

High-dimensional Linear Regression Problems via Graphical Models



Saeed Aldahmani

A Thesis Submitted for the Degree of
Doctor of Philosophy

Department of Mathematical Sciences

University of Essex

March 2017

Abstract

This thesis introduces a new method for solving the linear regression problem where the number of observations n is smaller than the number of variables (predictors) v . In contrast to existing methods such as ridge regression, Lasso and Lars, the proposed method uses the idea of graphical models and provides unbiased parameter estimates under certain conditions. In addition, the new method provides a detailed graphical conditional correlation structure for the predictors, whereby the real causal relationship between predictors can be identified.

Furthermore, the proposed method is extended to form a hybridisation with the idea of ridge regression to improve efficiency in terms of computation and model selection. In the extended method, less important variables are regularised by a ridge type penalty, and a search for models in the space is made for important covariates. This significantly reduces computational cost while giving unbiased estimates for the important variables as well as increasing the efficiency of model selection. Moreover, the extended method is used in dealing with the issue of portfolio selection within the Markowitz mean-variance framework, with $n < v$.

Various simulations and real data analyses were conducted for comparison between the two novel methods and the aforementioned existing methods. Our experiments indicate that the new methods outperform all the other methods when $n < v$.

Declaration

This research was carried out at the Department of Mathematical Sciences, University of Essex, United Kingdom. I declare that this thesis contains my original work, except where otherwise stated with reference to the literature. No part of this thesis has been submitted elsewhere for any other degree or qualification.

Publications

- Aldahmani, S. and Dai, H. (2015). Unbiased Estimation for Linear Regression When $n < v$. *International Journal of Statistics and Probability*, 4(3), p61.
- Aldahmani, S. Dai, H. and Zhang, Q., Hybrid Graphical Least Square Estimation and its application in Portfolio Selection, submitted.

Acknowledgements

First and foremost, I thank Almighty Allah for blessing my time and effort to complete this piece of research.

My sincere thanks and heartfelt gratitude go to my supervisor Dr. Hongsheng Dai for his benevolence and great support throughout my PhD study at the University of Essex. He has never hesitated to offer immense guidance and insightful feedback at all times. My thanks also go to my Supervisory Board Chair Dr. Aris Perperoglou for his valuable suggestions and recommendations and to all the academic and administrative staff at the Department of mathematical Sciences, University of Essex for their kindness and assistance.

I am extremely grateful to the United Arab Emirates University for funding this research and his highness Sheikh Mohammad Bin Zayed for his sincere efforts to highly educate and develop the country.

I am deeply indebted to my wife for sharing with me the happy and hard times during my study and to my two sons for being the jewels of my life and my best wishes go to my friends for their unforgettable company.

Last but not least, I dedicate this work to the memory of my mother and father, may their souls rest in Paradise.

Contents

Abstract	ii
Declaration	iii
Publications	iv
Acknowledgements	v
1 Introduction	1
1.1 Contributions	3
1.2 Thesis outline	4
2 Regression with $n < v$	6
2.1 Introduction	6
2.2 Ridge regression	7
2.2.1 Cross-validation for choosing tuning parameters λ	9
2.2.2 Problems with ridge	10
2.3 The Lasso	11
2.3.1 Problems with Lasso	13
2.4 Lars	16
2.4.1 Problems with Lars	17
2.5 The other related methods	20

Contents	vii
2.5.1 The Dantzig selector	20
2.5.2 Elastic-net	21
2.5.3 Adaptive Lasso	23
2.5.4 The grouped Lasso	24
2.6 Conclusion	25
3 Graphical models	27
3.1 Graph theory and notation	28
3.1.1 Graph theory	28
3.1.1.1 Marked graph	30
3.1.1.2 Decomposition of a graph	30
3.1.2 Matrices and vectors notation	32
3.2 Conditional independence and Markov properties	33
3.2.1 Conditional independence	33
3.2.2 Markov properties on undirected graphs	35
3.3 Gaussian graphical models	37
3.3.1 The likelihood	39
3.3.2 Model selection for undirected graph	42
3.3.2.1 Score-based search	43
3.3.2.2 Neighborhood selection method	44
3.4 Conclusion	44
4 Graphical least square estimation	45
4.1 Introduction	45

Contents	viii
4.2 The Graphical Least Squares estimates	47
4.2.1 The basic idea of Graphical Least Squares estimates	47
4.2.2 The general form of Graphical Least Squares Estimates . .	51
4.2.3 Variance estimation for the GLSE	55
4.2.4 Discussion on the unbiasedness and Condition 4.2.2	56
4.3 Graphical model selection	57
4.4 Simulation studies	62
4.4.1 Scenario 1	63
4.4.2 Scenario 2	67
4.4.3 Scenario 3	69
4.4.4 Scenario 4	71
4.4.5 Scenario 5	74
4.4.6 Scenario 6	77
4.5 Data analysis	78
4.6 Conclusion	83
5 Hybridisation of graphical least squares and ridge regression	85
5.1 Introduction	85
5.2 The hybridised GLSE (HGLSE)	86
5.2.1 HGLSE with two cliques	86
5.2.2 The general form of HGLSE	89
5.2.3 Bias and variance of HGLSE	90
5.3 Graph structure selection	90

Contents	ix
5.4 Portfolio optimisation	92
5.4.1 Applying HGLSE to portfolio optimisation	95
5.4.2 Financial data analysis	96
5.5 Genes expression data analysis	99
5.6 Simulation study	104
5.6.1 First simulation set	105
5.6.1.1 Scenario 1	106
5.6.1.2 Scenario 2	108
5.6.1.3 Scenario 3	111
5.6.1.4 Scenario 4	113
5.6.2 Second simulation set	115
5.6.2.1 Scenario 5	115
5.6.2.2 Scenario 6	119
5.7 Conclusion	122
6 Conclusion and future directions	123
6.1 Limitations and future direction	126
6.1.1 Discrete variables and robustness of the methods	127
6.1.2 Sparse portfolio applications	128
6.1.3 Elliptical distributions	128
Appendix	140
A Concentration matrix for Chapter 2 simulations	140

Contents	x
B Data generation procedure in Chapter 4	143
B.1 Scenario 3	143
B.2 Scenario 4	147
B.3 Scenario 5	153
C Data generated procedure in Chapter 5	161
C.1 Scenario 3	161
C.2 Scenario 4	164
C.3 Scenario 6	170
D GLSE simulation for more than one binary predictor variable in the data	178
D.1 Scenario 1	178
D.2 Data generation procedure in Scenario 1 Appendix D	181
E R package	196

CHAPTER 1

Introduction

Linear regression analysis is one of the most important classical techniques of statistics, modelling the relationship between a response variable and one or more predictor variables. It is now an essential part of many research fields such as finance, medicine, economics and psychology. The main interest is in exploring how changes in any independent variable lead to a variation in the dependent variable. Nonetheless, in today's research, applying linear regression analysis is not without some obstacles and limitations.

In modern real world applications of regression analysis, it is common to find that the number of measured variables (predictors) is larger than the sample size. Such datasets, when they exist, are usually called $n < v$ or high-dimensional problems, where n and v stand for the number of observations and variables, respectively. High-dimensional data is often regarded as one of the most complicated issues to handle in modern statistics (Fan et al. 2014, Fan & Lv 2010, Ferraty 2010). Examples of such datasets can be found in studies relating to gene expression (Carvalho et al. 2012) and finance (Bai & Shi 2011).

