Spring 2023

Bayesian Dependence Structure Analysis for Ordinal Data

Yang He

Bayesian Dependence Structure Analysis For Ordinal Data

by

Yang He

Bachelor of Science
Central South University, China 2012

Master of Science
West Virginia University, 2017

Master of Science
West Virginia University, 2018

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Statistics

College of Arts and Sciences

University of South Carolina

2023

Accepted by:

Xiaoyan Lin, Major Professor

Lianming Wang, Committee Member

Ray Bai, Committee Member

Xuemei Sui , Committee Member

Cheryl Addy, Interim Vice Provost and Dean of the Graduate School

Abstract

This dissertation explores different methods to study the dependence structure among many ordinal variables under the Bayesian framework.

Chapter 1 introduces ordinal data analysis methods, and the related literature works are briefly reviewed. An outline of the dissertation is put forward.

In Chapter 2, Gaussian copula graphical models with different priors of graphical Lasso, adaptive graphical Lasso, and spike-and-slab Lasso on the precision matrix are assessed and compared. The proposed models are well illustrated via simulations and a real ordinal survey data analysis.

In Chapter 3, adaptive spike-and-slab Lasso prior is proposed as an extension of Chapter 2. The developed adaptive spike-and-slab prior yields good results based on the simulation study. An improved simulation setting is utilized in this chapter compared to that in Chapter 2. Thus better guidance is achieved for the dependence structure learning in real data analysis.

Chapter 4 applies a Bayesian factor analysis model with Gaussian copula to the TSCC (Trauma Symptom Checklist for Children) ordinal data. The variable structure is investigated with a global-factor-local shrinkage before the factor loading matrix. The results with different numbers of factors are compared, and the corresponding estimates of the covariance matrix are obtained and compared with those obtained using the Gaussian copula graphical models in Chapters 2 and 3.

In Chapter 5, a summary of the studies in the previous chapters is presented, and at the same time, we put forward some ideas for future work.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Ordinal Data

Ordinal data is frequently encountered across numerous domains such as psychology, education, medicine, economics, consumer choice, and various other fields (Jamieson 2004; Carifio and Perla 2007; Vickers 1999; Spranca, Minsk, and Baron 1991; Clason and Dormody 1994; Hui and Bateson 1991; Feldman and Audretsch 1999). An ordinal variable assumes ordered categories as its possible values, where the categories have an ordering, but the distance between categories is hard to determine or measure. The ubiquity of ordinal data is mainly due to the widespread use of Likert-style response items (Likert 1932). A Likert item typically refers to a single question for which the response is indicated on a discrete ordered scale ranging from one qualitative endpoint to another qualitative endpoint, such as from "strongly disagree", "disagree", "no opinion", "agree", to "strongly agree". Response options for Likert items are typically discrete and range from 5 to 11.

In many social science disciplines, it is a common practice to "convert" ordinal Likert or Likert style scale data into interval data by assigning numbers, such as "1" for "very good", "2" for "good", and so on. Although the response options might look numerically labeled, the numerals only indicate the order and do not indicate equal intervals between levels. For example, if the response items include "2" = "sometimes", "3" = "very often", and "4" = "always", we cannot assume that the

increment in the frequency of occurrence from "2" to "3" is the same as the increment in the frequency of occurrence from "3" to "4".

There is a large amount of literature analyzing ordinal data. A popular model to analyze such Likert items is the polychotomous Rasch model (Davier and Carstensen 2007) that obtains interval-level estimates on a continuum—an idea we borrow in this dissertation. Ordinal response variables in regression analysis result in variations of the classic linear model, such as the proportional odds model (Walker and Duncan 1967; Peter McCullagh 1980), the partial proportional odds model (Peterson and Harrell Jr 1990), the probit model (Bliss 1935; Albert and Chib 1993; Chib and Greenberg 1998), Etc. McCullagh and Nelder (1989) and O'Connell (2006) provided a comprehensive overview of ordinal regression.

Ordinal data are an integral part of survey data, where respondents rate items or express a level of agreement/disagreement on topics under consideration. For example, each survey respondent could be asked to respond to a question using categories such as "strongly disagree", "disagree", "undecided", "agree", and "strongly agree". In this dissertation, we analyze survey data about the present state of the physiological and psychological health of current undergraduate students in China. The survey includes six parts: Social Demographic Information, STD/AIDS-Related Knowledge, Sexual Health Behavior, Other Behaviors, Drug Related Knowledge and Life Satisfaction Survey. Most questions are ordinal.

1.2 Gaussian Copula

Consider ordinal variables $\boldsymbol{Y_V}$ with joint distribution F, where there exists an ordering for the possible values of each variable $Y_v,\ v \in \boldsymbol{V}$. We can adopt the Gaussian copula (Nelsen 2007) to model the multivariate dependence among variables $Y_v,\ v \in \boldsymbol{V}$. Assume a latent vector $\boldsymbol{Z_V} \sim \mathbf{N}_p(\mathbf{0}, \boldsymbol{\Gamma})$ with correlation matrix $\boldsymbol{\Gamma}$, such that the two

random vectors are connected as follows:

$$Y_v = F_v^{-1}(\Phi(Z_v)),\ v \in \boldsymbol{V}, \tag{1.1}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution and F_v is the pseudo marginal CDF for each Y_v. In this way, the joint distribution F of $\boldsymbol{Y_V}$ is a function of the correlation matrix $\boldsymbol{\Gamma}$ and the univariate distributions F_v of $Y_v,\ v \in \boldsymbol{V}$:

$$Pr(Y_1 \leq y_1, ..., Y_p \leq y_p) = F(y_1, ..., y_p|\boldsymbol{\Gamma}, F_1, ..., F_p) \tag{1.2}$$

$$= \Phi_p\Big(\Phi^{-1}\Big(F_1(y_1)\Big), ..., \Phi^{-1}\Big(F_p(y_p)\Big)|\boldsymbol{\Gamma}\Big), \tag{1.3}$$

$$\triangleq C(F_1(y_1), ..., F_p(y_p)|\boldsymbol{\Gamma}) \tag{1.4}$$

where $\Phi_p(\cdot|\boldsymbol{\Gamma})$ is the distribution function of $\mathbf{N}_p(\mathbf{0}, \boldsymbol{\Gamma})$. The Gaussian copula is applied later to the ordinal data using the Gaussian graphical models in Chapters 2 and 3 and the factor analysis model in Chapter 4.

1.3 Bayesian Gaussian Graphical Models

Gaussian graphical models (GGMs) have been widely studied to study the dependence structure of continuous variables. Here we only review some of the Bayesian methods. Marlin and Murphy (2009) suggested a Bayesian model and a variational Bayes algorithm for Gaussian graphical models (GGMs) with a block structure. They used the stochastic block model as a prior, which does not require the blocks or groups to be specified a priori. The resulting problem is no longer convex, but they devised an efficient variational Bayes algorithm to solve it. Different priors were studied for the precision matrix by different researchers. Among these priors, G-Wishart priors were most commonly used (Carvalho and Scott 2009; Dobra, Lenkoski, and Rodriguez 2011; Wang and Li 2012; Mohammadi and Wit 2015). Wang (2012) presented a Bayesian approach to graphical Lasso models and their associated posterior computation algorithms, providing a fully Bayesian treatment. Using data augmentation, he developed a block Gibbs sampler for efficiently sampling covariance matrices.

He also generalized the Bayesian graphical Lasso to the Bayesian adaptive graphical Lasso, which showed the top overall performance among a range of frequentist and Bayesian methods in terms of both covariance matrix estimation and graphical structure learning. A Bayesian approach with mixture prior distributions that incorporate a point-mass and a Laplace distribution was examined by Banerjee and Ghosal (2015). They derived posterior consistency results and presented a computational approach using Laplace approximation. Except for the work of Banerjee and Ghosal (2015), the theoretical properties of Bayesian approaches for sparse precision matrix estimation have not been investigated. The results of Banerjee and Ghosal (2015) are on the estimation error rate in Frobenius norm similar to those of Rothman et al. (2008), but assume the underlying distribution to be Gaussian. Gan, Naveen N Narisetty, and Liang (2019) presented a novel Bayesian technique for estimating and recovering the structure of GGMs. They utilized a continuous spike-and-slab prior as a mixture of two Laplace distributions. Their theoretical results hold beyond GGMs despite the Gaussian likelihood being used in their Bayesian formulation. They proposed a fast EM algorithm that produces a maximum a posteriori (MAP) estimate of the precision matrix. Their algorithm was guaranteed to generate an estimator that is both symmetric and positive definite, in contrast to numerous existing estimators. In Chapters 2 and 3 of this dissertation, we will couple the Gaussian copula with the Bayesian Gaussian graphical models to analyze the dependence structure of ordinal variables. We investigate the performance of the Bayesian Gaussian copula graphical model with different priors on the estimation of the precision matrix and the graphical structure learning of ordinal variables.

1.4 Bayesian Factor Analysis Models

Factor analysis models have attracted much attention since they can perform exploratory analyses of the latent linear structure in high-dimensional data and con-

duct dimension reduction (Bernardo et al. 2003; Carvalho et al. 2008; Engelhardt and Stephens 2010).

The loading matrix $\mathbf{\Lambda}$ is of significant importance in a factor model. In high-dimensional data applications where the sample size n is much smaller than the variable dimension p (Bernardo et al. 2003), regularization on the loading matrix is crucial because the optimization problem is under-constrained with $n \ll p$ and has many equivalent solutions that optimize the data likelihood. The statistics literature employs priors or penalties to regularize the elements of the loading matrix and induce sparsity. Elementwise sparsity corresponds to dimension reduction, where a latent factor contributes to variation in only a subset of the observed variables, generating interpretable results (Bernardo et al. 2003; Carvalho et al. 2008; Knowles and Ghahramani 2011).

Elementwise sparsity has been imposed in latent factor models through regularization via l_1 type penalties (Zou, Hastie, and Tibshirani 2006; Witten, Tibshirani, and Hastie 2009; Salzmann et al. 2010). In the last ten years, Bayesian shrinkage methods with sparsity-inducing priors have been introduced for latent factor models (Carvalho et al. 2008; Archambeau and Bach 2008; Virtanen et al. 2011; Bhattacharya and D. B. Dunson 2011; Klami, Virtanen, and Kaski 2013). The spike-and-slab prior (Mitchell and Beauchamp 1988) and the classic two-groups Bayesian sparsity-inducing prior have been used for sparse Bayesian latent factor models (Carvalho et al. 2008). More sophisticated structured regularization approaches have been studied in classical statistics (Zou and Hastie 2005; Kowalski and Torrésani 2009; Huang, Zhang, and Metaxas 2009; Jenatton, Audibert, and Bach 2011).

1.5 Computing Techniques

Performing Bayesian inferences requires using the joint posterior distribution over a set of parameters. Practically speaking, closed-form solutions for the posterior

distribution do not exist, and the intractable integrals are challenging to calculate in most scenarios. Rather than resolving analytical equations, Markov Chain Monte Carlo (MCMC) employs sampling to provide a solution for statistical inferences.

The most commonly used sampling techniques are the Metropolis-Hastings (MH) sampling and the Gibbs sampling, where the latter is a special case of the former (Gelman 1993). Metropolis et al. (1953) were the first to develop the MH algorithm, which was later expanded upon by Hastings (1970). Because of its importance and usefulness, its application (Müller 1991; Chib and Greenberg 1996) is steadily increasing in the literature.

The Gibbs sampler is a special form of MH (Gelman 1993), mainly used to generate random variables from a (marginal) distribution to avoid the direct calculation of density. S. Geman and D. Geman (1984) used this method to study image processing. Gelfand and Smith (1990) reviewed and compared several sampling-based density calculation methods. Casella and George (1992) exploited simple cases to demystify the working mechanism behind the Gibbs Sampler. Since then, the Gibbs sampler approach has been widely used for posterior sampling. The basic idea is to sequentially draw samples for each random variable from the conditional distribution, with the remaining variables fixed to the current values. For a random vector $\boldsymbol{X} \in \mathbb{R}^{K\times 1}$, the algorithm proceeds as follows:

1. Initialize $x^{(0)} \sim q(x)$, where $q()$ is usually a prior distribution;

2. for iteration $i = 1, 2, ...$

$x_1^{(i)} \sim p(X_1 = x_1 | X_2 = x_2^{(i-1)}, X_3 = x_3^{(i-1)}, ..., X_K = x_K^{(i-1)})$

$x_2^{(i)} \sim p(X_2 = x_2 | X_1 = x_1^{(i)}, X_3 = x_3^{(i-1)}, ..., X_K = x_K^{(i-1)})$

...

$x_K^{(i)} \sim p(X_K = x_K | X_1 = x_1^{(i)}, X_2 = x_2^{(i)}, ..., X_{K-1} = x_{K-1}^{(i)})$

3. end.

Chapters 2, 3, and 4 of this dissertation use the Gibbs sampler approach for posterior inferences and the corresponding sampling details are described in each chapter.

Chapter 2

Bayesian Gaussian Copula Graphical Models with Graphical Lasso Priors and Spike-and-Slab Lasso Prior

2.1 Introduction

Over the past several decades, modeling ordinal data has become essential in various fields, especially in social and economic sciences, where natural ordering data commonly appears. For example, survey subjects are often asked to respond to a question using categories such as strongly disagree, disagree, undecided, agree, and strongly agree. Users of an online service could be asked to rate their experience from one star to five stars. These examples clearly show that ordinal data are pervasive in many real-world applications. Graphical models are widely used to analyze the conditional and marginal dependence for continuous variables, but there has been limited work for analyzing ordinal data. Our work will use the Gaussian copula graphical model from a Bayesian perspective to analyze ordinal data. We compare the performance of three different priors on estimating covariance and precision matrices using simulation studies. These priors are utilized to analyze survey data about the physiological and psychological health of current undergraduate students in China.

2.2 Models

2.2.1 Gaussian Graphical Models (GGMs)

Suppose $\boldsymbol{Y_V}$ with $\boldsymbol{V} = \{1, 2, ..., p\}$ is a random vector of continuous random variables Y_v with a joint distribution $p(\boldsymbol{Y_V})$. The conditional dependence relationships among $\{Y_v : v \in \boldsymbol{V}\}$ under $p(\boldsymbol{Y_V})$ can be summarized in a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, where each vertex $v \in \boldsymbol{V}$ corresponds to a random variable Y_v and $\boldsymbol{E} \subset \boldsymbol{V} \times \boldsymbol{V}$ are the undirected edges. In the undirected graphical model $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ (which is also called Markov Random Fields (MRFs) or Markov networks), the absence of an edge between Y_g and Y_h corresponds to the conditional independence of these two random variables given the remaining variables under $p(\boldsymbol{Y_V})$ and is denoted by

$$Y_g \perp\!\!\!\perp Y_h | \boldsymbol{Y}_{\boldsymbol{V}\setminus\{g,h\}},$$

where $g, h \in \boldsymbol{V}$.

Assume that $\boldsymbol{Y_V}$ follows a p-dimensional multivariate normal distribution $\mathbf{N}_p(\mathbf{0}, \boldsymbol{\Omega}^{-1})$ with precision matrix $\boldsymbol{\Omega} = (\omega_{gh})_{1 \le g,h \le p}$. Let $\boldsymbol{y} = (\boldsymbol{y_1}, ..., \boldsymbol{y_n})^T$ be the observed data of n independent samples of $\boldsymbol{Y_V}$. Then the likelihood function of $\boldsymbol{y}$ is

$$p(\boldsymbol{y}|\boldsymbol{\Omega}) \propto (\det(\boldsymbol{\Omega}))^{n/2} \exp\left\{ -\frac{1}{2}\mathrm{tr}(\mathbf{S}\boldsymbol{\Omega}) \right\}, \tag{2.1}$$

where $\mathbf{S} = \sum_{i=1}^{n} \boldsymbol{y_i}\boldsymbol{y_i}^T$. Without loss of generality, data $\{\boldsymbol{y_i}\}_{i=1,...,n}$ are centered and scaled so that each Y_v has a sample mean zero and sample variance of one. A graphical model $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ with $\boldsymbol{Y_V} \sim \mathbf{N}_p(\mathbf{0}, \boldsymbol{\Omega}^{-1})$ is called a Gaussian Graphical Model (GGM). In this paper, we denote $\boldsymbol{\mathcal{G}_V}$ as the set of all $2^{p(p-1)/2}$ undirected graphs with vertices $\boldsymbol{V}$. In the Gaussian graphical model, two variables Y_g and Y_h are conditionally independent given the remaining variables if and only if $\omega_{gh} = 0$.

To learn a sparse graphical structure, regularization methods have been utilized by imposing an l_1 penalty on the $\boldsymbol{\Omega}$ matrix and maximizing the penalized log-likelihood

function

$$\log((\det(\mathbf{\Omega})) - \mathbf{tr}(\frac{\mathbf{S}}{n}\mathbf{\Omega}) - \rho||\mathbf{\Omega}||_1, \tag{2.2}$$

over the space of positive definite matrices $\boldsymbol{M}^+$ where $\rho \geq 0$ is a regularization parameter and $||\mathbf{\Omega}||_1 = \sum_{1 \leq i,j \leq p} |\omega_{ij}|$ is the l_1 norm of $\mathbf{\Omega}$.

In conducting model selection for Gaussian graphical models, the standard approach is greedy stepwise forward-selection or backward-deletion, and then parameter estimation is based on the selected model. At each step, the decision of selecting or deleting edges is usually made based on hypothesis testing with a significance level of α. As was mentioned by Edwards (2000), it has long been recognized that this procedure does not correctly take account of the multiple comparisons involved. Another drawback of this standard stepwise procedure is its computational complexity. To remedy these problems, Drton and Perlman (2004) proposed a method that produces conservative simultaneous confidence intervals for the entire set of partial correlations, leading to a simple method for model selection that controls the overall error rate for incorrect edge inclusion. Meinshausen and Bühlmann (2006) proposed a computationally attractive method for covariance selection that can be used for huge Gaussian graphs. They performed neighborhood selection with the Lasso for each node in the graph and combined the results to learn the structure of a Gaussian concentration graph model. They demonstrated that their approach is consistent for sparse high-dimensional graphs. However, all the methods mentioned above perform model selection and parameter estimation separately. The estimator's instability often arises due to the discrete nature of such procedures, where even minor changes in the data can result in significantly different estimates (Breiman 1996). Yuan and Lin (2007) proposed a penalized-likelihood method that simultaneously does model selection and parameter estimation in the Gaussian concentration graph model. The efficiency of their method is demonstrated by utilizing the efficient maxdet algorithm

developed in convex optimization. They also proposed a BIC-type criterion for selecting the tuning parameter in the penalized likelihood methods.

Compared with the frequentist methods, precision matrix estimation in Gaussian graphical models has been less studied under the Bayesian framework, possibly due to the high computational cost associated with MCMC when the dimension of variables is large. Marlin and Murphy (2009) suggested a Bayesian model and a variational Bayes algorithm for Gaussian graphical models (GGMs) with a block structure. They used the stochastic block model as a prior, which does not require the blocks or groups to be specified a priori. The resulting problem is no longer convex, but they devised an efficient variational Bayes algorithm to solve it. Different priors were studied for the precision matrix by different researchers. Among these priors, G-Wishart priors are most commonly used (Carvalho and Scott 2009; Dobra, Lenkoski, and Rodriguez 2011; Wang and Li 2012; Mohammadi and Wit 2015). Wang (2012) presented a Bayesian approach to graphical Lasso models and their associated posterior computation algorithms, providing a fully Bayesian treatment. Using data augmentation, he developed a block Gibbs sampler for efficiently sampling covariance matrices. He also generalized the Bayesian graphical Lasso to the Bayesian adaptive graphical Lasso, which showed the top overall performance among a range of frequentist and Bayesian methods regarding both covariance matrix estimation and graphical structure learning. A Bayesian approach with mixture prior distributions that incorporate a point-mass and a Laplace distribution was examined by Banerjee and Ghosal (2015). They derived posterior consistency results and presented a computational approach using Laplace approximation. Except for the work of Banerjee and Ghosal (2015), the theoretical properties of Bayesian approaches for sparse precision matrix estimation have not been investigated. The findings of Banerjee and Ghosal (2015) regarding the estimation error rate in Frobenius norm are comparable to those of Rothman et al. (2008) but assume the underlying distribution to be Gaussian. Gan, Naveen N Narisetty,

and Liang (2019) presented a novel Bayesian technique for estimating and recovering the structure of GGMs. They utilized a continuous spike-and-slab prior as a mixture of two Laplace distributions. Their theoretical results hold beyond GGMs despite the Gaussian likelihood being used in their Bayesian formulation. They proposed a fast EM algorithm that produces a maximum a posteriori (MAP) estimate of the precision matrix. Unlike many existing estimators, their algorithm was guaranteed to produce a symmetric and positive definite estimator.

All the literature reviewed deals with continuous variables. In this project, we focus on ordinal variables. We will compare the performance of three different priors of graphical Lasso, adaptive graphical Lasso, and spike-and-slab Lasso on the estimation of the latent precision matrix and the graphical structure learning among the ordinal variables.

2.2.2 Gaussian Copula Graphical Models (GCGMs) with Regularization

Now consider variables $\boldsymbol{Y_V} = (Y_1, \ldots, Y_p)$ are discrete with joint distribution F, where there exists an ordering for the possible values of each variable $Y_v,\ v \in \boldsymbol{V}$. We adopt the Gaussian copula (Nelsen 2007) to model the multivariate dependence among variables $Y_v,\ v \in \boldsymbol{V}$. Assume a latent vector $\boldsymbol{Z_V} \sim \mathbf{N}_p(\mathbf{0}, \boldsymbol{\Gamma})$ with correlation matrix $\boldsymbol{\Gamma}$, such that the two random vectors are:

$$Y_v = F_v^{-1}(\Phi(Z_v)),\ v \in \boldsymbol{V}, \tag{2.3}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution and F_v is the pseudo marginal CDF for each of Y_v. In this way, the joint distribution F of $\boldsymbol{Y_V}$ is a function of the correlation matrix $\boldsymbol{\Gamma}$ and the univariate distributions F_v of $Y_v,\ v \in \boldsymbol{V}$:

$$\begin{aligned} Pr(Y_1 \leq y_1, ..., Y_p \leq y_p) &= F(y_1, ..., y_p | \boldsymbol{\Gamma}, F_1, ..., F_p) && (2.4) \\ &= \Phi_p\Big(\Phi^{-1}\Big(F_1(y_1)\Big), ..., \Phi^{-1}\Big(F_p(y_p)\Big) | \boldsymbol{\Gamma}\Big), && (2.5) \\ &\triangleq C(F_1(y_1), ..., F_p(y_p) | \boldsymbol{\Gamma}) && (2.6) \end{aligned}$$

where $\Phi_p(\cdot|\boldsymbol{\Gamma})$ is the distribution function of $\mathbf{N}_p(\mathbf{0}, \boldsymbol{\Gamma})$. This is a Gaussian copula graphical model (GCGM) with correlation matrix $\boldsymbol{\Gamma}$ of dimension $p \times p$ for $\boldsymbol{Y_V}$.

Similar to that in the Gaussian graphical models, in the Gaussian copula graphical model (2.5), the precision matrix $\boldsymbol{\Omega}$ $(= \boldsymbol{\Gamma}^{-1})$ contains information on the conditional dependence between variables $\boldsymbol{Y_V}$, where two ordinal variables Y_g and Y_h are conditionally independent given the remaining variables if and only if $\omega_{gh} = 0$.

To make the Gaussian copula model generalizable to the ordinal variables, the Gaussian copula model was coupled with the extended rank likelihood method (Hoff 2007) so that inference on the association parameters is purely based on the rank of data without explicitly modeling the marginal distribution functions. This makes use of the fact that for each underlying latent variable Z_v, the order is consistent with the corresponding observed ordinal variable Y_v, and a straightforward parametric expression allows for the inference of association parameters based on the latent variables that are "rank-based". Therefore, there is no need to specify the marginal distributions F_v directly when making inferences on the precision matrix $\boldsymbol{\Omega}$. For each observed data $\boldsymbol{y}_v = \{y_{1v}, ..., y_{nv}\}$, the corresponding latent data $\boldsymbol{z}_v = \{z_{1v}, ..., z_{nv}\}$ must lie in the set: $\left\{\boldsymbol{z}_v : \max_k \left\{z_{kv} : y_{kv} < y_{iv}\right\} < z_{iv} < \min_k \left\{z_{kv} : y_{kv} > y_{iv}\right\}\right\}$, where the index $k = 1, ..., n$ runs through all the observations. This same idea is adopted in our Bayesian sampling for the latent variables Z_v.

2.2.3 Bayesian Regularization of GCGM with Adaptive Graphical Lasso Prior

The graphical Lasso has a Bayesian interpretation. The graphical Lasso estimator is equivalent to the maximum a posteriori estimator under the graphical Lasso prior. Knowing $\boldsymbol{\Omega}$ determines the graphical structure, $\boldsymbol{\Omega}$ is the parameter of interest to estimate for $\boldsymbol{Y_V}$ in model (2.1).

Assuming a common shrinkage parameter λ, Wang (2012) introduced the graphical Lasso prior for $\mathbf{\Omega}$ as follows:

$$p(\mathbf{\Omega}|\lambda) = C^{-1} \prod_{i<j} \left\{ \text{DE}(\omega_{ij}|\lambda) \right\} \prod_{i=1}^{p} \left\{ \text{EXP}(\omega_{ii}|\frac{\lambda}{2}) \right\} 1_{\mathbf{\Omega} \in \boldsymbol{M}^+}, \tag{2.7}$$

where C is a normalizing constant not involving λ, $\text{DE}(\cdot|\lambda)$ and $\text{EXP}(\cdot|\frac{\lambda}{2})$ represent double exponential and exponential distributions with rate parameters λ and $\frac{\lambda}{2}$, respectively. The model with this prior is denoted as the $Bgla$ model ($Bgla$).

By comparison, instead of assuming a common shrinkage parameter λ, Wang (2012) proposed the adaptive graphical Lasso prior on the precision matrix $\mathbf{\Omega}$ as follows:

$$p(\mathbf{\Omega}|\boldsymbol{\lambda}) = C^{-1} \prod_{i<j} \left\{ \text{DE}(\omega_{ij}|\lambda_{ij}) \right\} \prod_{i=1}^{p} \left\{ \text{EXP}(\omega_{ii}|\frac{\lambda_{ii}}{2}) \right\} 1_{\mathbf{\Omega} \in \boldsymbol{M}^+}, \tag{2.8}$$

$$p(\boldsymbol{\lambda}) = C \prod_{i<j} \text{GA}(s, t), \tag{2.9}$$

where $\boldsymbol{\lambda} = \{\lambda_{ij},\ i, j = 1, ..., p\}$ is a vector with each element λ_{ij} associated with the corresponding element ω_{ij} in the $\mathbf{\Omega}$ matrix, and GA stands for a Gamma distribution, which is the prior distribution for each λ_{ij} $(i \neq j)$ with hyperparameters s and t. For each λ_{ii} $(i = 1, ..., p)$, the value is fixed and set to 1. This prior (2.7) is called adaptive graphical Lasso prior since it allows for different shrinkage effects on each unique element ω_{ij} in $\mathbf{\Omega}$ so that it addresses the shortcomings of the double exponential prior, which has a tendency to excessively shrink larger parameters while inadequately shrinking smaller ones.

