T_j\zeta_j - \mu_{(-\zeta_j)})\Big) \\
&\quad \times \left(\pi_e (2\pi\sigma^2\tau_{ej}^2)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{ej}^2}\zeta_j^2\Big)\mathbf{I}_{\{\zeta_j \neq 0\}} + (1-\pi_e)\delta_0(\zeta_j) \right)
\end{aligned}
$$

Let $\Sigma_{\zeta_j} = (T_j^\top T_j + \frac{1}{\tau_{ej}^2})^{-1}$, we have

$$
\begin{aligned}
l_{ej} &= \pi(\zeta_j \neq 0|\text{rest}) \\
&= \frac{\pi_e}{\pi_e + (1-\pi_e)(\tau_{ej}^2)^{\frac{1}{2}}(\Sigma_{\zeta_j})^{-\frac{1}{2}} \exp\Big(-\frac{\Sigma_{\zeta_j}}{2\sigma^2}\|(Y - \mu_{(-\zeta_j)})^\top T_j\|_2^2\Big)}
\end{aligned}
$$

Hence, the full conditional distribution of ζ_j is a spike and slab distribution

$$
\zeta_j|\text{rest} \sim l_{ej} N(\mu_{\zeta_j},\, \sigma^2\Sigma_{\zeta_j}) + (1 - l_{ej})\delta_0(\zeta_j)
$$

where

$$
\mu_{\zeta_j} = \Sigma_{\zeta_j} T_j^\top (Y - \mu_{(-\zeta_j)})
$$

Now, we derive the full conditional distribution for τ_{vj}^2, τ_{cj}^2 and τ_{ej}^2.

$$
\begin{aligned}
&\pi(\tau_{vj}^2|\text{rest}) \\
&\quad \propto \pi(\tau_{vj}^2|\lambda_v)\pi(\gamma_{j*}|\tau_{vj}^2, \sigma^2) \\
&\quad \propto (\tau_{vj}^2)^{\frac{L+1}{2}-1} \exp\Big(-\tau_{vj}^2 \frac{L\lambda_v^2}{2}\Big) \\
&\quad \times \left(\pi_v (2\pi\sigma^2\tau_{vj}^2)^{-\frac{L}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{vj}^2}\gamma_{j*}^\top\gamma_{j*}\Big)\mathbf{I}_{\{\gamma_{j*} \neq 0\}} + (1-\pi_v)\delta_0(\gamma_{j*}) \right)
\end{aligned}
\tag{B.2}
$$

When $\gamma_{j*} = 0$, B.2 is equal to

$$(1-\pi_v)(\tau_{vj}^2)^{\frac{L+1}{2}-1}\exp\Big(-\tau_{vj}^2\frac{L\lambda_v^2}{2}\Big)$$

Therefore, when $\gamma_{j*} = 0$, the posterior distribution for $(\tau_{vj}^2)^{-1}$ is Inverse-Gamma($\frac{L+1}{2}$, $\frac{L\lambda_v^2}{2}$). When $\gamma_{j*} \neq 0$, B.2 is equal to

$$\begin{aligned}
&\pi(\tau_{vj}^2|\text{rest}) \\
&\quad\propto (1-\pi_v)(2\pi\sigma^2\tau_{vj}^2)^{-\frac{L}{2}}(\tau_{vj}^2)^{\frac{L+1}{2}-1}\exp\Big(-\tau_{vj}^2\frac{L\lambda_v^2}{2}\Big)\exp\Big(-\frac{1}{2\sigma^2\tau_{vj}^2}\gamma_{j*}^\top\gamma_{j*}\Big) \\
&\quad\propto (1-\pi_v)(2\pi\sigma^2)^{-\frac{L}{2}}(\tau_{vj}^2)^{-\frac{1}{2}}\exp\left(-\tau_{vj}^2\frac{L\lambda_v^2}{2}-\frac{\|\gamma_{j*}\|_2^2}{2\sigma^2\tau_{vj}^2}\right)
\end{aligned}$$

Therefore, when $\gamma_{j*} \neq 0$, the posterior distribution for $(\tau_{vj}^2)^{-1}$ is Inverse-Gaussian($L\lambda_v^2$, $\sqrt{\frac{L\lambda_v^2\sigma^2}{\|\gamma_{j*}\|_2^2}}$). Together

$$(\tau_{vj}^2)^{-1}|\text{rest} \sim \begin{cases} \text{Inverse-Gamma}(\frac{L+1}{2},\ \frac{L\lambda_v^2}{2}) & \text{if } \gamma_{j*}=0 \\ \text{Inverse-Gaussian}(L\lambda_v^2,\ \sqrt{\frac{L\lambda_v^2\sigma^2}{\|\gamma_{j*}\|_2^2}}) & \text{if } \gamma_{j*}\neq 0 \end{cases}$$

Similarly, the posterior distribution for $(\tau_{cj}^2)^{-1}$ is

$$\begin{aligned}
&\pi(\tau_{cj}^2|\text{rest}) \\
&\quad\propto \pi(\tau_{cj}^2|\lambda_c)\pi(\gamma_{j1}|\tau_{cj}^2,\sigma^2) \\
&\quad\propto \frac{\lambda_c^2}{2}\exp\Big(-\tau_{cj}^2\frac{\lambda_c^2}{2}\Big) \\
&\quad\times\left(\pi_c(2\pi\sigma^2\tau_{cj}^2)^{-\frac{1}{2}}\exp\Big(-\frac{1}{2\sigma^2\tau_{cj}^2}\gamma_{j1}^2\Big)\mathbf{I}_{\{\gamma_{j1}\neq 0\}}+(1-\pi_c)\delta_0(\gamma_{j1})\right)
\end{aligned} \tag{B.3}$$

When $\gamma_{j1} = 0$, B.3 is equal to

$$(1-\pi_c)\frac{\lambda_c^2}{2}\exp\Big(-\tau_{cj}^2\frac{\lambda_c^2}{2}\Big)$$

Therefore, when $\gamma_{j1} = 0$, the posterior distribution for $(\tau_{cj}^2)^{-1}$ is Inverse-Gamma(1, $\frac{\lambda_c^2}{2}$). When $\gamma_{j1} \neq 0$, B.3 is equal to

$$\begin{aligned}
&\pi(\tau_{cj}^2|\text{rest}) \\
&\quad\propto (1-\pi_c)(2\pi\sigma^2\tau_{cj}^2)^{-\frac{1}{2}}\frac{\lambda_c^2}{2}\exp\Big(-\tau_{cj}^2\frac{\lambda_c^2}{2}\Big)\exp\Big(-\frac{1}{2\sigma^2\tau_{cj}^2}\gamma_{j1}^2\Big) \\
&\quad\propto (\tau_{cj}^2)^{-\frac{1}{2}}\exp\left(-\tau_{cj}^2\frac{\lambda_c^2}{2}-\frac{\gamma_{j1}^2}{2\sigma^2\tau_{cj}^2}\right)
\end{aligned}$$

Therefore, when $\gamma_{j1} \neq 0$, the posterior distribution for $(\tau_{cj}^2)^{-1}$ is Inverse-Gaussian(λ_c^2, $\sqrt{\frac{\lambda_c^2\sigma^2}{\gamma_{j1}^2}}$). Together

$$(\tau_{cj}^2)^{-1}|\text{rest} \sim \begin{cases} \text{Inverse-Gamma}(1,\ \frac{\lambda_c^2}{2}) & \text{if } \gamma_{j1} = 0 \\ \text{Inverse-Gaussian}(\lambda_c^2,\ \sqrt{\frac{\lambda_c^2\sigma^2}{\gamma_{j1}^2}}) & \text{if } \gamma_{j1} \neq 0 \end{cases}$$

The posterior distribution $(\tau_{ej}^2)^{-1}$

$$\begin{aligned}
&\pi(\tau_{ej}^2|\text{rest}) \\
&\quad\propto \pi(\tau_{ej}^2|\lambda_e)\pi(\zeta_j|\tau_{ej}^2,\sigma^2) \\
&\quad\propto \frac{\lambda_e^2}{2}\exp\Big(-\tau_{ej}^2\frac{\lambda_e^2}{2}\Big) \\
&\quad\times\left(\pi_e(2\pi\sigma^2\tau_{ej}^2)^{-\frac{1}{2}}\exp\Big(-\frac{1}{2\sigma^2\tau_{ej}^2}\zeta_j^2\Big)\mathbf{I}_{\{\zeta_j\neq 0\}} + (1-\pi_e)\delta_0(\zeta_j)\right)
\end{aligned}$$

Following the similar arguments, we have

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

Now, we derive the full conditional distribution for λ_v^2 and τ_{cj}^2. The posterior distribution for λ_v^2:

$$\begin{aligned} &\pi(\lambda_v^2|\text{rest}) \\ &\quad\propto \pi(\lambda_v^2)\prod_{j=1}^{p}\pi(\tau_{vj}^2|\lambda_v^2) \\ &\quad\propto (\lambda_v^2)^{a_v-1}\exp(-b_v\lambda_v^2)\prod_{j=1}^{p}\left(\frac{L\lambda_v^2}{2}\right)^{\frac{L+1}{2}}\exp\left(-\frac{L\lambda_v^2}{2}\tau_{vj}^2\right) \\ &\quad\propto (\lambda_v^2)^{a_v+\frac{p(L+1)}{2}-1}\exp\left(-(b_v+\frac{L\sum_{j=1}^{p}\tau_{vj}^2}{2})\lambda_v^2\right) \end{aligned}$$

the posterior distribution for λ_v^2 is Gamma($a_v+\frac{p(L+1)}{2}$, $b_v+\frac{L\sum_{j=1}^{p}\tau_{vj}^2}{2}$).

$$\begin{aligned} &\pi(\lambda_c^2|\text{rest}) \\ &\quad\propto \pi(\lambda_c^2)\prod_{j=1}^{p}\pi(\tau_{cj}^2|\lambda_c^2) \\ &\quad\propto (\lambda_c^2)^{a_c-1}\exp(-b_c\lambda_c^2)\prod_{j=1}^{p}\frac{\lambda_c^2}{2}\exp\left(-\frac{\lambda_c^2}{2}\tau_{cj}^2\right) \\ &\quad\propto (\lambda_c^2)^{a_c+p-1}\exp\left(-(b_c+\frac{\sum_{j=1}^{p}\tau_{cj}^2}{2})\lambda_c^2\right) \end{aligned}$$

the posterior distribution for λ_c^2 is Gamma(a_c+p, $b_c+\frac{\sum_{j=1}^{p}\tau_{cj}^2}{2}$). Similarly, the full conditional distribution for λ_e^2 is Gamma(a_e+p, $b_e+\frac{\sum_{j=1}^{p}\tau_{ej}^2}{2}$).