Ordinary least squares (OLS) regression is the commonly used method for linear regression problems. However, it is inapplicable when estimating regression coefficients for high-dimensional datasets ($n < v$). Therefore, in order to improve OLS, great efforts have been exerted and many statistical approaches have been utilised to overcome the high-dimensional data problem, drawing on the regularisation framework. Examples of such methods include ridge re-

gression, proposed by Hoerl & Kennard (1970), least absolute shrinkage and selection operator (Lasso) proposed by Tibshirani (2011) and least angle regression (Lars) proposed by Efron et al. (2004). These methods use the idea of penalised/regularised estimation to solve $n < v$ problems. Although these methods do not have the same estimation difficulty as that of OLS when $n < v$, they have their own limitations. Biased estimates are one of the first and most important limitations of these methods. The bias in coefficients is towards zero as some of these methods shrink the variables in the direction of zero. Another limitation is that some of them, Lasso and Lars for example, cannot include all the variables in the model if necessary; they can only select n variables at most (Hastie et al. 2009, Zou & Hastie 2005), which can be a shortcoming in the model when it has to include more or even all variables. Ridge regression, on the other hand, can include all the variables in the model, but this method lacks a proper justification for deciding on the significance level of the estimated coefficient due to high bias. This can also be difficult to interpret when the number of variables in the model is large. More information about the advantages and the limitations of the existing regularisation methods is provided in Chapter 2.

This thesis aims to solve the above regression problems by proposing some novel unbiased methods based on graphical models. Graphical models are natural tools used in statistics for problems covering a wide range of practical applications in areas such as genetics, physics and many others, depicting the relationships among variables. There are two main constituents in a graphical model $g = (V, \mathcal{E})$: the vertices V corresponding to random variables and the edges $\mathcal{E}$ providing information about conditional dependence among the variables. If the variables are conditionally independent, given all the other variables, there will be no edge between the two nodes. Therefore, a graph g is a compact representation of the conditionally independent structure of the

variables.

The primary importance of this study lies in the fact that it used graphical models to provide unbiased estimates of regression coefficients in high dimensional $n < v$ problems which are needed in many real world applications, such as finance. The study also provides a more attractive and interpretable way of explaining interactions between predictors, unlike the usual methods. In a nutshell, this study is a significant contribution to the literature on $n < v$ regression analyses. The main contributions of this thesis are explicitly stated in the next section.

1.1 Contributions

The main contributions of this thesis can be detailed as follows:

- A new estimator called graphical least squares estimation (GLSE) was developed, which can provide unbiased estimates of regression coefficients in high dimensional $n < v$ problems. The proposed idea is closely related to the theory of graphical models where least squares estimation in conjunction with undirected Gaussian graphical models is implemented. Due to the graphical models' inherited property of using the conditional correlation of variables, our method can thus find a graph structure (particularly in genes expression), which no existing methods can find. This makes it easy to consider the causal interpretation between predictors, which may help to identify the true regression model. Furthermore, undirected Gaussian graphical models assume that the observed variables follow a multivariate normal distribution; however, the GLSE method still works if the variables do not follow a multivariate normal distribution. The unbiasedness of the proposed methods has been shown theoretically and via

simulation.

- A hybridised form of graphical least squares estimation (called HGLSE) was developed by combining the ideas of ridge regression and GLSE. The HGLSE introduces a ridge type penalty on some variables and maintains the unbiasedness of the rest. This is very helpful in reducing the computational cost in cases where datasets have a large number of predictors, by focusing only on finding significant predictor variables, thus decreasing the scope of search for the best model.
- The portfolio optimisation problem in the financial sector is solved using the HGLSE method, which increases the portfolio's Sharpe ratio and expected returns and reduces the portfolio's risk, which has positive implications in financial applications.
- Finally, An R-package called high-dimensional linear regression via graphical models was created to implement all the work done in this thesis.

1.2 Thesis outline

The remainder of this thesis comprises 5 further chapters.

Chapter 2 gives a literature review of some established methods used for solving $n < v$ regression problems. Ridge regression (Hoerl & Kennard 1970), Lasso (Tibshirani 2011) and Lars (Efron et al. 2004) are introduced in this chapter along with a brief description of the Dantzig selector (Candes & Tao 2007), elastic net (Zou & Hastie 2005), adaptive Lasso (Zou 2006) and grouped Lasso (Yuan & Lin 2006)

Chapter 3 presents the graphical models, focusing on Gaussian graphical models. It also introduces necessary theorems in support of the models, in

addition to associated notations used throughout the thesis.

Chapter 4 introduces the idea of graphical least squares estimation (GLSE) proposed for solving linear regression problems. The proposed estimators proved to be unbiased. The chapter also provides a newly developed algorithm to find which graph g is the best fit for the data under a regression framework, and to find the GLSE. Four simulation scenarios along with an analysis of a real dataset are also reported in this chapter.

Chapter 5 consists of an extension of the proposed method where hybridisation of graphical least squares estimation with ridge regression is considered. This method (HGLSE) is useful when it is necessary to determine which variables are important and it reduces the computational cost of finding the optimal graph. The chapter also describes how to search for models in the important variables, while imposing a ridge type penalty on the graph structure associated with the less-important variables. The chapter is supported by real and simulated data analyses as well as theoretical work, demonstrating unbiased estimates for important variables. In addition, Chapter 5 presents the application of HGLSE to financial datasets to construct portfolios.

Chapter 6 concludes the thesis and it presents a summary of the main results of the experiments. It also outlines some of the limitations of the proposed estimators and includes suggestions for further research.

CHAPTER 2

Regression with $n < v$

2.1 Introduction

This chapter presents some of existing research work on linear regression problem when the number of variables v is greater than the number of observations n .

Consider a linear regression model with a univariate response, v covariates and n independent and identically distributed (i.i.d.) observations. Let $\mathbf{y} = (y_1, \dots, y_n)'$ and $X_j = (x_{1j}, \dots, x_{nj})'$ ($j = 1, \dots, v$) be the observations for the response and the j th covariate, respectively. Denote the matrix of the covariate observations by $\mathbf{x} = (X_1, \dots, X_v)$. Assume the following linear regression model representing the relationship of the response and the covariates,

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.1}$$

where the elements of $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)'$ are i.i.d. random errors with mean 0 and variance σ^2 .

This problem ($n < v$) has recently received a significant interest, especially in studies related to genomics and several other areas of computational biology (Carvalho et al. 2012, Hastie et al. 2009). As mentioned in Chapter 1, the OLS method, which uses the saturated model assumption (i.e. any pair of covariates are correlated conditional on the other covariates), does not work for solving

this regression problem because the associated design matrix is not invertible. Several methods have been proposed in the literature to avoid this problem such as ridge regression, Lasso and Lars. Other similar methods which are closely related to the aforementioned methods have been developed by other studies, including the Dantzig selector (Candes & Tao 2007), Elastic-net (Zou & Hastie 2005), adaptive Lasso (Zou 2006), grouped Lasso (Yuan & Lin 2006) and smoothly clipped absolute deviation (SCAD) (Fan & Li 2001).

Most of these methods select a parsimonious model (subset of predictors for the model) by constraining/shrinking the coefficients. This might lead to a more interpretable model with possibly high prediction accuracy. Some of the methods, such as principal component regression (Hastie et al. 2009), reduce the dimensions of the feature space before running the corresponding algorithm.

The rest of this chapter is arranged as follows. Section 2.2 describes ridge regression, while Section 2.3 introduces Lasso. Lars is explained in Section 2.4, and Section 2.5 gives an overview of the Dantzig selector, Elastic-net, adaptive Lasso and grouped Lasso. The chapter ends with a summary in Section 2.6.

2.2 Ridge regression

Ridge regression, proposed by Hoerl & Kennard (1970), is one the most popular shrinkage methods that shrinks coefficients in proportion to their sizes using L_2 penalty.

The ridge regression here has the following form:

$$\hat{\beta}_{Ridge} = \arg \min_{\beta \in \mathbb{R}^v} \left(\|\mathbf{y} - \mathbf{x}\beta\|_2^2 + \lambda \sum_{j=1}^v \beta_j^2 \right), \quad (2.2)$$

where v is the number of variables and $\lambda \geq 0$ is the shrinkage/tuning parameter

that controls the amount of shrinkage. For large values of λ , the amount of shrinkage will be large. There are several ways to choose the optimal value of the tuning parameter λ ; the most well-known of these is cross-validation.