For computational advantage, the double exponential distribution can be expressed as a scale mixture of normals (Andrews and Mallows 1974; West 1987). For the adaptive Lasso prior, this is achieved by introducing latent scale parameters $\boldsymbol{\tau} = \{\tau_{ij}\}_{i<j}$ as follows:

$$\omega_{ij}|\tau_{ij} \sim \text{N}(0, \text{variance} = \tau_{ij}), \tag{2.10}$$

$$\tau_{ij}|\lambda_{ij} \sim \text{EXP}(\frac{\lambda_{ij}^2}{2}),\ 1 \leq i < j \leq p. \tag{2.11}$$

So the resulting adaptive graphical Lasso prior can be re-expressed as follows:

$$p(\boldsymbol{\omega}|\boldsymbol{\tau},\boldsymbol{\lambda}) = C_{\boldsymbol{\tau}}^{-1}\prod_{i<j}\left\{\mathrm{N}\big(\omega_{ij}|0,\tau_{ij}\big)\right\}\prod_{i=1}^{p}\left\{\mathrm{EXP}(\omega_{ii}|\frac{\lambda_{ii}}{2})\right\}1_{\boldsymbol{\Omega}\in\boldsymbol{M}^{+}}, \tag{2.12}$$

$$p(\boldsymbol{\tau}|\boldsymbol{\lambda}) \propto C_{\boldsymbol{\tau}}\prod_{i<j}\left\{\mathrm{EXP}(\tau_{ij}|\frac{\lambda_{ij}^{2}}{2})\right\}, \tag{2.13}$$

$$p(\boldsymbol{\lambda}) = \prod_{i<j}\mathrm{GA}(s,t), \tag{2.14}$$

where $\boldsymbol{\omega} = \{\omega_{ij}\}_{i\leq j}$ is the vector of the upper off-diagonal and diagonal entries of $\boldsymbol{\Omega}$, $C_{\boldsymbol{\tau}}$ is a normalizing term depending on $\boldsymbol{\tau}$ and is analytically intractable, and $\mathrm{N}(\cdot|0,\tau_{ij})$ represents Normal distribution with mean 0 and variance τ_{ij}. The model with the adaptive graphical Lasso prior is denoted as the *Bada* model (*Bada*).

2.2.4 Bayesian Regularization of GCGM with Spike-and-Slab Lasso Prior

The spike-and-slab Lasso prior was developed in a series of work by Ročková (2018); Ročková and George (2016); Ročková and George (2018), which is specified as follows:

$$p(\omega_{ij}|\mathrm{r}_{ij}) = \mathrm{r}_{ij}\mathrm{DE}(\omega_{ij}|\frac{1}{\nu_1}) + (1-\mathrm{r}_{ij})\mathrm{DE}(\omega_{ij}|\frac{1}{\nu_0}) \tag{2.15}$$

Where ν_0 and ν_1 are the scale parameters in double exponential distributions with ν_1 being large compared with ν_0, and r_{ij} is an indicator variable which takes the value 1 if ω_{ij} is allocated to the slab component, and value 0 if ω_{ij} is allocated to the spike component. In the spike-and-slab Lasso prior, the spike component, which concentrates its mass at values close to zero, allows shrinkage of small ω_{ij} to zero. In contrast, the slab component distributes its mass across a broad spectrum of potential values.

The conventional spike-and-slab prior features "spike" element that is positioned as a zero point mass, which aligns with our $\nu_0 = 0$ setting. George and McCulloch (1993) introduced continuous spike-and-slab priors utilizing normal components in the context of linear regression, and their efficacy in high-dimensional shrinkage was subsequently examined by Ishwaran and Rao (2005); Naveen Naidu Narisetty and

He (2014). Ročková and George (2018) considered the spike-and-slab Lasso prior given by (2.14) for linear regression and studied the adaptive shrinkage property of such priors as well as various asymptotic properties concerning the posterior mode. Continuous spike-and-slab priors offer the benefit of continuous prior distributions on ω_{ij}, facilitating the use of efficient algorithms that eliminate the need to switch the active parameter dimension.

We utilize the spike-and-slab Lasso prior given by (2.14) for each off-diagonal element in the precision matrix $\mathbf{\Omega}$. After using the normal mixture presentation of the double exponential distribution, it can be re-expressed as follows:

$$p(\mathbf{\Omega}|\boldsymbol{\tau},\lambda) \propto \prod_{i<j}\left\{\mathrm{N}\big(\omega_{ij}|0,\tau_{ij}\big)\right\}\prod_{i=1}^{p}\left\{\mathrm{EXP}(\omega_{ii}|\frac{\lambda}{2})\right\}1_{\mathbf{\Omega}\in\boldsymbol{M}^{+}}, \tag{2.16}$$

$$p(\boldsymbol{\tau}|\mathbf{r}) = \prod_{i<j}\left\{\mathrm{EXP}(\tau_{ij}|\frac{1}{2\nu_0^2})1_{\{\mathrm{r}_{ij}=0\}} + \mathrm{EXP}(\tau_{ij}|\frac{1}{2\nu_1^2})1_{\{\mathrm{r}_{ij}=1\}}\right\}, \tag{2.17}$$

$$p(\mathbf{r}|\eta) = \prod_{i<j}\mathrm{BERN}(\eta), \tag{2.18}$$

$$p(\eta) = \mathrm{BETA}(a,b), \tag{2.19}$$

$$p(\lambda) = \mathrm{GA}(s,t). \tag{2.20}$$

where $\mathbf{r} = \{\mathrm{r}_{ij}\}_{i<j}$ and each indicator variable r_{ij} has a prior inclusion probability η which follows a Beta distribution with parameters a and b. The model with spike-and-slab Lasso prior is denoted as Bss model (Bss).

2.3 Block Gibbs Sampling Algorithm

From the Bayesian perspective, the posterior distribution of the precision matrix $\mathbf{\Omega}$ and the latent variables $\boldsymbol{Z_V}$ is as follows:

$$p(\mathbf{\Omega}, \boldsymbol{Z_V}|\boldsymbol{Y_V}) \propto p(\boldsymbol{Y_V}|\boldsymbol{Z_V}) \times p(\boldsymbol{Z_V}|\mathbf{\Omega}) \times p(\mathbf{\Omega}), \tag{2.21}$$

where $p(\mathbf{\Omega})$ is the prior distribution of $\mathbf{\Omega}$. Given the prior, we adopt the block Gibbs sampler idea from Wang (2012) for posterior sampling. The block Gibbs sampler

with the three different priors is summarized as follows:

Block Gibbs Sampler. Given the current value $\boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{\Gamma}, \boldsymbol{\Omega}$ $(= \boldsymbol{\Gamma}^{-1}) \in \boldsymbol{M}^{+}$, $\boldsymbol{\tau}$ and $\mathbf{r}$,

1. Sample the latent variables $\mathrm{Z}_{iv},\ i = 1, ..., n;\ v = 1, ..., p$.

$\mathrm{Z}_{iv} \sim \mathrm{TN}\Big(\boldsymbol{\Gamma_{v,-v}}(\boldsymbol{\Gamma_{-v,-v}})^{-1}(\boldsymbol{Z_{i,-v}})', \boldsymbol{\Gamma_{v,v}} - \boldsymbol{\Gamma_{v,-v}}(\boldsymbol{\Gamma_{-v,-v}})^{-1}\boldsymbol{\Gamma_{-v,v}}\Big)$, which is a truncated normal distribution with lower bound as $l_{iv} = \max_k\Big\{z_{kv} : y_{kv} < y_{iv}\Big\}$, and upper bound as $u_{iv} = \min_k\Big\{z_{kv} : y_{kv} > y_{iv}\Big\}$, for $k = 1, ..., n$. Here $\boldsymbol{\Gamma_{v,-v}}$ is the v^{th} row of $\boldsymbol{\Gamma}$ with the v^{th} column removed, $\boldsymbol{\Gamma_{-v,-v}}$ is the sub-matrix of $\boldsymbol{\Gamma}$ with the v^{th} row and the v^{th} column removed, $\boldsymbol{Z_{i,-v}}$ is the i^{th} row of $\mathbf{Z}$ with the the v^{th} column removed, $\boldsymbol{\Gamma_{v,v}}$ is the element of $\boldsymbol{\Gamma}$ on the v^{th} row and the v^{th} column, and $\boldsymbol{\Gamma_{-v,v}}$ is the v^{th} column of $\boldsymbol{\Gamma}$ with the v^{th} row removed.

2-1 *Bgla* Model

2-1(a). Sample $\lambda \sim$ GA (shape= $p(p+1)/2 + s$, rate = $||\boldsymbol{\Omega}||/2 + t$).

2-1(b). For $1 \leq i < j \leq p$, sample $u_{ij} \sim$ INV-GAU (μ', λ') where $\mu' = \sqrt{(\lambda^2/\omega_{ij}^2)}$ and $\lambda' = \lambda^2$, and then update $\tau_{ij} = 1/u_{ij}$.

2-2 *Bada* Model

2-2(a). For $1 \leq i < j \leq p$, sample $\lambda_{ij} \sim$ GA (shape= $1 + s$, rate = $|\omega_{ij}| + t$).

2-2(b). For $1 \leq i < j \leq p$, sample $u_{ij} \sim$ INV-GAU (μ', λ') where $\mu' = \sqrt{(\lambda_{ij}^2/\omega_{ij}^2)}$ and $\lambda' = \lambda_{ij}^2$, and then update $\tau_{ij} = 1/u_{ij}$.

2-3 *Bss* Model

2-3(a). Sample $\lambda \sim$ GA (shape $= p + s$, rate $= t + \sum_{i=1}^{p} \frac{\omega_{ii}}{2}$).

2-3(b). Sample $\eta \sim$ Beta $(\sum_{i<j} \mathrm{r}_{ij} + a, \frac{p(p-1)}{2} - \sum_{i<j} \mathrm{r}_{ij} + b)$.

2-3(c). For $1 \leq i < j \leq p$, sample $\mathrm{r}_{ij} \sim$ Bernoulli $(p = \frac{A}{A+B})$, where $A =$

$\frac{\eta}{2\nu_1}\exp\Big\{-\frac{|\omega_{ij}|}{\nu_1}\Big\}$, and $B=\frac{1-\eta}{2\nu_0}\exp\Big\{-\frac{|\omega_{ij}|}{\nu_0}\Big\}$.

2-3(d). For $1\le i<j\le p$, sample $u_{ij}\sim$ INV-GAU $(\mu^{'},\lambda')$ where $\mu^{'}=\frac{1}{\nu_0|\omega_{ij}|}$ and $\lambda^{'}=\frac{1}{\nu_0^2}$, if $\mathrm{r}_{ij}=0$, and $\mu^{'}=\frac{1}{\nu_1|\omega_{ij}|}$ and $\lambda^{'}=\frac{1}{\nu_1^2}$, if $\mathrm{r}_{ij}=1$. Then update $\tau_{ij}=1/u_{ij}$.

3. For $i=1,...,p$,

(a) Partition $\boldsymbol{\Omega}$, $\boldsymbol{S}$ and $\boldsymbol{\tau}$ into blocks as follows,

$$\boldsymbol{\Omega}=\begin{pmatrix}\boldsymbol{\Omega}_{11}, & \boldsymbol{\omega}_{12}\\ \boldsymbol{\omega}_{12}^{'}, & \omega_{22}\end{pmatrix},\ \boldsymbol{S}=\begin{pmatrix}\boldsymbol{S}_{11}, & \boldsymbol{s}_{12}\\ \boldsymbol{s}_{12}^{'}, & s_{22}\end{pmatrix},\ \boldsymbol{\Upsilon}=\begin{pmatrix}\boldsymbol{\Upsilon}_{11}, & \boldsymbol{\tau}_{12}\\ \boldsymbol{\tau}_{12}^{'}, & 0\end{pmatrix},$$

where $\boldsymbol{S}=\boldsymbol{Z}^{'}\boldsymbol{Z}$, $\boldsymbol{\Omega}_{11}$, $\boldsymbol{S}_{11}$, and $\boldsymbol{\Upsilon}_{11}$ are the $(p-1)\times(p-1)$ sub-matrices of $\boldsymbol{\Omega}$, $\boldsymbol{S}$ and $\boldsymbol{\Upsilon}$ with the last row and last column removed, $\boldsymbol{\omega}_{12}$, $\boldsymbol{s}_{12}$ and $\boldsymbol{\tau}_{12}$ are the $(p-1)\times 1$ column vectors, ω_{22} and s_{22} are the elements of $\boldsymbol{\Omega}$ and $\boldsymbol{S}$ on the last row and last column.

(b) Sample $\gamma\sim$ GA (shape $=n/2+1$, rate $=(s_{22}+\lambda_{22})/2$) and $\boldsymbol{\beta}\sim$ N($-\mathbf{C}\boldsymbol{s}_{12}$,$\mathbf{C}$) where $\mathbf{C}=\{(s_{22}+\lambda_{22})(\boldsymbol{\Omega}_{11})^{-1}+\boldsymbol{D}_{\boldsymbol{\tau}}^{-1}\}^{-1}$ with $\boldsymbol{D}_{\boldsymbol{\tau}}=\mathbf{diag}(\boldsymbol{\tau}_{12})$ and $\lambda_{22}=1$.

(c) Update $\boldsymbol{\omega}_{12}=\boldsymbol{\beta},\boldsymbol{\omega}_{21}=\boldsymbol{\beta}^{'},\omega_{22}=\gamma+\boldsymbol{\beta}^{'}(\boldsymbol{\Omega}_{11})^{-1}\boldsymbol{\beta}$.

4. Rescale $\boldsymbol{\Omega}$ into the inverse matrix of the correlation matrix $\boldsymbol{\Gamma}$.

$\widetilde{\boldsymbol{\Gamma}}=\boldsymbol{\Omega}^{-1}$;

$\boldsymbol{\Gamma}_{g,h}=\widetilde{\boldsymbol{\Gamma}}_{g,h}/\sqrt{\widetilde{\boldsymbol{\Gamma}}_{g,g}\widetilde{\boldsymbol{\Gamma}}_{h,h}}=(\boldsymbol{\Omega}^{-1})_{g,h}/\sqrt{(\boldsymbol{\Omega}^{-1})_{g,g}(\boldsymbol{\Omega}^{-1})_{h,h}},\ g,h=1,...,p.$

$\boldsymbol{\Omega}=\boldsymbol{\Gamma}^{-1}$.

2.4 Simulation Studies

2.4.1 Simulation settings

This simulation experiment is designed to evaluate the performance of the Gaussian copula graphical model in terms of parameter estimation and structure learning using the three different priors on the precision matrix $\mathbf{\Omega}$. We consider a block model in our simulation with the following two scenarios:

- Scenario 1: A block model of dimension $p = 10$ with correlation matrix $\Sigma_{p\times p} = (\sigma_{ij})$ where
  $\sigma_{ij} = 0.5$ for $1 \leq i \neq j \leq 4$, $\sigma_{ij} = 0.6$ for $5 \leq i \neq j \leq 7$, $\sigma_{ij} = 0.1$ for $\{1 \leq i \leq 4, 5 \leq j \leq 7\} \cup \{5 \leq i \leq 7, 1 \leq j \leq 4\}$ and $\sigma_{ij} = 0$ otherwise.

- Scenario 2: A block model with dimension $p = 50$ with correlation matrix $\Sigma_{p\times p} = (\sigma_{ij})$ where
  $\sigma_{ij} = 0.8$ for $1 \leq i \neq j \leq 18$, $\sigma_{ij} = 0.5$ for $19 \leq i \neq j \leq 34$, $\sigma_{ij} = 0.1$ for $\{1 \leq i \leq 18, 19 \leq j \leq 34\} \cup \{19 \leq i \leq 34, 1 \leq j \leq 18\}$ and $\sigma_{ij} = 0$ otherwise.

We generated 200 samples of size $n = 200$ for Scenario 1 and size $n = 1000$ for Scenario 2. For each generated sample, we fit the Gaussian copula graphical model with the three different priors. For the shrinkage parameter λ in the $Bgla$ and Bss models and each individual λ_{ij} in the $Bada$ model, the hyper parameters are set as $s = 10^{-2}$ and $t = 10^{-6}$. Within the Bss model with the spike-and-slab Lasso prior, to investigate the effects of hyper-parameter ν_0 on the performance of the model, three different values of ν_0 of 10^{-1}, 10^{-3}, 10^{-5} are chosen, and they are labeled as Bss_1, Bss_2 and Bss_3, respectively. The hyperparameter ν_1 in the spike-and-slab Lasso prior is fixed at 10 for all the Bss models. In all three Bss models, the hyperparameters for the prior inclusion probability η are set as $a = b = 1$ since we set the prior

distribution of η as the standard uniform distribution. The Bayesian estimates were based on 4000 iterations of the MCMC samples after 1000 burn-in iterations.

2.4.2 Simulation results

Matrix Estimation

The posterior means of covariance matrix $\mathbf{\Gamma}$, precision matrix $\mathbf{\Omega}$, and partial correlation matrix are estimated based on their MCMC samples, and their corresponding averages over the 200 replicates are used as the final estimate for each matrix. Figure 2.1 and 2.2 show the estimation results of the covariance matrix as well as the partial correlation matrix using all five models of $Bgla$, $Bada$, Bss_1, Bss_2 and Bss_3 for Scenario 1 (p = 10). In Figure 2.1, compared with the true covariance matrix, all five models show good results in covariance matrix estimation, which implies that all these models perform well in identifying the marginal relationship between the ten ordinal variables. It is a similar case in partial correlation matrix estimation in Figure 2.2. All five models perform well when the five model estimates are compared with the true partial correlation matrix.

Figure 2.3 and 2.4 show the estimation results of the covariance matrix as well as the partial correlation matrix using all five models for Scenario 2 (p = 50). Similar to the results for Scenario 1, all five models also give good estimates of the covariance and partial correlation matrix when the number of ordinal variables is increased to 50. It is worth mentioning that in Scenario 2 (p = 50), the true partial correlation is much smaller than that in Scenario 1 (p = 10), making the estimation more challenging for all five models.

The performance of the precision matrix estimation can also be evaluated using the Frobenius Loss (FL) and Stein's Entropy Loss (EL). The Frobenius Loss and

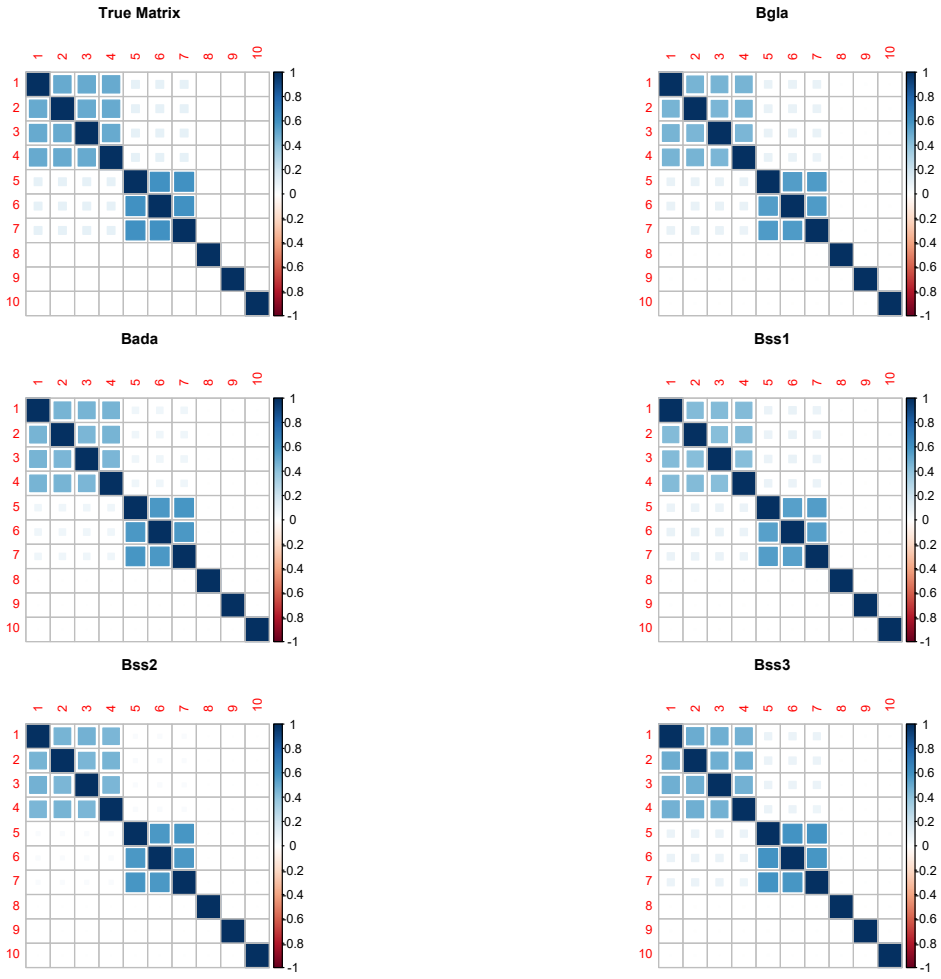

Figure 2.1 Covariance matrix estimation for Scenario 1 (p=10).

Stein's Entropy Loss for the precision matrix $\mathbf{\Omega}$ are defined as follows:

$$FL(\hat{\mathbf{\Omega}}, \mathbf{\Omega}) = \frac{||\mathbf{\Omega} - \hat{\mathbf{\Omega}}||_F^2}{||\mathbf{\Omega}||_F^2}; \tag{2.22}$$

$$EL(\hat{\mathbf{\Omega}}, \mathbf{\Omega}) = \mathbf{tr}(\hat{\mathbf{\Omega}}\mathbf{\Omega}^{-1}) - \log\{\det(\hat{\mathbf{\Omega}}\mathbf{\Omega}^{-1})\} - p, \tag{2.23}$$

where $\mathbf{\Omega}$ is the true precision matrix and $\hat{\mathbf{\Omega}}$ is the estimator of $\mathbf{\Omega}$, which is the posterior mean estimator in all models.

Table 2.1 compares the calculated mean, median, and standard deviation of Frobenius Loss and Stein's Entropy Loss for the precision matrix $\mathbf{\Omega}$ for Scenario 1 and 2. Based on the results, the *Bada* model yields the best Stein's Entropy Loss perfor-

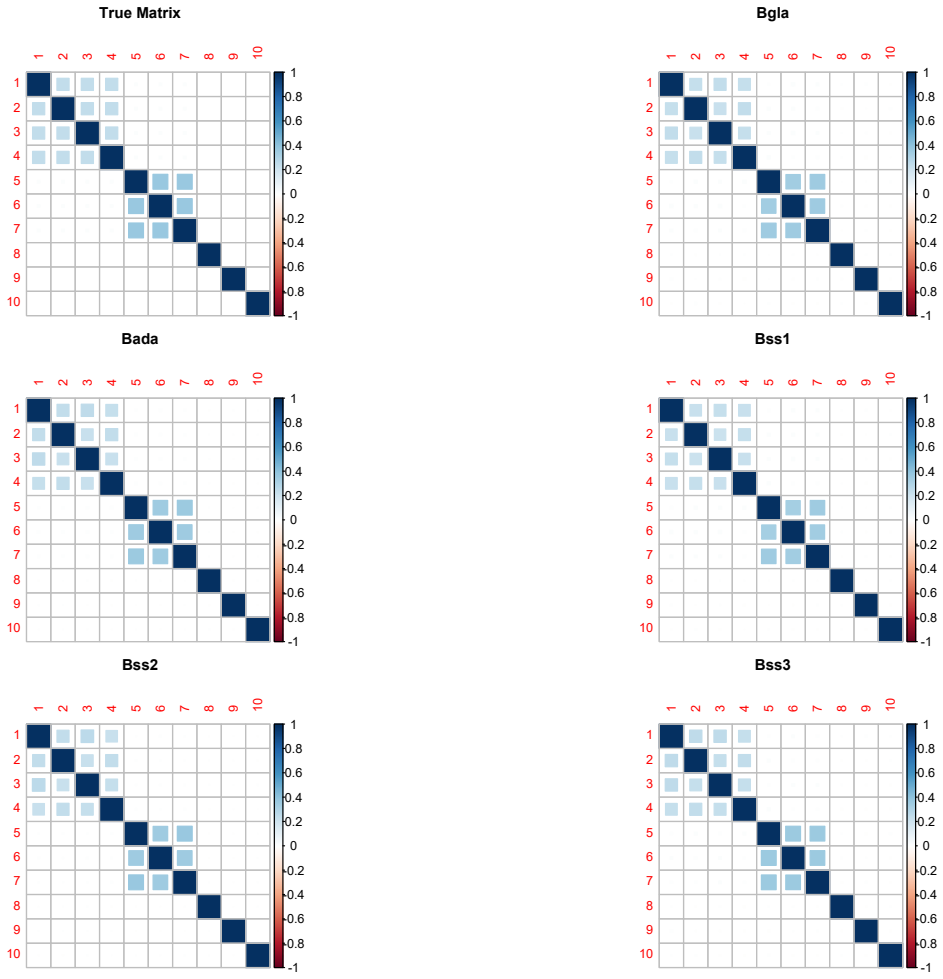

Figure 2.2 Partial correlation matrix estimation for Scenario 1 (p=10).

mance for both Scenario 1 (p = 10) and 2 (p = 50). Regarding the performance in Frobenius Loss, the Bss_1 model gives the best performance for both scenarios. Overall, the results are similar among all five models, although the Bss_2 and Bss_3 models generally have relatively larger Frobenius Loss and Stein's Entropy Loss.