Next, we derive the full conditional distribution for π_v, π_c and π_e. The posterior distri-

bution for π_v

$$
\begin{aligned}
&\pi(\pi_v|\text{rest}) \\
&\quad \propto \pi(\pi_v) \prod_{j=1}^{p} \pi(\gamma_{j*}^2|\pi_v, \tau_{vj}^2, \sigma^2) \\
&\quad \propto \pi_v^{r_v-1}(1-\pi_v)^{w_v-1} \\
&\quad \times \prod_{j=1}^{p} \left(\pi_v (2\pi\sigma^2\tau_{vj}^2)^{-\frac{L}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{vj}^2}\gamma_{j*}^\top\gamma_{j*}\Big)\mathbf{I}_{\{\gamma_{j*}\neq 0\}} + (1-\pi_v)\delta_0(\gamma_{j*}) \right) \\
&\quad \propto \pi_v^{r_v+\sum_{j=1}^{p}\mathbf{I}_{\{\gamma_{j*}\neq 0\}}-1}(1-\pi_v)^{w_v+\sum_{j=1}^{p}\delta_0(\gamma_{j*})-1}
\end{aligned}
$$

the posterior distribution for π_v is Beta($r_v + \sum_{j=1}^{p}\mathbf{I}_{\{\gamma_{j*}\neq 0\}}$, $w_v + \sum_{j=1}^{p}\delta_0(\gamma_{j*})$).

$$
\begin{aligned}
&\pi(\pi_c|\text{rest}) \\
&\quad \propto \pi(\pi_c) \prod_{j=1}^{p} \pi(\gamma_{j1}^2|\pi_c, \tau_{cj}^2, \sigma^2) \\
&\quad \propto \pi_c^{r_c-1}(1-\pi_c)^{w_c-1} \\
&\quad \times \prod_{j=1}^{p} \left(\pi_c (2\pi\sigma^2\tau_{cj}^2)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{cj}^2}\gamma_{j1}^2\Big)\mathbf{I}_{\{\gamma_{j1}\neq 0\}} + (1-\pi_c)\delta_0(\gamma_{j1}) \right) \\
&\quad \propto \pi_c^{r_c+\sum_{j=1}^{p}\mathbf{I}_{\{\gamma_{j1}\neq 0\}}-1}(1-\pi_c)^{w_c+\sum_{j=1}^{p}(\delta_0(\gamma_{j1}))-1}
\end{aligned}
$$

the posterior distribution for π_c is Beta($r_c + \sum_{j=1}^{p} \mathbf{I}_{\{\gamma_{j1}\neq 0\}}$, $w_c + \sum_{j=1}^{p} \delta_0(\gamma_{j1})$).

$$
\begin{aligned}
&\pi(\pi_e|\text{rest}) \\
&\quad \propto \pi(\pi_e) \prod_{j=1}^{p} \pi(\zeta_j^2|\pi_e, \tau_{ej}^2, \sigma^2) \\
&\quad \propto \pi_e^{r_e-1}(1-\pi_e)^{w_e-1} \\
&\quad \times \prod_{j=1}^{p} \left(\pi_e (2\pi\sigma^2\tau_{ej}^2)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{ej}^2}\zeta_j^2 \Big) \mathbf{I}_{\{\zeta_j\neq 0\}} + (1-\pi_e)\delta_0(\zeta_j) \right) \\
&\quad \propto \pi_e^{r_e+\sum_{j=1}^{p} \mathbf{I}_{\{\zeta_j\neq 0\}}-1} (1-\pi_e)^{w_e+\sum_{j=1}^{p}(\delta_0(\zeta_j))-1}
\end{aligned}
$$

the posterior distribution for π_e is Beta($r_e + \sum_{j=1}^{p} \mathbf{I}_{\{\zeta_j\neq 0\}}$, $w_e + \sum_{j=1}^{p} \delta_0(\zeta_j)$). Last, the full conditional distribution for σ^2

$$
\begin{aligned}
&\pi(\sigma^2|\text{rest}) \\
&\quad \propto \pi(\sigma^2)\pi(y|\cdot) \prod_{j=1}^{p} \pi(\gamma_{j1}|\pi_c, \tau_{cj}^2, \sigma^2)\pi(\gamma_{j*}|\pi_v, \tau_{vj}^2, \sigma^2)\pi(\zeta_j|\pi_e, \tau_j^2, \sigma^2) \\
&\quad \propto (\sigma^2)^{-s-1} \exp(-\frac{h}{\sigma^2})(\sigma^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma^2}(Y-\mu)^\top (Y-\mu) \Big) \\
&\qquad \times \sum_{j=1}^{p} \left(\pi_c (2\pi\sigma^2\tau_{cj}^2)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{cj}^2}\gamma_{j1}^2 \Big) \mathbf{I}_{\{\gamma_{j1}\neq 0\}} + (1-\pi_c)\delta_0(\gamma_{j1}) \right) \\
&\qquad \times \sum_{j=1}^{p} \left(\pi_v (2\pi\sigma^2\tau_{vj}^2)^{-\frac{L}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{vj}^2}\gamma_{j*}^\top\gamma_{j*} \Big) \mathbf{I}_{\{\gamma_{j*}\neq 0\}} + (1-\pi_v)\delta_0(\gamma_{j*}) \right) \\
&\qquad \times \sum_{j=1}^{p} \left(\pi_e (2\pi\sigma^2\tau_{ej}^2)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2\tau_{ej}^2}\zeta_j^2 \Big) \mathbf{I}_{\{\zeta_j\neq 0\}} + (1-\pi_e)\delta_0(\zeta_j) \right) \\
&\quad \propto (\sigma^2)^{-(s+\frac{n+L\sum \mathbf{I}_{\{\gamma_{j*}\neq 0\}}+\sum \mathbf{I}_{\{\zeta_j\neq 0\}}+\sum \mathbf{I}_{\{\gamma_{j1}\neq 0\}}}{2})-1} \\
&\qquad \times \exp\left(-\frac{1}{\sigma^2}\Big(h + \frac{(Y-\mu)^\top(Y-\mu) + \sum_{j=1}^{p}(\tau_{cj}^2)^{-1}\gamma_{j1}^2 + (\tau_{vj}^2)^{-1}\gamma_{j*}^\top\gamma_{j*} + (\tau_j^2)^{-1}\zeta_j^2}{2} \Big) \right)
\end{aligned}
$$

the posterior distribution for σ^2 is Inverse-Gamma(μ_{σ^2}, Σ_{σ^2}) where

$$\mu_{\sigma^2} = s + \frac{n + L\sum_{j=1}^{p} \mathbf{I}_{\{\gamma_{j*}\neq 0\}} + \sum_{j=1}^{p} \mathbf{I}_{\{\zeta_j \neq 0\}} + \sum_{j=1}^{p} \mathbf{I}_{\{\gamma_{j1}\neq 0\}}}{2}$$

$$\Sigma_{\sigma^2} = h + \frac{(Y-\mu)^\top (Y-\mu) + \sum_{j=1}^{p} (\tau_{cj}^2)^{-1}\gamma_{j1}^2 + (\tau_{vj}^2)^{-1}\gamma_{j*}^\top \gamma_{j*} + (\tau_j^2)^{-1}\zeta_j^2}{2}$$

B.11 Posterior inference for the BSSVC method

B.11.1 Priors

$$Y|\eta, \gamma_1, \ldots, \gamma_p, \alpha_1, \ldots, \alpha_q, \zeta_0, \zeta_1, \ldots, \zeta_p, \sigma^2 \propto (\sigma^2)^{-\frac{n}{2}} \exp\Big\{ -\frac{1}{2\sigma^2}(Y-\mu)^\top (Y-\mu)\Big\}$$

$$\eta \sim \mathrm{N}_{q_n}(0,\, \Sigma_{\eta 0})$$

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

$$\zeta_0 \sim \mathrm{N}(0,\, \sigma_{\zeta_0}^2)$$

$$\gamma_j|\pi_c, \tau_{vj}^2, \sigma^2 \sim \pi_v \mathrm{N}_{q_n}(0,\, \mathrm{Diag}(\sigma^2\tau_{vj}^2, \ldots, \sigma^2\tau_{vj}^2)) + (1-\pi_v)\delta_0(\gamma_j), \quad j = 1, \ldots, p$$

$$\tau_{vj}^2|\lambda_v \sim \mathrm{Gamma}(\frac{q_n+1}{2},\, \frac{q_n\lambda_v^2}{2}), \quad j = 1, \ldots, p$$

$$\zeta_j|\pi_e, \tau_{ej}^2, \sigma^2 \sim \pi_e \mathrm{N}(0,\, \sigma^2\tau_{ej}^2) + (1-\pi_e)\delta_0(\zeta_j), \quad j = 1, \ldots, p$$

$$\tau_{ej}^2|\lambda_e \sim \frac{\lambda_e^2}{2}\exp(-\frac{\lambda_e^2\tau_{ej}^2}{2}), \quad j = 1, \ldots, p$$

$$\sigma^2 \sim (\sigma^2)^{-s-1}\exp(-\frac{h}{\sigma^2})$$

Consider the following conjugate gamma priors for λ_v^2 and λ_e^2

$$\lambda_v^2 \sim \mathrm{Gamma}(a_v,\, b_v) \quad \text{and} \quad \lambda_e^2 \sim \mathrm{Gamma}(a_e,\, b_e)$$

and conjugate beta priors for π_v and π_e

$$\pi_v \sim \text{Beta}(r_v,\, w_v) \quad \text{and} \quad \pi_e \sim \text{Beta}(r_e,\, w_e)$$

B.11.2 Posterior distribution

$\pi(\eta|\text{rest}) \sim \text{N}_{q_n}(\mu_\eta,\, \Sigma_\eta)$ where

$$\mu_\eta = \Big(\Sigma_{\eta 0}^{-1} + \frac{1}{\sigma^2} B_0^\top B_0\Big)^{-1} \Big(\frac{1}{\sigma^2}(Y - \mu_{(-\eta)})^\top B_0\Big)^\top$$
$$\Sigma_\eta = \Big(\Sigma_{\eta 0}^{-1} + \frac{1}{\sigma^2} B_0^\top B_0\Big)^{-1}$$

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

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

$\pi(\zeta_0|\text{rest}) \sim \text{N}(\mu_{\zeta_0}, \Sigma_{\zeta_0})$ where

$$\mu_{\zeta_0} = \Big(\sigma_{\zeta 0}^{-1} + \frac{1}{\sigma^2}\sum_{i=1}^n E_i^2\Big)^{-1} \Big(\frac{1}{\sigma^2}\sum_{i=1}^n (y_i - \mu_{(-\zeta_0)}) E_i\Big)$$
$$\Sigma_{\zeta 0} = \Big(1/\sigma_{\zeta_0}^2 + \frac{1}{\sigma^2}\sum_{i=1}^n E_i^2\Big)^{-1}$$