An alternative way of expressing the ridge regression estimator is as follows:

$$\begin{aligned} \hat{\beta}_{Ridge} &= \arg \min_{\beta \in \mathbb{R}^v} \|\mathbf{y} - \mathbf{x}\beta\|_2^2 \\ &\text{subject to } \sum_{j=1}^v \beta_j^2 \leq t, \end{aligned} \quad (2.3)$$

where the value of t directly depends on the amount of shrinkage λ . The above formalisation makes the size restriction on the parameters clear.

Ridge regression does not make the coefficients exactly zero but shrinks them toward zero and they become close to each other. This method does not carry out variables selection like other methods, such as Lasso (described in the next section). To demonstrate how the ridge method works, consider a two-dimensional regression problem, where the estimation is graphically shown in Figure 2.1.

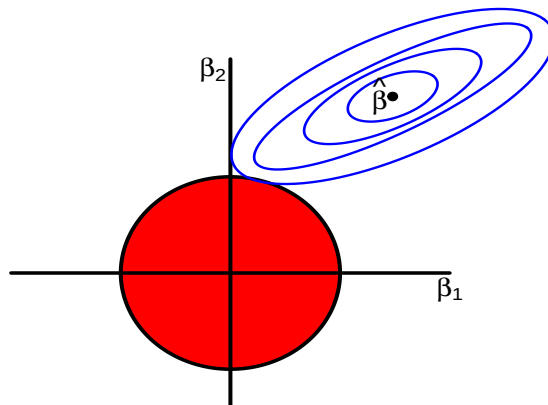


Figure 2.1: Estimation picture of the two-dimensional linear regression problem for ridge.

In the graph, the L_2 constraint region, $\beta_1^2 + \beta_2^2 \leq t$, is shown by the ball. The ellipses show the contours of the error function centered at the ordinary (full) least square estimates. The solution of ridge takes place at the point where the contours hit the ball (constraint region).

2.2.1 Cross-validation for choosing tuning parameters λ

The tuning parameter λ influences the bias in estimating the coefficients of ridge. It is usually chosen through different ways which all fit the model to a sequence of λ values. Then, based on a suitable criterion, the set of candidate models is refined and the final model is selected.

Cross-validation is one of the methods used to find the optimal λ , which includes minimising an estimate of the prediction error. A single run of cross-validation splits a given sample dataset into subgroups. The subgroups from which the model is created are called the training subgroups and the subgroup used for validating the analysis of the model is called the validation or testing subgroup. For reducing variability, multiple runs of cross-validation are conducted using different partitions of the given sample, and the results obtained from validations are averaged over the runs to obtain a single estimate of the prediction error. There are various variants of cross-validation, for example, 2-fold cross-validation, 10-fold cross-validation, or generally, N -fold cross-validation. In N -fold cross-validation, the given sample data are randomly divided into N subgroups of equal sizes. Of them, one subgroup is used as the validation data for testing purposes, leaving $N - 1$ subgroups as training data for training the model. This process is repeated for all the N -folds and a single average of the estimates is obtained from all the folds. This completes a single iteration of the cross validation. In the next iteration the given data is again split randomly into N -folds and another estimate of the desired measure is obtained. This process is

repeated multiple times, facilitated by available computational resources (James et al. 2013).

It is noteworthy that cross-validation is used with all the regularisation methods adopted in this thesis to find λ and the optimal models.

2.2.2 Problems with ridge

One of the major limitations of this technique is that it does not shrink regression coefficients exactly to zero, which leads to a non-sparse model that is difficult to interpret when the number of variables is extremely large (Yuan et al. 2007). This problem makes it difficult to apply ridge to genes expression datasets where the number of variables is huge and it is believed that only a few variables have significant effects on the response. This limitation is manifest when analysing genes expression data in Chapters 4 and 5.

Another issue with ridge regression is that all of its estimates are biased giving emphasis to the famous notion of “bias-variance tradeoff”, which might lead to over-shrinking some of the coefficients; this can in turn cause a problem in portfolio optimisation when ridge is used to deal with high dimensional data (i.e. the number of assets is greater than the number of their returns). This problem is further examined in Chapter 5.

Finally, the choice of the tuning parameter λ is a judgmental one and has no objective basis (Uslu et al. 2014). While several formal procedures have been proposed for making this choice, such as cross-validation, they suffer from their own limitations (Uslu et al. 2014). For instance, one of the drawbacks of cross-validation is that there may be an overlap in the validation subsets owing to the fact that some observations may be selected several times while other subsets may not be selected at all (Maimon & Rokach 2005).

2.3 The Lasso

The Lasso method was proposed by Tibshirani (2011). It is a penalised regression method that works in a similar fashion as the ridge regression. It shrinks the coefficients and variable selection simultaneously using L_1 penalty on the coefficients. The Lasso regression has the following estimator:

$$\hat{\beta}_{Lasso} = \arg \min_{\beta \in \mathbb{R}^v} \left(\|\mathbf{y} - \mathbf{x}\beta\|_2^2 + \lambda \sum_{j=1}^v |\beta_j| \right) \quad (2.4)$$

where $\lambda \geq 0$ is the shrinkage/tuning parameter that controls the amount of shrinkage. For large values of λ , the amount of shrinkage will be large. The optimal value of the tuning parameter λ can be chosen in a similar fashion as in the case of the ridge regression.

Another way of expressing the Lasso regression estimator is as follows:

$$\begin{aligned} \hat{\beta}_{Lasso} &= \arg \min_{\beta \in \mathbb{R}^v} \|\mathbf{y} - \mathbf{x}\beta\|_2^2 \\ &\text{subject to } \sum_{j=1}^v |\beta_j| \leq t, \end{aligned} \quad (2.5)$$

where the value of t directly depends on the amount of shrinkage λ .

Due to the nature of the above constraint, the Lasso can make the coefficients of several variables exactly zero (unlike the ridge regression) by making t small enough. This makes the Lasso a variable selection method which is very useful in high dimensional settings.

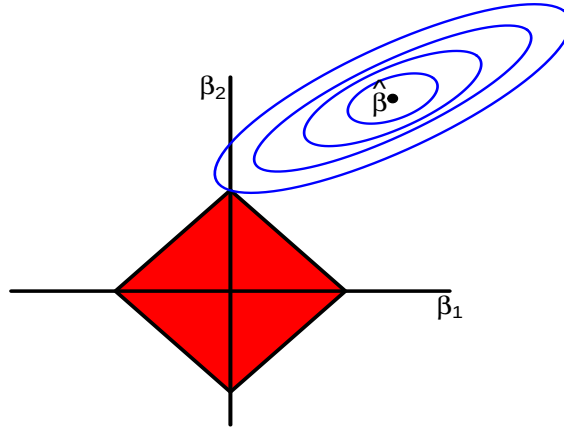


Figure 2.2: Estimation picture of the two-dimensional linear regression problem for the Lasso.

To illustrate how the Lasso method leads some regression coefficients to be exactly zero and see how it is different from ridge regression, consider a two-dimensional regression problem where the estimation picture by the Lasso is shown in Figure 2.2. The L_1 constraint region, $|\beta_1| + |\beta_2| \leq t$, is shown by the diamond while the ellipses are the contours from the OLS error function. Similar to the case of ridge, the solution of the Lasso occurs at the point where the contours hit the constraint region. The L_1 constraint region shown by the square (diamond) has its corner on the coordinate axes. It follows that at least one of the parameters (in our case β_1) is zero if the solution occurs at a corner of the diamond. As t controls the amount of shrinkage in the estimates, smaller values of t lead to more shrinkage and force some of the estimates to be equal to zero. This makes the Lasso solution sparse in that some of the regression coefficients are always shrunk to zero, which helps in interpreting the results. On the contrary, in ridge regression, there are no edges in the ball, making it unlikely for the contour lines to touch a corner, which results in the lack of regression coefficients that are equal to zero.