Graphical Structure Determination

To assess the performance of the graphical structure learning of all models, we compute specificity, sensitivity, and Matthews Correlation Coefficient(MCC), which were

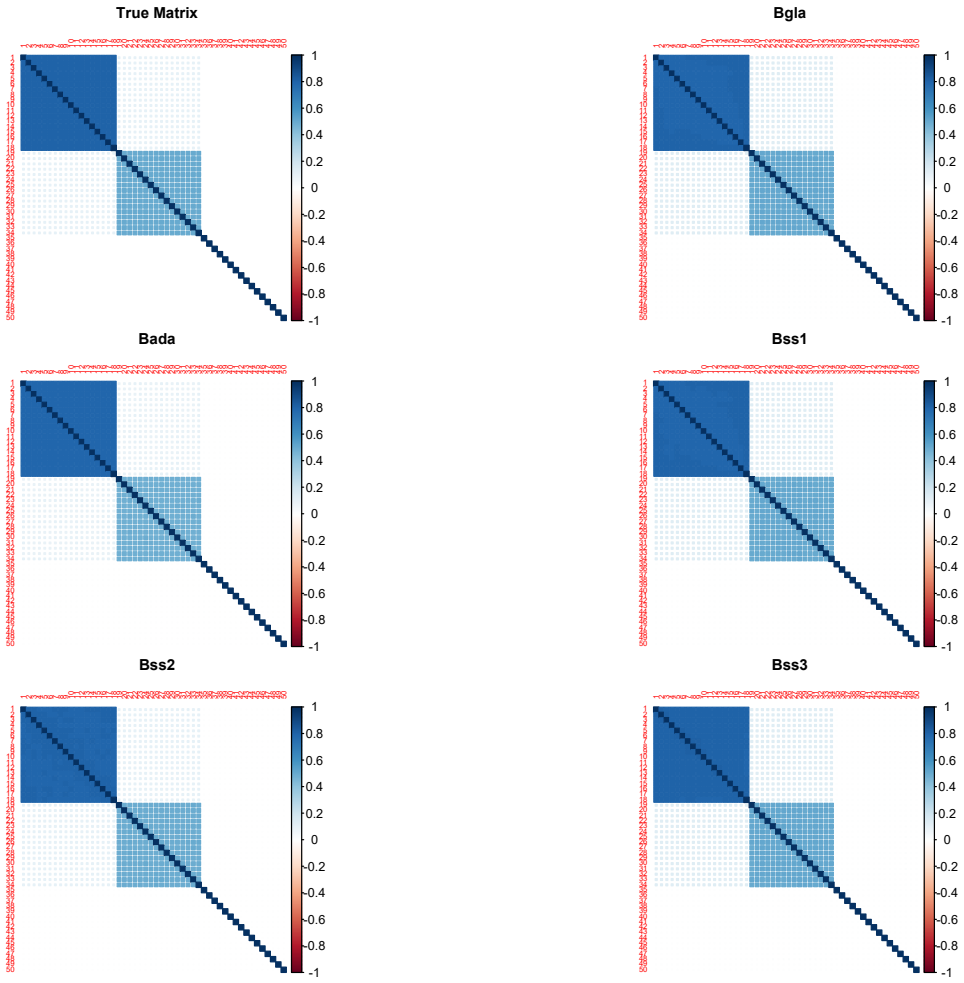

Figure 2.3 Covariance matrix estimation for Scenario 2 (p=50).

used in Fan, Feng, and Wu (2009), and are defined as follows:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}}, \tag{2.24}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}}, \tag{2.25}$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP+FP})(\text{TP+FN})(\text{TN+FP})(\text{TN+FN})}}, \tag{2.26}$$

where TP, TN, FP, and FN are the number of true positives, true negatives, false positives, and false negatives, respectively. MCC is generally regarded as a balanced classification measure because it considers TP, TN, FP, and FN. For all three metrics, the larger the values are, the better the classification is.

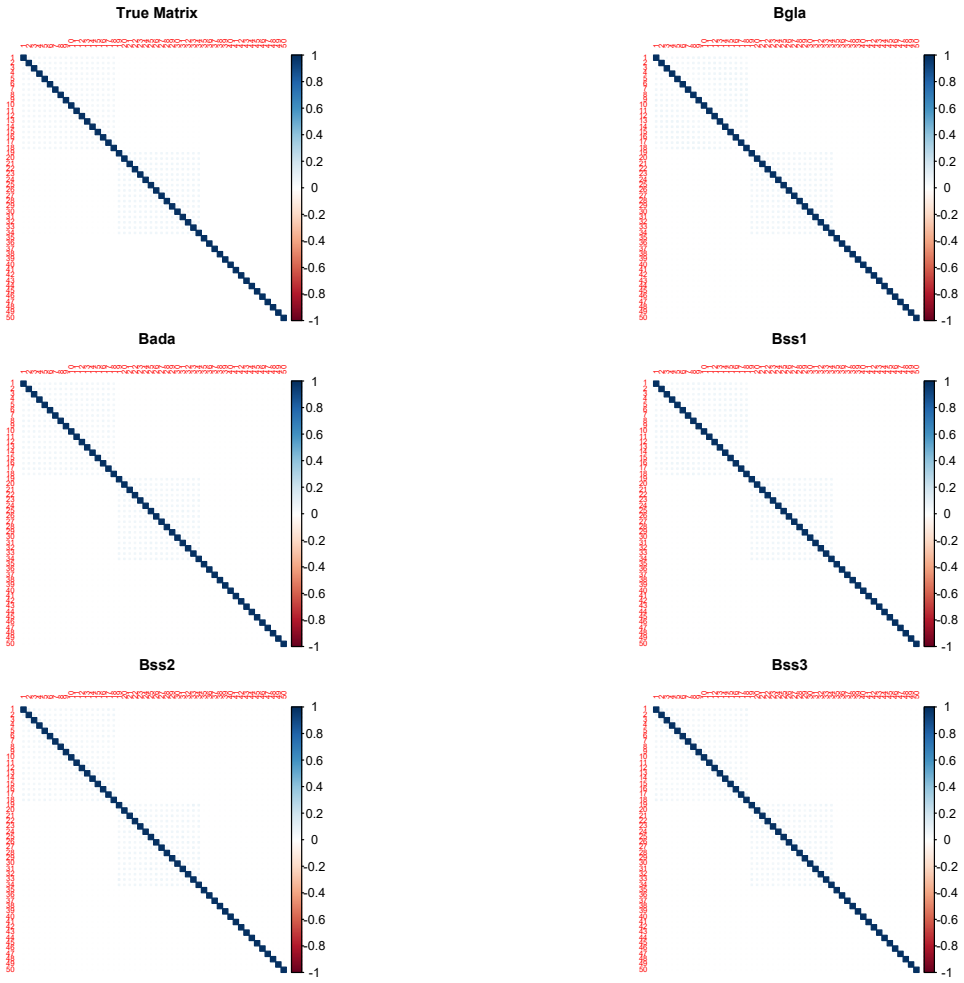

Figure 2.4 Partial correlation matrix estimation for Scenario 2 (p=50).

Table 2.2 is an overall summary of the average specificity, sensitivity, and Matthews Correlation Coefficient (MCC) of the partial correlation matrix for Scenarios 1 and 2 using different cutoff values based on 200 replicates for all five models. Five cutoff values of 0.005, 0.008, 0.01, 0.02, and 0.05, and the mean and standard deviation of the corresponding average specificity, sensitivity, and MCC are reported. The left side panel shows results for Scenario 1 (p = 10). For $Bgla$, $Bada$, Bss_1, and Bss_3 models, the average MCC increases drastically as the cutoff value increases. That can be explained by the relatively stable high average sensitivity when the average specificity increases dramatically with the cutoff value. For the Bss_2 model, the av-

Table 2.1 Mean, median and standard deviation of Frobenius Loss (FL) and Stein's Entropy Loss (EL) for the precision matrix $\mathbf{\Omega}$ for Scenario 1 (p=10) and Scenario 2 (p=50).

$\boldsymbol{p = 10}$						
Models	FL_Mean	FL_Median	FL_sd	EL_Mean	EL_Median	EL_sd
$Bgla$	0.04768	0.04356	**0.01796**	0.26303	0.25831	**0.06146**
$Bada$	0.05547	0.04906	0.02869	**0.18518**	**0.18064**	0.06877
Bss_1	**0.04497**	**0.03732**	0.02619	0.20702	0.20061	0.06223
Bss_2	0.10233	0.09088	0.05216	0.30682	0.28276	0.12164
Bss_3	0.06937	0.05974	0.03610	0.26610	0.24492	0.09556
$\boldsymbol{p = 50}$						
Models	FL_Mean	FL_Median	FL_sd	EL_Mean	EL_Median	EL_sd
$Bgla$	0.05973	0.05678	0.01601	1.54730	1.54739	0.15047
$Bada$	0.07002	0.06728	**0.01540**	**0.97745**	**0.95988**	**0.09971**
Bss_1	**0.05890**	**0.05371**	0.02392	1.25489	1.23132	0.17076
Bss_2	0.17740	0.17381	0.03164	2.04801	2.02992	0.16966
Bss_3	0.12554	0.12151	0.02947	1.91220	1.89546	0.23758

erage MCC is overall very high compared with the other four models, although the amount of increase is barely unnoticeable as the cutoff value increases.

The right side panel shows results for Scenario 2 (p = 50). With the same set of cutoff values, $Bgla$, Bss_1, and Bss_3 show a similar trend in the average MCC values with that of Scenario 1. Whereas for the $Bada$ model, the highest average MCC value is obtained when the cutoff values are at 0.01 and 0.02, and the overall performance of the $Bada$ model is the best among the five models in terms of average MCC. Interestingly, the average MCC is decreased slightly for the Bss_2 model as the cutoff value increases. For Scenario 2 with p = 50, we consider a cutoff value of 0.05 as a reasonable threshold to get an overall good performance in average MCC. In the next section, we will also use this cutoff value in graphical structure determination in real data analysis.

To illustrate the results in more detail, boxplots of sensitivity and specificity over 200 replicates of simulation for Scenarios 1 and 2 are shown in Figure 2.5 and 2.6,

Table 2.2 Average specificity (SP%), sensitivity (SE%) and MCC (MCC%) of the partial correlation matrix for Scenario 1 (p=10) and Scenario 2 (p=50).

$p = 10$				$p = 50$			
		Bgla				***Bgla***	
Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)	Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)
0.005	7.3(3.9)	100(0)	12.5(3.3)	0.005	15.8(1.2)	98.3(0.8)	17.7(1.4)
0.008	11.9(5.3)	100(0)	15.9(4.0)	0.008	25.1(1.4)	97.0(1.0)	22.9(1.4)
0.01	15.0(6.1)	100(0)	18.3(4.3)	0.01	31.1(1.5)	96.1(1.2)	26.1(1.6)
0.02	28.5(8.2)	100(0)	27.1(5.3)	0.02	56.3(1.8)	90.1(1.6)	38.7(2.0)
0.05	62.0(8.0)	99.8(1.4)	49.9(6.6)	0.05	92.1(1.0)	57.6(1.9)	52.8(2.4)
		Bada				***Bada***	
Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)	Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)
0.005	56.4(10.3)	99.8(1.6)	45.7(8.0))	0.005	81.6(1.6)	91.0(1.7)	63.6(2.2)
0.008	71.4 (8.8)	99.4 (2.8)	58.0 (8.7)	0.008	90.3(1.1)	85.3(2.1)	71.3(2.3)
0.01	76.7 (8.4)	99.4 (2.8)	63.4 (9.2)	0.01	93.0(0.9)	82.0(2.2)	73.5(2.2)
0.02	90.1(5.2)	97.8 (4.6)	79.6 (9.2)	0.02	97.8(0.5)	68.8(2.3)	73.6(2.1)
0.05	97.5 (2.5)	93.5 (7.9)	90.2 (7.3)	0.05	100(0.2)	45.7(1.7)	61.7(1.5)
		***Bss*$_1$**				***Bss*$_1$**	
Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)	Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)
0.005	12.4(5.8)	100(0)	16.3(4.4)	0.005	21.4(1.2)	99.2(0.6)	22.9(1.1)
0.008	20.5(7.1)	100(0)	22.0(4.8)	0.008	33.1(1.6)	98.5(0.7)	30.0(1.2)
0.01	25.3(7.8)	100(0)	25.1(7.0)	0.01	40.3(1.7)	97.9(0.9)	34.1(1.4)
0.02	47.4(9.1)	100(0)	39.4(6.2)	0.02	68.1(1.8)	91.7(1.8)	50.0(2.0)
0.05	84.4(5.8)	99.9(1.1)	72.7(7.8)	0.05	96.0(0.6)	54.4(2.5)	58.5(2.3)
		***Bss*$_2$**				***Bss*$_2$**	
Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)	Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)
0.005	96.3(3.5)	80.4(9.3)	78.8(9.2)	0.005	99.5(0.2)	42.5(1.8)	58.6(1.6)
0.008	97.4(2.9)	79.2(8.7)	80.3(8.2)	0.008	99.7(0.2)	41.0(1.7)	58.0(1.5)
0.01	97.8(2.7)	78.7(8.5)	81.0(8.0)	0.01	99.7(0.1)	40.3(1.6)	57.8(1.5)
0.02	98.7(2.1)	76.8(8.6)	81.8(7.2)	0.02	99.9(0.1)	38.2(1.4)	56.5(1.2)
0.05	99.3(1.4)	74.3(8.5)	81.8(6.8)	0.05	100.0(0.0)	34.8(1.0)	54.1(0.9)
		***Bss*$_3$**				***Bss*$_3$**	
Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)	Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)
0.005	53.2(9.6)	96.7(5.7)	40.6(8.2)	0.005	67.9(2.0)	72.1(1.9)	33.8(2.2)
0.008	59.3(9.6)	96.4(5.7)	45.1(8.7)	0.008	73.5(1.9)	68.8(2.0)	36.8(2.3)
0.01	62.4(9.4)	96.2(5.8)	47.5(8.9)	0.01	76.2(1.9)	67.1(2.0)	38.2(2.4)
0.02	71.9(8.3)	95.1(6.6)	55.0(9.4)	0.02	84.2(1.7)	60.7(1.8)	42.8(2.6)
0.05	83.8(6.6)	92.8(7.6)	66.8(10.3)	0.05	92.8(1.2)	50.8(1.5)	48.5(2.7)

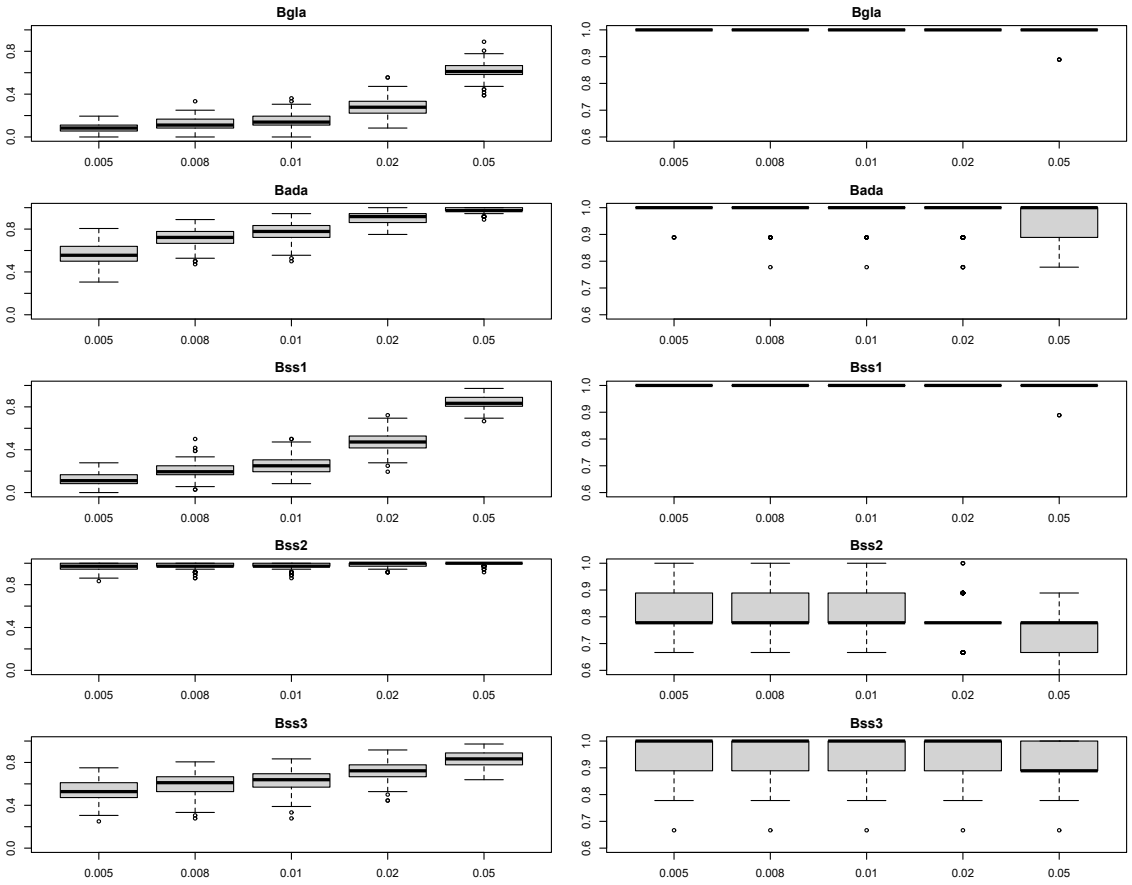

Figure 2.5 Boxplots of specificity (left panel) and sensitivity (right panel) for Scenario 1 (p=10).

respectively. In each figure, the left panel represents the specificity, and the right panel represents the sensitivity. Figure 2.5 shows the boxplots of Scenario 1 (p=10). Among the five models, $Bgla$ and Bss_1 show great average sensitivity with small variability under all five cutoff values. The average specificity increases significantly when these two models' cutoff value increases to 0.05. $Bada$ model also shows great overall average sensitivity under all cutoffs but with much better specificity levels than $Bgla$ and Bss_1. Like the $Bada$ model, Bss_3 gives relatively good specificity levels but more significant sensitivity variability under all cutoffs. Bss_2 model shows the highest level of average specificity with small variability under all cutoff values, whereas its sensitivity has relatively large variability.

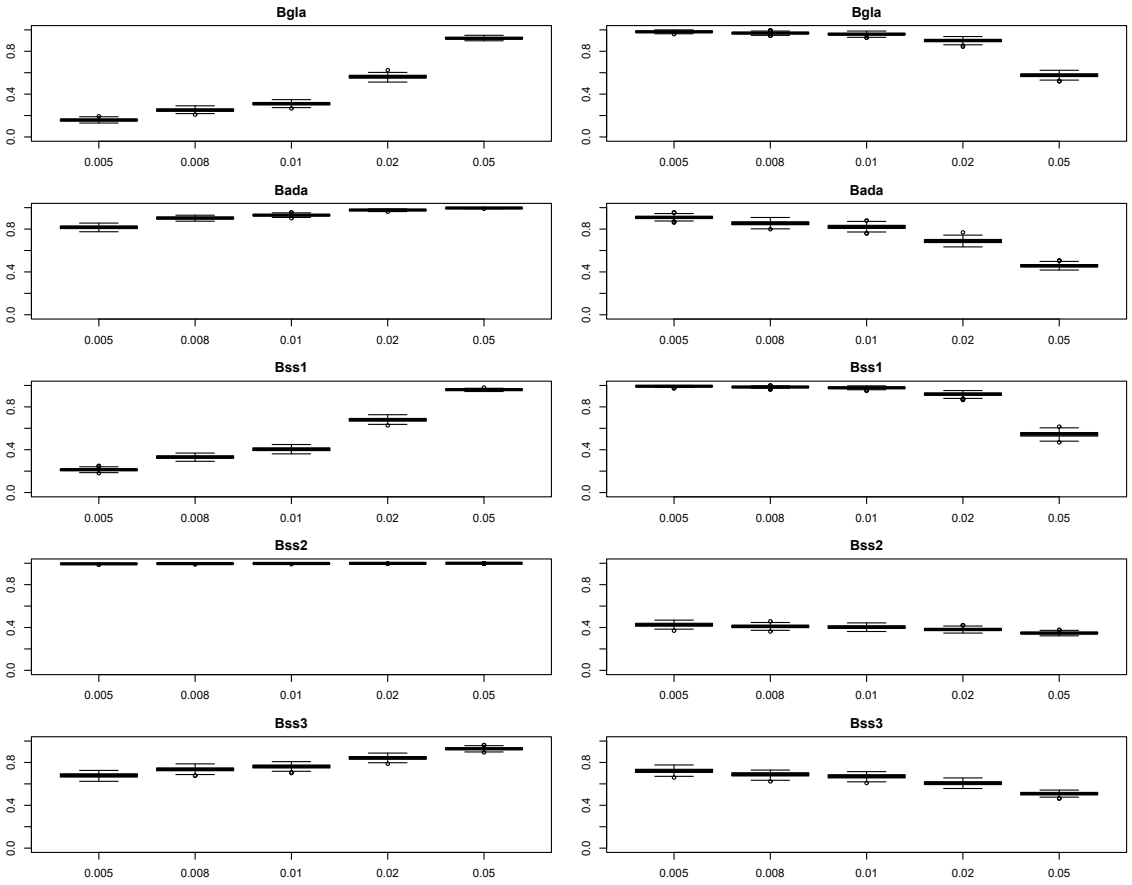

Figure 2.6 Boxplots of specificity (left panel) and sensitivity (right panel) for Scenario 2 (p=50).

Figure 2.6 shows the boxplots of Scenario 2 (p=50). For average specificity on the left panel, the trend of each model is the same as that in Scenario 1, but with a smaller magnitude of variability under each cutoff value. For the average sensitivity on the right panel, $Bgla$, $Bada$, Bss_1, and Bss_3 models show a relatively significant decrease when the cutoff value increases to 0.05. In comparison, both average sensitivity and specificity levels are relatively stable under different cutoff values for Bss_2, although the overall average sensitivity is relatively low (roughly 40% under all cutoff values).

To investigate the effect of the hyper-parameter ν_0 on the performance of the Bss models in terms of graphical structure learning, the average specificity, sensitivity, and MCC of the precision matrix $\mathbf{\Omega}$ are compared for the three different values of ν_0

for both scenarios. In Scenarios 1 and 2, the true precision matrix $\mathbf{\Omega}$ has the following format:

- Scenario 1 (p = 10) : $\omega_{ij} = \omega_1 = -0.3978$ for $1 \leq i \neq j \leq 4$, $\omega_{ij} = \omega_2 = -0.6784$ for $5 \leq i \neq j \leq 7$, $\omega_{ij} = \omega_3 = -0.0186$ for $\{1 \leq i \leq 4, 5 \leq j \leq 7\} \cup \{5 \leq i \leq 7, 1 \leq j \leq 4\}$ and $\omega_{ij} = 0$ otherwise.

- Scenario 2 (p = 50): $\omega_{ij} = \omega_1 = -0.2739$ for $1 \leq i \neq j \leq 18$, $\omega_{ij} = \omega_2 = -0.1175$ for $19 \leq i \neq j \leq 34$, $\omega_{ij} = \omega_3 = -0.0008$ for $\{1 \leq i \leq 18, 19 \leq j \leq 34\} \cup \{19 \leq i \leq 34, 1 \leq j \leq 18\}$ and $\omega_{ij} = 0$ otherwise.

Since the magnitude of ω_3 is very small for both scenarios, we treat ω_3 as 0 in the true precision matrix for both Scenario 1 and 2. To calculate the specificity and sensitivity for each data set, each off-diagonal entry $I(\omega_{ij} = 0)$ can be estimated according to $p(\mathrm{r}_{ij} = 1|Y)$, which can be estimated by the mean of r_{ij} among 4000 iterations, as follows:

- If $p(\mathrm{r}_{ij} = 1|Y) > 0.5$, then $\omega_{ij} \neq 0$;

- If $p(\mathrm{r}_{ij} = 1|Y) \leq 0.5$, then $\omega_{ij} = 0$.

The simulation results of average specificity, sensitivity, and MCC for the 200 data replicates are shown in Table 2.3. For Scenario 1 with p = 10, Bss_2 model with $\nu_0 = 10^{-3}$ yields the highest average MCC. It is a similar situation for Scenario 2 with p = 50, although the overall average MCC under these three Bss models is relatively low compared with that for Scenario 1. The performance of these three Bss models in graphical structure determination using precision matrix in Table 2.3 is consistent with that using partial correlation matrix in Table 2.2. In the real data analysis, we analyze the graphical structure of all models directly via both partial correlation matrix and r_{ij}.

Table 2.3 Average specificity (SP%), sensitivity (SE%) and MCC (MCC%) of the precision matrix for Scenario 1 (p=10) and Scenario 2 (p=50) based on the proportion of $r_{ij} = 1$ given observed data Y.

$p = 10$				$p = 50$			
ν_0	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)	ν_0	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)
10^{-1}	100(0)	9.22(9.65)	35.69(7.86)	10^{-1}	100(0)	0.90(0.99)	9.22(3.51)
10^{-3}	99.64(0.98)	70.44(8.44)	79.90(6.37)	10^{-3}	99.94(0.07)	33.66(0.83)	52.91(0.77)
10^{-5}	78.15(7.81)	91.56(7.97)	59.04(0.45)	10^{-5}	83.26(1.82)	60.60(1.92)	41.41(2.41)

2.5 Real Data Analysis

2.5.1 Dataset Description

A survey was conducted to study the present state of the physiological and psychological health of current undergraduate students in China, and the results from 2369 respondents were collected. This survey is constructed in Chinese and includes six parts: Social-Demographic Information, STD/AIDS-Related Knowledge, Sexual Health Behavior, Other Behaviors, Drug-Related Knowledge, and Life Satisfaction Survey. There is a total of 84 questions, and some questions have multiple sub-questions, which makes a total of 401 sub-questions. Almost all the questions are categorical, among which most are ordinal.

In this work, we focus on Part VI - Life Satisfaction Survey (LSS) and call this dataset the LSS dataset in short. Within the LSS dataset are 15 questions, with 54 and 10 sub-questions within questions 14 and 15, respectively. The 54 sub-questions for Question No. 14 are the Trauma Symptom Checklist for Children (TSCC), initially designed as a self-report measure of posttraumatic distress and related psychological symptomatology among children 8-16 years of age (Briere 1996). In this survey, these questions were translated into Chinese and asked to check whether the respondents had any of these behaviors within the past six months. The 10 sub-

questions within No. 15 were asked to identify the current self-assessment status of the respondents. In this work, we only work on questions 14 and 15, which makes Survey Data No. 1 with 2369 observations and 64 questions.

2.5.2 Analysis Methodology

We apply the Gaussian copula graphical model with all three priors of graphical Lasso ($Bgla$), adaptive graphical Lasso ($Bada$), and spike-and-slab Lasso (Bss) priors to the dataset to analyze the association between these 64 questions. We treat each question in the dataset as a node in the graph and responses of individuals to these questions as a sample drawn from the graph. There are $p = 64$ questions entered into the model, and all of them are ordinal variables with four ordinal levels of 1, 2, 3, and 4. Among these 64 questions, the first 54 come from Question No. 14, and levels 1, 2, 3, and 4 represent "Never", "Sometimes", "Very Often", and "Always", respectively. Questions 55 to 64 are the ten sub-questions from Question No. 15, and levels 1, 2, 3, and 4 represent "Strongly Disagree", "Disagree", "Agree", and "Strongly Agree", respectively.