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

$$\mu_{\gamma_j} = \Sigma_{\gamma_j} U_j^\top (Y - \mu_{(-\gamma_j)})$$
$$\Sigma_{\gamma_j} = (U_j^\top U_j + \frac{1}{\tau_{vj}^2}\mathbf{I}_{q_n})^{-1}$$
$$l_{vj} = \frac{\pi_v}{\pi_v + (1 - \pi_v)(\tau_{vj}^2)^{\frac{q_n}{2}} |\Sigma_{\gamma_j}|^{-\frac{1}{2}} \exp\Big(-\frac{1}{2\sigma^2}\|\Sigma_{\gamma_j}^{\frac{1}{2}} U_j^\top (Y - \mu_{(-\gamma_j)})\|_2^2\Big)}$$

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

$$\mu_{\zeta_j} = \Sigma_{\zeta_j} T_j^\top (Y - \mu_{(-\zeta_j)})$$

$$\Sigma_{\zeta_j} = (T_j^\top T_j + \frac{1}{\tau_{ej}^2})^{-1}$$

$$l_{ej} = \frac{\pi_e}{\pi_e + (1-\pi_e)(\tau_{ej}^2)^{\frac{1}{2}}(\Sigma_{\zeta_j})^{-\frac{1}{2}} \exp\Big(- \frac{\Sigma_{\zeta_j}}{2\sigma^2} \|(Y - \mu_{(-\zeta_j)})^\top T_j\|_2^2 \Big)}$$

At the gth iteration, the values of $\phi_{vj}^{(g)}$ and $\phi_{ej}^{(g)}$ can be determined by whether the $\gamma_j^{(g)}$ and $\zeta_j^{(g)}$ are set to 0 or not, respectively.

$$(\tau_{vj}^2)^{-1}|\text{rest} \sim \begin{cases} \text{Inverse-Gamma}(\frac{q_n+1}{2},\ \frac{q_n\lambda_v^2}{2}) & \text{if } \gamma_j = 0 \\ \text{Inverse-Gaussian}(q_n\lambda_v^2,\ \sqrt{\frac{q_n\lambda_v^2\sigma^2}{\|\gamma_j\|_2^2}}) & \text{if } \gamma_j \neq 0 \end{cases}$$

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

λ_v and λ_e all have Gamma posterior distributions

$$\lambda_v^2 \sim \text{Gamma}(a_v + \frac{p(q_n+1)}{2},\, b_v + \frac{q_n \sum_{j=1}^p \tau_{vj}^2}{2})$$
$$\lambda_e^2 \sim \text{Gamma}(a_e + p,\, b_e + \frac{\sum_{j=1}^p \tau_{ej}^2}{2})$$

π_v and π_e have beta posterior distributions

$$\pi_v \sim \text{Beta}(r_v + \sum_{j=1}^p \mathbf{I}_{\{\gamma_j \neq 0\}},\, w_v + \sum_{j=1}^p \delta_0(\gamma_j))$$
$$\pi_e \sim \text{Beta}(r_e + \sum_{j=1}^p \mathbf{I}_{\{\zeta_j \neq 0\}},\, w_e + \sum_{j=1}^p \delta_0(\zeta_j))$$

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

$$\mu_{\sigma^2} = s + \frac{n + q_n \sum_{j=1}^{p} \mathbf{I}_{\{\gamma_j \neq 0\}} + \sum_{j=1}^{p} \mathbf{I}_{\{\zeta_j \neq 0\}}}{2}$$

$$\Sigma_{\sigma^2} = h + \frac{(Y-\mu)^\top (Y-\mu) + \sum_{j=1}^{p} (\tau_{vj}^2)^{-1} \gamma_j^\top \gamma_j + (\tau_{cj}^2)^{-1} \zeta_j^2}{2}$$

B.12 Posterior inference for the BVC-SI method

B.12.1 Priors

$$Y|\eta, \gamma_{11}, \ldots, \gamma_{p1}, \gamma_{1*}, \ldots, \gamma_{p*}, \alpha_1, \ldots, \alpha_q, \zeta_0, \zeta_1, \ldots, \zeta_p, \sigma^2$$
$$\propto (\sigma^2)^{-\frac{n}{2}} \exp\Big\{ -\frac{1}{2\sigma^2}(Y-\mu)^\top (Y-\mu) \Big\}$$

$$\eta \sim \mathrm{N}_{q_n}(0, \Sigma_{\eta 0})$$

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

$$\zeta_0 \sim \mathrm{N}(0,\, \sigma_{\zeta_0}^2)$$

$$\gamma_{j1}|\tau_{cj}^2, \sigma^2 \sim \mathrm{N}(0,\, \sigma^2 \tau_{cj}^2), \quad j = 1, \ldots, p$$

$$\tau_{cj}^2|\lambda_c \sim \frac{\lambda_c^2}{2} \exp(-\frac{\lambda_c^2 \tau_{cj}^2}{2}), \quad j = 1, \ldots, p$$

$$\gamma_{j*}|\tau_{vj}^2, \sigma^2 \sim \mathrm{N}_L(0,\, \mathrm{diag}(\sigma^2 \tau_{vj}^2, \ldots, \sigma^2 \tau_{vj}^2)), \quad j = 1, \ldots, p$$

$$\tau_{vj}^2|\lambda_v \sim \mathrm{Gamma}(\frac{L+1}{2},\, \frac{L\lambda_v^2}{2}), \quad j = 1, \ldots, p$$

$$\zeta_j|\tau_{ej}^2, \sigma^2 \sim \mathrm{N}(0,\, \sigma^2 \tau_{ej}^2), \quad j = 1, \ldots, p$$

$$\tau_{ej}^2|\lambda \sim \frac{\lambda_e^2}{2} \exp(-\frac{\lambda_e^2 \tau_{ej}^2}{2}), \quad j = 1, \ldots, p$$

$$\sigma^2 \sim (\sigma^2)^{-s-1} \exp(-\frac{h}{\sigma^2})$$

Consider the following conjugate gamma priors for λ_c^2, λ_v^2 and λ_e^2

$$\lambda_c^2 \sim \text{Gamma}(a_c,\, b_c), \quad \lambda_v^2 \sim \text{Gamma}(a_v,\, b_v) \quad \text{and} \quad \lambda_e^2 \sim \text{Gamma}(a_e,\, b_e)$$

B.12.2 Gibbs Sampler

$\pi(\eta|\text{rest}) \sim \text{N}_{q_n}(\mu_\eta,\, \Sigma_\eta)$ where

$$\mu_\eta = \Big(\Sigma_{\eta 0}^{-1} + \frac{1}{\sigma^2} B_0^\top B_0\Big)^{-1} \Big(\frac{1}{\sigma^2}(Y - \mu_{(-\eta)})^\top B_0\Big)^\top$$

$$\Sigma_\eta = \Big(\Sigma_{\eta 0}^{-1} + \frac{1}{\sigma^2} B_0^\top B_0\Big)^{-1}$$

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

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

$$\Sigma_\alpha = \Big(\Sigma_{\alpha 0}^{-1} + \frac{1}{\sigma^2} W^\top W\Big)^{-1}$$

$\pi(\zeta_0|\text{rest}) \sim \text{N}(\mu_{\zeta_0}, \Sigma_{\zeta_0})$ where

$$\mu_{\zeta_0} = \Big(1/\sigma_{\zeta 0}^2 + \frac{1}{\sigma^2}\sum_{i=1}^n E_i^2\Big)^{-1} \Big(\frac{1}{\sigma^2}\sum_{i=1}^n (y_i - \mu_{(-\zeta_0)}) E_i\Big)$$

$$\Sigma_{\zeta 0} = \Big(1/\sigma_{\zeta_0}^2 + \frac{1}{\sigma^2}\sum_{i=1}^n E_i^2\Big)^{-1}$$

The full conditional distribution of γ_j

$$\pi(\gamma_{j*}|rest)$$

$$\propto \pi(\gamma_{j*}|\tau_{vj}^2, \sigma^2)\pi(y|\cdot)$$

$$\propto \exp\left(-\frac{1}{2}(\sigma^2\tau_{vj}^2)^{-1}\gamma_{j*}^\top\gamma_{j*} - \frac{1}{2\sigma^2}(Y - U_j\gamma_{j*} - \mu_{(-\gamma_{j*})})^\top(Y - U_j\gamma_{j*} - \mu_{(-\gamma_{j*})})\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\Big((\tau_{vj}^2)^{-1}\gamma_{j*}^\top\gamma_{j*} + \gamma_{j*}^\top U_j^\top U_j\gamma_{j*} - 2(Y - \mu_{(-\gamma_{j*})})^\top(U_j\gamma_{j*})\Big)\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\Big(\gamma_{j*}^\top\Big((\tau_{vj}^2)^{-1} + U_j^\top U_j\Big)\gamma_{j*} - 2(Y - \mu_{(-\gamma_{j*})})^\top U_j\gamma_{j*}\Big)\right)$$

where $U_j = (U_{1j}, \ldots, U_{nj})^\top$. Hence, the full conditional distribution of γ_{j*} is multivariate normal with mean

$$\mu_{\gamma_{j*}} = \Big((\tau_{vj}^2)^{-1}\mathbf{I}_L + U_j^\top U_j\Big)^{-1}\Big((Y - \mu_{(-\gamma_{j*})})^\top U_j\Big)^\top$$

and variance

$$\Sigma_{\gamma_{j*}} = \sigma^2\Big((\tau_{vj}^2)^{-1}\mathbf{I}_L + U_j^\top U_j\Big)^{-1}$$

Similarly, the full conditional distribution of γ_{j1} is normal distribution with mean

$$\mu_{\gamma_{j1}} = \Big((\tau_{cj}^2)^{-1} + X_j^\top X_j\Big)^{-1}\Big((y_i - \mu_{(-\gamma_{j1})})^\top X_j\Big)$$

and variance

$$\Sigma_{\gamma_{j1}} = \sigma^2\Big((\tau_{cj}^2)^{-1} + X_j^\top X_j\Big)^{-1}$$