2.3.1 Problems with Lasso

The major limitation of Lasso is that in the case of $n < v$ (high dimensional situation), Lasso can select n predictors at most (Zou & Hastie 2005). This is due to the nature of the convex optimisation problem that Lasso tries to minimise, which limits the application of Lasso in some areas, such as the construction of a non-sparse portfolio in the financial sector when the number of assets (variables) is greater than the number of asset returns (observations). This shortcoming is further illustrated in the simulation implemented later in this section.

Another drawback of Lasso lies in the issue of its stability, especially with high-dimensional data (Zou & Zhang 2009). Stability is considered as the main feature for any variables selection method, where the method is regarded as stable if it selects almost the same variables when used in various random arrangements of the given data. When Lasso is used for variables selection, it selects different sets of variables at each run. In other words, the group of variables selected by Lasso does not consistently comprise the true set of important variables (Guo et al. n.d., Mordellet et al. 2013). This makes it difficult to identify which set of variables is vital in regulating the response; hence, this lack of stability casts doubt on the relevance of the subset of variables Lasso selects. Likewise, this weakness will be shown experimentally in the simulations carried out below and in Chapter 5, where the model size suffers from high variation.

Furthermore, in the case where the variables are correlated, i.e. observed most of the time in real life datasets, Lasso selects a single variable from a bunch of correlated variables arbitrarily (Zou & Hastie 2005), which is not usually preferred as this may not lead to the production of the optimal model.

To address the first limitation of Lasso, which is its ineffectiveness and lack

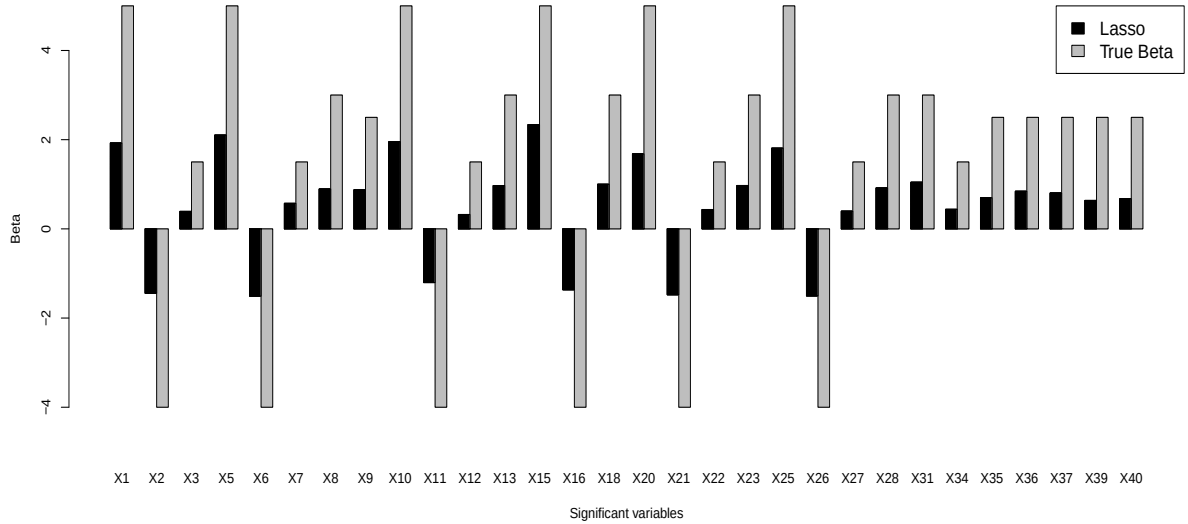
of consistency in variable selection when the number of variables that have significant effects on the response is greater than the number of observations, we consider the following scenario. The generated dataset consists of a total of $v = 40$ variables, 30 of which have a significant impact on the response. A total of $n = 25$ samples are generated. The observations on the predictor variables are generated from multivariate normal distribution with mean $\mathbf{0}$ and variance covariance matrix Σ (given in Appendix A). Then the response y is generated from model (2.1), where the random errors ϵ are normally distributed with mean 0 and standard error $\sigma = 1$. Cross-validation is used to obtain the penalty parameter λ . The main focus is on how many of the significant variables are chosen by Lasso based on the following 3 criteria (Li & Shao 2015):

- Sensitivity: the proportion of true non-zero regression coefficients being estimated as non-zero (the higher the better);
- Specificity: the proportion of true zero regression coefficients being estimated as zero (the higher the better);
- Model size: the size of the chosen model (the closer it is to the true model size, the better, but without affecting the other two criteria).

Another goal of this scenario is to assess the performance of Lasso in estimating the true regression coefficients for the significant variables. The average estimated results from 500 realisations of the data are given in Table 2.1 and Figure 2.3.

Table 2.1: *Results of the Scenario, $n = 25, v = 40$.*

Method	Sensitivity	Specificity	Model Size
Lasso	44.05% (0.220)	74.76% (0.186)	15.74(7.946)

**Figure 2.3:** *Comparison between the true coefficients and the Lasso ones.*

The results from Table 2.1 indicate that despite the fact that Lasso has acceptable specificity, it has low sensitivity and the model size is very small with high simulation standard deviation (given in parentheses) compared to the size of the true model. While the true model size (number of significant variables) is 30, Lasso, which is supposed to select up to 25 variables, is not able to select more than 15 on average. Lasso is also not consistent in selecting variables as can be seen from the high standard deviation (given in parenthesis) of the model size. In addition, Figure 2.3 shows that Lasso over-shrinks the coefficients, corroborating Radchenko et al. (2011), which is problematic, especially in portfolio optimisation, when the estimated coefficients are used to find the weights of each asset on the portfolio (Britten-Jones 1999).

2.4 Lars

Lars, proposed by Efron et al. (2004), uses a very similar strategy to build a regression model via stepwise addition of variables into the model. In stepwise regression model selection, the best variable is identified and included in the model, which results in updating the least square fit. In this way, all the potentially useful variables are added in the model. Similarly, least angle regression adds variables in the model one by one starting from the variable that is most correlated with the response. Instead of completely fitting this variable, Lars decreases the value of the correlation by moving the coefficient of the variable towards its least square coefficient. This gives a chance to another variable to join the model if its correlation with the response becomes equal to the correlation of the variable identified in the first step. Now the coefficients of both variables are moved in a way to give a chance to a third variable to join and so on. Lars takes a maximum of v steps as the variables are added sequentially. The algorithm takes n steps if and only if $v > n$. The algorithm steps are as follows:

Algorithm 1 Least angle regression

-
- 1: Let the starting residual be $\hat{r} = \mathbf{y}$ and coefficients $\beta_1, \dots, \beta_v = 0$.
 - 2: Identify a predictor X_j that has the highest amount of correlation with the residual $\hat{r}$. Let the coefficient of this variable be β_j .
 - 3: To decrease the correlation, β_j is moved from 0 in the direction of its least square coefficient of the current residual $\langle X_j, \hat{r} \rangle$. This is done continuously until another variable $X_{j'}$, with coefficient $\beta_{j'}$, has the same amount of correlation with the current residual vector $\hat{r}$.
 - 4: To further reduce the correlation and give the chance to another variable, β_j and $\beta_{j'}$ are both moved in the direction of their joint least square coefficient of the residual until another variable joins the active set.
 - 5: The above process is continued until all the v predictors are added. The full least square solution arrives after $\min(v, n)$ steps.
-

A simple modification of the Lars algorithm can give the whole Lasso path. The modification is that if one of the non-zero coefficients at any stage of the algorithm, except the first one, becomes zero, its variable is removed from the active set variables and the current combined direction of the least squares is recalculated. The Lars algorithm is extremely efficient in that it requires the same order of computation as that required for a single least squares fit using the v variables (Hastie et al. 2009). To reach the full least squares estimates, Lars always takes v steps, whereas the Lasso path might have more than v steps in spite of the fact that the two often exhibit very similar performance (Hastie et al. 2009).