For each of the five models, we run the block Gibbs sampling algorithm for a total of 10,000 iterations, with the first 2,000 iterations as the burn-in period. For the common shrinkage parameter λ in $Bgla$ and shrinkage parameters λ_{ij} in $Bada$ models, The hyper-parameters were chosen to be $s = 10^{-2}$ and $t = 10^{-6}$. For the spike-and-slab Lasso prior (Bss), the hyper-parameter ν_0 was 10^{-1}, 10^{-3} and 10^{-5} for Bss_1, Bss_2 and Bss_3, respectively. For Bss_3 model with hyper-parameter $\nu_0 = 10^{-5}$, the block Gibbs sampling algorithm stopped after the first 169 iterations due to the small magnitude of ν_0, so we collect results from the four models of $Bgla$, $Bada$, Bss_1 and Bss_2. The posterior distribution of the covariance matrix $\mathbf{\Gamma}$, precision matrix $\mathbf{\Omega}$, and partial correlation matrix are obtained. To perform matrix estimation, we use

the posterior mean as the estimator for all three matrices. We utilize the R packages *qgragh* and *igraph* to visualize the graphical structure of these 64 ordinal variables.

2.5.3 Results

Matrix Estimation

Figure 2.7 and 2.8 show the plots of the estimated posterior mean of the covariance matrix and partial correlation matrix, respectively, using the four models of $Bgla$, $Bada$, Bss_1 and Bss_2. Figure 2.7 shows the marginal relationship of these 64 questions via covariance matrix estimation. Marginally, within Question No. 14 (the first 54 questions), the questions are highly positively correlated (dark blue color) and thus form a large block with dimension 54×54 (Block 1). Within Question No. 15 (questions 55-64), we can observe two additional blocks of questions 55-62 (Block 2) and questions 63-64 (Block 3). Between blocks 1 and 2, the questions are marginally highly negatively correlated, whereas, between Blocks 1 and 3, the questions are marginally positively correlated. Among these 64 questions, question 59 in Block 2 has a weak covariance with the other questions expect that of questions 63 and 64 (Block 3).

Figure 2.8 shows the conditional relationship of these 64 questions via partial correlation matrix estimation. All four models show a relatively low overall correlation between these 64 questions, except for some specific question pairs. The magnitude of the overall positive partial correlation (blue color) is much greater than that of the negative partial correlation (orange color). Unlike the results in marginal correlation, conditionally, these 64 questions do not show obvious large block structures. Instead, relatively high conditional correlations (greater than 0.3) exist only between some particular question pairs, especially between the 10 questions within Question No. 15.

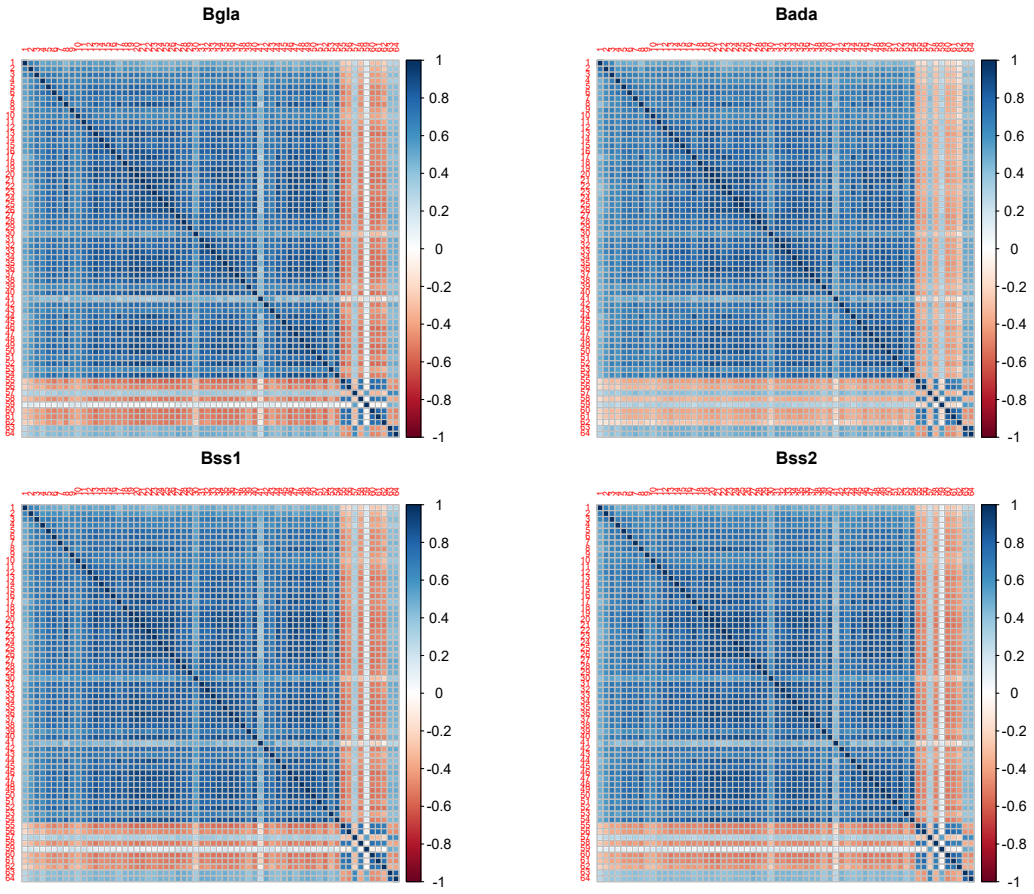

Figure 2.7 The posterior mean estimate of the covariance matrix for Survey Data No. 1.

Graphical Structure Determination

The estimated graphical structure of these 64 ordinal variables for all models of $Bgla$, $Bada$, Bss_1 and Bss_2 are plotted using the R packages *qgragh* and *igraph* and are shown in Figure 2.9 and 2.10. To easier identify Questions No. 14 and 15, these two large groups are labeled as group "A" and "B", respectively. "$A1$" to "$A54$" represent the first 54 questions and "$B1$" to "$B10$" represent questions 55-64. In the graphs, each circle represents one question, with blue and yellow colors representing Questions No. 14 and 15, respectively. The edges between circle pairs represent a partial correlation between the corresponding question pairs, which contain information on the conditional relationship. Figure 2.9 shows the graphs based on partial correlation

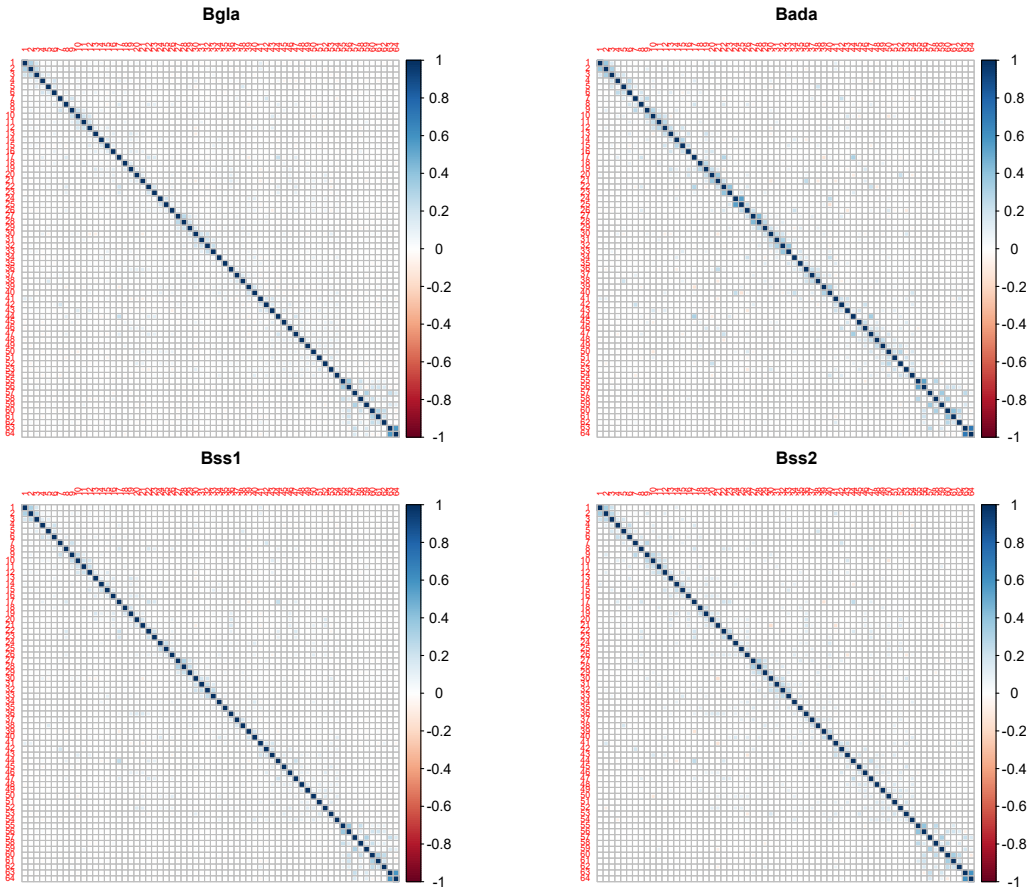

Figure 2.8 The posterior mean estimate of the partial correlation matrix for Survey Data No. 1.

matrix estimation. In Figure 2.9, the edge colors of green and red represent positive and negative partial correlations, respectively, and edge thickness is proportional to the magnitude of the partial correlation.

In the simulation section, for both Scenario 1 (p = 10) and 2 (p = 50), the performance of specificity, sensitivity, and MCC of each scenario using different thresholds are compared. The results for Scenario 2 (p = 50) show that the threshold of 0.05 gives the best performance overall. Since the dimension of our Survey Data No. 1 with p=64 is closer to that of p = 50, we use the same threshold value of 0.05 in constructing the estimated graphical structure for Survey Data No. 1.

Comparing the graphical structures of $Bgla$ and $Bada$ models, the $Bada$ model shows apparent block structures within Question No. 14 (Group A) and 15 (Group B), but the $Bgla$ model fails to separate these two groups. It is similar when we compare the graphical structures of Bss_1 and Bss_2 models. Compared with $\nu_0 = 10^{-3}$ in Bss_2, the relatively large value of $\nu_0 = 10^{-1}$ in Bss_1 is not able to specify the block structures of Questions No. 14 and 15, and thus fail to distinguish between these two groups. Comparing all four models, the $Bada$ and Bss_2 models give similar block structures in graphical structure determination, although the details may differ after careful examination.

We examined pairs of questions exhibiting the strongest positive and negative partial correlations for all four models, and the results are shown in Table 2.4 with the corresponding question description listed in Table 2.5 and 2.6. The left side panel of Table 2.4 shows the strongest positive partial correlations. For all four models, most question pairs come from Question No. 15. The question pair $B9-B10$ has the strongest positive partial correlation for all four models, and the value is much greater than the other pairs within each model, with the largest of 0.71 in $Bada$. These two questions are both in Question No. 15 and are very close in meaning, suggesting that we may reduce one of these questions in the survey to make it more concise. The other question pairs with relatively high partial correlation include $B1-B2$, $B3-B5$, and $B6-B7$ pairs, and we can also consider removing one question from each pair to reduce the size of the survey. From Question No. 14 (TSCC), The question pair $A1-A2$ show a relatively high positive partial correlation for three models of $Bgla$, Bss_1 and Bss_2, whereas for $Bada$ model, question pair $A24-A25$ shows a large partial correlation of 0.54. $A24$ and $A25$ discuss whether the respondent is scared of men or women. A further study can be conducted to check whether this conditional dependence is related to the gender of the respondents or not by analyzing the partial correlation of this question pair for male and female respondents separately.

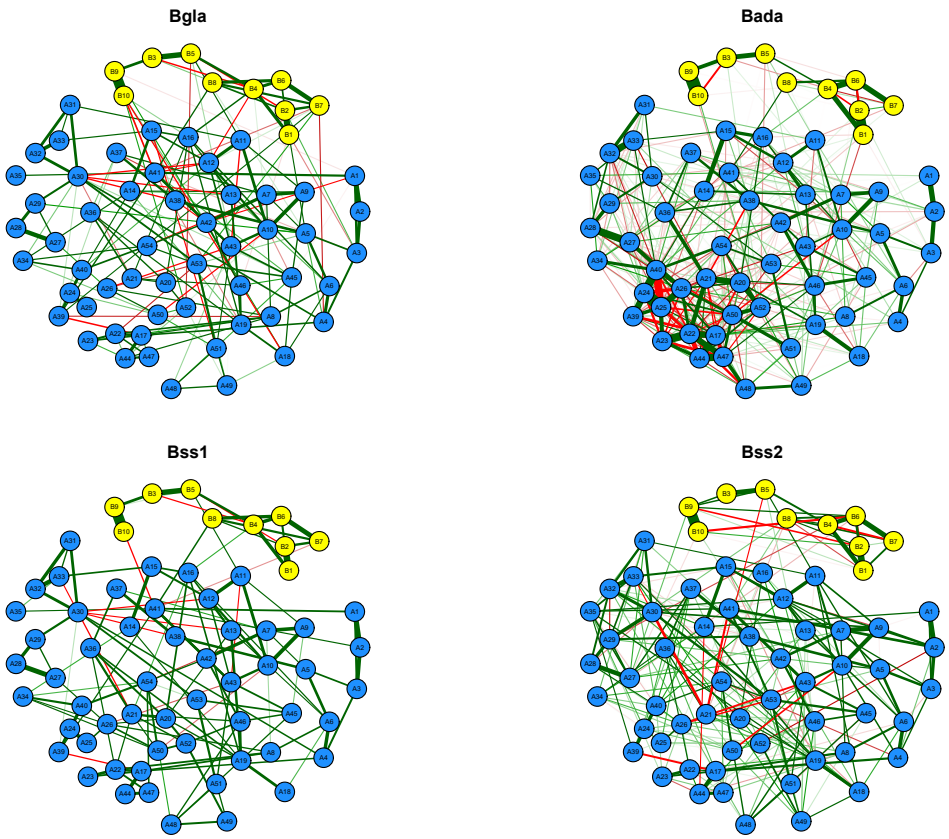

Figure 2.9 The estimated graph structure for Survey Data No. 1 based on the partial correlation matrix.

Analogously, the right side panel of Table 2.4 lists the question pairs exhibiting the strongest negative partial correlations. For all four models, all the top five pairs belong to Question No. 14 (TSCC) and are much smaller in magnitude than that of the positive partial correlation pairs in Table 2.4. With such small values of negative partial correlations in Question No. 14, it is confirmed that the 54-item Trauma Symptom Checklist for Children (TSCC) is a good design even for undergraduate students who are generally 17-24 years old.

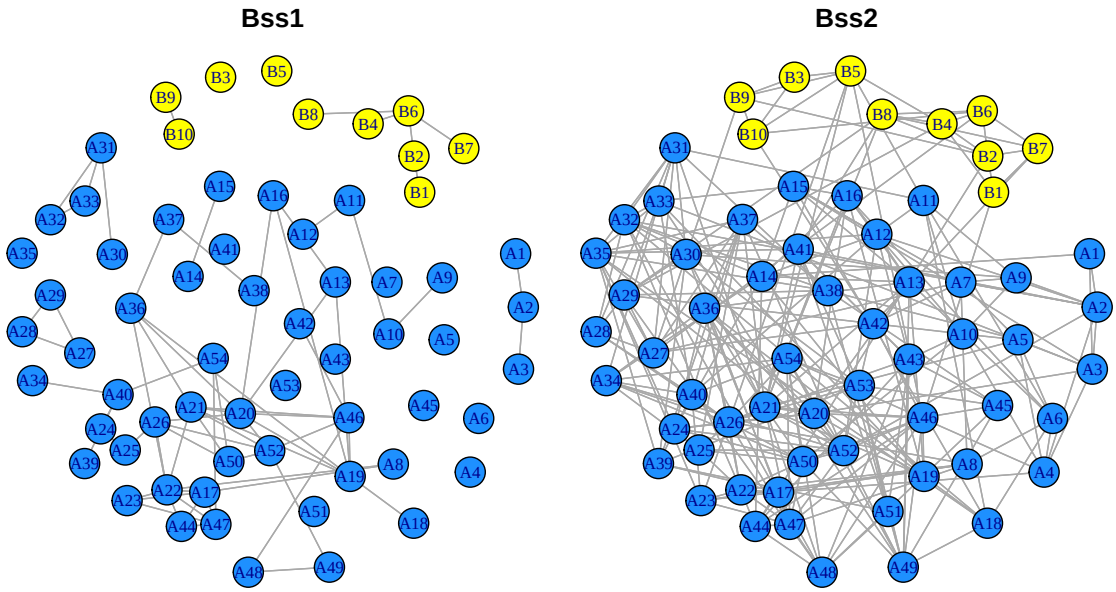

Figure 2.10 The estimated graph structure for Survey Data No. 1 based on the proportion with a threshold value of 0.5.

2.6 Discussion

In the simulation study, for both scenarios, we compared the performance of the three different priors of graphical Lasso ($Bgla$), adaptive graphical Lasso ($Bada$), and spike-and-slab Lasso (Bss). All three priors give good results in matrix estimation of correlation, partial correlation, and precision matrix. Performance in graphical structure determination was also compared among the three priors. In the Bss model, the effect of hyper-parameter ν_0 was also investigated. Based on the simulation study, the adaptive graphical Lasso prior ($Bada$) and spike-and-slab Lasso prior with $\nu_0 = 10^{-3}$ (Bss_2) show much better performance than the graphical Lasso ($Bgla$), which specifies the adaptive performance of $Bada$ and Bss. Within the Bss model, the hyper-parameter $\nu_0 = 10^{-3}$ shows the best performance among the selected three values, which confirms that the performance is sensitive to the selection of ν_0.

Table 2.4 List of pairs of questions with strongest positive (left panel) and negative (right panel) partial correlations for Survey Data No. 1.

Positive					**Negative**				
Pairs	$Bgla$	$Bada$	Bss_1	Bss_2	Pairs	$Bgla$	$Bada$	Bss_1	Bss_2
B9-B10	0.569	0.712	0.588	0.595	A21-A30	-0.076	-0.105	-0.078	-0.157
B1-B2	0.354	0.520	0.373	0.393	A21-A41	-0.062	-0.089	-0.055	-0.135
A1-A2	0.307	0.389	0.312	0.337	A17-A39	-0.089	-0.134	-0.076	-0.134
B3-B5	0.305	0.352	0.307	0.319	A10-A50	-0.07	-0.115	-0.056	-0.112
B6-B7	0.300	0.408	0.314	0.330	A26-A43	-0.071	-0.102	-0.057	-0.112
A27-A28	0.268	0.444	0.291	0.308	B6-B10	-0.046	-0.045	-0.041	-0.099
A2-A3	0.246	0.328	0.264	0.27	A26-A53	-0.057	-0.095	-0.039	-0.094
B6-B8	0.239	0.276	0.246	0.248	A14-A44	-0.044	-0.023	-0.041	-0.09
A17-A44	0.224	0.312	0.235	0.247	A2-A52	-0.045	-0.059	-0.033	-0.088
A32-A33	0.222	0.388	0.248	0.251	A21-B5	-0.043	-0.026	-0.031	-0.084
B3-B9	0.222	0.253	0.221	0.226	A13-A30	-0.073	-0.044	-0.059	0
A9-A10	0.216	0.288	0.225	0.254	A12-A30	-0.07	-0.069	-0.059	-0.025
B1-B4	0.209	0.271	0.215	0.252	A8-A39	-0.062	-0.049	-0.055	0
B4-B6	0.208	0.314	0.219	0.249	A19-A41	-0.061	-0.03	-0.049	0
A31-A32	0.208	0.316	0.234	0.235	A7-A54	-0.061	-0.077	-0.047	-0.025
A24-A25	0.183	0.543	0.216	0.107	A25-A40	0.015	-0.137	0.013	0
A22-A23	0.165	0.399	0.197	0.227	A21-A38	-0.001	-0.097	0	0.002
A20-A21	0.112	0.363	0.145	0.157	A22-A48	-0.01	-0.092	-0.003	0
A44-A47	0.165	0.33	0.192	0.225	A20-A44	-0.008	-0.087	-0.003	-0.002
A7-A9	0.189	0.261	0.2	0.264	A39-A44	-0.055	-0.006	-0.051	-0.003

Ordinal survey data were analyzed using the Gaussian copula graphical model from a Bayesian perspective with the three priors on the precision matrix $\mathbf{\Omega}$. Based on the results, the three priors of graphical Lasso, adaptive graphical Lasso, and spike-and-slab Lasso give similar results in the matrix estimation of both covariance and partial correlation matrices. In graphical structure determination via partial correlation matrix, adaptive graphical Lasso prior ($Bada$) and spike-and-slab Lasso prior with hyper-parameter $\nu_0 = 10^{-3}$ (Bss_2) give similar results, and both of them show the block structure of Question No. 14 and 15. In comparison, the graphical Lasso

Table 2.5 List of questions description for Survey Data No. 1 (Part 1).

A1	Nightmares.
A2	I feared that bad things would happen.
A3	Scary thoughts or images come to mind suddenly.
A7	Loneliness.
A9	Sadness.
A10	I remembered things that happened that I did not like.
A12	I remembered things I was afraid of.
A13	I Wanted to yell loudly and break things.
A14	Crying.
A17	I wanted to have sex.
A19	I wanted to yell at people.
A20	Desire to physically hurt myself.
A21	Desire to physically hurt others.
A22	I thought of touching someone's private parts.
A23	Sexual feelings when I do not want to have them.
A24	Feared of men.
A25	Feared of women.
A26	I kept cleaning myself because I felt dirty.
A27	I felt myself stupid or bad.
A28	I felt like I have done something wrong.
A30	I forgot things.
A31	I felt lost.
A32	I felt nervous or jittery.
A33	I felt scared.
A38	I imagined that I am in somewhere else.
A39	I was afraid of the dark.
A40	I felt scared or upset when I thought about sex.
A41	I worried about things.
A43	I thought of things I do not want to remember.
A44	Feeling sexual arousal.
A47	I cannot stop thinking about sex.
A48	I tried not to feel anything.
A50	I was afraid someone would kill me.
A52	I had suicidal ideation.
A53	I was often in a daze.
A54	I got upset when other people talked about sex.
B1	I think I am as useful as everyone else.
B2	I think I have many good qualities.
B3	Generally speaking, I tend to feel like a loser.
B4	I can get things done as most people do.
B5	I do not feel much proud of myself.
B6	I have a positive attitude towards myself.

Table 2.6 List of questions description for Survey Data No. 1 (Part 2).

B7	Generally speaking, I am satisfied with myself.
B8	I wish I could earn more respect.
B9	I often feel that I am useless.
B10	I often think that I am useless.

prior ($Bgla$) and spike-and-slab Lasso prior with hyper-parameter $\nu_0 = 10^{-1}$ (Bss_1) do not show the pattern that obviously. All three priors yield low partial correlation within Question No. 14, which confirms that this 54-item Trauma Symptom Checklist for Children (TSCC), originally designed for children 8-16 years old, is a good design for undergraduate students with age 17-24 as well. As to Question No. 15, all three priors show a relatively high partial correlation between the sub-questions, which could give us instructions on reducing some sub-questions and thus make the survey more concise.

Chapter 3

Bayesian Gaussian Copula Graphical Models with Adaptive Spike-and-Slab Lasso Prior

3.1 Introduction

In Chapter 2, we compared the performance of Gaussian copula graphical models with the three different priors of graphical Lasso, adaptive graphical Lasso, and spike-and-slab Lasso. For the spike-and-slab Lasso prior, we chose three values of ν_0 and found that Bss_2 model with $\nu_0 = 10^{-3}$ had the best performance, which is comparable to that of the $Bada$ model with adaptive graphical Lasso prior. The best performance was achieved by deliberately and carefully choosing the ν_0 value. For this, we must redo the whole procedure each time we choose a different ν_0 value, which is very time-consuming. In this chapter, we want to allow the data to help choose the ν_0 value. Instead of treating ν_0 as a fixed hyper-parameter, we adaptively update its value and thus automatically yield the best performance without repeating the procedure multiple times. This improved model is labeled as Bss_4. Section 3.2 presents the Bss_4 model and the corresponding block Gibbs sampling algorithm. Section 3.3 covers a specifically designed simulation study to evaluate the performance of the Bss_4 model. Section 3.4 analyzes a real dataset with the Bss_4 model and all other models used in the previous chapter. The results of the simulation study and the real data analysis are briefly summarized in the discussion section of 3.5.

3.2 Model

In the Bss models of Chapter 2, the hyper-parameter ν_0 was set as a fixed value. The performance of the overall Bss model can be improved if we allow ν_0 to be updated dynamically. Ročková (2018) induced a prior distribution on the prior inclusion probability η and set ν_0 deterministically to $\eta/\ 1-\eta$. By tying ν_0 with η, putting a prior only on η would be enough to obtain the desired adaptivity. Based on Ročková (2018), the beta prior $\pi(\eta) \sim Beta(a, b)$ was considered, where $a \ll b$, so that η is small with high probability. Following the settings of Ročková (2018), we set $a = 1$ and $b = 4n$.

The block Gibbs sampling algorithm in Section 2.3 is also applied to the Bss_4 model. The only modification is done in step 2 of the algorithm; the other steps are the same. Here below is step 2 for the Bss_4 model:

2-4 Bss_4 Model with adaptive ν_0

2-4(a). Sample $\lambda \sim$ GA (shape $= p + s$, rate $= t + \sum_{i=1}^{p} \frac{\omega_{ii}}{2}$).

2-4(b). Sample $\eta \sim$ Beta $(\sum_{i<j} \mathrm{r}_{ij} + a, \frac{p(p+1)}{2} - \sum_{i<j} \mathrm{r}_{ij} + b)$, where $a = 1$ and $b = 4n$.

2-4(c). For $1 \leq i < j \leq p$, sample $\mathrm{r}_{ij} \sim$ Bernoulli $(p = \frac{A}{A+B})$, where $A = \frac{\eta}{2\nu_1}\exp\Big\{ - \frac{|\omega_{ij}|}{\nu_1}\Big\}$, and $B = \frac{1-\eta}{2\nu_0}\exp\Big\{ - \frac{|\omega_{ij}|}{\nu_0}\Big\}$, where $\nu_0 = \frac{\eta}{1-\eta}$.

2-4(d). For $1 \leq i < j \leq p$, sample $u_{ij} \sim$ INV-GAU $(\mu^{'}, \lambda')$ where $\mu^{'} = \frac{1}{\nu_0|\omega_{ij}|}$ and $\lambda^{'} = \frac{1}{\nu_0^2}$, if $\mathrm{r}_{ij} = 0$, and $\mu^{'} = \frac{1}{\nu_1|\omega_{ij}|}$ and $\lambda^{'} = \frac{1}{\nu_1^2}$, if $\mathrm{r}_{ij} = 1$. Then update $\tau_{ij} = 1/u_{ij}$.