Let $D_{\tau_e} = \text{diag}(\tau_{e1}^2, \ldots, \tau_{ep}^2)$. The full conditional distribution of $\zeta = (\zeta_1, \ldots, \zeta_p)^\top$

$$
\begin{aligned}
&\pi(\zeta|rest) \\
&\quad \propto \pi(\zeta|\tau_{e1}^2, \ldots, \tau_{ep}^2, \sigma^2)\pi(y|\cdot) \\
&\quad \propto \exp\left(-\frac{1}{2}(\sigma^2 D_{\tau_e})^{-1}\zeta^\top\zeta - \frac{1}{2\sigma^2}(Y - T\zeta - \mu_{(-\zeta)})^\top(Y - T\zeta - \mu_{(-\zeta)})\right) \\
&\quad \propto \exp\left(-\frac{1}{2\sigma^2}\Big(D_{\tau_e}^{-1}\zeta^\top\zeta + \zeta^\top T^\top T\zeta - 2(Y - \mu_{(-\zeta)})^\top T\zeta\Big)\right) \\
&\quad \propto \exp\left(-\frac{1}{2\sigma^2}\Big(\zeta^\top(D_{\tau_e}^{-1} + T^\top T)\zeta - 2(Y - \mu_{(-\zeta)})^\top T\zeta\Big)\right)
\end{aligned}
$$

where $T = (T_1, \ldots, T_n)^\top$. The full conditional is $\text{N}_p(\mu_\zeta, \sigma^2\Sigma_\zeta)$ with

$$\mu_\zeta = \Big(D_{\tau_e}^{-1} + T^\top T\Big)^{-1}\Big((Y - \mu_{(-\zeta)})^\top T\Big)^\top$$

and variance

$$\Sigma_\zeta = \Big(D_{\tau_e}^{-1} + T^\top T\Big)^{-1}$$

Now, we derive the full conditional distribution for τ_{ej}^2 and λ_e^2.

$$
\begin{aligned}
&\pi(\tau_{vj}^2|rest) \\
&\quad \propto \pi(\tau_{vj}^2|\lambda_v)\pi(\gamma_{j*}|\tau_{vj}^2, \sigma^2) \\
&\quad \propto (\tau_{vj}^2)^{\frac{L+1}{2}-1}\exp\left(-\tau_{vj}^2\frac{L\lambda_v^2}{2}\right)(\tau_{vj}^2)^{-\frac{L}{2}}\exp\left(-\frac{1}{2}(\sigma^2\tau_{vj}^2)^{-1}\gamma_{j*}^\top\gamma_{j*}\right) \\
&\quad \propto (\tau_{vj}^2)^{-\frac{1}{2}}\exp\left(-\tau_{vj}^2\frac{L\lambda_v^2}{2} - \frac{||\gamma_{j*}||_2^2}{2\sigma^2\tau_{vj}^2}\right)
\end{aligned}
$$

the posterior distribution for $(\tau_{vj}^2)^{-1}$ is Inverse-Gaussian($L\lambda_v^2$, $\sqrt{\frac{L\lambda_v^2\sigma^2}{||\gamma_{j*}||_2^2}}$). Similarly, the posterior distribution for $(\tau_{cj}^2)^{-1}$ is Inverse-Gaussian(λ_c^2, $\sqrt{\frac{\lambda_c^2\sigma^2}{\gamma_{j1}^2}}$), and the posterior distribution

for $(\tau_{ej}^2)^{-1}$ is Inverse-Gaussian(λ_e^2, $\sqrt{\frac{\lambda_e^2\sigma^2}{\zeta_j^2}}$).

$$
\begin{aligned}
&\pi(\lambda_v^2|rest) \\
&\quad\propto \pi(\lambda_v^2)\prod_{j=1}^{p}\pi(\tau_{vj}^2|\lambda_v^2) \\
&\quad\propto (\lambda_v^2)^{a_v-1}\exp(-b_v\lambda_v^2)\prod_{j=1}^{p}\left(\frac{L\lambda_v^2}{2}\right)^{\frac{L+1}{2}}\exp\left(-\frac{L\lambda_v^2}{2}\tau_{vj}^2\right) \\
&\quad\propto (\lambda_v^2)^{a_v+\frac{p(L+1)}{2}-1}\exp\left(-(b_v+\frac{L\sum_{j=1}^{p}\tau_{vj}^2}{2})\lambda_v^2\right)
\end{aligned}
$$

the posterior distribution for λ_c^2 is Gamma(a_c+p, $b_c+\frac{\sum_{j=1}^{p}\tau_{cj}^2}{2}$). Similarly, the full conditional distribution for λ_e^2 is Gamma(a_e+p, $b_e+\frac{\sum_{j=1}^{p}\tau_{ej}^2}{2}$). Last, the full conditional distribution for σ^2

$$
\begin{aligned}
&\pi(\sigma^2|rest) \\
&\propto \pi(\sigma^2)\pi(y|\cdot)\prod_{j=1}^{p}\pi(\gamma_{j1}|\pi_c,\tau_{cj}^2,\sigma^2)\pi(\gamma_{j*}|\pi_v,\tau_{vj}^2,\sigma^2)\pi(\zeta_j|\pi_e,\tau_{ej}^2,\sigma^2) \\
&\propto (\sigma^2)^{-s-1}\exp(-\frac{h}{\sigma^2})(\sigma^2)^{-\frac{n}{2}}\exp\Big(-\frac{1}{2\sigma^2}(Y-\mu)^\top(Y-\mu)\Big) \\
&\quad\times(\sigma^2)^{-\frac{p}{2}}\exp\Big(-\frac{1}{2\sigma^2}\sum_{j=1}^{p}(\tau_{cj}^2)^{-1}\gamma_{j1}^2\Big) \\
&\quad\times(\sigma^2)^{-\frac{pL}{2}}\exp\Big(-\frac{1}{2\sigma^2}\sum_{j=1}^{p}(\tau_{vj}^2)^{-1}\gamma_{j*}^\top\gamma_{j*}\Big) \\
&\quad\times(\sigma^2)^{-\frac{p}{2}}\exp\Big(-\frac{1}{2\sigma^2}\sum_{j=1}^{p}(\tau_{ej}^2)^{-1}\zeta_j^2\Big) \\
&\propto (\sigma^2)^{-(s+\frac{n+2p+pL}{2})-1} \\
&\quad\times\exp\left(-\frac{1}{\sigma^2}\Big(h+\frac{(Y-\mu)^\top(Y-\mu)+\sum_{j=1}^{p}(\tau_{cj}^2)^{-1}\gamma_{j1}^2+(\tau_{vj}^2)^{-1}\gamma_{j*}^\top\gamma_{j*}+(\tau_{ej}^2)^{-1}\zeta_j^2}{2}\Big)\right)
\end{aligned}
$$

the posterior distribution for σ^2 is Inverse-Gamma(μ_{σ^2}, Σ_{σ^2}) where

$$\mu_{\sigma^2} = s + \frac{n + 2p + pL}{2}$$

$$\Sigma_{\sigma^2} = h + \frac{(Y-\mu)^\top (Y-\mu) + \sum_{j=1}^{p} (\tau_{cj}^2)^{-1} \gamma_{j1}^2 + (\tau_{vj}^2)^{-1} \gamma_{j*}^\top \gamma_{j*} + (\tau_{ej}^2)^{-1} \zeta_j^2}{2}$$

B.13 Posterior inference for the BVC method

B.13.1 Priors

$$Y|\eta, \gamma_1, \ldots, \gamma_p, \alpha_1, \ldots, \alpha_q, \zeta_0, \zeta_1, \ldots, \zeta_p, \sigma^2 \propto (\sigma^2)^{-\frac{n}{2}} \exp\Big\{ -\frac{1}{2\sigma^2}(Y-\mu)^\top (Y-\mu) \Big\}$$

$$\eta \sim \mathrm{N}_{q_n}(0, \Sigma_{\eta 0})$$

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

$$\zeta_0 \sim \mathrm{N}(0,\, \sigma_{\zeta_0}^2)$$

$$\gamma_j|\tau_{vj}^2, \sigma^2 \sim \mathrm{N}_{q_n}(0,\, \mathrm{diag}(\sigma^2\tau_{vj}^2, \ldots, \sigma^2\tau_{vj}^2)), \quad j = 1, \ldots, p$$

$$\tau_{vj}^2|\lambda_v \sim \mathrm{Gamma}(\frac{q_n+1}{2},\, \frac{q_n\lambda_v^2}{2}), \quad j = 1, \ldots, p$$

$$\zeta_j|\tau_{ej}^2, \sigma^2 \sim N(0,\, \sigma^2\tau_{ej}^2), \quad j = 1, \ldots, p$$

$$\tau_{ej}^2|\lambda_e \sim \frac{\lambda_e^2}{2} \exp(-\frac{\lambda_e^2\tau_{ej}^2}{2}), \quad j = 1, \ldots, p$$

$$\sigma^2 \sim (\sigma^2)^{-s-1} \exp(-\frac{h}{\sigma^2})$$

Consider the following conjugate gamma priors for λ_v^2 and λ_e^2

$$\lambda_v^2 \sim \mathrm{Gamma}(a_v,\, b_v) \quad \text{and} \quad \lambda_e^2 \sim \mathrm{Gamma}(a_e,\, b_e)$$

B.13.2 Gibbs Sampler

$\pi(\eta|\text{rest}) \sim \text{N}_{q_n}(\mu_\eta, \Sigma_\eta)$ where

$$\mu_\eta = \Big(\Sigma_{\eta 0}^{-1} + \frac{1}{\sigma^2} B_0^\top B_0\Big)^{-1} \Big(\frac{1}{\sigma^2}(Y - \mu_{(-\eta)})^\top B_0\Big)^\top$$

$$\Sigma_\eta = \Big(\Sigma_{\eta 0}^{-1} + \frac{1}{\sigma^2} B_0^\top B_0\Big)^{-1}$$

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

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

$$\Sigma_\alpha = \Big(\Sigma_{\alpha 0}^{-1} + \frac{1}{\sigma^2} W^\top W\Big)^{-1}$$

$\pi(\zeta_0|\text{rest}) \sim \text{N}(\mu_{\zeta_0}, \Sigma_{\zeta_0})$ where

$$\mu_{\zeta_0} = \Big(1/\sigma_{\zeta 0}^2 + \frac{1}{\sigma^2}\sum_{i=1}^n E_i^2\Big)^{-1} \Big(\frac{1}{\sigma^2}\sum_{i=1}^n (y_i - \mu_{(-\zeta_0)}) E_i\Big)$$

$$\Sigma_{\zeta_0} = \Big(1/\sigma_{\zeta_0}^2 + \frac{1}{\sigma^2}\sum_{i=1}^n E_i^2\Big)^{-1}$$

$\gamma_j|\text{rest} \sim \text{N}_{q_n}(\mu_{\gamma_j}, \sigma^2\Sigma_{\gamma_j})$ where

$$\mu_{\gamma_j} = \Sigma_{\gamma_j} U_j^\top (Y - \mu_{(-\gamma_j)})$$

$$\Sigma_{\gamma_j} = (U_j^\top U_j + \frac{1}{\tau_{vj}^2}\mathbf{I}_{q_n})^{-1}$$

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

$$\mu_{\zeta_j} = \Sigma_{\zeta_j} T_j^\top (Y - \mu_{(-\zeta_j)})$$

$$\Sigma_{\zeta_j} = (T_j^\top T_j + \frac{1}{\tau_{ej}^2})^{-1}$$

The posterior distribution for $(\tau_{vj}^2)^{-1}$ is Inverse-Gaussian$(q_n\lambda_v^2,\ \sqrt{\frac{q_n\lambda_v^2\sigma^2}{||\gamma_j||_2^2}})$. Similarly, the posterior distribution for $(\tau_{ej}^2)^{-1}$ is Inverse-Gaussian$(\lambda_e^2,\ \sqrt{\frac{\lambda_e^2\sigma^2}{\zeta_j^2}})$. λ_v and λ_e all have Gamma posterior distributions

$$\lambda_v^2 \sim \text{Gamma}(a_v + \frac{p(q_n+1)}{2},\ b_v + \frac{q_n\sum_{j=1}^p \tau_{vj}^2}{2})$$
$$\lambda_e^2 \sim \text{Gamma}(a_e + p,\ b_e + \frac{\sum_{j=1}^p \tau_{ej}^2}{2})$$

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

$$\mu_{\sigma^2} = s + \frac{n+p+pq_n}{2}$$
$$\Sigma_{\sigma^2} = h + \frac{1}{2}\Big((Y-\mu)^\top(Y-\mu) + \sum_{j=1}^p (\tau_{vj}^2)^{-1}\gamma_j^\top\gamma_j + (\tau_{ej}^2)^{-1}\zeta_j^2\Big)$$

Appendix C

Appendices for Chapter 4

C.1 Summary of methods.