2.4.1 Problems with Lars

Lars appears to share most of the limitations associated with Lasso. The major shortcoming of Lars is that it can take at most n covariates (Hastie et al. 2009).

This is problematic in some areas where more than n or even all covariates have to be included in the model. For instance, in finance, like Lasso, this method is ineffective when building a non-sparse portfolio where the number of assets is larger than the number of observations. This limitation is further discussed in the simulations presented below.

A simulation is presented here to prove the main limitation of Lars, i.e. the fact that it is ineffective when the number of significant variables which affect the response is greater than the number of observations. The same procedures of generating the data in the Lasso simulation are also applied in this scenario. Cross-validation is used to find the number of steps that Lars requires to find the optimal model. The main focus is on how many of the significant variables are chosen by Lars based on the same 3 criteria used in the Lasso simulation (2.3.1). One further objective of this scenario is to evaluate the performance of Lars in estimating the true regression coefficients for the significant variables. The average estimated results from 500 realisations of the data are given in Table 2.2 and Figure 2.4.

Table 2.2: Results for the Scenario, $n = 25, v = 40$.

Method	Sensitivity	Specificity	Model Size
Lars	30.00% (0.170)	82.04% (0.156)	10.80(6.083)

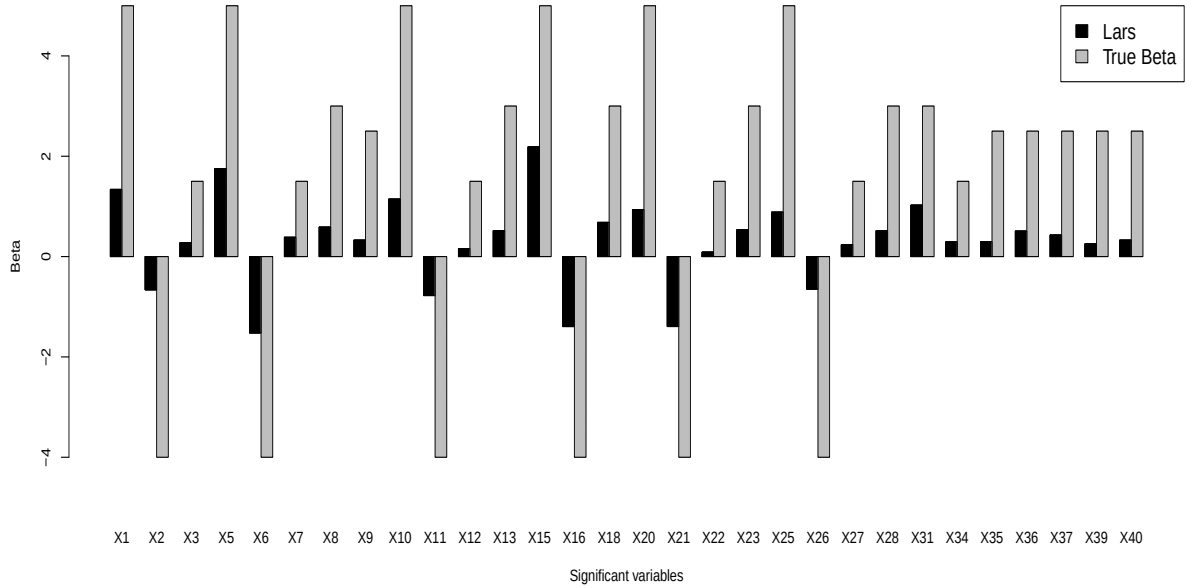
**Figure 2.4:** Comparison between the true coefficients and the Lars ones.

Table 2.2 shows that although Lars has acceptable specificity, it has very low sensitivity and the model size is extremely small with high standard deviation compared to the size of the true model. In addition, while it is supposed to select up to 25 variables, it only selects 10 variables on average out of the 30 significant variables. Moreover, it appears to lack consistency in variable selection as can be seen from the standard deviation in the model size. Besides, Figure 2.4 displays that Lars, to a similar or even greater extent than Lasso, over-shrinks the coefficients.

2.5 The other related methods

Other regression methods have been utilised in the literature for high-dimensional data ($n < v$), a summary of which is presented in this section.

2.5.1 The Dantzig selector

Candes & Tao (2007) proposed a method, called the Dantzig selector, which takes the following form:

$$\min_{\beta} \|\beta\|_1 \quad \text{subject to} \quad \|\mathbf{x}^T(\mathbf{y} - \mathbf{x}\beta)\|_{\infty} \leq \lambda. \quad (2.6)$$

An equivalent expression for the above Dantzig selector is given as follows:

$$\min_{\beta} \|\mathbf{x}^T(\mathbf{y} - \mathbf{x}\beta)\|_{\infty} \quad \text{subject to} \quad \|\beta\|_1 = \sum_{j=1}^v |\beta_j| \leq t. \quad (2.7)$$

In the above expression, $\|\cdot\|_{\infty}$ represents the L_{∞} norm, which is the maximum absolute value of the vector components (Hastie et al. 2009). In the latter of the above two forms, the Dantzig selector replaces the squared error loss of the Lasso by the maximum absolute gradient value, which makes a resemblance between the two approaches. The Dantzig selector, like the Lasso, gives the L1 solution if the value of t is large for $n > v$. However when t is small, the Dantzig selector produces a different path of solutions than those produced by the Lasso (Hastie et al. 2009).

One of the limitations of the Dantzig selector is that when using the Lars algorithm to find the Lasso path and compare the result with that of the Dantzig selector, the latter can include a variable in the model that has smaller correlation with the active set residual than with those of the non-active variables (Efron et al.

2007). This appears to be unsatisfactory and may sometimes lead to weakness in prediction accuracy. Furthermore, since the regularisation parameter λ is varied, the Dantzig selector can give inconsistent coefficient paths (Efron et al. 2007).

2.5.2 Elastic-net

Due to the nature of Lasso penalty, Lasso cannot select more variables than n observations. To fix this problem, and do a reconciliation between ridge and Lasso, Zou & Hastie (2005) proposed a modified form of penalty for the Lasso. They called this penalty the Elastic-net penalty which has the following form:

$$\lambda \sum_{j=1}^v \left((1 - \varrho) |\beta_j| + \varrho \beta_j^2 \right), \quad (2.8)$$

where ϱ determines the amount of share from the ridge and Lasso penalties usually chosen by cross validation or some other tuning parameter selection method.

Variable selection via the Elastic-net is similar to that of the Lasso; however, like ridge regression, it shrinks the coefficients of correlated variables at the same time. The first term in the above penalty inspires a sparse solution in terms of coefficients (similar to Lasso) while the second part of the penalty encourages the highly correlated variables toward average (resembling ridge penalty). This might be helpful in applications where strong correlations in the predictor variables are present. For instance, genes in genomic data are considered to have high correlations among themselves that operate in specific molecular pathways. In such data Elastic-net might be more suitable to use than Lasso.

Another advantage of using the Elastic-net method rather than Lasso is that

the former can produce more than n non-zero coefficient estimates in situations when $n < v$.

One of the major shortcomings of the Elastic-net is that it lacks the oracle property (Zou & Zhang 2009), which means that it may not be consistent in variable selection when two-dimensional cross-validation is used to find the tuning parameter, as demonstrated in the simulation below. Another drawback of the elastic net is that it tends to over-shrink the final coefficients (Croiseau & Cordell 2009).