3.3 Simulation Study

3.3.1 Simulation Setting

This simulation experiment is designed to compare the performance of the Bss_4 model with that of the five models in Chapter 2 regarding parameter estimation and

structure learning. The simulation setting for this chapter is motivated by the results of Chapter 2. Section 2.4 considered the two-block covariance structure of simulation Scenario 1 and 2. For both scenarios, we investigated a series of cutoff values of the partial correlation, and the value 0.01 gave the best performance in terms of graphical structure determination. This cutoff value of 0.01 was then applied to the real data analysis in Section 2.5. Since the dependence structures of both scenarios in the simulation study were very different from that of the real data, applying this optimal cutoff value to the real data analysis is not convincing. Based on this, we made improvements in designing the simulation settings of this chapter. In this chapter, the simulation setting is closely related to the real data structure in Section 3.4. In Section 3.4, we obtained the estimated posterior mean of the covariance matrix of Survey Data No. 2 using the Bss_4 model. This estimate was then used in this simulation study as the true covariance matrix and is labeled as Scenario 3. The true covariance and the corresponding partial correlation matrix are shown later in the simulation results.

We generated 200 samples of size $n = 1000$ for Scenario 3. For each generated sample, we fit each model of $Bgla$, $Bada$, Bss_1, Bss_2, Bss_3, and Bss_4. The settings of the hyper-parameters for the five models of $Bgla$, $Bada$, Bss_1, Bss_2, and Bss_3 are the same as those in Chapter 2. As to the Bss_4 model, the hyper-parameters of the prior inclusion probability η are set as $a = 1$ and $b = 4n$, respectively.

3.3.2 Simulation results

Matrix Estimation

For each of the 200 samples, the posterior distribution of the covariance matrix $\mathbf{\Gamma}$, precision matrix $\mathbf{\Omega}$, and partial correlation matrix are sampled using each of the six models. We ran 5000 iterations using the block Gibbs sampler, and the first 1000 were burn-in iterations. The posterior mean was then calculated based on the 4000

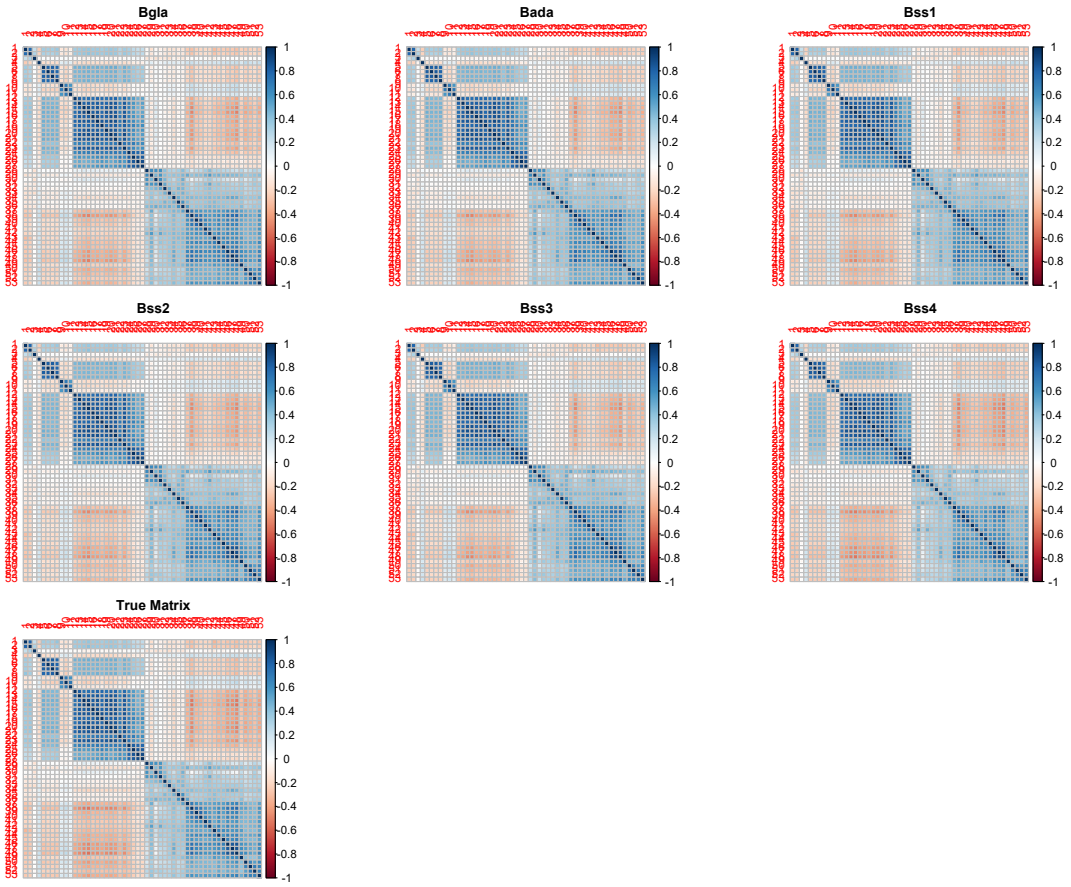

Figure 3.1 Covariance matrix estimation for Scenario 3 (p=53).

iterations, and the mean over these 200 samples was then used as the estimate for each matrix.

Figure 3.1 and 3.2 show the estimation results of the covariance matrix as well as the partial correlation matrix using all six models of $Bgla$, $Bada$, Bss_1, Bss_2, Bss_3 and Bss_4 for Scenario 3 (p = 53). In Figure 3.1, the Bss_4 model shows good results in covariance matrix estimation compared with the true covariance matrix, which is similar to that of the previous five models. This means that the Bss_4 also performs well in identifying the marginal relationship between the 53 ordinal variables. It is a similar case in partial correlation matrix estimation in Figure 3.2, where the Bss_4 model performs well compared to the true partial correlation matrix.

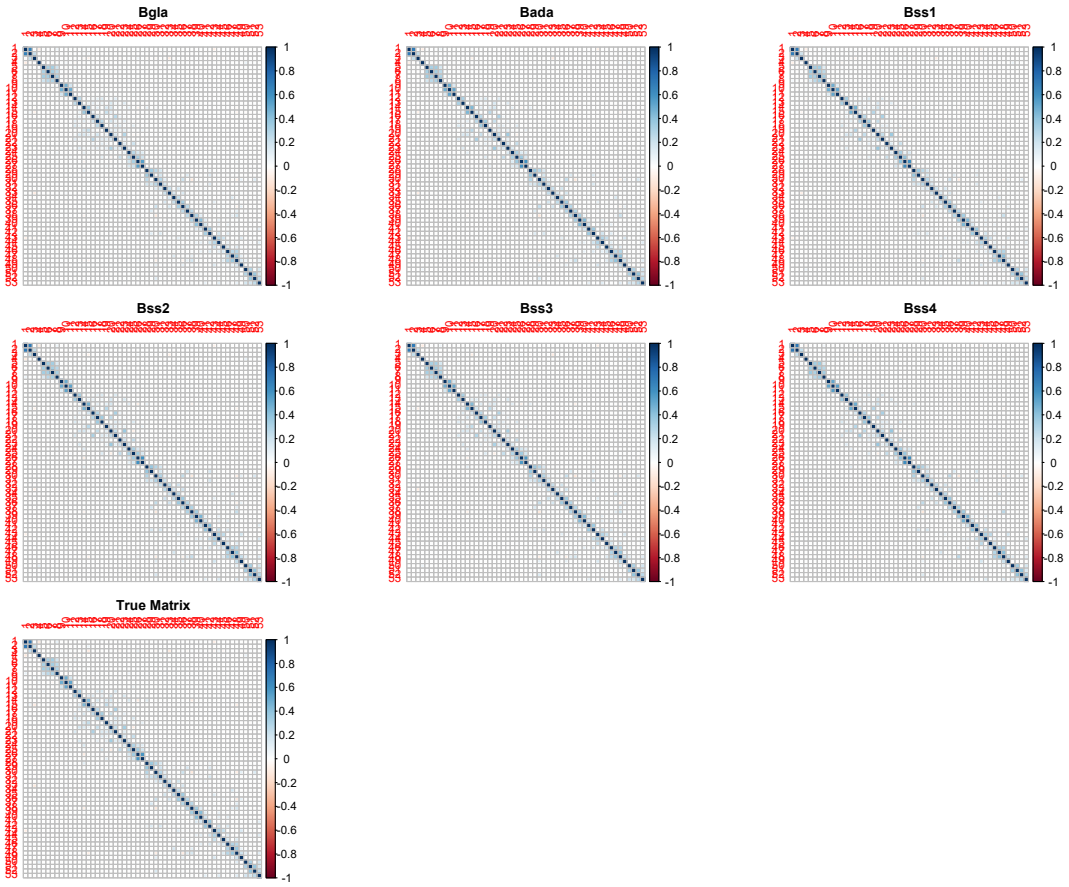

Figure 3.2 Partial correlation matrix estimation for Scenario 3 (p=53).

The performance of the precision matrix estimation is also evaluated for the Bss_4 model. Table 3.1 compares the calculated mean, median, and standard deviation of Frobenius Loss and Stein's Entropy Loss for the precision matrix $\mathbf{\Omega}$ for Scenario 3. Regarding the Frobenius Loss, the adaptive Bss_4 model did not yield the best performance among the six models. However, this adaptive model performed best within the four Bss models regarding Stein's Entropy Loss. Overall, the results are similar among all six models, with the $Bada$ model performing best regarding both the Frobenius Loss and Stein's Entropy Loss.

Table 3.1 Mean, median and standard deviation of Frobenius Loss (FL) and Stein's Entropy Loss (EL) for the precision matrix $\mathbf{\Omega}$ for Scenario 3 (p=53).

Models	FL_Mean	FL_Median	FL_sd	EL_Mean	EL_Median	EL_sd
$Bgla$	0.0422	0.0415	**0.0047**	2.0552	2.0428	0.1679
$Bada$	**0.0356**	**0.0333**	0.0107	**0.99**	**0.978**	**0.1187**
Bss_1	0.0516	0.0493	0.0141	1.4623	1.4502	0.1364
Bss_2	0.0593	0.0563	0.0169	1.3937	1.3714	0.1507
Bss_3	0.104	0.1000	0.0244	2.5133	2.5002	0.2787
Bss_4	0.0672	0.063	0.0224	1.2415	1.2336	0.1357

Graphical Structure Determination

To assess the performance of the graphical structure learning of the Bss_4 model, we computed the specificity, sensitivity, and Matthews Correlation Coefficient (MCC). Table 3.2 summarizes the average specificity, sensitivity, and Matthews Correlation Coefficient (MCC) of the partial correlation matrix for Scenario 3 using different cutoff values based on the 200 samples for all six models. Six cutoff values of 0.01, 0.02, 0.05, 0.1, 0.2, and 0.3 are selected, and the mean and standard deviation of the corresponding average specificity, sensitivity, and MCC are reported. The cutoff value of 0.2 gives the best MCC performance for each model. In comparison, the optimal cutoff value of 0.05 in Section 2.4 also performs relatively well for Scenario 3 for each model. Similar to that of the $Bada$ and Bss_2 models, the Bss_4 model has a very high average MCC among different cutoff values, which is illustrated in Figure 3.3.

Except for the partial correlation matrix, we also calculated the average specificity, sensitivity, and MCC for the precision matrix $\mathbf{\Omega}$ for the four Bss models. The calculation includes two parts. In the first part, we did the calculation based on the posterior probability of $\mathrm{r}_{ij} = 1$ given the observed data Y. For each of the 200 samples, the simulation estimate of each off-diagonal entry ω_{ij} is assigned as 1 or 0

Table 3.2 Average specificity (SP%), sensitivity (SE%) and MCC (MCC%) of the partial correlation matrix for Scenario 3 (p=53).

		Bgla				***Bada***	
Cutoff	SP%	SE%	MCC%	Cutoff	SP%	SE%	MCC%
	Mean(sd)	Mean(sd)	Mean(sd)		Mean(sd)	Mean(sd)	Mean(sd)
0.01	33.7(1.5)	85.8(1.9)	17(2.1)	0.01	91.6(0.9)	56.5(1.8)	50.1(2.2)
0.02	59.5(1.6)	89.9(1.9)	32.4(1.6)	0.02	96.4(0.5)	76.4(2.2)	72.2(2.3)
0.05	93.1(0.7)	87.9(2.1)	67.6(2.6)	0.05	99.1(0.3)	81.1(2.4)	84.6(2.2)
0.1	99.4(0.2)	79(2.8)	84(2.3)	0.1	99.5(0.2)	81.1(2.8)	85.9(2.2)
0.2	99.7(0.1)	83.6(3.8)	86(3.2)	0.2	99.4(0.2)	93.5(3.2)	88(2.9)
0.3	99.9(0.1)	70.1(7.2)	79.5(5.4)	0.3	99.7(0.1)	85.1(6.9)	84.1(5.6)
		***Bss*$_1$**				***Bss*$_2$**	
Cutoff	SP%	SE%	MCC%	Cutoff	SP%	SE%	MCC%
	Mean(sd)	Mean(sd)	Mean(sd)		Mean(sd)	Mean(sd)	Mean(sd)
0.01	58.5(1.6)	78.3(2.2)	29.5(2.1)	0.01	97.9(0.4)	43(1.5)	53.8(2)
0.02	83.3(1.1)	83.8(2.3)	50.9(2.2)	0.02	98.2(0.3)	66.2(2.3)	71.5(2.5)
0.05	98.5(0.3)	76.3(2.5)	78.3(2.4)	0.05	98.9(0.3)	77.7(2.7)	80.9(2.7)
0.1	99.8(0.1)	70.7(2.9)	81.5(2.1)	0.1	99(0.2)	83.3(2.8)	83.8(2.5)
0.2	99.5(0.1)	87.2(4.2)	86.2(3.4)	0.2	99(0.2)	95.5(2.7)	84.5(2.7)
0.3	99.7(0.1)	86.2(6.6)	83.8(5.2)	0.3	99.6(0.1)	87.5(6.5)	81.8(5.2)
		***Bss*$_3$**				***Bss*$_4$**	
Cutoff	SP%	SE%	MCC%	Cutoff	SP%	SE%	MCC%
	Mean(sd)	Mean(sd)	Mean(sd)		Mean(sd)	Mean(sd)	Mean(sd)
0.01	60.7(2)	63(2.3)	19.2(2.4)	0.01	97.3(0.5)	46(2)	54.6(2)
0.02	72.8(1.7)	77.6(2.6)	35.1(2.4)	0.02	99.2(0.3)	62.2(2.4)	72.7(2.3)
0.05	87.8(1.3)	85.5(2.7)	55.1(3)	0.05	99.7(0.2)	68.4(2.7)	79.5(2.2)
0.1	97.8(0.4)	86.1(2.8)	79.2(2.8)	0.1	99.5(0.2)	73.7(2.6)	81.4(2.2)
0.2	99.4(0.2)	92.4(3.1)	87.9(3.1)	0.2	98.9(0.2)	96.1(2.7)	83.5(2.7)
0.3	99.7(0.1)	83.9(6.5)	83.4(4.9)	0.3	99.5(0.2)	90.2(5.7)	81.4(4.8)

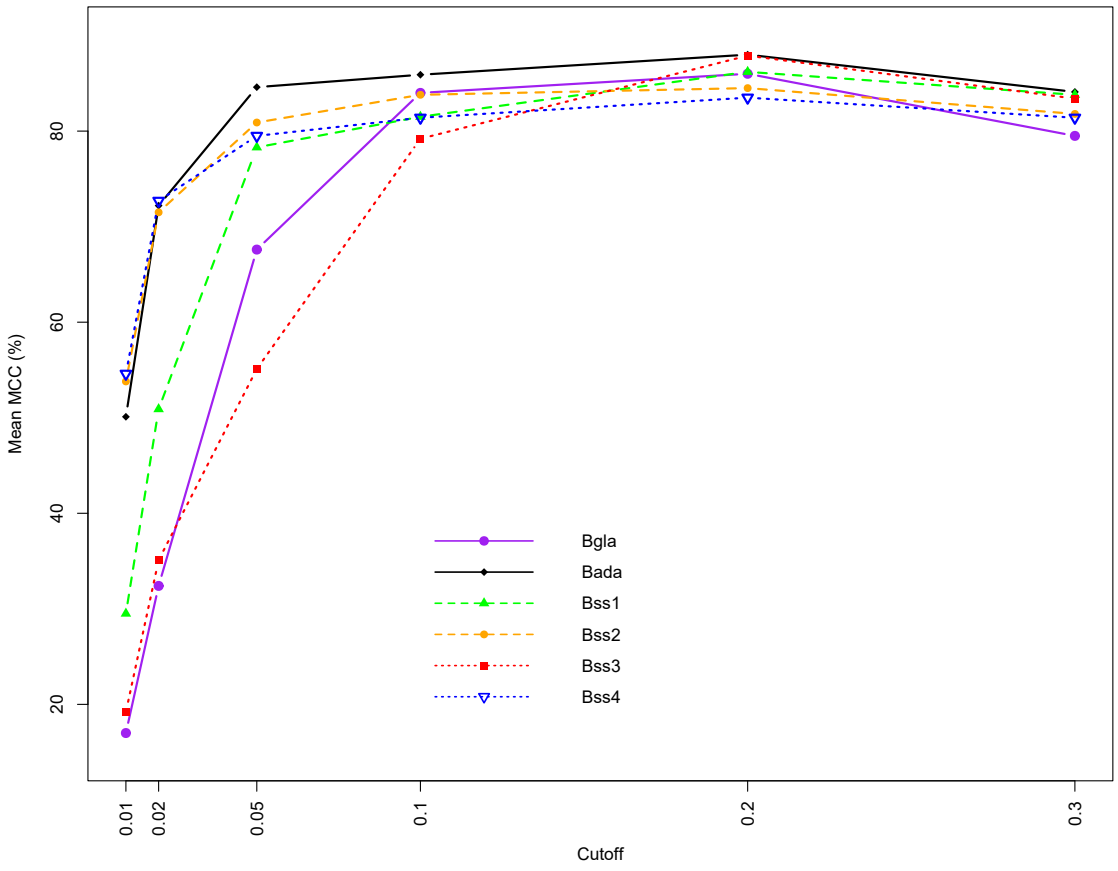

Figure 3.3 Average MCC (%) of the partial correlation matrix with different cutoff values for Scenario 3 (p=53).

according to $p(\mathrm{r}_{ij} = 1|Y)$, which is the estimated posterior mean of r_{ij} among 4000 iterations as follows:

- If $p(\mathrm{r}_{ij} = 1|Y) > 0.5$, then ω_{ij}=1;
- If $p(\mathrm{r}_{ij} = 1|Y) \leq 0.5$, then ω_{ij}=0.

The result of this part is shown in Table 3.3. In the second part of the calculation for the precision matrix, the procedure is the same as that of the partial correlation matrix, where we set a list of cutoff values for each entry in the precision matrix and calculate the corresponding average sensitivity, specificity, and MCC. The results are shown in Table 3.4. Based on the results in Table 3.3, the Bss_4 model with adaptive

Table 3.3 Average specificity (SP%), sensitivity (SE%) and MCC(MCC%) of the precision matrix $\mathbf{\Omega}$ for Scenario 3 (p=53) based on the proportion of $r_{ij} = 1$ given observed data Y.

Model	ν_0	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)
Bss_1	10^{-1}	100(0)	33.2(2.1)	55.8(1.9)
Bss_2	10^{-3}	98(0.3)	80.9(2.6)	78(2.5)
Bss_3	10^{-5}	64.8(2)	92(2.5)	32.7(1.9)
Bss_4	Adaptive	99.7(0.1)	70.8(2.7)	80.7(2.3)

ν_0 yields the best performance in terms of structure determination within the four Bss_4 models, and it is slightly better than that of the Bss_2 model. This result is consistent with that of Table 3.4, which shows that Bss_4 and Bss_2 give comparably high performance in terms of graphical structure determination based on the precision matrix.

The latent graphical structure of Scenario 3 based on the partial correlation matrix with a threshold value of 0.05 is shown in Figure 3.4. The difference between the six models is not that obvious in this figure. However, the superiority of the Bss_4 model is shown in Figure 3.5, where the graphical structure is determined based on the proportion of $r_{ij} = 1$ given the observed data Y. Based on the results, the Bss_4 model yields the best graphical structure, most close to the true graphical structure. In comparison, the Bss_1 model is too sparse that it misses the dependence between some variable pairs. In contrast, the Bss_3 model is too dense to show the independence between many variable pairs. The Bss_2 model is much better than the Bss_1 and Bss_3 model but with a slightly denser structure than that of the Bss_4 model and the true model. However, we have to keep in mind that for the Bss_2 model, the fixed hyper-parameter value ν_0 was set carefully so that it would not be too big or too small, and this primarily rely on the experience of the researcher conducting the

Table 3.4 Average specificity (SP%), sensitivity (SE%) and MCC (MCC%) of the precision matrix $\mathbf{\Omega}$ for Scenario 3 (p=53) based on different cutoff values.

$\boldsymbol{Bss_1}$				$\boldsymbol{Bss_2}$			
Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)	Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)
0.01	100(0)	14.4(0.9)	34.3(1.2)	0.01	98.9(0.3)	39.8(1.4)	54.4(1.9)
0.02	100(0)	23.7(1.5)	46.2(1.6)	0.02	98.7(0.3)	63.6(2.2)	71.6(2.4)
0.05	100(0)	29.7(1.9)	52.5(1.8)	0.05	98.5(0.3)	77.5(2.7)	79.1(2.7)
0.1	99.9(0.1)	37.7(2.5)	59.1(2.2)	0.1	97.4(0.3)	85.9(2.7)	77.2(2.4)
0.2	99.3(0.1)	70.8(4)	72.6(3.3)	0.2	94(0.3)	99.5(1.1)	57.1(1)
0.3	98.5(0.2)	91.7(4.1)	67.4(3)	0.3	92.6(0.3)	99.9(0.6)	40.7(0.7)
$\boldsymbol{Bss_3}$				$\boldsymbol{Bss_4}$			
Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)	Cutoff	SP% Mean(sd)	SE% Mean(sd)	MCC% Mean(sd)
0.01	64.5(2.1)	58.7(2.3)	19(2.4)	0.01	99.8(0.1)	31.4(1.2)	50.8(1.4)
0.02	65(2)	77.5(2.7)	28.4(2.3)	0.02	99.8(0.1)	51.6(2)	68.4(1.9)
0.05	65.1(2)	89.1(2.7)	32.8(2.1)	0.05	99.8(0.1)	64.4(2.5)	77.3(2.1)
0.1	64.3(2)	95.1(2.1)	32(1.7)	0.1	99.2(0.2)	76(2.6)	80.7(2.2)
0.2	61.8(1.9)	99.9(0.4)	21.9(0.8)	0.2	96.5(0.2)	98.7(1.7)	67.3(1.6)
0.3	60.8(1.8)	100(0)	15.6(0.6)	0.3	95(0.2)	99.6(1.5)	48.2(1.1)

simulation. Instead, the Bss_4 model with adaptive ν_0 is more convenient since we do not need deliberately choose a ν_0 value to guarantee a good performance.