Table C.1: *Summary of the proposed and alternative methods.*

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

Note: We constructed the models that can be applied to G×E settings based on the references.

C.2 Hyper-parameters sensitivity analysis

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

Table C.2: *Sensitivity analysis for RBSG-SS using Example 1. mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.*

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

C.3 Assessment of the convergence of MCMC chains

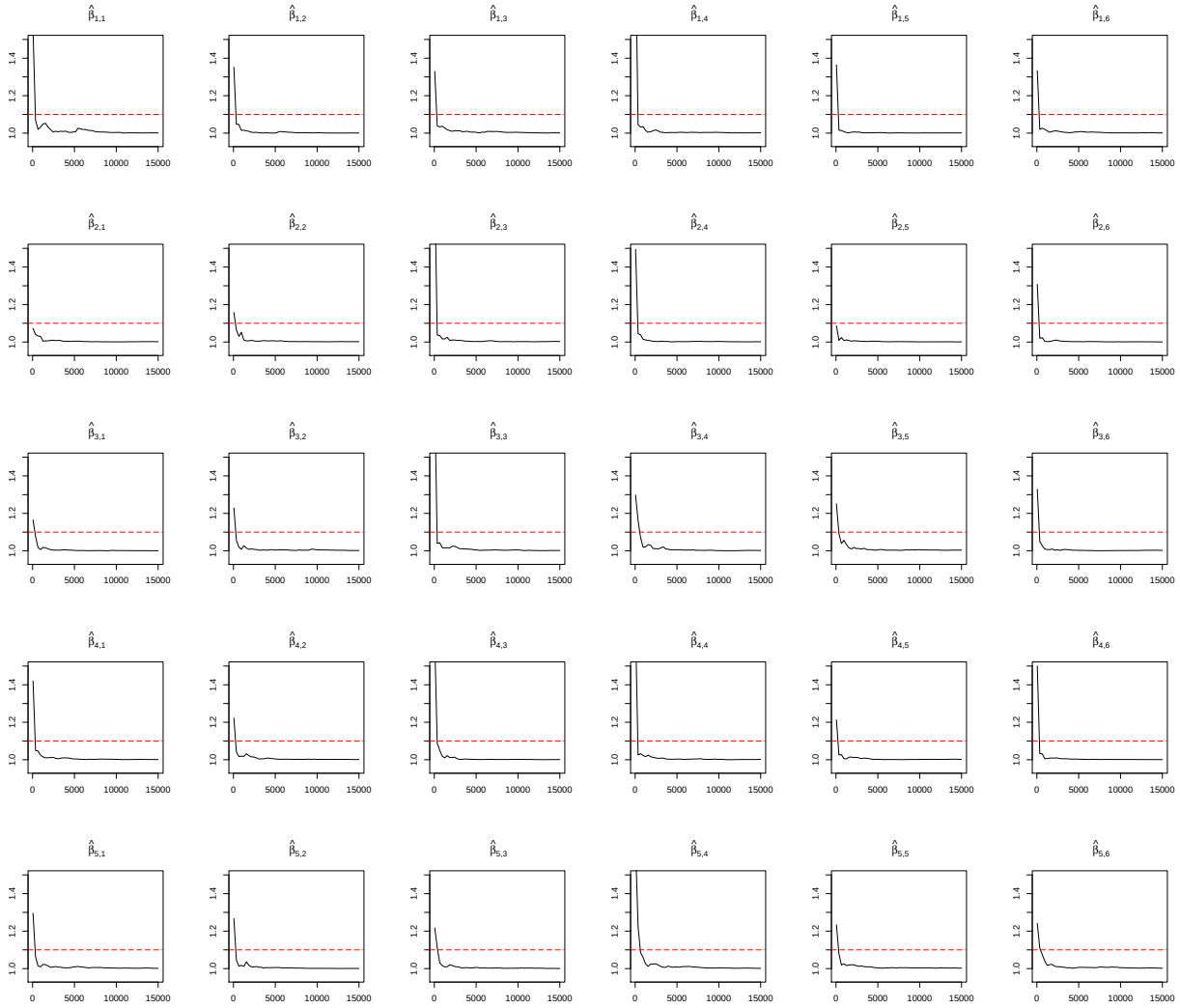

Figure C.1: *Potential scale reduction factor (PSRF) against iterations for the first five groups of coefficients in Example 1. Black line: the PSRF. Red line: the threshold of 1.1. The $\hat{\beta}_{j1}$ to $\hat{\beta}_{j6}$ represent the six estimated coefficients for the main and interaction effects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

C.4 Additional simulation results

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

		RBSG-SS	RBG-SS	RBL-SS	BSG-SS	BG-SS	BL-SS
Error 1	TP	24.87(0.35)	25.00(0.00)	24.53(0.51)	24.83(0.38)	25.00(0.00)	24.53(0.51)
N	FP	1.63(1.16)	31.40(3.38)	2.30(1.86)	1.13(1.04)	29.20(1.10)	0.60(0.85)
	Pred	0.85(0.03)	0.86(0.03)	0.86(0.03)	1.09(0.06)	1.13(0.06)	1.10(0.07)
Error 2	TP	22.23(1.76)	24.67(0.76)	19.23(1.72)	19.97(1.63)	24.47(0.90)	15.27(1.91)
L	FP	1.90(1.30)	35.73(7.83)	2.10(1.40)	2.33(1.42)	34.13(7.44)	1.73(1.39)
	Pred	2.24(0.14)	2.18(0.11)	2.38(0.16)	10.21(1.27)	9.13(0.94)	11.30(1.83)
Error 3	TP	21.50(1.48)	25.00(0.00)	17.43(2.13)	18.73(2.02)	25.00(0.00)	13.10(1.54)
Mix.L	FP	2.13(1.14)	35.20(6.77)	1.90(1.37)	2.90(1.71)	34.00(6.88)	1.37(0.96)
	Pred	2.39(0.18)	2.29(0.11)	2.52(0.22)	12.46(1.67)	10.40(0.94)	13.04(1.35)
Error 4	TP	23.58(1.49)	25.00(0.00)	21.04(2.29)	15.94(5.34)	23.18(3.50)	12.08(4.60)
t2	FP	0.80(0.93)	30.32(3.27)	0.78(1.07)	7.46(27.02)	53.50(58.05)	3.56(8.91)
	Pred	1.85(0.16)	1.82(0.13)	1.92(0.17)	25.65(55.13)	25.63(67.60)	30.67(87.77)
Error 5	TP	24.12(1.00)	25.00(0.00)	21.82(1.90)	18.04(3.64)	24.24(1.88)	13.12(2.99)
logNor	FP	0.90(1.02)	29.48(1.64)	0.82(0.90)	2.72(1.75)	36.12(12.21)	1.48(1.25)
	Pred	1.81(0.13)	1.82(0.12)	1.89(0.15)	14.85(6.53)	12.87(5.94)	15.19(6.43)

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

		RBSG	RBG	RBL	BSG	BG	BL
Error 1	TP	24.20(0.61)	25.00(0.00)	24.23(0.57)	24.33(0.61)	25.00(0.00)	24.30(0.60)
N	FP	2.93(1.86)	54.80(17.34)	3.30(1.97)	1.87(1.61)	56.20(10.30)	5.77(2.56)
	Pred	1.14(0.05)	1.32(0.06)	1.15(0.05)	1.74(0.12)	2.13(0.14)	2.01(0.15)
Error 2	TP	14.00(2.27)	22.20(2.33)	13.63(2.66)	13.70(2.29)	23.63(1.61)	14.20(1.97)
L	FP	0.60(0.85)	31.40(12.07)	0.83(1.05)	1.33(1.18)	62.77(24.90)	8.80(4.54)
	Pred	2.57(0.13)	2.77(0.14)	2.58(0.14)	12.18(1.15)	14.42(1.40)	14.91(1.43)
Error 3	TP	12.40(2.03)	22.47(1.17)	12.27(1.87)	12.43(1.77)	23.20(1.49)	13.37(2.13)
Mix.L	FP	0.57(0.77)	29.33(5.54)	0.60(0.93)	1.47(1.31)	59.80(13.17)	8.17(3.04)
	Pred	2.69(0.11)	2.86(0.11)	2.69(0.10)	13.59(1.05)	16.32(1.46)	16.93(1.60)
Error 4	TP	15.98(2.92)	23.04(2.78)	16.10(3.12)	10.20(5.31)	20.52(5.81)	11.08(5.00)
t2	FP	0.26(0.53)	27.36(6.21)	0.30(0.65)	2.34(3.56)	65.04(30.70)	9.38(6.26)
	Pred	2.19(0.16)	2.35(0.16)	2.21(0.17)	26.27(53.26)	34.08(78.47)	34.95(79.04)
Error 5	TP	16.48(2.69)	23.48(1.58)	16.30(2.63)	11.96(3.66)	22.26(3.70)	12.70(3.50)
logNor	FP	0.32(0.59)	28.72(5.89)	0.34(0.59)	1.40(1.21)	62.34(20.52)	8.34(3.77)
	Pred	2.20(0.14)	2.41(0.14)	2.20(0.13)	15.79(5.97)	18.90(6.61)	19.53(6.65)