A simulation is presented here to examine whether the Elastic-net method, which can solve one of the main limitation of Lasso (the fact that Lasso can take at most n covariates), can also be consistent in terms of variable selection. The same procedures of generating the data in the Lasso and Lars simulations are also applied here. Cross validation on a two-dimensional surface is carried out for the two tuning parameters of Elastic-net. The objective is to discover how many significant variables are selected by Elastic-net based on the same 3 criteria (2.3.1) used in the Lasso simulation (sensitivity, specificity and model size). Another objective of this simulation is to assess the performance of the Elastic-net method in true regression coefficient estimation for the significant variables. Table 2.3 and Figure 2.5 show the average estimations from the 500 runs of the data.

Table 2.3: Results for the Scenario, $n = 25, v = 40$.

Method	Sensitivity	Specificity	Model Size
Elastic-net	83.43% (0.206)	28.10% (0.317)	32.22(9.100)

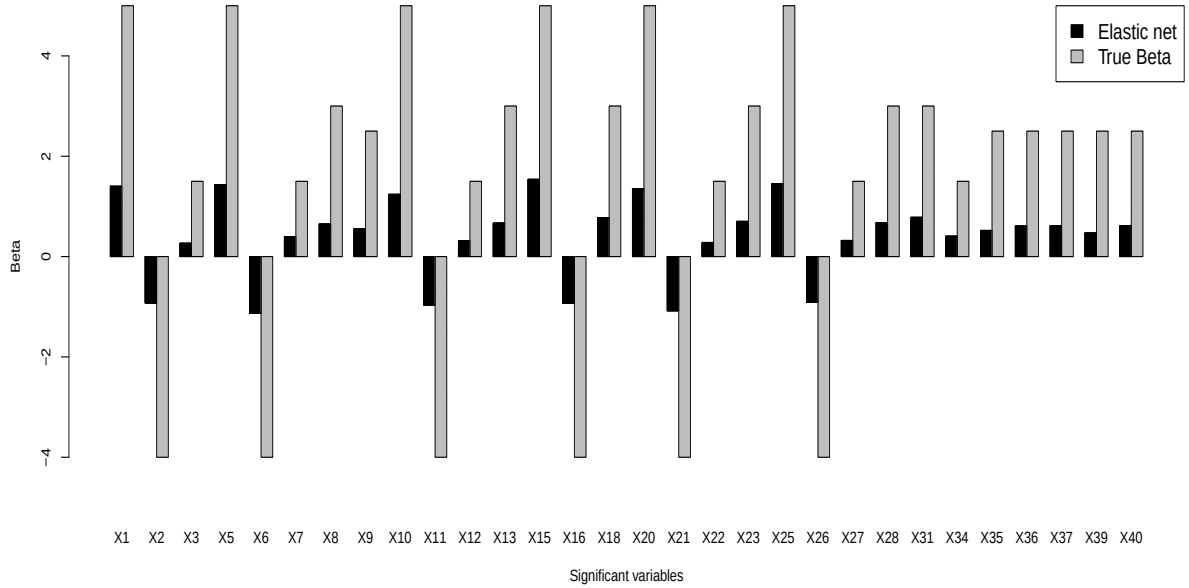
**Figure 2.5:** Comparison between the true coefficients and the Elastic-net ones.

Table 2.3 displays that although Elastic-net has acceptable sensitivity, it has very low specificity and the model size has high simulation standard deviation (given in parentheses). This suggests that Elastic-net is not consistent in variable selection and is even worse than Lasso, as indicated by the standard deviation in the model size. In addition, Figure 2.5 shows that Elastic-net, like Lars and Lasso, over-shrinks the coefficients, confirming the proposition of Croiseau & Cordell (2009).

2.5.3 Adaptive Lasso

Adaptive Lasso, proposed by Zou (2006), is an extension of the Lasso method. The main reason for proposing this method was to favour those predictors that

have high marginal correlation with the response and to avoid spurious and noisy selection of variables. In the adaptive Lasso method, a relatively higher penalty is used for zero coefficients while a lower penalty is used for nonzero coefficients, hoping to minimise the estimation bias and obtain better variable selection accuracy compared to the standard Lasso approach. This approach uses a weighted form of penalty given as:

$$\sum_{j=1}^p \tau_j |\beta_j|, \quad (2.9)$$

where the weights τ_j are calculated as

$$\tau_j = \frac{1}{|\hat{\beta}_j|^\vartheta}. \quad (2.10)$$

In the above expression, ϑ is a positive number and $\hat{\beta}_j$ is the ordinary least square estimate.

Despite the above mentioned advantage, adaptive Lasso still shares Lasso's main shortcoming, whereby it can only select n variables at most when $n < v$. Moreover, the adaptive Lasso method's selection at the group level is empirically shown to be unsatisfactory (Seetharaman 2013).

2.5.4 The grouped Lasso

The grouped Lasso generalised by Yuan & Lin (2006) might be useful in situations when it is desirable to shrink and select variables that belong to a particular group. The existence of predefined groups of variables can be observed in several problems, such as gene expression data, where some genes might belong to the same biological pathway.

To state the grouped Lasso criterion formally, assume that the v variables can

be divided in L groups, where there are v_l variables in the l th group represented by matrix $\mathbf{x}_l$. Let the associated coefficients for the variables $\mathbf{x}_l$ be $\boldsymbol{\beta}_l$. Then the grouped Lasso tries to minimise the following convexity criterion:

$$\min_{\boldsymbol{\beta} \in R^v} \left(\left\| \mathbf{y} - \sum_{l=1}^L \mathbf{x}_l \boldsymbol{\beta}_l \right\|_2^2 + \lambda \sum_{l=1}^L \sqrt{v_l} \|\boldsymbol{\beta}_l\|_2 \right), \quad (2.11)$$

where $\sqrt{v_l}$ and $\|\cdot\|_2$ stand for the varying group size and Euclidean norm (not squared), respectively (Friedman et al. 2010). This approach encourages a sparse solution in groups and individually. The use of the above criterion might also be useful when the levels of a categorical variable are treated as groups where dummies are introduced for the levels.

The grouped Lasso shares some of Lasso's drawbacks. For instance, it lacks consistency when it selects variables and it tends to over-shrink large coefficients due to the fact that its rate of penalisation remains the same with large coefficients (Breheny & Huang 2015).

2.6 Conclusion

In this chapter, it has been highlighted that the OLS, which is the most popular method in linear regression, utilises the saturated model assumption when $n > v$. However, this method does not work for datasets where there are more predictors than observations (i.e. $n < v$). To overcome the linear regression problems when $n < v$, many alternative methods have been used in the literature, such as ridge, Lasso and Lars, but these methods do not take conditional correlation into consideration when $n < v$ and consequently may not be able to identify the true model. In addition, some recently proposed methods have also been briefly introduced in this chapter, which include the Dantzig selector,

Elastic-net, adaptive Lasso and the grouped Lasso. However, these methods still do not make use of the conditional correlation. For example, adaptive Lasso has considered the marginal correlation between the response and predictors, but high marginal correlation does not mean high conditional correlation as some predictors may be independent of the response given a certain subset of other predictors. The next chapter provides further discussion on the concept of conditional correlation and the following chapters introduce methods that can make effective use of conditional correlation when $n < v$.

CHAPTER 3

Graphical models

Graphical models, which have their origin in genetics (Wright 1923) and several other fields such as economics (Wold 1954) and social sciences (Blalock 1971), are probabilistic models that graphically represent the conditional dependencies by edges among multivariate random variables. These models have emerged as some of the most attractive methods in that they provide an intuitively appealing framework/representation of the real-world scenarios (Jordan 1998). Edwards (2000) mentioned a situation where focusing only on pairwise association between variables is inadequate, in which sexual discrimination was observed in university admissions of female applicants against males. The data showed a 41.6% admission rate for females against 58.4% for males for an equal number of male and female applicants. However, when the given figures were considered on departmental levels, there was no such discrimination in that some of the departments had a large number of seats and only a few females applied to those departments. It followed that the number of admissions for males was independent from the number of admissions for females given the department. Hence, it is suggested that we must study the conditional association of variables given the rest, and not only their marginal correlation.