3.4 Real Data Analysis

3.4.1 Dataset Description

In this section, we analyze ordinal survey data labeled Survey Data No. 2. This dataset is part of the Life Satisfaction Survey (LSS) dataset described in Section 2.5. It consists of questions 1, 2, 3, 4, 6, 7, and 9 from the LSS dataset. There are 7, 16, and 26 sub-questions within questions 6, 7, and 9, respectively. In this dataset, questions 1, 2, 3, and 4 are four single questions describing the overall satisfactory status

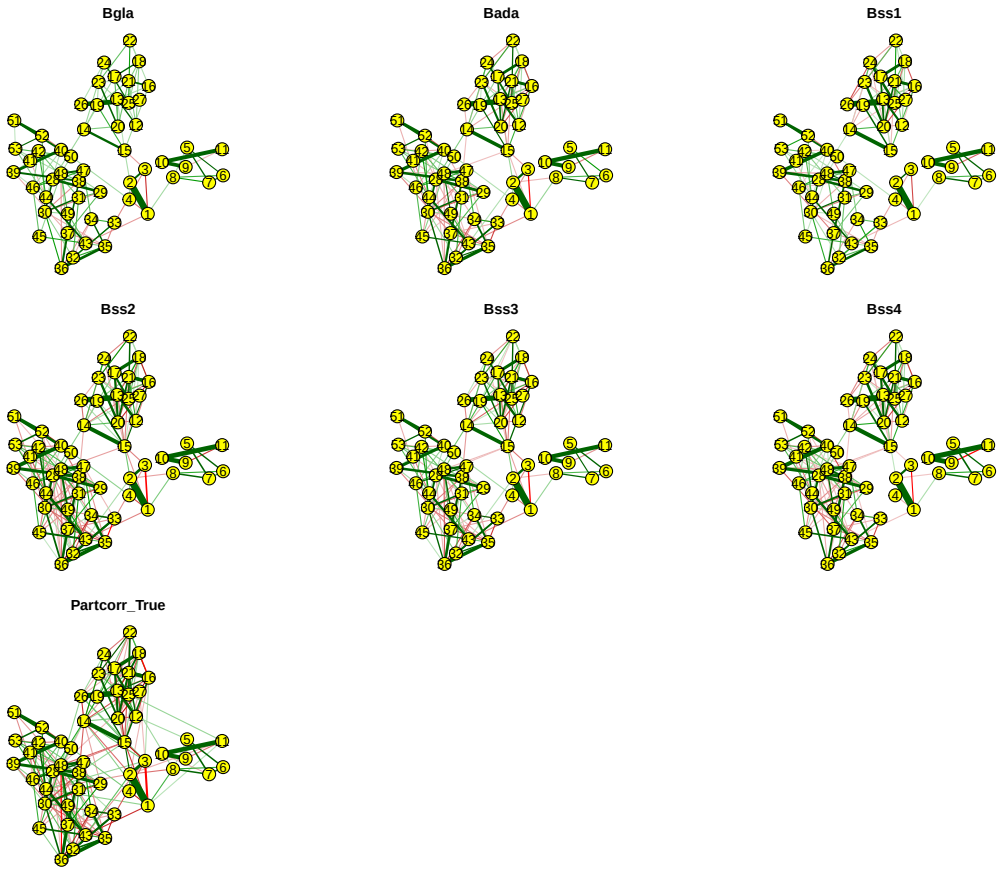

Figure 3.4 The estimated graphical structure of Scenario 3 (p=53) based on partial correlation matrix with cutoff value 0.05.

of the respondents. Question No. 6 is a group of sub-questions about respondents' confidence in their future lives. The 16 sub-questions in Question No. 7 ask whether they received any help when the respondents were in a hard time in life. Question No. 9 asks about general situations that the respondents face in life. Survey Data No. 2 has 2369 observations and 53 variables (questions). Questions 1, 2, and 3 have five ordinal levels of 1-5 representing "Very Poor", "Poor", "Acceptable", "Good" and "Very Good", respectively. For question No. 4, these 5 levels represent "Very Good", "Good", "Acceptable", "Poor", and "Very Poor", respectively. Question No. 6 has four ordinal levels, with 1-4 representing "Strongly Disagree", "Disagree", "Agree", and "Strongly Agree", respectively. Question No. 7 has five ordinal levels, with 1-5

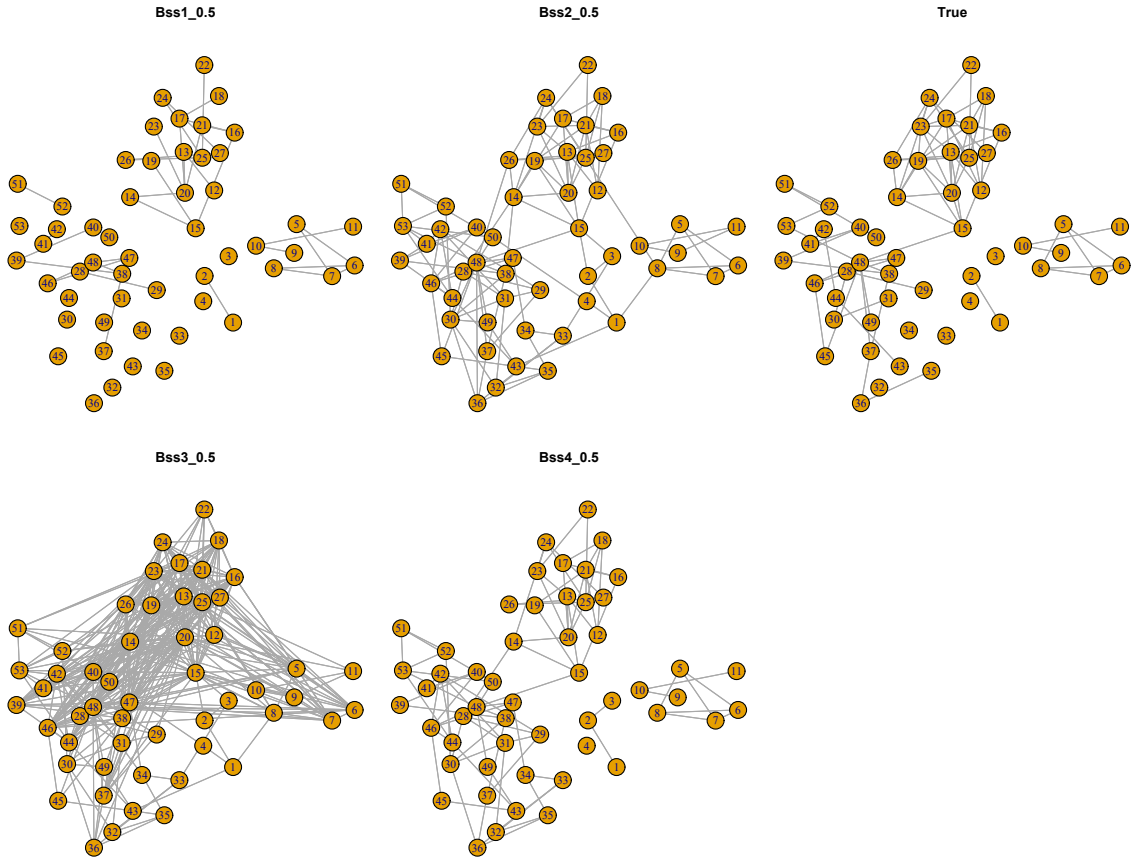

Figure 3.5 The estimated graphical structure of Scenario 3 (p=53) for the Bss models based on proportion with threshold value 0.5.

representing "Strongly Disagree", "Disagree", "Undecided", "Agree", and "Strongly Agree", respectively. In question No. 9, there are five ordinal levels with 1-5 representing "Never Happened", "Happened, with no influence on me", "Happened, with slight influence on me", "Happened, with moderate influence on me", and "Happened, with severe influence on me", respectively.

3.4.2 Analysis Methodology

We apply the six Gaussian graphical copula models to Survey Data No. 2 and analyze the association between the questions. For each model, we run the block Gibbs sampling algorithm for a total of 10,000 iterations, with the first 2,000 iterations

as the burn-in period. Similar to that in Section 3.3 for the simulation study, in analyzing Survey Data No. 2, the settings of the hyper-parameters for the five models of $Bgla$, $Bada$, Bss_1, Bss_2, and Bss_3 are the same as that in Chapter 2. As to the Bss_4 model, the hyper-parameters of the prior inclusion probability η are set as $a = 1$ and $b = 4n$, respectively. For the $Bgla$ and $Bada$ models, the block Gibbs sampling algorithm stopped after the first 200 iterations due to the computational instability of these two models. We collect results from the four Bss models and compare the performance of these four models later in the results analysis part.

3.4.3 Results

Matrix Estimation

Figure 3.6 and 3.7 show the plots of the estimated posterior mean of the covariance matrix and partial correlation matrix, respectively, using all four models of Bss_1, Bss_2, Bss_3 and Bss_4. Figure 3.6 shows the marginal relationship of these 53 questions via covariance matrix estimation, and the Bss_4 model gives similar results as that of the previous three Bss models. Marginally, we can see the block structure within questions 7 and 9. Moreover, within Question No. 6, we can observe two blocks of questions $f601 - f604$ (Block 1) and questions $f605 - f607$ (Block 2) with a strong positive correlation within each block and a moderate negative correlation between these two blocks.

Figure 3.7 shows the conditional relationship of these 53 questions via partial correlation matrix estimation. The Bss_4 model result is similar to the previous three models. All four models show a relatively low overall correlation between these 53 questions, except for some specific question pairs. The magnitude of the overall positive partial correlation (blue color) is much greater than that of the negative partial correlation (orange color). Unlike the results in marginal correlation, conditionally, these 53 questions do not show apparent large block structures. Instead,

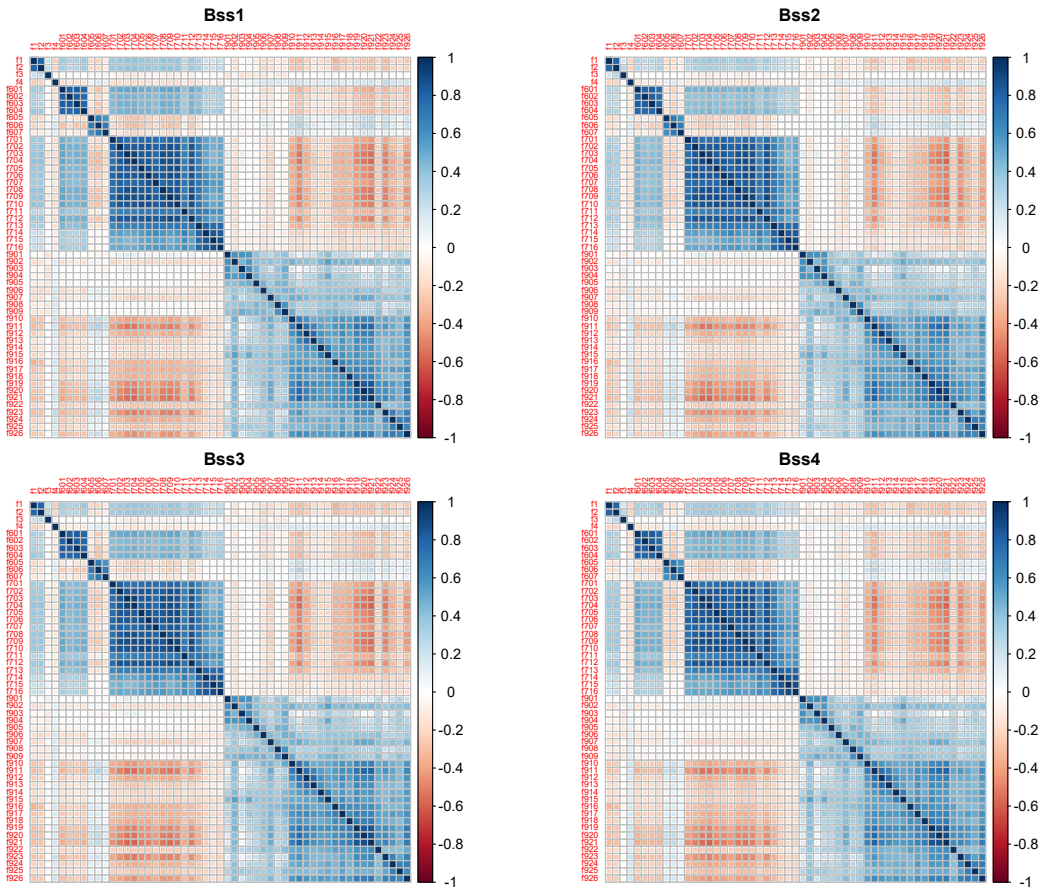

Figure 3.6 The posterior mean estimate of the covariance matrix for Survey Data No. 2.

relatively high conditional correlations (greater than 0.3) exist only between some particular question pairs, especially between the neighboring questions like $f1 - f2$ and $f601 - f604$.

Graphical Structure Determination

The estimated latent graphical structure of Survey Data No. 2 for all four *Bss* models are plotted using the R packages *qgragh* and *igraph* and are shown in Figure 3.8 and 3.9. Figure 3.8 shows the estimated graphical structure based on the partial correlation matrix, where we utilized the cutoff value of 0.05 for this dataset. Similar to Section 2.5, each circle in the graph represents one question, with different colors

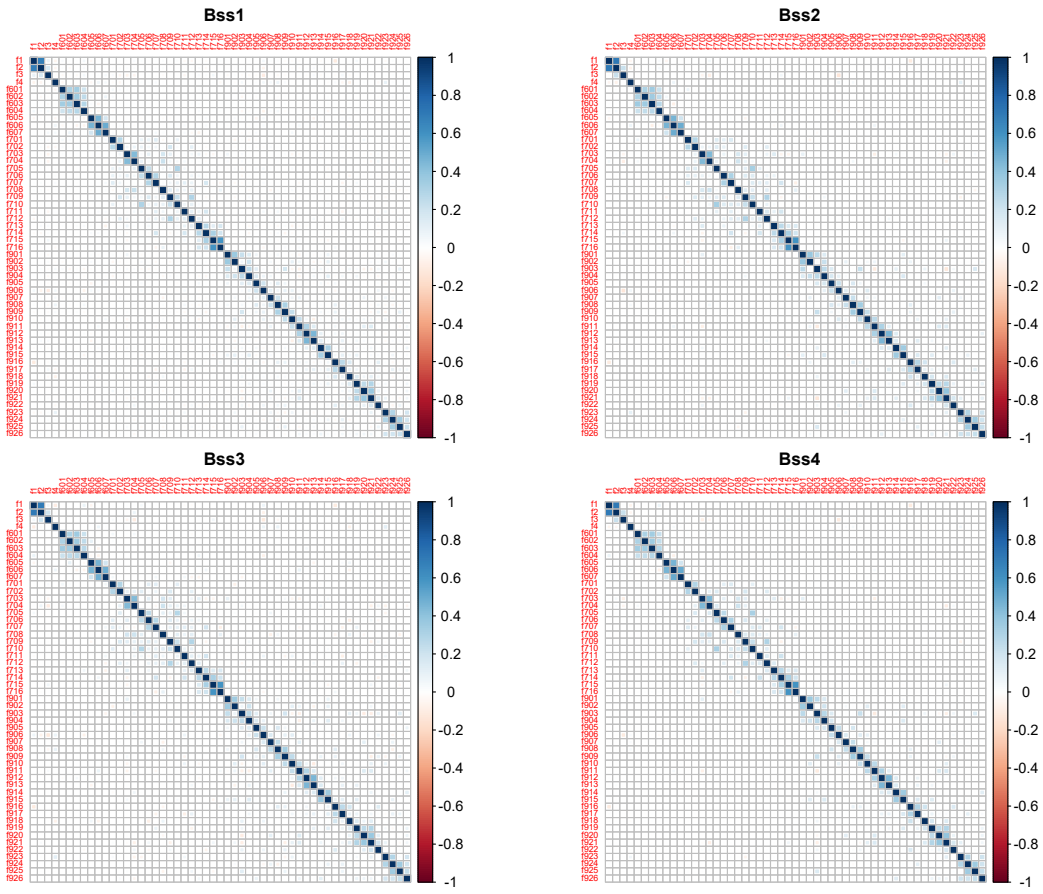

Figure 3.7 The posterior mean estimate of the partial correlation matrix for Survey Data No. 2.

representing different questions. The edges between circle pairs represent a partial correlation between the corresponding question pairs, which contain information on the conditional relationship. Edge colors of green and red represent positive and negative partial correlations, respectively, and edge thickness is proportional to the magnitude of the partial correlation. Based on Figure 3.8, it is not easy to tell the difference between the estimated graphical structures of the four models. However, we can indeed identify the blocking structure of Questions No. 6, 7, and 9 for the Bss_4 model.

To check the results more thoroughly, we examined pairs of questions exhibiting the strongest positive and negative partial correlations for all four models. The results

are shown in Table 3.5 with the corresponding question description listed in Table 3.6 and 3.7. The left side panel of Table 3.5 shows the strongest twenty positive partial correlations. The top seven pairs with the highest positive partial correlation are the same for all four models. The question pair with the strongest positive partial correlation is $f1 - f2$ asking the respondents their feelings about their current living conditions and whether they are satisfied with their current life. All question pairs come from the same block (for example, the two questions in the $f715 - f716$ pair are both from the Question No. 7 block).

The right side panel of Table 3.5 lists the question pairs exhibiting the strongest negative partial correlations. Most pairs come from between-block questions, except that of $f910 - f913$, $f903 - f911$, $f903 - f921$, $f903 - f920$, and $f909 - f921$, where the questions come from the same block of Question No. 9, and that of $f603 - f606$ and $f707 - f716$, where the questions come from the same block of Question No.6 and 7, respectively. Overall, the negative partial correlations are small compared with the strongest positive partial correlations.

We also analyzed the graphical structure based on the proportion with the threshold value of 0.5, and the results are shown in Figure 3.9. Based on the results, the Bss_1 model shows the simplest structure, and the block structure of Question No. 9 can hardly be identified, with only a few pairs detected within this question. For Bss_2 and Bss_3 models, the block structure of Questions No. 7 and 9 are not that obvious, and we could barely separate these two blocks using the Bss_3 model. The Bss_4 model shows the best block identification performance.

3.5 Discussion

This chapter is an extension of Chapter 2. In Chapter 2, we mainly analyzed the ordinal data with Bayesian Gaussian copula graphical models with different priors of graphical Lasso ($Bgla$), adaptive graphical Lasso ($Bada$), and spike-and-slab Lasso

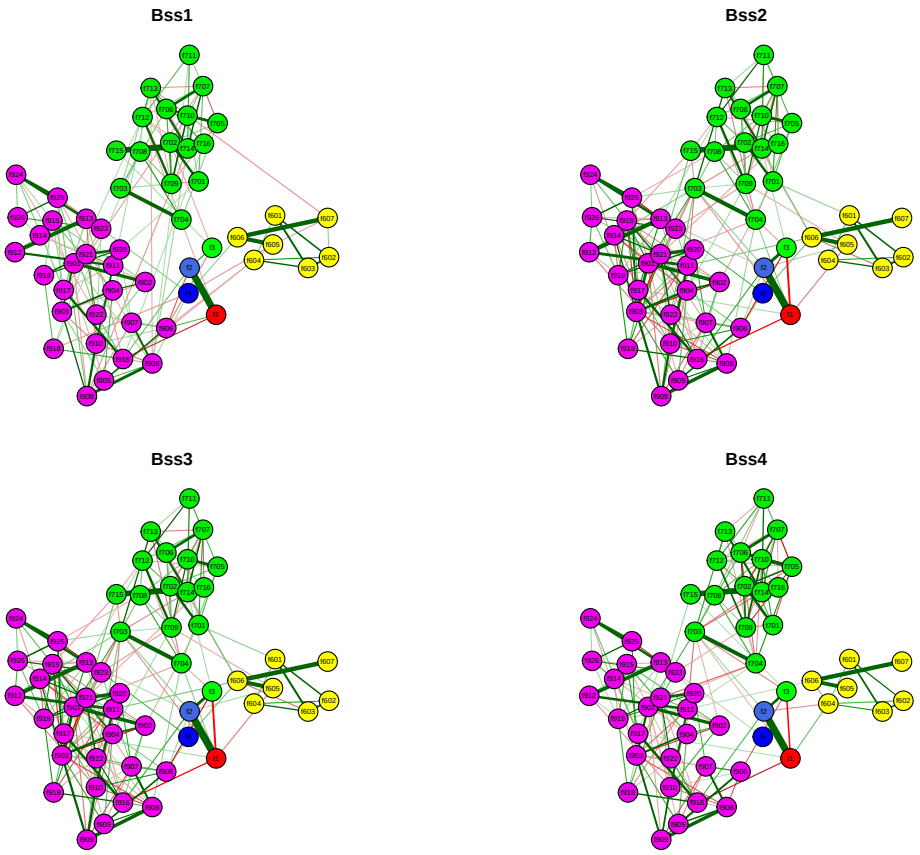

Figure 3.8 The estimated graph structure for Survey data No. 2 based on the partial correlation matrix.

(Bss) priors. Within the Bss models, we chose three values for the hyper-parameter ν_0 and found that the Bss_2 model with $\nu_0 = 10^{-3}$ gave the best performance. The Bss_2 model with this carefully selected ν_0 yielded comparable performance to that of the $Bada$ model. This motivated us to continue working on the Bss model to explore the best ν_0 value. Instead of setting ν_0 as a fixed value in this chapter, we tied ν_0 with the prior inclusion probability η. We set $\nu_0 = \eta/\ 1 - \eta$ and set a $Beta$ prior distribution to η. By doing so, we adaptively updated ν_0 and thus saved the time of running the whole procedure for each chosen ν_0 value, which makes this adaptive Bss_4 model convenient and efficient.

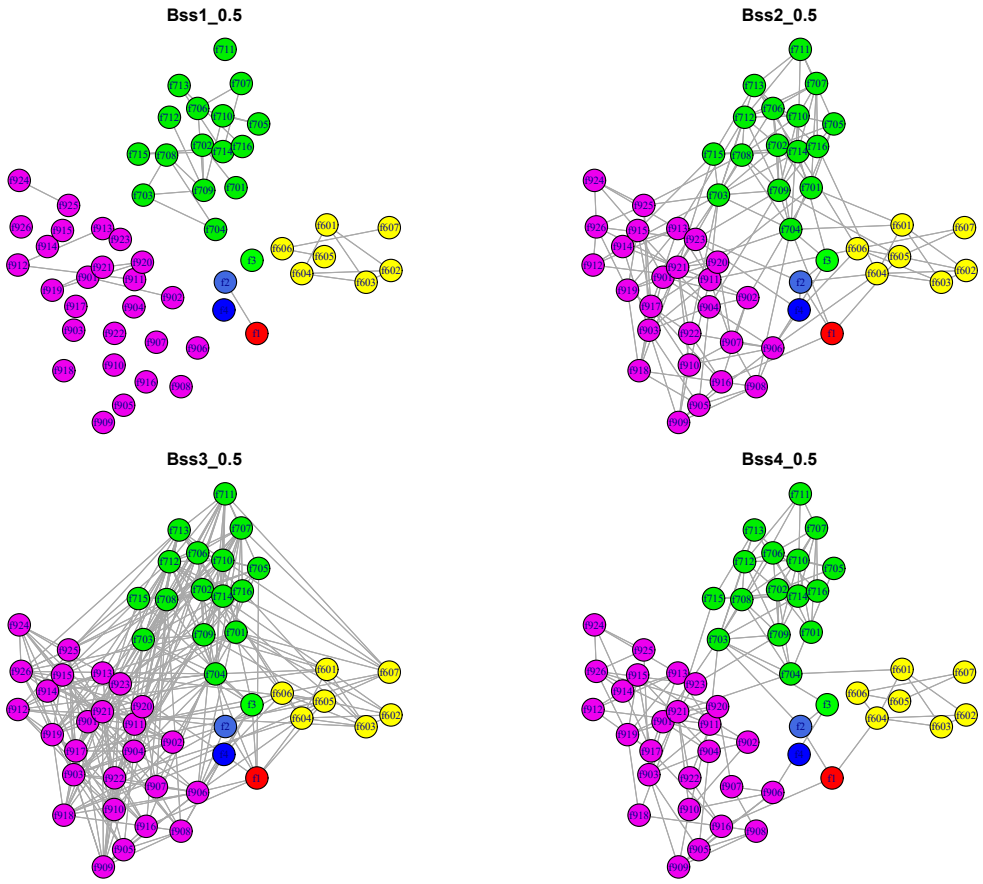

Figure 3.9 The estimated latent graph structure for Survey data No. 2 based on proportion with threshold value 0.5.

To assess the performance of this adaptive Bss_4 model in terms of both matrix estimation and graphical structure determination, we conducted a simulation study with the simulation setting resembling the structure of Survey Data No. 2. Compared with the simulation settings of Scenarios 1 and 2 in Chapter 2, this improved version of simulation Scenario 3 gives better guidance in constructing the graphical structure for the real ordinal data of Survey Data No. 2. Comparing the results of the Bss_4 model and the other five models investigated in Chapter 2, Bss_4 yields similar results as that of the other five models in terms of matrix estimation for both the partial correlation matrix and the covariance matrix. In terms of graphical structure determination, Bss_4 has comparable performance to that of the $Bada$ and Bss_2 models. Within the

Table 3.5 List of pairs of questions with strongest positive (left panel) and negative (right panel) partial correlations for Survey Data No. 2.

Positive					**Negative**				
Pairs	Bss_1	Bss_2	Bss_3	Bss_4	Pairs	Bss_1	Bss_2	Bss_3	Bss_4
f1-f2	0.690	0.709	0.700	0.707	f1-f916	-0.101	-0.108	-0.106	-0.109
f715-f716	0.598	0.597	0.595	0.597	f3-f906	-0.093	-0.131	-0.117	-0.105
f606-f607	0.493	0.499	0.486	0.500	f910-f913	-0.070	-0.009	-0.091	-0.008
f605-f606	0.463	0.461	0.467	0.464	f903-f911	-0.07	-0.107	-0.073	-0.108
f912-f913	0.451	0.453	0.455	0.453	f4-f918	-0.060	-0.001	-0.065	-0.012
f703-f704	0.432	0.447	0.433	0.443	f903-f921	-0.059	-0.129	-0.082	-0.074
f920-f921	0.378	0.397	0.375	0.392	f604-f906	-0.058	-0.041	-0.061	-0.024
f901-f902	0.365	0.375	0.353	0.370	f607-f713	-0.058	0.000	-0.055	-0.004
f602-f603	0.353	0.358	0.355	0.351	f903-f920	-0.057	0.000	-0.094	-0.036
f924-f925	0.350	0.355	0.354	0.359	f2-f906	-0.054	0.000	-0.063	-0.007
f601-f603	0.346	0.351	0.351	0.350	f3-f704	-0.049	-0.089	-0.099	-0.041
f914-f915	0.328	0.354	0.354	0.359	f707-f716	-0.047	-0.077	-0.073	-0.052
f714-f715	0.323	0.327	0.325	0.330	f603-f606	-0.045	-0.072	-0.057	-0.012
f601-f602	0.314	0.315	0.313	0.311	f703-f911	-0.026	-0.072	-0.055	-0.074
f908-f909	0.312	0.344	0.332	0.345	f2-f4	-0.047	-0.066	-0.018	-0.041
f709-f712	0.304	0.302	0.294	0.307	f704-f921	-0.032	-0.062	-0.067	-0.085
f705-f710	0.302	0.335	0.284	0.321	f1-f4	-0.054	-0.041	-0.096	-0.053
f911-f912	0.301	0.318	0.296	0.311	f706-f918	-0.039	0.000	-0.079	-0.004
f701-f702	0.288	0.295	0.299	0.294	f909-f921	-0.032	-0.004	-0.074	-0.030
f706-f707	0.297	0.310	0.281	0.312	f711-f917	-0.046	0.000	-0.074	-0.008

four Bss models using the spike-and-slab Lasso prior, Bss_4 has the best performance in terms of graphical structure determination based on r_{ij}.

The Bss_4 model was then applied to the real ordinal data of Survey Data No. 2, and the result was compared with the other models investigated in Chapter 2. Since there is a high similarity between the structure of simulation Scenario 3 and the real ordinal Survey Data No. 2, and Bss_4 has good performance in the simulation study, the real data analysis results of Survey Data No. 2 using Bss_4 is highly reliable. We thus obtained a convincing graphical structure for Survey Data No. 2.

Table 3.6 List of questions description for Survey Data No. 2 (Part 1).

f1	How do you think of your current living condition?
f2	Are you satisfied with your current life?
f3	How is your current school life compare with what you imagined?
f4	How was your academic performance in the past semester?
f601	My future is largely in my own hands.
f602	I can do what I decide to do.
f603	My future is created by myself.
f604	I am confident about the future.
f605	Sometimes I feel like I have nothing to look forward to in my future.
f606	I muddle alone.
f607	It is useless to worry about the future because it is what it is.
f701	Someone will offer to help me when I am in trouble.
f702	I can share my joys and sorrows with others.
f703	My family is trying to help me.
f704	I can get the care and support I need from my family.
f705	There is someone who can comfort me.
f706	My friends are trying to help me.
f707	When things go bad, I can count on my friends.
f709	I have friends with whom I can share my joys and sorrows.
f710	There is someone who cares about my feelings.
f711	My relatives are willing to help me make decisions.
f712	I can discuss my problems with my friends.
f713	My teachers are trying to help me.
f714	When things go bad, I can count on my teachers.
f715	I can share my joys and sorrows with my teachers.
f716	I can discuss the problems I have with my teachers.
f901	I was being misunderstood.
f902	I was being discriminated against.
f903	My exams failed or were not as good as expected.
f906	I did not like my major.
f908	I was away from my family for a long time and unable to reunite with them.
f909	I had a heavy study load.
f910	I had a tense relationship with my teachers.
f911	I was seriously ill.
f912	My family members or friends are seriously ill.
f913	My family members or friends pass away.
f914	I had something stolen or lost.
f915	I lost face in public.
f916	My family was financially struggling.

Table 3.7 List of questions description for Survey Data No. 2 (Part 2).

f917	There was a conflict within my family.
f918	I did not obtain scholarships as expected.
f920	I transferred schools or took a leave of absence.
f921	I got fined.
f924	I was scolded by my parents.
f925	My family put pressure on me to study.