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

		RBSG-SS	RBG-SS	RBL-SS	BSG-SS	BG-SS	BL-SS
Error 1	TP	24.00(0.91)	25.00(0.00)	22.13(1.57)	24.33(0.66)	25.00(0.00)	22.83(1.37)
N	FP	1.85(1.46)	33.40(5.21)	1.87(1.36)	1.27(1.34)	29.60(1.83)	0.77(1.04)
	Pred	0.86(0.03)	0.86(0.03)	0.89(0.03)	1.11(0.07)	1.13(0.07)	1.17(0.10)
Error 2	TP	17.63(2.37)	24.73(0.69)	14.37(2.54)	15.00(2.32)	23.73(1.46)	11.00(1.95)
L	FP	2.50(1.41)	33.27(5.14)	2.67(1.99)	2.60(1.43)	30.67(6.69)	1.87(1.53)
	Pred	2.33(0.12)	2.15(0.10)	2.40(0.17)	10.37(1.01)	9.01(0.81)	10.71(0.94)
Error 3	TP	17.23(1.77)	24.80(0.61)	14.47(2.21)	15.10(2.29)	23.67(1.77)	11.03(1.38)
Mix.L	FP	2.27(1.78)	32.20(6.94)	1.63(1.43)	2.13(1.70)	30.93(5.48)	1.17(1.34)
	Pred	2.39(0.13)	2.24(0.10)	2.45(0.13)	11.98(1.45)	10.32(1.04)	12.37(1.41)
Error 4	TP	23.63(1.19)	24.67(0.92)	20.13(2.19)	15.07(4.69)	22.67(3.68)	11.40(4.01)
t2	FP	1.30(1.12)	29.13(2.67)	1.17(0.95)	3.37(1.88)	29.93(9.48)	2.37(1.97)
	Pred	1.48(0.13)	1.45(0.11)	1.55(0.14)	12.66(12.40)	10.10(8.77)	12.75(11.83)
Error 5	TP	24.80(0.48)	25.00(0.00)	23.57(1.43)	20.30(2.83)	24.87(0.51)	15.87(2.32)
logNor	FP	0.33(0.55)	29.60(1.83)	0.40(1.04)	3.00(1.66)	32.93(5.86)	2.33(1.63)
	Pred	1.19(0.10)	1.21(0.10)	1.21(0.11)	6.055(1.77)	5.47(1.73)	6.54(1.75)

Table C.6: *Simulation results in Example 3.* $(n, q, k, p) = (500, 2, 5, 100)$. *mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.*

		RBSG	RBG	RBL	BSG	BG	BL
Error 1	TP	21.27(1.17)	25.00(0.00)	21.30(1.06)	22.23(0.94)	25.00(0.00)	22.13(1.28)
N	FP	1.97(1.56)	45.40(11.68)	2.03(1.56)	1.23(1.33)	43.60(10.41)	3.37(2.03)
	Pred	1.08(0.04)	1.21(0.05)	1.07(0.04)	1.57(0.12)	1.87(0.13)	1.79(0.13)
Error 2	TP	17.63(2.37)	24.73(0.69)	14.37(2.54)	15.00(2.32)	23.73(1.46)	11.00(1.95)
L	FP	2.50(1.41)	33.27(5.14)	2.67(1.99)	2.60(1.43)	30.67(6.69)	1.87(1.53)
	Pred	2.33(0.12)	2.15(0.10)	2.40(0.17)	10.37(1.01)	9.01(0.81)	10.71(0.94)
Error 3	TP	8.43(2.18)	16.70(3.29)	8.70(2.00)	7.97(2.04)	18.60(3.07)	8.27(1.78)
Mix.L	FP	0.33(0.71)	17.70(4.60)	0.43(0.73)	0.60(0.72)	33.60(10.63)	3.70(2.34)
	Pred	2.54(0.11)	2.69(0.12)	2.55(0.11)	12.33(1.15)	14.30(1.40)	14.55(1.40)
Error 4	TP	13.77(2.18)	21.20(2.06)	13.67(2.04)	9.67(3.74)	20.60(4.79)	9.77(3.76)
t2	FP	0.43(0.63)	22.80(3.98)	0.57(0.63)	1.03(1.13)	38.00(12.21)	4.47(2.50)
	Pred	1.73(0.12)	1.85(0.13)	1.73(0.13)	11.78(9.05)	13.94(11.84)	14.22(12.41)
Error 5	TP	19.10(1.86)	24.87(0.73)	19.10(1.60)	15.27(2.94)	24.07(1.70)	15.07(2.88)
logNor	FP	0.20(0.48)	31.13(5.96)	0.23(0.57)	1.10(1.16)	43.93(11.82)	3.83(2.21)
	Pred	1.45(0.08)	1.61(0.09)	1.46(0.08)	6.13(1.13)	7.19(1.24)	7.16(1.29)

Table C.7: *Simulation results in Example 4.* $(n, q, k, p) = (500, 2, 5, 100)$. *mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.*

		RBSG-SS	RBG-SS	RBL-SS	BSG-SS	BG-SS	BL-SS
Error 1	TP	24.93(0.37)	25.00(0.00)	24.93(0.25)	25.00(0.00)	25.00(0.00)	24.90(0.31)
N	FP	1.33(0.99)	30.60(3.84)	1.47(1.25)	1.00(1.02)	29.20(1.10)	0.33(0.61)
	Pred	0.84(0.02)	0.88(0.02)	0.85(0.03)	1.10(0.04)	1.20(0.06)	1.11(0.05)
Error 2	TP	20.80(2.65)	23.60(1.47)	17.24(2.96)	18.58(3.46)	23.04(1.64)	14.08(3.26)
L	FP	1.32(1.22)	30.76(4.93)	1.66(1.29)	1.98(1.53)	27.72(5.49)	1.42(1.25)
	Pred	2.25(0.11)	2.22(0.08)	2.37(0.12)	10.32(1.25)	9.53(0.75)	11.43(1.20)
Error 3	TP	20.56(2.73)	23.69(1.38)	16.53(3.20)	17.56(3.49)	22.80(1.65)	12.67(3.39)
Mix.L	FP	1.40(1.30)	30.04(5.46)	1.78(1.82)	1.76(1.28)	27.60(5.24)	1.22(1.43)
	Pred	2.38(0.13)	2.35(0.10)	2.51(0.16)	12.04(1.40)	11.12(0.96)	13.32(1.44)
Error 4	TP	24.60(0.93)	24.67(0.92)	23.77(1.57)	20.10(6.38)	22.27(5.10)	15.63(6.69)
t2	FP	0.40(0.56)	29.13(2.97)	0.47(0.73)	1.83(1.90)	28.13(9.22)	1.17(1.15)
	Pred	1.48(0.09)	1.52(0.09)	1.51(0.11)	11.54(6.94)	11.33(6.95)	12.64(6.74)
Error 5	TP	23.16(1.68)	24.96(0.28)	19.60(2.14)	15.64(3.76)	23.44(1.83)	11.60(2.75)
logNor	FP	1.08(1.16)	29.16(1.33)	0.72(0.83)	2.20(1.83)	30.56(7.43)	1.48(1.43)
	Pred	1.56(0.14)	1.53(0.13)	1.63(0.15)	10.98(5.80)	9.45(5.38)	11.38(6.03)

Table C.8: *Simulation results in Example 4.* $(n, q, k, p) = (500, 2, 5, 100)$. *mean(sd) of true positives (TP), false positives (FP) and prediction errors (Pred) based on 100 replicates.*

		RBSG	RBG	RBL	BSG	BG	BL
Error 1	TP	21.47(1.87)	24.40(1.07)	21.67(1.81)	22.70(1.64)	24.87(0.51)	22.53(1.85)
N	FP	3.17(2.51)	56.00(20.03)	3.33(2.59)	2.30(1.66)	66.33(14.04)	6.57(2.62)
	Pred	1.26(0.06)	1.44(0.07)	1.27(0.06)	2.40(0.27)	2.83(0.24)	2.75(0.30)
Error 2	TP	9.08(2.54)	19.38(3.12)	9.20(2.60)	9.68(2.41)	20.82(2.93)	10.80(2.65)
L	FP	0.78(0.86)	30.30(10.26)	0.84(0.89)	2.18(1.48)	65.94(19.60)	8.62(3.38)
	Pred	2.67(0.08)	2.89(0.09)	2.67(0.09)	13.54(0.87)	16.38(1.19)	16.87(1.27)
Error 3	TP	8.51(2.31)	18.71(3.37)	8.62(2.33)	9.02(2.33)	20.60(2.76)	10.58(2.50)
Mix.L	FP	0.56(0.69)	25.29(7.94)	0.56(0.66)	1.87(1.36)	56.87(15.55)	7.38(2.91)
	Pred	2.79(0.11)	3.00(0.12)	2.79(0.12)	15.34(1.29)	18.66(1.69)	19.30(1.88)
Error 4	TP	13.30(3.32)	21.93(2.72)	13.47(3.33)	10.93(4.30)	20.97(4.76)	11.97(4.43)
t2	FP	0.50(0.57)	29.07(9.28)	0.40(0.50)	1.70(1.39)	60.03(20.56)	7.90(3.92)
	Pred	2.03(0.12)	2.22(0.12)	2.03(0.12)	15.20(7.98)	18.61(12.19)	19.41(13.39)
Error 5	TP	14.38(2.64)	22.36(2.22)	14.40(2.70)	10.12(3.53)	20.84(3.74)	10.56(3.39)
logNor	FP	0.30(0.58)	25.16(4.36)	0.22(0.46)	0.88(1.12)	36.52(12.60)	3.70(2.57)
	Pred	1.84(0.15)	1.99(0.17)	1.84(0.15)	11.23(5.54)	13.02(5.80)	13.18(5.90)

C.5 Estimation results for data analysis

Table C.9: Analysis of the NHS T2D data using RBSG-SS.

SNP	Gene		chol	act	gl	ceraf	alcohol
			3.503	-3.447	0.752	-3.364	-2.639
rs10741150	DOCK1		-0.948				
rs10765059	TCERG1L	-0.531					0.877
rs10786611	RF00019	0.668	0.723		0.530		
rs10884466	RNA5SP326	-0.466		0.643			
rs10885423	NRG3				-0.715		
rs10886442	GRK5		0.805				
rs11196539	NRG3				-0.608		-0.801
rs11198590	CACUL1	-0.494				0.994	-0.687
rs11259039	FRMD4A	1.016					
rs1194657	THAP12P3				0.798		
rs1219508	RPS15AP5	-0.742					
rs12265854	SLC16A12	0.397					
rs12414552	TCERG1L	0.667		0.470	0.585		-0.690
rs12767723	SLC25A18P1	0.820				-0.515	
rs12772559	TACR2				0.938	0.510	
rs12774333	LRMDA	-0.599					
rs12775160	FOXI2	-0.651	-0.501				0.647
rs16916794	SLC39A12	-0.552	0.511	0.455			
rs16920092	PLXDC2						-0.843
rs17094114	GFRA1	-0.615					

Continued on the next page

Table C.9: Continued from the previous page.