In this thesis, the main focus is on Gaussian graphical models, which are used to solve linear regression problems in high dimensional datasets. A significant amount of literature is available on graphical models in general, including major studies by Edwards (2000), Lauritzen (1996) and Whittaker (2009) and the

references therein.

The theory behind these models is summarised in this chapter, which is organised as follows. Section 3.1 provides the graph theory and the notations used in this thesis. Section 3.2 gives an introduction to conditional independence and Markov properties while Section 3.3 introduces Gaussian graphical models. Finally, Section 3.4 concludes the chapter.

3.1 Graph theory and notation

3.1.1 Graph theory

A graph g consists of two sets, a set V and a set $\mathcal{E}$. The set V denotes the vertices representing variables and $\mathcal{E}$ is the set of edges (a subset of $V \times V$) connecting the vertices. The elements in the set V are usually natural numbers, i.e. $V = 1, 2, \dots, v$, representing the labels of random variables. If a graph with v vertices has $\binom{v}{2}$ edges then it is complete; otherwise, the graph is incomplete. The pair (i, j) denotes the edge between vertex i and vertex j if the two vertices are connected by a line, in which case they are said to be adjacent or neighbours. The set of neighbours for a vertex i is represented by $ne(i)$. If the vertices i and j are not connected by a line, the i and j are non-adjacent. This kind of graph is called an undirected graph.

If $A \subseteq V$, the subset A induces a subgraph $g_A = (A, \mathcal{E}_A)$, where $\mathcal{E}_A = \mathcal{E} \cap (A \times A)$. A graph is complete if all vertices are joined by an edge and a subset is complete if it induces a complete subgraph from g . This subgraph is maximal if it cannot be extended by including one more neighbouring vertex. A complete subset that is maximal is called a clique. A path in a graph g is a sequence of edges that connect a sequence of vertices $i_0, \dots, i_v$. If the sequence of edges in the given path starts and ends at the same vertex, then it is called a cycle.

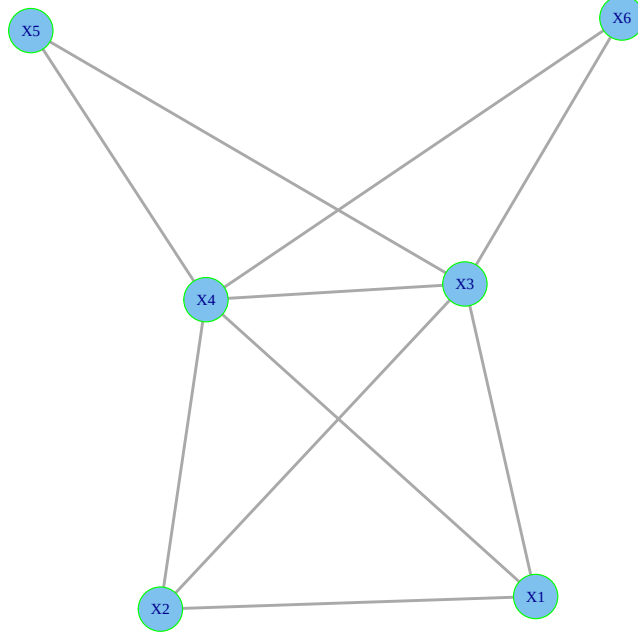


Figure 3.1: An example of undirected graphs.

Example 3.1.1 The diagram of an undirected graph is given in Figure 3.1. The graph has a set of vertices

$V = \{X_1, \dots, X_6\}$ and edge set $\mathcal{E} = \{(X_1, X_2), (X_1, X_3), (X_1, X_4), (X_2, X_3), (X_2, X_4), (X_3, X_4), (X_3, X_5), (X_3, X_6), (X_4, X_5), (X_4, X_6), \cup (X_2, X_1), (X_3, X_1), (X_4, X_1), (X_3, X_2), (X_4, X_2), (X_4, X_3), (X_5, X_3), (X_6, X_3), (X_5, X_4), (X_6, X_4)\}$,

and many paths can be seen in the given graph from vertex X_1 to X_6 . The neighbourhood of vertex X_1 is $ne(X_1) = \{X_2, X_3, X_4\}$; similarly the neighbourhood of vertex X_5 is $ne(X_5) = \{X_3, X_4\}$. The neighbourhood of the set $\{X_1, X_2\}$ is $ne(X_1, X_2) = \{X_3, X_4\}$. The cliques in the given graph are

$$C = \{\{X_1, X_2, X_3, X_4\}, \{X_3, X_4, X_5\}, \{X_3, X_4, X_6\}, \}.$$

An undirected graph is said to be triangulated if all the cycles having lengths $v \geq 4$ possess two non-consecutive vertices that are neighbours; that is, they should have a chord (short cut).

If A, B and C are three subsets of the vertices set V , we say that B is separating A from C if all possible paths from A to C are blocked by B .

3.1.1.1 Marked graph

A graph that has two types of vertices is referred to as a marked graph (Lauritzen 1996). The vertex set in a marked graph has the following structure:

$$V = \Delta \cup \Psi \text{ such that } \Delta \cap \Psi = \Phi,$$

where Φ is the empty set, Δ is the set of vertices for qualitative variables and Ψ denotes the set of quantitative variables. On the other hand, a graph is said to be *pure* if it consists of the same type of vertices.

3.1.1.2 Decomposition of a graph

A triple (A, B, C) of disjoint subsets of the vertex set V of an undirected graph g is said to form a *decomposition* of g if $V = A \cup B \cup C$ and the following conditions hold (Lauritzen 1996):

- B separates A from C ;
- B is a complete subset of V ;
- $B \subseteq \Delta$ OR $C \subseteq \Psi$.

If $B \subseteq \Delta$ OR $C \subseteq \Psi$ does not hold, then the undirected graph g is said to form a (weak) decomposition of g .

An undirected graph g is considered as decomposable if it holds one of the following:

- Graph g is complete.
- There is a proper decomposition (A, B, C) into decomposable subgraphs g_{AB} and g_{BC} where B separates A from C .

Example 3.1.2 *Graphs with strong and weak decomposition are given in Figure 3.2.*