Chapter 4

Bayesian Factor Analysis for Ordinal Data

4.1 Introduction

In this chapter, we analyze the ordinal data from a different perspective. In Chapters 2 and 3, we focus on exploring the dependence structure of the ordinal variables. We have found some question pairs with relatively large partial correlations in the ordinal survey data sets. This motivates us to investigate whether we can remove some ordinal questions to make the survey more concise. Dimension reduction can be achieved by factor analysis. In this chapter, we apply the Bayesian factor analysis model to the Trauma Symptom Checklist for Children (TSCC) data and explore its dimension reduction performance.

Factor analysis was first introduced by Spearman (1904) on general intelligence and the one-factor model based on p variables and n objects can be written as

$$y_{ij} = \mu_j + \beta_j g_i + \varepsilon_{ij} \tag{4.1}$$

for $i = 1, ..., n,\ j = 1, ..., p$, where y_{ij} is the observed data, μ_j is the mean of variable j, g_i is the factor for object i, β_j is the loading of variable j onto the factor g, and ε_{ij} is the random error term for object i and variable j.

Three decades later, multiple-factor analysis models were introduced. Thurstone (1935, 1947) and Lawley (1940, 1942) proposed estimation methods via the centroid method and maximum likelihood, respectively.

Factor analysis models have attracted much attention since they performed exploratory analyses of the latent linear structure in high-dimensional data (Bernardo

et al. 2003; Carvalho et al. 2008; Engelhardt and Stephens 2010). With a latent factor model, high-dimensional data $\boldsymbol{y}_i \in \mathbb{R}^{p\times 1}$ with p variables can be represented by a low-dimensional latent factor $\boldsymbol{x}_i \in \mathbb{R}^{k\times 1}$ for $i = 1, ..., n$. A sample in the low-dimensional space is linearly projected to the original high-dimensional space via a loading matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{p\times k}$ with Gaussian noise $\boldsymbol{\varepsilon}_i \in \mathbb{R}^{p\times 1}$:

$$\boldsymbol{y}_i = \boldsymbol{\Lambda}\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i \tag{4.2}$$

for $i = 1, ..., n$. It is generally assumed that $\boldsymbol{x}_i$ follows a $\mathbf{N}_k(\mathbf{0}; \mathbf{I}_k)$ distribution, where $\boldsymbol{I}_k$ is the identity matrix of dimension k, and $\boldsymbol{\varepsilon}_i \sim \mathbf{N}_p(\mathbf{0}; \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a $p \times p$ diagonal covariance matrix with σ_j^2 for $j = 1, ..., p$ on the diagonal. The number of latent factors k is generally much smaller than the number of variables p and the sample size n. The covariance matrix of $\boldsymbol{y}_i$, $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times p}$, is estimated via low-rank estimation and has the following format:

$$\boldsymbol{\Gamma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Sigma} = \sum_{h=1}^{k} \boldsymbol{\lambda}_{\cdot h}\boldsymbol{\lambda}_{\cdot h}^T + \boldsymbol{\Sigma} \tag{4.3}$$

where $\boldsymbol{\lambda}_{\cdot h}$ is the h^{th} column of $\boldsymbol{\Lambda}$. This factorization suggests that each of the k factors contributes to the covariance of the sample through its corresponding loading $\boldsymbol{\lambda}_{\cdot h}$.

The loading matrix $\boldsymbol{\Lambda}$ is of significant importance in a factor model. In high-dimensional data applications where the sample size n is much smaller than the variable dimension p (Bernardo et al. 2003), regularization on the loading matrix is crucial because the optimization problem is under-constrained with $n \ll p$ and has many equivalent solutions that optimize the data likelihood. The statistics literature employs priors or penalties to regularize the elements of the loading matrix and induce sparsity. Elementwise sparsity corresponds to dimension reduction, where a latent factor contributes to variation in only a subset of the observed variables, generating interpretable results (Bernardo et al. 2003; Carvalho et al. 2008; Knowles and Ghahramani 2011).

Elementwise sparsity has been imposed in latent factor models through regularization via l_1 type penalties (Zou, Hastie, and Tibshirani 2006; Witten, Tibshirani, and Hastie 2009; Salzmann et al. 2010). In the last ten years, Bayesian shrinkage methods with sparsity-inducing priors have been introduced for latent factor models (Carvalho et al. 2008; Archambeau and Bach 2008; Virtanen et al. 2011; Bhattacharya and D. B. Dunson 2011; Klami, Virtanen, and Kaski 2013). The spike-and-slab prior (Mitchell and Beauchamp 1988) and the classic two-groups Bayesian sparsity-inducing prior have been used for sparse Bayesian latent factor models (Carvalho et al. 2008). More sophisticated structured regularization approaches have been studied in classical statistics (Zou and Hastie 2005; Kowalski and Torrésani 2009; Huang, Zhang, and Metaxas 2009; Jenatton, Audibert, and Bach 2011).

Zhao et al. (2016) developed a group factor analysis (GFA) model using Bayesian shrinkage with a hierarchical structure that encourages both element-wise and column-wise sparsity. The structured sparsity in their model is achieved with the multi-scale application of the three-parameter beta prior ($\mathcal{TPB}$), a hierarchical sparsity-inducing prior that has a computationally tractable representation (Armagan, Clyde, and D. Dunson 2011; Gao, Brown, and Engelhardt 2013). With the global-factor-local shrinkage, the model of Zhao et al. (2016) has the following features: i) The model globally shrinks the loading matrix by eliminating factors that lack sufficient support in the data; ii) the model shrinks the loading columns, effectively separating latent spaces from arbitrary subsets of observations; iii) permitting factor loadings to have either a sparse or non-sparse prior at the element-wise level, the model combines dimension reduction with interpretability.

In this chapter, we follow the idea of Zhao et al. (2016) and apply the global-factor-local shrinkage to our ordinal data analysis. Section 4.2 details the model and its associated Gibbs sampling algorithm. Section 4.3 applies this Bayesian factor

analysis model to our real ordinal survey data and discusses the results in Section 4.4.

4.2 Bayesian Factor Analysis Model

4.2.1 Latent factor models on ordinal data

Similar to that in Chapters 2 and 3, for ordinal variables $\boldsymbol{Y_V}$ with joint distribution F, we adopt the Gaussian copula so that the Bayesian factor model can be used on the latent normal variables. The full model is presented below:

$$\mathbf{F}_j(y_{ji}) = \Phi(\ z_{ji});\ j = 1, ...p;\ i = 1, ..., n; \tag{4.4}$$

$$\boldsymbol{z}_i = \boldsymbol{\Lambda}\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i;\ \boldsymbol{\Lambda} = \{\lambda_{jh}\}, j = 1, ..., p;\ h = 1, ..., k; \tag{4.5}$$

$$\boldsymbol{x}_i \sim \mathbf{N}_k(\mathbf{0}, \mathbf{I}_k); \tag{4.6}$$

$$\boldsymbol{\varepsilon}_i \sim \mathbf{N}_p(\mathbf{0}, \boldsymbol{\Sigma}). \tag{4.7}$$

where y_{ji} is the ordinal variable j on object i and z_{ji} is the corresponding continuous variable which is used in the Bayesian factor analysis model in vector format of $\boldsymbol{z}_i$. In the Bayesian factor analysis model, $\boldsymbol{\Lambda}$ is the $p \times k$ loading matrix with entries λ_{jh}, $\boldsymbol{x}_i$ is the latent factor assumed to follow a $\mathbf{N}_k(\mathbf{0}; \mathbf{I}_k)$ distribution with $\mathbf{I}_k$ as the identity matrix of dimension k, and $\boldsymbol{\varepsilon}_i \sim \mathbf{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ is the Gaussian noise, where $\boldsymbol{\Sigma}$ is the covariance matrix with σ_j^2, for $j = 1, ..., p$ on the diagonal.

4.2.2 Bayesian sparsity-inducing priors and three-parameter beta prior

Due to their flexible and interpretable solutions, Bayesian shrinkage priors are extensively utilized in latent factor models (Bernardo et al. 2003; Carvalho et al. 2008; Polson and Scott 2010; Knowles and Ghahramani 2011; Bhattacharya and D. B. Dunson 2011). Chapters 2 and 3 use graphical Lasso prior and spike-and-slab Lasso prior on the precision matrix to induce sparsity. This chapter utilizes the three-parameter beta before loading matrix entries λ_{jh}. The three-parameter beta ($\mathcal{TPB}$) distribution

for a random variable $\varphi \in (0,1)$ has the following density (Armagan, Clyde, and D. Dunson 2011):

$$f(\varphi; a, b, \nu) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \nu^b \varphi^{b-1} (1-\varphi)^{a-1} \{1 + (\nu - 1)\varphi\}^{-(a+b)}, \tag{4.8}$$

where $a,\ b,\ \nu > 0$. When $0 < \mathrm{a} < 1$ and $0 < \mathrm{b} < 1$, the distribution is bimodal, with modes at 0 and 1. The variance parameter ν provides the distribution with freedom: with fixed a and b, smaller values of ν enable greater probability on $\varphi = 1$, while larger values of ν move the probability mass towards $\varphi = 0$ (Armagan, Clyde, and D. Dunson 2011). In a special case, where $\nu = 1$, this distribution is simplified to the beta distribution $Beta(b, a)$.

Suppose λ is the parameter to which we apply sparsity-inducing regularization. The following $\mathcal{TPB}$ normal scale mixture distribution, $\mathcal{TPBN}$, is assigned to λ:

$$\lambda \sim \mathrm{N}(0, \frac{1}{\varphi} - 1),\ \text{with } \varphi \sim \mathcal{TPB}(a, b, \nu), \tag{4.9}$$

where the shrinkage parameter φ adheres to a $\mathcal{TPB}$ distribution. When $a = b = 1/2$ and $\nu = 1$, this prior becomes the horseshoe prior (Carvalho, Polson, and Scott 2010; Armagan, Clyde, and D. Dunson 2011; Gao, Brown, and Engelhardt 2013). The bimodal nature of φ creates two distinct shrinkage behaviors: the mode near one drives $\frac{1}{\varphi} - 1$ towards zero and results in strong shrinkage on λ; the mode near zero makes $\frac{1}{\varphi} - 1$ large and generates a diffuse prior on λ. Further decreasing the variance parameter ν yields stronger shrinkage (Armagan, Clyde, and D. Dunson 2011; Gao, Brown, and Engelhardt 2013). If we define $\theta = \frac{1}{\varphi} - 1$, then this mixture has the following hierarchical representation:

$$\lambda \sim \mathrm{N}(0, \theta),\ \text{with } \theta \sim \mathrm{GA}(a, \delta),\ \delta \sim \mathrm{GA}(b, \nu), \tag{4.10}$$

where θ is the variance of the Normal distribution which follows a *Gamma* distribution with shape parameter a and rate parameter δ, and δ further follows a *Gamma* distribution with shape parameter b and rate parameter ν.

4.2.3 Global-factor-local shrinkage

The $\mathcal{TPB}$ prior has been used in the Bayesian factor analysis models due to their flexible representation. Zhao et al. (2016) extended the $\mathcal{TPB}$ prior to three levels of regularization on the loading matrix (Gao, Brown, and Engelhardt 2013):

$$\lambda_{jh} \sim \mathrm{N}(0, \frac{1}{\varphi_{jh}} - 1), \tag{4.11}$$

$$\varphi_{jh} \sim \mathcal{TPB}(a, b, \frac{1}{\zeta_h} - 1), \ \text{(Local)} \tag{4.12}$$

$$\zeta_h \sim \mathcal{TPB}(c, d, \frac{1}{\varrho} - 1), \quad \text{(Factor-specific)} \tag{4.13}$$

$$\varrho \sim \mathcal{TPB}(e, f, \nu), \qquad \text{(Global)}. \tag{4.14}$$

A $\mathcal{TPB}$ distribution is employed at all three levels to induce sparsity through its estimated variance parameter. The global shrinkage parameter, ϱ, imposes significant shrinkage across the k columns of the loading matrix and adjusts the support of column-specific parameter, ζ_h, $h \in \{1, ..., k\}$, towards either zero or one. This induces sufficient shrinkage across the loading columns to retrieve the number of factors supported by the observed data. When ζ_h approaches one, all elements of column h are near zero, effectively eliminating the h^{th} component. Alternatively, as ζ_h approaches zero, the factor-specific regularization parameter adjusts the shrinkage applied to each element of the h^{th} loading column by borrowing strength across all the elements in the column, resulting in column-wise shrinkage estimation. Meanwhile, the local shrinkage parameter, φ_{jh}, enables element-wise sparsity in the loading matrix through a $\mathcal{TPBN}$. The three levels of global-factor-local shrinkage allow for simultaneous modeling of column-wise and element-wise shrinkage and provide the model with nonparametric behavior in the number of factors through model selection.

Let $\theta_{jh} = \frac{1}{\varphi_{jh}} - 1$, $\phi_h = \frac{1}{\zeta_h} - 1$, and $\eta = \frac{1}{\varrho} - 1$ for $j = 1, ..., p$ and $h = 1, ..., k$, this global-factor-local shrinkage prior can be equivalently written as (Armagan, Clyde,

and D. Dunson 2011; Gao, Brown, and Engelhardt 2013):

$$\lambda_{jh} \sim \text{N } (0, \theta_{jh}), \tag{4.15}$$

$$\text{Local } \begin{cases} \theta_{jh} & \sim \text{GA } (a, \delta_{jh}) \\ \delta_{jh} & \sim \text{GA } (b, \phi_h) \end{cases} \tag{4.16}$$

$$\text{Factor-specific } \begin{cases} \phi_h & \sim \text{GA } (c, \tau_h) \\ \tau_h & \sim \text{GA } (d, \eta) \end{cases} \tag{4.17}$$

$$\text{Global } \begin{cases} \eta & \sim \text{GA } (e, \gamma) \\ \gamma & \sim \text{GA } (f, \nu) \end{cases}. \tag{4.18}$$

Zhao et al. (2016) extended this prior to jointly modeling sparse and dense components by assigning a two-component mixture distribution to the local shrinkage parameter (Gao, Brown, and Engelhardt 2013):

$$\theta_{jh} \sim \pi \text{GA } (a, \delta_{jh}) + (1 - \pi)\delta_{\phi_h}(\cdot) \tag{4.19}$$

where $\delta_{\phi_h}(\cdot)$ is the Dirac delta function centered at ϕ_h. According to Zhao et al. (2016), a two-component mixture allows the prior on the loading to select between element-wise sparsity or column-wise sparsity. Element-wise sparsity is encouraged via the $\mathcal{TPBN}$ prior. Column-wise sparsity applies a joint regularization on each element of the column using a shared variance term, denoted by ζ_h. Modeling each λ_{jh} using a shared regularized variance term ζ_h has two possible behaviors: i) When ζ_h is close to 1, the entire column is strongly shrunk towards zero, effectively eliminating this factor. ii) When ζ_h is close to zero, all elements of the column have a shared Gaussian distribution, inducing only non-zero elements in that loading. Such factors with only non-zero elements are referred to as dense factors.

Indicator variables $\omega_h,\ h = 1, ..., k$, are introduced to indicate which mixture component each θ_h is generated from, where $\omega_h = 1$ means $\theta_h \sim \text{GA } (a, \delta_{jh})$ and

$\omega_h = 0$ means $\theta_h \sim \delta_{jh}(\cdot)$ (Zhao et al. 2016). Therefore, when $\omega_h = 1$, a component represents a sparse factor, and when $\omega_h = 0$, it either represents a dense factor or is eliminated. We let $\boldsymbol{\omega} = [\omega_1, ..., \omega_k]$ and put a Bernoulli distribution with parameter π on ω_h. We further let π have a flat beta distribution $Beta(1, 1)$. Using this setting, we can determine the posterior probability of each factor h generated from each mixture component type through the parameter ω_h.

In this chapter, the hyper-parameters of the global-factor-local $\mathcal{TPB}$ prior were set to $a = b = c = d = e = f = 0.5$, which results in a horseshoe prior at all three levels of the hierarchy. The hyper-parameters for the error variances, a_σ and b_σ, are set to 1 and 0.3, respectively, to allow relatively wide support of variances (Bhattacharya and D. B. Dunson 2011).

4.2.4 Full Model

Based on the discussion above, we summarize the full model as follows:

$$
\begin{aligned}
\mathbf{F}_j(y_{ji}) &= \Phi(\ z_{ji});\ j = 1, ...p;\ i = 1, ..., n; \\
\boldsymbol{z}_i &= \boldsymbol{\Lambda}\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i;\ \boldsymbol{\Lambda} = \{\lambda_{jh}\}, j = 1, ..., p;\ h = 1, ..., k; \\
\boldsymbol{x}_i &\sim \mathbf{N}_k(\mathbf{0}, \mathbf{I}_k); \\
\boldsymbol{\varepsilon}_i &\sim \mathbf{N}_p(\mathbf{0}, \boldsymbol{\Sigma}); \\
\lambda_{jh} &\sim \mathrm{N}(0, \theta_{jh}); \\
\text{Local} &\begin{cases} \theta_{jh} & \sim \pi\mathrm{GA}\ (a, \delta_{jh}) + (1-\pi)\delta_{\phi_h}(\cdot) \\ \delta_{jh} & \sim \mathrm{GA}\ (b, \phi_h) \end{cases} \\
\text{Factor-specific} &\begin{cases} \phi_h & \sim \mathrm{GA}\ (c, \tau_h) \\ \tau_h & \sim \mathrm{GA}\ (d, \eta) \end{cases} \\
\text{Global} &\begin{cases} \eta & \sim \mathrm{GA}\ (e, \gamma) \\ \gamma & \sim \mathrm{GA}\ (f, \nu) \end{cases}, a = b = c = d = e = f = 0.5 \\
\pi &\sim Beta\ (1, 1), \\
\sigma_j^{-2} &\sim \mathrm{GA}\ (a_\sigma, b_\sigma),\ a_\sigma = 0.1,\ b_\sigma = 0.3
\end{aligned}
$$

4.3 Gibbs Sampling Algorithm

The Gibbs sampler is summarized as follows:

Given the current value $\mathbf{Z}$, $\mathbf{X}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Theta}$, $\boldsymbol{\Delta}$, $\boldsymbol{\omega}$, $\boldsymbol{\phi}$, $\boldsymbol{\tau}$, η, γ, and π,

1. Sample the latent variables Z_{ji}, $j = 1, ..., p$; $i = 1, ...n$.

$\mathrm{Z}_{ji} \sim \mathbf{TN}\Big(\boldsymbol{\Gamma}_{j,-j}(\boldsymbol{\Gamma}_{-j,-j})^{-1}\mathbf{Z}_{-j,i}, \boldsymbol{\Gamma}_{j,j} - \boldsymbol{\Gamma}_{j,-j}(\boldsymbol{\Gamma}_{-j,-j})^{-1}\boldsymbol{\Gamma}_{-j,j}\Big)$, which is a truncated normal distribution with lower bound as $Z_{ji}(lb) = \max\Big\{z_{jh} : y_{jh} < y_{ji}\Big\}$, and upper bound as $Z_{ji}(ub) = \min\Big\{z_{jh} : y_{jh} > y_{ji}\Big\}$, for $i, h = 1, ..., n$ within each row $j = 1, ..., p$, where $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Sigma}$.

2. For $i = 1, ..., n$, sample latent factors $\mathbf{X}_i$.

$\mathbf{X}_i \sim \mathbf{N}\Big((\mathbf{\Lambda}^T\mathbf{\Sigma}^{-1}\mathbf{\Lambda} + \mathbf{I})^{-1}\mathbf{\Lambda}^T\mathbf{\Sigma}^{-1}\mathbf{Z}_i, (\mathbf{\Lambda}^T\mathbf{\Sigma}^{-1}\mathbf{\Lambda} + \mathbf{I})^{-1}\Big)$.

3. Sample loading matrix $\mathbf{\Lambda}$.

For $j = 1, ..., p$, $\boldsymbol{\lambda}_{j\cdot}^T \sim \mathbf{N}\Big((\sigma_j^{-2}\mathbf{X}\mathbf{X}^T + \mathbf{D}_j^{-1})^{-1}\sigma_j^{-2}\mathbf{X}\mathbf{Z}_{j\cdot}^T, (\sigma_j^{-2}\mathbf{X}\mathbf{X}^T + \mathbf{D}_j^{-1})^{-1}\Big)$,

where $\mathbf{D}_j^{-1}$= diag $\Big(\theta_{j1}^{\mathbf{I}(\omega_1=1)} \cdot \phi_1^{\mathbf{I}(\omega_1=0)}), ..., \ \theta_{jk}^{\mathbf{I}(\omega_k=1)} \cdot \phi_k^{\mathbf{I}(\omega_k=0)}\Big)$.

4. Sample $\mathbf{\Theta}$, $\mathbf{\Delta}$, and $\boldsymbol{\phi}$.

For $h = 1, ..., k$,

- If $\boldsymbol{\omega}_h = 1$,

4.1 For $j = 1, ..., p$, $\theta_{jh} \sim \mathcal{GIG}$ $(lambda = a - \frac{1}{2},\ psi = 2\delta_{jh},\ chi = \lambda_{jh}^2)$;

4.2 For $j = 1, ..., p$, $\delta_{jh} \sim$ GA $(a + b,\ \theta_{jh} + \phi_h)$;

4.3 $\phi_h \sim$ GA $(bp + c,\ \sum_{j=1}^p \delta_{jh} + \tau_h)$.

- If $\boldsymbol{\omega}_h = 0$,

4.1 For $j = 1, ..., p$, $\theta_{jh}^{\text{iter}} = \theta_{jh}^{\text{iter}-1}$;

4.2 For $j = 1, ..., p$, $\delta_{jh}^{\text{iter}} = \delta_{jh}^{\text{iter}-1}$;

4.3 $\phi_h \sim \mathcal{GIG}$ $(lambda = c - \frac{p}{2},\ psi = 2\tau_h,\ chi = \sum_{j=1}^p \lambda_{jh}^2)$.

5. For $h = 1, ..., k$, $\tau_h \sim$ GA $(c + d,\ \phi_h + \eta)$.

6. $\eta \sim$ GA $(dk + e,\ \sum_{h=1}^k \tau_h + \gamma)$

7. $\gamma \sim$ GA $(e + f,\ \eta + \nu)$.

8. $\pi \sim$ Beta $(1 + \sum_{h=1}^{k} \omega_h,\ 1 + k - \sum_{h=1}^{k} \omega_h)$.

9. For $h = 1, ..., k$, $\omega_h \sim$ Bernoulli $(p = \frac{A}{A+B})$, with $A = \pi \prod_{j=1}^{p} \{N(\lambda_{jh}; 0, \theta_{jh}) \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\theta_{jh}^{a-1} \cdot \phi_h^b}{(\theta_{jh}+\phi_h)^{a+b}}\}$, and $B = (1-\pi) \prod_{j=1}^{p} \{N(\lambda_{jh}; 0, \phi_h)\}$.

10. Sample $\mathbf{\Sigma}$.

For $j = 1, ..., p$, $\sigma_j^{-2} \sim$ GA $(\frac{n}{2} + a_\sigma,\ \frac{1}{2}(\mathbf{Z}_{j\cdot} - \boldsymbol{\lambda}_{j\cdot}\mathbf{X})(\mathbf{Z}_{j\cdot} - \boldsymbol{\lambda}_{j\cdot}\mathbf{X})^T + b_\sigma)$, with $a_\sigma = 0.1$, and $b_\sigma = 0.3$.

4.4 Real Data Analysis

4.4.1 Dataset Description

Question No. 14 of the LSS dataset in Chapter 2 is analyzed in this section. As was described in Section 2.4.1, there are 54 sub-questions within Question No. 14, and these sub-questions are the Trauma Symptom Checklist for Children (TSCC), which was initially designed as a self-report measure of posttraumatic distress and related psychological symptomatology among children 8-16 years of age (Briere 1996). In this survey, these questions are translated into Chinese and asked to check whether the respondents have had any of these behaviors within the past six months. In later analysis, this dataset with 2369 observations and 54 variables (sub-questions) is called the TSCC dataset.

4.4.2 Analysis Methodology

The Bayesian factor analysis model is applied to the TSCC dataset to explore the dimension reduction performance among these 54 variables. The 54 sub-questions are entered into the model, and all are ordinal variables with four ordinal levels of 1, 2, 3, and 4.

The Gibbs sampling algorithm is utilized for a total of 10,000 iterations, with the first 2,000 iterations as the burn-in period. The hyper-parameters of $a,\ b,\ c,\ d,\ e,$ and f for the $\mathcal{TPB}$ distribution are set as 0.5. The hyperparameters of the prior distribution of σ_j^{-2} are set as $a_\sigma = 0.1$, and $b_\sigma = 0.3$. Regarding k, the number of factors and different values, such as 4, 6, 8, and 10, are tried. For each of them, the posterior mean of the loading matrix $\mathbf{\Lambda}$, the error term covariance matrix $\mathbf{\Sigma}$, and thus the latent covariance matrix $\mathbf{\Gamma}$ ($\mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Sigma}$) are estimated.

4.4.3 Results

The posterior mean of the loading matrix $\mathbf{\Lambda}$ is estimated using the 8,000 iterations after 2,000 burn-in iterations. Figure 4.1 shows the plot of the estimated posterior mean of $\mathbf{\Lambda}$ for $k = 4,\ 6,\ 8,$ and 10. For all k values, the magnitudes of most of the loading matrix entries are within [-2.5, 2.5]. For $k = 6$, the loading matrix entries for sub-questions No. 10, 19, 36, and 43 are all zero. Similarly, for $k = 10$, the loading matrix entries for sub-questions No. 24 and 31 are all zero. This implies that these sub-questions are not related to any of the factors.

The estimated posterior mean of the latent covariance matrix, which is labeled as $\mathbf{\Gamma}$, is also calculated for all k values via the equation $\mathbf{\Gamma} = \mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Sigma}$. Figure 4.2 shows the plot of the estimated posterior mean of $\mathbf{\Gamma}$ for $k = 4,\ 6,\ 8,$ and 10, respectively. These estimates are compared with that of the Bayesian Gaussian copula graphical models (GCGMs) in Figure 4.3 and the Spearman estimation in Figure 4.4. Based on the results, the Bayesian factor analysis result of $k = 8$ is similar to that of the GCGMs, but with a slightly smaller magnitude (the color is lighter when the magnitude is smaller). In addition, the Bayesian factor analysis result of $k = 8$ is also similar to that of the Spearman estimate but with a slightly larger magnitude.