SNP	Gene		chol	act	gl	ceraf	alcohol
rs2492664	OR6L1P	0.695			-0.737		
rs2784767	PLAC9	-0.540					
rs2814322	GRID1				-0.830		
rs3740063	ABCC2				-0.966		
rs3763722	LARP4B	0.332	-1.156				0.866
rs4411238	PRKG1	0.537					
rs4578341	CHST15		-0.822			0.602	
rs4747517	ITIH5		-1.468			0.920	
rs4749926	IL2RA	-0.840	-0.815				
rs4917817	PYROXD2	-0.624		0.594			
rs4918904	XRCC6P1					0.997	
rs6482836	DOCK1	-0.957		1.067			
rs7070789	GPAM				-1.245		-0.791
rs7072255	ANTXRLP1		0.800				
rs7077721	SNRPD2P1	0.858		0.774			
rs7896554	NACAP2	0.840		-0.630	-0.565		
rs7897847	LGI1						0.962
rs870753	CFAP58				-0.783		
rs881726	GFRA1						1.001

Table C.10: Analysis of the NHS T2D data using RBL-SS.

SNP	Gene		chol	act	gl	ceraf	alcohol
			1.354	-0.430	-0.778	-2.424	-4.000
rs1041168	PLPP4	0.463					
rs10741150	DOCK1		-1.126				
rs10786611	RF00019	0.632					
rs10794069	ADAM12	0.524					
rs10824802	MBL2	0.553					
rs10884466	RNA5SP326	-0.439		0.503			
rs10885423	NRG3				-1.060		
rs10886047	MIR3663HG						-0.410
rs10886442	GRK5		1.087				
rs10998780	ATP5MC1P7				0.150		
rs11003665	RNA5SP318		0.632				
rs11013740	KIAA1217						0.852
rs11196539	NRG3						-0.624
rs11198590	CACUL1	-0.628					
rs11202221	BMPR1A				0.815		
rs11259039	FRMD4A	1.021					
rs11595123	AKR1E2			1.079			
rs11813505	KIAA1217					1.301	
rs1194657	THAP12P3				0.663		
rs1219508	RPS15AP5	-0.886					
rs12265854	SLC16A12	0.596					

Continued on the next page

Table C.10: Continued from the previous page.

SNP	Gene		chol	act	gl	ceraf	alcohol
rs12269237	RF00017					0.884	
rs12414552	TCERG1L	0.594					
rs12414627	PNLIPRP1				-0.551		
rs12767723	SLC25A18P1	0.962					
rs12772559	TACR2				0.906		
rs12774333	LRMDA	-0.449					
rs12775160	FOXI2	-0.560					
rs1573137	SORCS3				0.615		
rs16916794	SLC39A12	-0.803	0.528				
rs16920092	PLXDC2						-0.655
rs17094114	GFRA1	-0.563					
rs2291314	PLPP4	0.536					
rs2420979	TACC2					-1.091	
rs2492664	OR6L1P	0.655			-0.363		
rs2664339	RNU6-543P	-0.501					
rs2666236	IATPR					0.689	
rs2784767	PLAC9	-0.452	0.730				
rs2814322	GRID1				-0.806		
rs2842129	DYNC1I2P1	-0.662					
rs2900814	SNRPD2P1	-0.643					
rs3740063	ABCC2				-0.885		
rs3763722	LARP4B						1.036
rs4411238	PRKG1	0.399					

Continued on the next page

Table C.10: Continued from the previous page.

SNP	Gene		chol	act	gl	ceraf	alcohol
rs4578341	CHST15		-0.582			0.479	
rs4747009	LRRC20						0.710
rs4747517	ITIH5		-0.905				
rs4749926	IL2RA	-0.607					
rs4752432	PLPP4		0.725				
rs4917817	PYROXD2	-0.506					
rs4934762	PCAT5	-0.560					
rs6482836	DOCK1	-0.709					
rs7070789	GPAM				-0.820		
rs7072255	ANTXRLP1		0.811				
rs7077721	SNRPD2P1	0.702					
rs7894809	PCGF5	0.501					
rs7896554	NACAP2	0.850		-0.953			
rs7897847	LGI1						0.929
rs7903853	FRMD4A		-1.185				
rs7920351	TCERG1L			-0.713			
rs881726	GFRA1						0.675
rs943213	DOCK1			-0.939			

Table C.11: Analysis of the NHS T2D data using BSG-SS.

SNP	Gene	chol	act	gl	ceraf	alcohol
		2.045	-2.049	-2.204	-1.796	-4.436

Continued on the next page

Table C.11: Continued from the previous page.

SNP	Gene		chol	act	gl	ceraf	alcohol
rs1041168	PLPP4	0.638					
rs10765059	TCERG1L			0.709			
rs10786611	RF00019	0.773	0.556				
rs10829671	EBF3	-0.505					
rs10884466	RNA5SP326	-0.563		0.625			
rs10886442	GRK5		1.038				
rs10998780	ATP5MC1P7				0.704		
rs11017821	TCERG1L						0.665
rs11198590	CACUL1	-0.698			0.905	0.568	
rs11200996	CCSER2					0.508	
rs11259039	FRMD4A	1.174					
rs1219508	RPS15AP5	-0.787					
rs12265854	SLC16A12	0.681			-0.494		
rs12269237	RF00017					0.684	
rs12414552	TCERG1L	0.480					
rs12764378	ARID5B	-0.420					
rs12767723	SLC25A18P1	0.638				-0.559	
rs12775160	FOXI2	-0.762					0.893
rs1361709	PCDH15			-0.852			
rs1395465	RN7SL63P	0.292					
rs16916794	SLC39A12	-0.614	0.580	0.622			
rs16920092	PLXDC2						-0.692
rs17094114	GFRA1	-0.676					

Continued on the next page

Table C.11: Continued from the previous page.

SNP	Gene		chol	act	gl	ceraf	alcohol
rs17469499	KIAA1217					-0.527	
rs2472737	RET	0.629					
rs2577356	GFRA1					0.875	
rs2784767	PLAC9	-0.569	0.537				
rs2792708	GPAM	0.488					
rs2900814	SNRPD2P1	-0.460					
rs2926458	RNU6-463P	-0.680					
rs3763722	LARP4B		-1.251				1.186
rs4411238	PRKG1	0.619					
rs4747517	ITIH5		-1.257				
rs4752432	PLPP4		0.956				
rs4917817	PYROXD2	-0.626		0.630			
rs4922535	GDF10			-0.601		-0.649	
rs4934762	PCAT5	-0.640					
rs4934858	NRP1	0.281					
rs6482836	DOCK1	-0.773					
rs7070789	GPAM				-0.642		
rs7072255	ANTXRLP1		0.723				
rs7085788	RHOBTB1	-0.720					
rs7086058	RN7SKP143	-0.507					
rs716168	VTI1A		-0.570				
rs7894809	PCGF5	0.642					
rs7895870	RN7SKP167		-0.867				

Continued on the next page

Table C.11: Continued from the previous page.

SNP	Gene		chol	act	gl	ceraf	alcohol
rs7896554	NACAP2	1.097		-0.477			
rs7917422	HTR7						0.794
rs881726	GFRA1						0.933

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

SNP	Gene		chol	act	gl	ceraf	alcohol
			3.095	-2.406	-2.373	-1.716	-3.721
rs1041168	PLPP4	0.670					
rs10508670	KIAA1217		0.773				
rs10765059	TCERG1L			0.445			
rs10829671	EBF3	-0.717					
rs10884466	RNA5SP326	-0.528					
rs10998780	ATP5MC1P7				1.195		
rs11017821	TCERG1L						0.307
rs11198590	CACUL1				1.273		
rs11200996	CCSER2					0.509	
rs11202221	BMPR1A				0.954		
rs11259039	FRMD4A	1.020					
rs11594070	ATE1-AS1				-0.401		
rs1194657	THAP12P3				0.681		
rs12248205	CDH23				-0.938		
rs12256982	ZMIZ1				0.152		
rs12265854	SLC16A12	0.661			-0.830		

Continued on the next page

Table C.12: Continued from the previous page.

SNP	Gene		chol	act	gl	ceraf	alcohol
rs12269237	RF00017					0.864	
rs12412976	RPLP1P10			0.592	-0.590		
rs12414552	TCERG1L	0.549					
rs12414627	PNLIPRP1				-0.572		
rs12764378	ARID5B	-0.564					
rs12767723	SLC25A18P1	1.062					
rs12775160	FOXI2	-0.636					
rs1361709	PCDH15			-0.729			
rs1395465	RN7SL63P	0.562					
rs1573137	SORCS3				0.869		
rs16916794	SLC39A12	-0.430	0.862				
rs16920092	PLXDC2						-0.508
rs17094114	GFRA1	-0.734					
rs17469499	KIAA1217					-0.680	
rs2384105	SNRPEP8						-0.738
rs2420979	TACC2	-0.629					
rs2472737	RET	0.553					
rs2577356	GFRA1					0.739	
rs2784767	PLAC9	-0.593	0.576				
rs2792708	GPAM	0.568					
rs2900814	SNRPD2P1					-0.157	
rs2926458	RNU6-463P	-0.527					
rs3763722	LARP4B		-1.002				1.151

Continued on the next page

Table C.12: Continued from the previous page.

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

Table C.13: Analysis of the TCGA SKCM data using RBSG-SS.

Gene		clark	stage	age	gender
		0.834	0.228	-0.116	-0.183
AHNAKRS	0.107				
ANKRD28	0.134		0.138		
ASH2L		-0.297			
BTD		-0.312			
C1ORF140	-0.002	0.246	-0.083	-0.022	0.092
CD44					0.070
CHP1	0.107	0.045			
CXCL6	0.126	-0.120	-0.095		
DLG6	0.113	-0.015	0.067	0.185	-0.142
DOK5				-0.066	
ETNK2	0.152				
FILIP1	-0.030				
JADE1	-0.147				
JPH4		0.115			
KBF2	-0.032	0.182		0.034	-0.026
LRRN2	-0.061				
MAGED4	-0.098		-0.020		
NHSL2	-0.088				
PITPNA	0.151	-0.051	-0.012	-0.033	0.008
SOX8	0.088		-0.212		
TMEM145					0.048

Continued on the next page

Table C.13: Continued from the previous page.

Gene		clark	stage	age	gender
TMEM159	0.160	-0.121	-0.042		0.189
WBSCR27		0.070		0.126	

Table C.14: Analysis of the TCGA SKCM data using RBL-SS.

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

Continued on the next page

Table C.14: Continued from the previous page.

Gene		clark	stage	age	gender
GAMT				0.082	
IL11RA	-0.087				
IQCK					-0.090
JADE1	-0.161				
JPH4		0.159			
KDM6B			-0.142		
LRFN2		0.096			
MAGED4	-0.130				
MAPE					-0.191
MPD1	-0.078				
NHSL2	-0.144	-0.306			
PAX1	0.171		0.217		
PBX2	0.141			0.130	
PITPNA	0.161			-0.056	
RNPEPL1			0.052		
SLC12A5					-0.081
SOX8	0.140		-0.091		
STPG1			0.184		
TMEM145					0.222
TMEM159	0.123				
TNFAIP1				0.283	
TP53TG1		0.102			-0.063
WBSCR27		0.090		0.126	

Table C.15: Analysis of the TCGA SKCM data using BSG-SS.

Gene		clark	stage	age	gender
		0.487	0.163	0.048	0.087
AHNAKRS	0.120				
ANKRD28	0.138				
ARMC9	0.008				
ASH2L	0.019	-0.194		-0.107	
BTD		-0.303		-0.138	
C14ORF2		0.251			
C1ORF140		0.100	0.024		0.029
CD44					0.125
CHP1	0.123				
CPXM1	-0.047				
CXCL6	0.032				
DLG6		0.093		0.204	-0.061
DOK5			-0.052		
ETNK2	0.094				
FILIP1	-0.049				
GAMT		-0.004			
IL11RA	-0.045				
JADE1	-0.149				
JPH4		0.110			
KBF2	-0.077				
LRRN2	-0.073				

Continued on the next page

Table C.15: Continued from the previous page.

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

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

Gene		clark	stage	age	gender
		0.545	0.308	0.080	0.047
AHNAKRS	0.102				
ANKRD28	0.180		0.134		
ASH2L				-0.185	
BTD		-0.386			

Continued on the next page

Table C.16: Continued from the previous page.

Gene		clark	stage	age	gender
C14ORF2		0.126			
C1ORF140		0.199			
CELSR2			0.112		
CHP1	0.080				
CPXM1	-0.067				
CSNK2A2	-0.026				
CYP1B1-AS1			0.104		
DAP				-0.139	
DLG6	0.088			0.236	
ETNK2	0.206				-0.089
FHL5		0.076			
FILIP1			-0.062		
GAMT				0.058	
IL11RA	-0.056				
IQCK					-0.098
JADE1	-0.203				
JPH4		0.101			
KBF2	-0.089				
KDM6B			-0.173		
LRFN2		0.109			
LRRN2	-0.091				
MAGED4	-0.113				
MAPE					-0.114

Continued on the next page

Table C.16: Continued from the previous page.

Gene		clark	stage	age	gender
MPD1	-0.100				
NHSL2	-0.035				
PAX1			0.050		
PBX2	0.126			0.072	
PHP1B					-0.054
PIP4K2C			-0.101		
PITPNA	0.193				
PTP4A3			-0.138		
RNPEPL1			0.171		
SAA2	0.021			-0.058	
SLC12A5					-0.112
SOX8	0.132		-0.084		
TIE1	-0.093				
TMEM145					0.188
TMEM159	0.174				0.181
TP53TG1		0.156			-0.030
WBSCR27		0.048		0.105	

C.6 Biological similarity analysis

We carried out an examination of the Gene Ontology (GO) biological processes which provide us with a deeper insight on the differences of the markers identified by different methods. We totally identified 77 unique genes using our proposed method along with three other methods for the NHS data. We conducted the GO enrichment analysis using the R package GOSim and found these genes involve in a total of 158 GO biological processes, the p-values of which

are smaller than 0.1 in the GO enrichment analysis. Then we divided the 158 processes into four categories: positive regulation (P), negative regulation (N), regulation (R, without a well-defined direction) and other (O). We computed the proportions of genes that involve in the four categories of processes for each of the four methods. Similarly for the TCGA SKCM data, 109 genes were identified by our method along with three other alternative methods. GO enrichment analysis showed that they involve in 183 biological processes, with p-values smaller than 0.1. The results for NHS and TCGA SKCM are provided in Figure C.2, which shows an obvious difference between our proposed method and the three alternatives in both datasets.

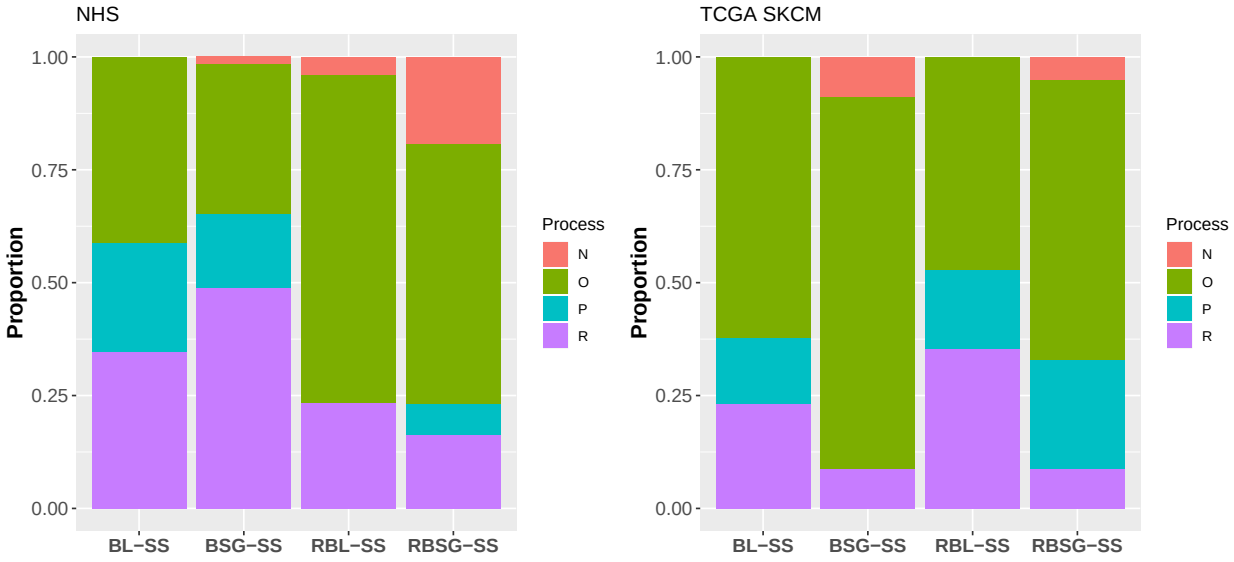

Figure C.2: *Gene Ontology (GO) analysis: proportions of genes that have the four categories of processes with different approaches. Left: NHS data. Right: TCGA SKCM data.*

C.7 Posterior inference

C.7.1 RBG-SS

Hierarchical model specification

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

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

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

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

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

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

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

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

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

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

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

Gibbs Sampler

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

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

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

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

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

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

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

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

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

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

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

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

- The posterior of s_j is

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

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

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

C.7.2 RBL-SS

Hierarchical model specification

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

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

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

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

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

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

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

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

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

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

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

Gibbs Sampler

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

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

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

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

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

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

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

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

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

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

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

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

- The posterior of s_{jl} is

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

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

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

C.7.3 RBSG

Hierarchical model specification

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

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

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

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

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

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

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

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

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

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

Gibbs Sampler

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

C.7.4 RBG

Hierarchical model specification

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

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

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

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

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

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

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

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

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

Gibbs Sampler

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

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

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

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

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

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

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

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

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

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

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

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

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

C.7.5 RBL

Hierarchical model specification

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

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

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

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

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

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

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

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

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

Gibbs Sampler

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

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

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

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

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

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

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

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

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

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

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

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

C.7.6 BSG-SS

Hierarchical model specification

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

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

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

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

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

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

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

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

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

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

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

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

Gibbs Sampler

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

- $\pi_1|\text{rest} \sim \text{Beta}\left(a_1 + \sum_{j,l} \mathbf{I}_{\{\omega_{jl} \neq 0\}},\ b_1 + \sum_{j,l} \mathbf{I}_{\{\omega_{jl}=0\}}\right)$
- η is estimated with the EM approach used in the proposed method. For the gth EM update $\eta^{(g)} = \frac{1}{E_{\eta^{(g-1)}}\left[\frac{1}{s^2}|Y\right]}$.
- $\sigma^2|\text{rest} \sim \text{Inv-Gamma}(\frac{n}{2},\ \frac{\sum_{i=1}^{n} \tilde{y_i}^2}{2})$, where $\tilde{y_i} = y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta$.

C.7.7 BGL-SS

Hierarchical model specification

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

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

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

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

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

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

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

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

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

Gibbs Sampler

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

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

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

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

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

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

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

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

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

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

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

- The posterior of s_j is

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

- $\pi_0|\text{rest} \sim \text{Beta}\left(a_0 + \sum_{j=1}^{p} \mathbf{I}_{\{\beta_j \neq 0\}},\ b_0 + \sum_{j=1}^{p} \mathbf{I}_{\{\beta_j = 0\}}\right)$
- $\eta|\text{rest} \sim \text{Gamma}\,(s_\eta,\, r_\eta)$, where $s_\eta = \frac{p+p\times L}{2} + d_1$ and the rate parameter $r_\eta = \frac{\sum_{j=1}^{p} s_j}{2} + d_2$.
- $\sigma^2|\text{rest} \sim \text{Inv-Gamma}(\frac{n+L\sum_{j=1}^{p} \mathbf{I}_{\{\beta_j \neq 0\}}}{2},\ \frac{\sum_{i=1}^{n} \tilde{y_i}^2 + \sum_{j=1}^{p} (s_j)^{-1} \beta_j^\top \beta_j}{2})$, where $\tilde{y_i} = y_i - W_i^\top \alpha - E_i^\top \theta - U_i^\top \beta$.

C.7.8 BL-SS

Hierarchical model specification

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

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

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

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

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

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

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

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

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

Gibbs Sampler

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

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

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

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

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

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

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

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

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

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

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

- The posterior of s_{jl} is

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

- $\pi_1|\text{rest} \sim \text{Beta}\left(a_1 + \sum_{j,l} \mathbf{I}_{\{\beta_{jl}\neq 0\}},\ b_1 + \sum_{j,l} \mathbf{I}_{\{\beta_{jl}=0\}}\right)$
- $\eta|\text{rest} \sim \text{Gamma}\,(s_\eta,\, r_\eta)$, where $s_\eta = p\times L + d_1$ and the rate parameter $r_\eta = \frac{\sum_{j,l} s_{jl}}{2} + d_2$.
- $\sigma^2|\text{rest} \sim \text{Inv-Gamma}(\frac{n+\sum_{j,l}\mathbf{I}_{\{\beta_{jl}\neq 0\}}}{2},\ \frac{\sum_{i=1}^n \tilde{y_i}^2 + \sum_{j,l}(s_{jl}^{-1})\beta_{jl}^2}{2})$, where $\tilde{y_i} = y_i - W_i^\top\alpha - E_i^\top\theta - U_i^\top\beta$.

C.7.9 BSG

Hierarchical model specification

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

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

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

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

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

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

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

Gibbs Sampler

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

C.7.10 BGL

Hierarchical model specification

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

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

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

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

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

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

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

Gibbs Sampler

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

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

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

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

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

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

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

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

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

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

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

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

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

C.7.11 BL

Hierarchical model specification

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

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

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

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

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

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

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

Gibbs Sampler

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

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

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

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

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

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

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

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

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

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

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

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

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