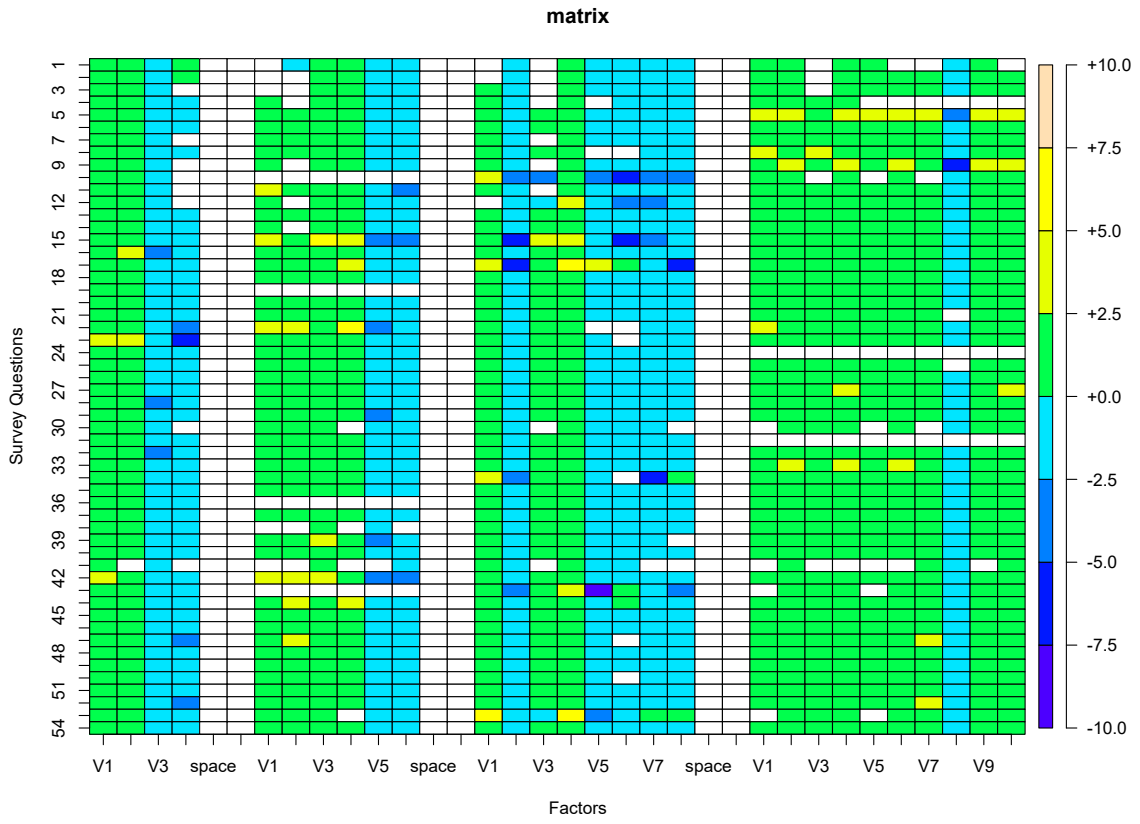

Figure 4.1 The posterior mean estimate of the loading matrix $\mathbf{\Lambda}$ for the TSCC data using the Bayesian factor analysis model with $k = 4,\ 6,\ 8$, and 10 (from left to the right)

4.5 Discussion

This chapter applies the Bayesian factor analysis model to the Trauma Symptom Checklist for Children (TSCC) data. To represent this observed 54×2369 ordinal data with the relatively low-dimensional latent factors, we select the factor number of $k = 4$, 6, 8, and 10 in the model. We cannot obtain a sparse loading matrix based on the results, which may be related to the choice of k. Optimizing the value of k may help gain a desirable sparse loading matrix. In estimating the covariance matrix $\mathbf{\Gamma}$, we compare the results with that in Chapter 2, where we used Bayesian Gaussian copula graphical models to do the estimation. We have found that the Bayesian factor

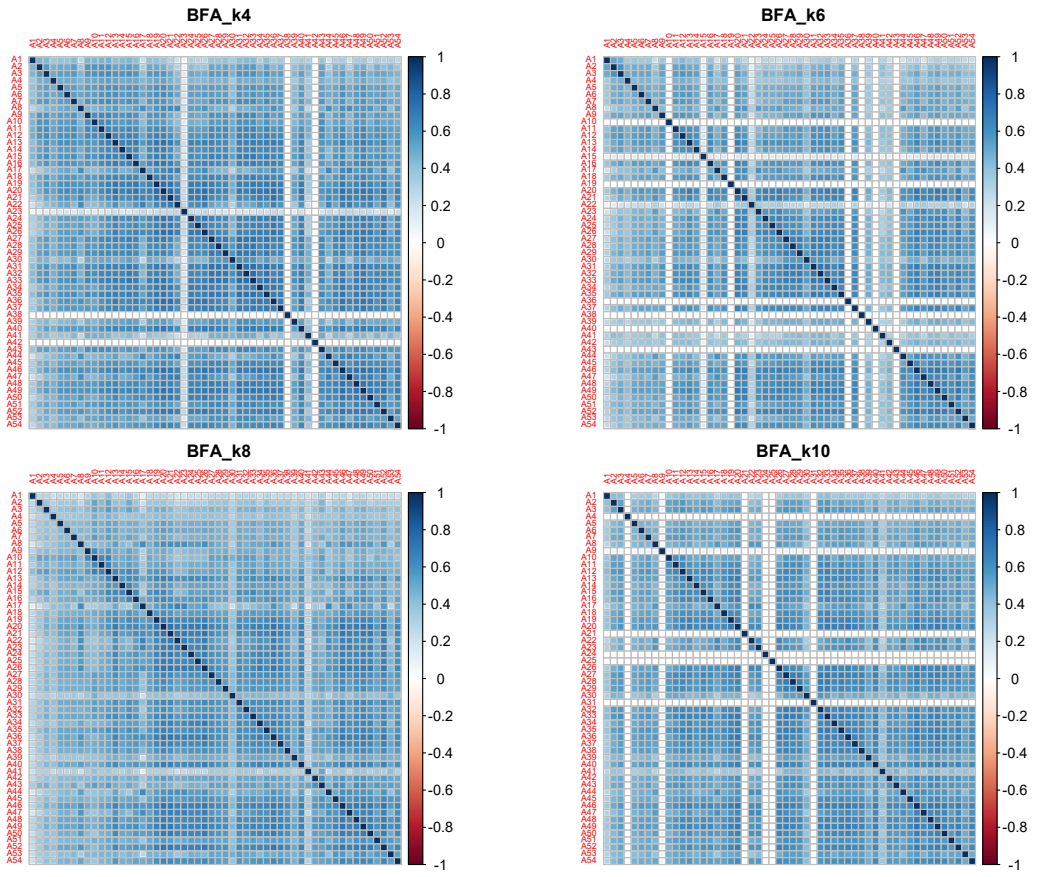

Figure 4.2 The posterior mean estimate of the covariance matrix $\mathbf{\Gamma}$ for the TSCC data using the Bayesian factor analysis model with $k = 4,\ 6,\ 8$, and 10.

analysis model with $k = 8$ yielded similar results to Chapter 2 but with a slightly smaller magnitude on the estimation. For future studies, one direction is to conduct a simulation study to investigate the optimization of k. To analyze the same dataset using different models of graphical models and factor analysis models, we also need to investigate model comparison via the simulation study.

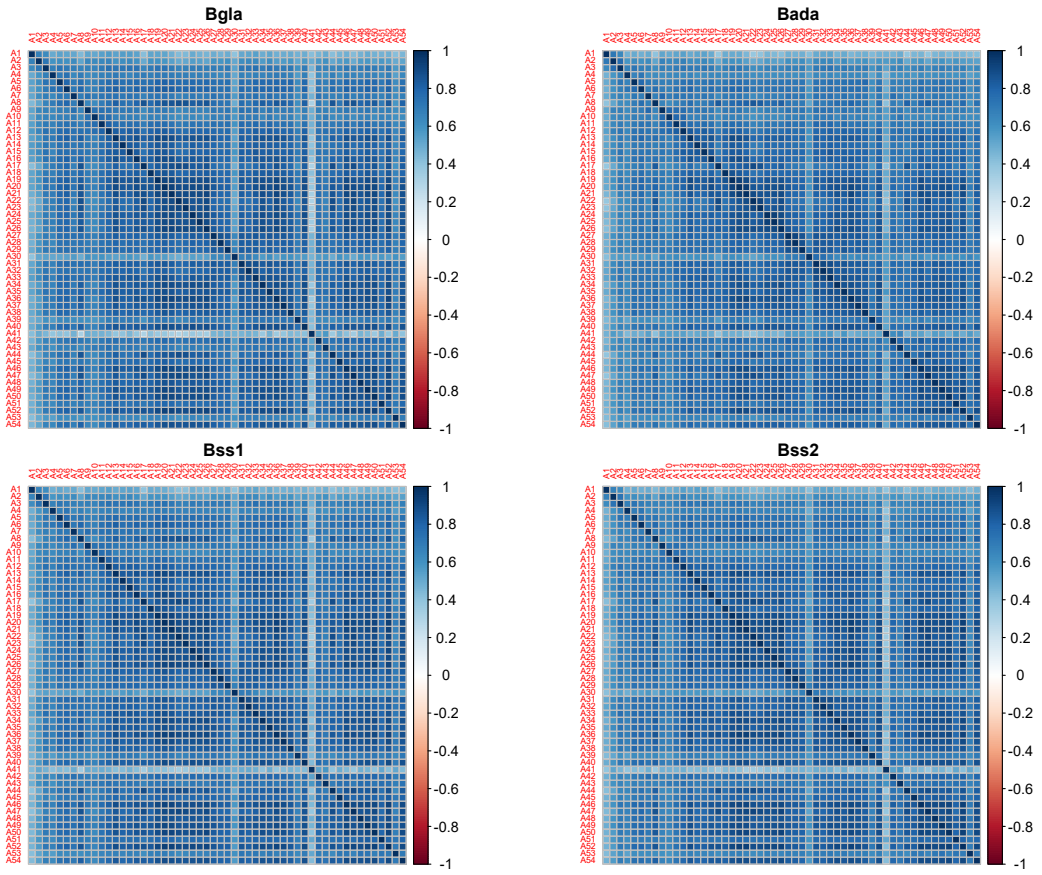

Figure 4.3 The posterior mean estimate of the covariance matrix $\mathbf{\Gamma}$ for the TSCC data using the Bayesian Gaussian copula graphical models.

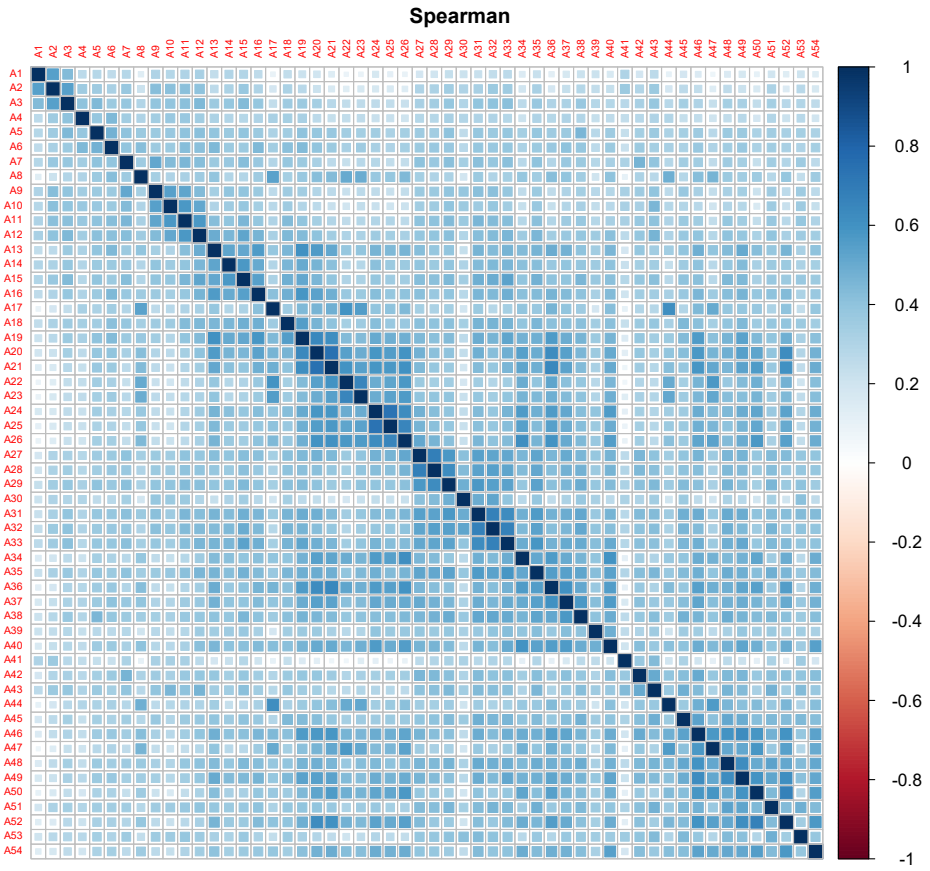

Figure 4.4 The Spearman estimate of the covariance matrix $\mathbf{\Gamma}$ for the TSCC data.

Chapter 5

Conclusion

In this dissertation, ordinal data are analyzed utilizing two different methods of Gaussian copula graphical models and factor analysis models under the Bayesian framework. Chapter 2 and Chapter 3 investigate the Gaussian copula graphical models with different prior distributions on the precision matrix of the ordinal data. Chapter 4 utilizes the factor analysis model to investigate the variable structure based on latent factors. All three chapters share a common approach in their implementation by utilizing MCMC sampling methods, and each algorithm is straightforward to implement and demonstrates effectiveness in the respective situations. In addition, they all apply a Gaussian copula on the ordinal data so that the Gaussian graphical and factor analysis models, commonly used on continuous variables, can be utilized for the ordinal data analysis.

In Chapter 2, the dependence structure of the ordinal variables is analyzed using the Gaussian copula graphical models with three different priors of graphical Lasso ($Bgla$), adaptive graphical Lasso ($Bada$) and spike-and-slab Lasso (Bss) on the precision matrix of these ordinal data. The performance of these different priors in terms of both matrix estimation of the covariance matrix, partial correlation matrix, and precision matrix and graphical structure determination are compared via simulation studies. Based on the results of simulation studies, the Bss_2 model with hyperparameter $\nu_0 = 10^{-3}$ works comparably well as the $Bada$ model. All the models are applied to the real ordinal Survey Data No. 1 and obtain the dependence structure of these ordinal survey questions. From the results of this real ordinal data analysis,

we have found that some question pairs have a relatively large partial correlation. For the future direction, we can explore some methods to do dimension reduction to make the survey more concise with fewer questions included.

The Bss model in Chapter 2 manually chooses the value of the hyper-parameter ν_0 for the spike-and-slab Lasso prior. The performance of the Bss model, especially on the structure learning, heavily relies on the choice of the value of ν_0. To overcome this problem, Chapter 3 proposes the adaptive spike-and-slab Lasso prior. The hyper-parameter ν_0 is tied with the prior inclusion probability η, and a prior is assigned on η. This enables automatically updating ν_0 when η is updated and essentially allows the data to decide the value of ν_0 instead of fixing its value. It is also worth mentioning that in this chapter, the simulation setting is carefully designed by making it resemble the scenario of the real ordinal Survey Data No. 2. In this way, we better assess the performance of the models using the simulation study and thus give better guidance in selecting the cutoff values for the structure learning for the real data analysis.

Chapter 4 analyzes the TSCC ordinal survey data using the Bayesian factor analysis model in an attempt to do dimension reduction. A novel sparsity-inducing global-factor-local shrinkage is applied to the loading matrix to achieve a sparse variable structure. Different values of k, the number of latent factors, are explored. The covariance matrix $\mathbf{\Gamma}$ is also estimated with each k value, and the results are compared to that using Bayesian Gaussian copula graphical models. For future research, one direction is to conduct a simulation study to investigate the optimization of k. To analyze the same dataset using different models of graphical models and factor analysis models, we also need to investigate model comparison via the simulation study.

Bibliography

Albert, James H and Siddhartha Chib (1993). "Bayesian analysis of binary and polychotomous response data". In: *Journal of the American statistical Association* 88.422, pp. 669–679.

Andrews, David F and Colin L Mallows (1974). "Scale mixtures of normal distributions". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36.1, pp. 99–102.

Archambeau, Cédric and Francis Bach (2008). "Sparse probabilistic projections". In: *Advances in neural information processing systems* 21.

Armagan, Artin, Merlise Clyde, and David Dunson (2011). "Generalized beta mixtures of Gaussians". In: *Advances in neural information processing systems* 24.

Banerjee, Sayantan and Subhashis Ghosal (2015). "Bayesian structure learning in graphical models". In: *Journal of Multivariate Analysis* 136, pp. 147–162.

Bernardo, JM et al. (2003). "Bayesian factor regression models in the "large p, small n" paradigm". In: *Bayesian statistics* 7, pp. 733–742.

Bhattacharya, Anirban and David B Dunson (2011). "Sparse Bayesian infinite factor models". In: *Biometrika*, pp. 291–306.

Bliss, Chester Ittner (1935). "The calculation of the dosage-mortality curve". In: *Annals of Applied Biology* 22.1, pp. 134–167.

Breiman, Leo (1996). "Heuristics of instability and stabilization in model selection". In: *The annals of statistics* 24.6, pp. 2350–2383.

Briere, John (1996). "Trauma symptom checklist for children". In: *Odessa, FL: Psychological Assessment Resources*, pp. 00253–8.

Carifio, James and Rocco J Perla (2007). "Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes". In: *Journal of social sciences* 3.3, pp. 106–116.

Carvalho, Carlos M, Nicholas G Polson, and James G Scott (2010). "The horseshoe estimator for sparse signals". In: *Biometrika* 97.2, pp. 465–480.

Carvalho, Carlos M and James G Scott (2009). "Objective Bayesian model selection in Gaussian graphical models". In: *Biometrika* 96.3, pp. 497–512.

Carvalho, Carlos M et al. (2008). "High-dimensional sparse factor modeling: applications in gene expression genomics". In: *Journal of the American Statistical Association* 103.484, pp. 1438–1456.

Casella, George and Edward I George (1992). "Explaining the Gibbs sampler". In: *The American Statistician* 46.3, pp. 167–174.

Chib, Siddhartha and Edward Greenberg (1996). "Markov chain Monte Carlo simulation methods in econometrics". In: *Econometric theory* 12.3, pp. 409–431.

Chib, Siddhartha and Edward Greenberg (1998). "Analysis of multivariate probit models". In: *Biometrika* 85.2, pp. 347–361.

Clason, Dennis L and Thomas J Dormody (1994). "Analyzing data measured by individual Likert-type items". In: *Journal of agricultural education* 35.4, p. 4.

Davier, Matthias von and Claus H Carstensen (2007). *Multivariate and mixture distribution Rasch models: Extensions and applications*. Springer.

Dobra, Adrian, Alex Lenkoski, and Abel Rodriguez (2011). "Bayesian inference for general Gaussian graphical models with application to multivariate lattice data". In: *Journal of the American Statistical Association* 106.496, pp. 1418–1433.

Drton, Mathias and Michael D Perlman (2004). "Model selection for Gaussian concentration graphs". In: *Biometrika* 91.3, pp. 591–602.

Edwards, David (2000). *Introduction to Graphical Modelling*. Springer Science & Business Media.

Engelhardt, Barbara E and Matthew Stephens (2010). "Analysis of population structure: a unifying framework and novel methods based on sparse factor analysis". In: *PLoS genetics* 6.9, e1001117.

Fan, Jianqing, Yang Feng, and Yichao Wu (2009). "Network exploration via the adaptive LASSO and SCAD penalties". In: *The annals of applied statistics* 3.2, p. 521.

Feldman, Maryann P and David B Audretsch (1999). "Innovation in cities:: Science-based diversity, specialization and localized competition". In: *European economic review* 43.2, pp. 409–429.

Gan, Lingrui, Naveen N Narisetty, and Feng Liang (2019). "Bayesian regularization for graphical models with unequal shrinkage". In: *Journal of the American Statistical Association* 114.527, pp. 1218–1231.

Gao, Chuan, Christopher D Brown, and Barbara E Engelhardt (2013). "A latent factor model with a mixture of sparse and dense factors to model gene expression data with confounding effects". In: *arXiv preprint arXiv:1310.4792*.

Gelfand, Alan E and Adrian FM Smith (1990). "Sampling-based approaches to calculating marginal densities". In: *Journal of the American statistical association* 85.410, pp. 398–409.

Gelman, Andrew (1993). "Iterative and non-iterative simulation algorithms". In: *Computing science and statistics*, pp. 433–433.

Geman, Stuart and Donald Geman (1984). "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images". In: *IEEE Transactions on pattern analysis and machine intelligence* 6, pp. 721–741.

George, Edward I and Robert E McCulloch (1993). "Variable selection via Gibbs sampling". In: *Journal of the American Statistical Association* 88.423, pp. 881–889.

Hastings, W Keith (1970). "Monte Carlo sampling methods using Markov chains and their applications". In.

Hoff, Peter D (2007). "Extending the rank likelihood for semiparametric copula estimation". In: *The Annals of Applied Statistics* 1.1, pp. 265–283.

Huang, Junzhou, Tong Zhang, and Dimitris Metaxas (2009). "Learning with structured sparsity". In: *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 417–424.

Hui, Michael K and John EG Bateson (1991). "Perceived control and the effects of crowding and consumer choice on the service experience". In: *Journal of consumer research* 18.2, pp. 174–184.

Ishwaran, Hemant and J Sunil Rao (2005). "Spike and slab variable selection: frequentist and Bayesian strategies". In: *The Annals of Statistics* 33.2, pp. 730–773.

Jamieson, Susan (2004). "Likert scales: How to (ab) use them?" In: *Medical education* 38.12, pp. 1217–1218.

Jenatton, Rodolphe, Jean-Yves Audibert, and Francis Bach (2011). "Structured variable selection with sparsity-inducing norms". In: *The Journal of Machine Learning Research* 12, pp. 2777–2824.

Klami, Arto, Seppo Virtanen, and Samuel Kaski (2013). "Bayesian Canonical correlation analysis." In: *Journal of Machine Learning Research* 14.4.

Knowles, David and Zoubin Ghahramani (2011). "Nonparametric Bayesian sparse factor models with application to gene expression modeling". In: *The Annals of Applied Statistics* 5.2B, pp. 1534–1552.

Kowalski, Matthieu and Bruno Torrésani (2009). "Structured sparsity: from mixed norms to structured shrinkage". In: *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*.

Likert, Rensis (1932). "A technique for the measurement of attitudes." In: *Archives of psychology*.

Marlin, Benjamin M and Kevin P Murphy (2009). "Sparse Gaussian graphical models with unknown block structure". In: *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 705–712.

McCullagh, P and JA Nelder (1989). "Binary data". In: *Generalized linear models*. Springer, pp. 98–148.

McCullagh, Peter (1980). "Regression models for ordinal data". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 42.2, pp. 109–127.

Meinshausen, Nicolai and Peter Bühlmann (2006). "High-dimensional graphs and variable selection with the lasso". In: *The annals of statistics* 34.3, pp. 1436–1462.

Metropolis, Nicholas et al. (1953). "Equation of state calculations by fast computing machines". In: *The journal of chemical physics* 21.6, pp. 1087–1092.

Mitchell, Toby J and John J Beauchamp (1988). "Bayesian variable selection in linear regression". In: *Journal of the american statistical association* 83.404, pp. 1023–1032.

Mohammadi, Abdolreza and Ernst C Wit (2015). "Bayesian structure learning in sparse Gaussian graphical models". In: *Bayesian Analysis* 10.1, pp. 109–138.

Müller, Peter (1991). "A generic approach to posterior integration and Gibbs sampling". In: *Tech-nical Report*, pp. 91–09.

Narisetty, Naveen Naidu and Xuming He (2014). "Bayesian variable selection with shrinking and diffusing priors". In: *The Annals of Statistics* 42.2, pp. 789–817.

Nelsen, Roger B (2007). *An introduction to copulas*. Springer Science & Business Media.

O'Connell, Ann A (2006). *Logistic regression models for ordinal response variables*. Vol. 146. sage.

Peterson, Bercedis and Frank E Harrell Jr (1990). "Partial proportional odds models for ordinal response variables". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 39.2, pp. 205–217.

Polson, Nicholas G and James G Scott (2010). "Shrink globally, act locally: Sparse Bayesian regularization and prediction". In: *Bayesian statistics* 9.501-538, p. 105.

Ročková, Veronika (2018). "Bayesian estimation of sparse signals with a continuous spike-and-slab prior". In: *The Annals of Statistics* 46.1, pp. 401–437.

Ročková, Veronika and Edward I George (2016). "Fast Bayesian factor analysis via automatic rotations to sparsity". In: *Journal of the American Statistical Association* 111.516, pp. 1608–1622.

Ročková, Veronika and Edward I George (2018). "The spike-and-slab lasso". In: *Journal of the American Statistical Association* 113.521, pp. 431–444.

Rothman, Adam J et al. (2008). "Sparse permutation invariant covariance estimation". In: *Electronic Journal of Statistics* 2, pp. 494–515.

Salzmann, Mathieu et al. (2010). "Factorized orthogonal latent spaces". In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, pp. 701–708.

Spearman, Charles (1904). "E. 1904."General Intelligence" Objectively Determined and Measured". In: *American Journal of Psychology* 15.2, pp. 201–293.

Spranca, Mark, Elisa Minsk, and Jonathan Baron (1991). "Omission and commission in judgment and choice". In: *Journal of experimental social psychology* 27.1, pp. 76–105.

Vickers, Andrew J (1999). "Comparison of an ordinal and a continuous outcome measure of muscle soreness". In: *International journal of technology assessment in health care* 15.4, pp. 709–716.

Virtanen, Seppo et al. (2011). "Bayesian group factor analysis". In: *arXiv preprint arXiv:1110.3204*.

Walker, Strother H and David B Duncan (1967). "Estimation of the probability of an event as a function of several independent variables". In: *Biometrika* 54.1-2, pp. 167–179.

Wang, Hao (2012). "Bayesian graphical lasso models and efficient posterior computation". In: *Bayesian Analysis* 7.4, pp. 867–886.

Wang, Hao and Sophia Zhengzi Li (2012). "Efficient Gaussian graphical model determination under G-Wishart prior distributions". In: *Electronic Journal of Statistics* 6, pp. 168–198.

West, Mike (1987). "On scale mixtures of normal distributions". In: *Biometrika* 74.3, pp. 646–648.

Witten, Daniela M, Robert Tibshirani, and Trevor Hastie (2009). "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis". In: *Biostatistics* 10.3, pp. 515–534.

Yuan, Ming and Yi Lin (2007). "Model selection and estimation in the Gaussian graphical model". In: *Biometrika* 94.1, pp. 19–35.

Zhao, Shiwen et al. (2016). "Bayesian group factor analysis with structured sparsity". In: *The Journal of Machine Learning Research*.

Zou, Hui and Trevor Hastie (2005). "Regularization and variable selection via the elastic net". In: *Journal of the royal statistical society: series B (statistical methodology)* 67.2, pp. 301–320.

Zou, Hui, Trevor Hastie, and Robert Tibshirani (2006). "Sparse principal component analysis". In: *Journal of computational and graphical statistics* 15.2, pp. 265–286.