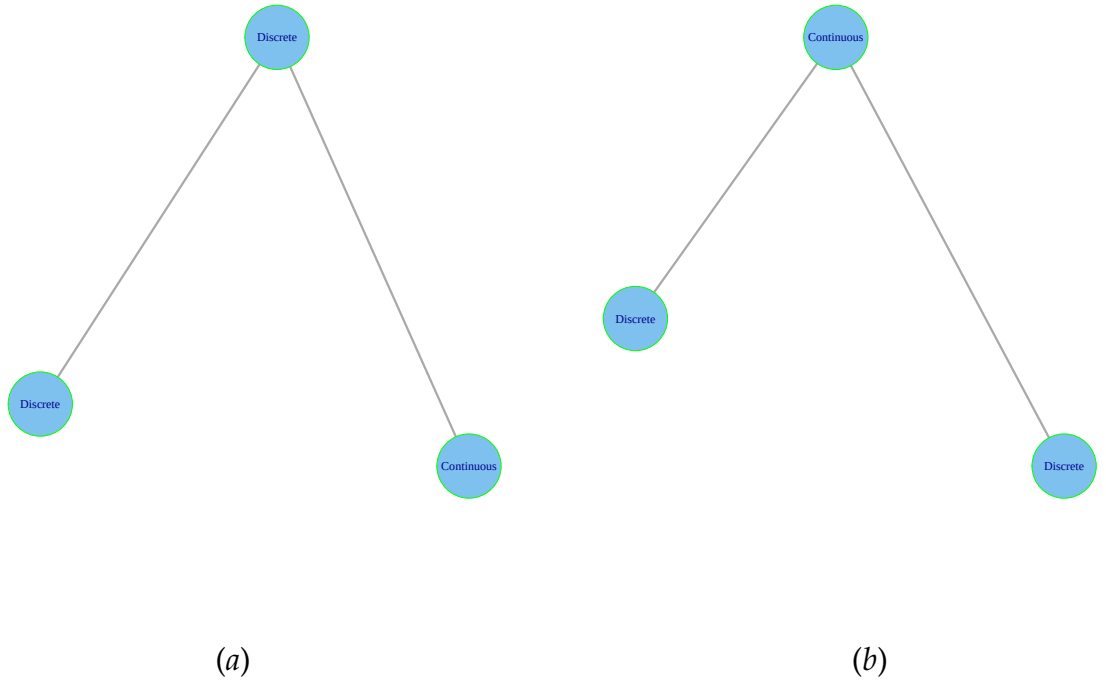


Figure 3.2: Example of (a) strong decomposable and (b) weak decomposable graphs.

Consider a sequence of sets $C_1, \dots, C_q$ that are the subsets of the vertex set V of an undirected graph g such that $C_1 \cup \dots \cup C_q = V$. If the following holds, then the given sequence is said to be a perfect sequence (Lauritzen 1996):

$$S_j = C_j \cap (C_1 \cup C_2 \cup \dots \cup C_{j-1}) \subseteq C_i,$$

where $j = 2, \dots, q$ and $i \in \{1, \dots, j-1\}$. The sets S_j are the separators. These orderings, if they exist, might not be unique.

Example 3.1.3 *The perfect ordering of the separators and cliques for the graph given in Figure 3.1 can be given as: $\mathcal{S} = \{S_2 = \{X_3, X_4\}, S_3 = \{X_3, X_4\}\}$ and $\mathcal{C} = \{C_1 = \{X_1, X_2, X_3, X_4\}, C_2 = \{X_3, X_4, X_5\}, C_3 = \{X_3, X_4, X_6\}\}$, respectively.*

An undirected graph g will be decomposable if and only if it admits a perfect ordering of the cliques (Golumbic 2004).

3.1.2 Matrices and vectors notation

In this and later chapters, we reserve the capital letters A, B, C and S to denote a subset of V and $|\cdot|$ to denote the number of elements for an edge or vertex set. The lower letters i, j, k are reserved for the notation of indices.

The following notations mainly follow the style in Lauritzen (1996). A vector $\mathbf{Z} \in \mathbb{R}^v$ can also be written as $(Z_j)_{j \in V}$ or $[Z_j]_{j \in V}$. The j th element Z_j can also be written as $(\mathbf{Z})_j$ or $[\mathbf{Z}]_j$. Denote $\mathbf{Z}_A$ as an a -dimensional subvector ($a = |A|$) of $\mathbf{Z}$, with $\mathbf{Z}_A = (Z_j)_{j \in A}$. Denote $[\mathbf{Z}_A]^\Gamma$ as a v -dimensional vector obtained by filling up 0s, with

$$([\mathbf{Z}_A]^\Gamma)_j = \begin{cases} Z_j & \text{if } j \in A \\ 0 & \text{otherwise.} \end{cases} \quad (3.1)$$

We define similar notations for matrices. A $v \times v$ matrix $\mathbf{z}$ can also be written as $(z_{kj})_{k,j \in V}$. We may also write $z_{kj} = (\mathbf{z})_{kj}$. Denote $\mathbf{z}_{AB} = (z_{kj})_{k \in A, j \in B}$, a submatrix of $\mathbf{z}$. Denote $[\mathbf{z}_{AB}]^\Gamma$ as a $v \times v$ -dimensional matrix obtained by filling up 0s, with

$$([\mathbf{z}_{AB}]^\Gamma)_{jk} = \begin{cases} z_{jk} & \text{if } j \in A, k \in B \\ 0 & \text{otherwise.} \end{cases} \quad (3.2)$$

With the above notation, we can use $\mathbf{x}_A$ to denote the covariate matrix only having covariates with indices in set A and $ssd_A = \mathbf{x}_A' \mathbf{x}_A$. The notation $[(ssd_A)^{-1}]^\Gamma$ denotes a $v \times v$ -dimensional matrix obtained by filling up 0s, with

$$\left([(ssd_A)^{-1}]^\Gamma\right)_{jk} = \begin{cases} ((ssd_A)^{-1})_{jk} & \text{if } j, k \in A \\ 0 & \text{otherwise.} \end{cases} \quad (3.3)$$

Clearly $[\mathbf{x}_B]^\Gamma [(ssd_A)^{-1}]^\Gamma = \mathbf{0}$ if $B \cap A = \phi$, the empty set.

3.2 Conditional independence and Markov properties

3.2.1 Conditional independence

The concept of conditional independence is one of the essential notions in the study of graphical models and is concerned with the relationship between random variables.

Definition 3.2.1 (Conditional independence (Whittaker 2009)). *Let X_1, X_2 , and X_3 be three random variables then we say the random variable X_1 is conditionally independent of X_2 given the random variable X_3 if and only if the joint probability density function f satisfies*

$$f(x_1, x_2 | x_3) = f(x_1 | x_3) f(x_2 | x_3) \quad \forall x_1, x_2, \quad \text{for all } x_3 \text{ where } f(x_3) > 0. \quad (3.4)$$

This relationship is denoted by $X_1 \perp\!\!\!\perp X_2 | X_3$.

In plain terms, conditional independence can be considered as describing the idea of irrelevance under a given context by interpreting the relationship

$X_1 \perp\!\!\!\perp X_2 \mid X_3$ as “if X_3 is known, then information about X_2 has no relevance for the knowledge of X_1 ” (Edwards 2000). This statement is helpful when we use graphs to report the conclusions of an analysis.

Another characterisation of $X_1 \perp\!\!\!\perp X_2 \mid X_3$ is that the joint density of (X_1, X_2, X_3) can be factorised in two functions as given in Whittaker (2009) and is known as the factorisation criterion, formalised in the following proposition.

Proposition 3.2.1 *The random variable X_1 is conditionally independent of X_2 given the random variable X_3 if and only if there exist two functions d and h such that*

$$f(x_1, x_2, x_3) = d(x_1, x_3) h(x_2, x_3) \quad \forall x_1, x_2, \quad \text{for all } x_3 \text{ for which } f(x_3) > 0. \quad (3.5)$$

Clearly, from the above equation, function d does not depend on X_2 and function h does not depend on X_1 .

In graph theory, a given graph is said to be a conditional independence graph if a pair of variables are not connected by a line; the two variables in that pair should always be independent from each other given the rest of the variables (Whittaker 2009).

Formally, a conditional independence graph of $\mathbf{x}$ is an undirected graph $g = (V, \mathcal{E})$, where $V = \{1, \dots, v\}$ and a pair (i, j) of vertices is not contained in the edge set $\mathcal{E}$ if and only if the following condition is satisfied:

$$X_i \perp\!\!\!\perp X_j \mid \mathbf{x}_{V \setminus \{i, j\}},$$

that is,

$$(i, j) \notin \mathcal{E} \Leftrightarrow X_i \perp\!\!\!\perp X_j \mid \mathbf{x}_{V \setminus \{i, j\}}.$$

The graph thus formed gives a diagrammatic representation of the structure of dependence/association between the variables.

3.2.2 Markov properties on undirected graphs

A graph is described as a conditional independence graph if the edge is missing between two vertices whenever the two variables are independent given all the other variables. However, associated with the undirected graph $g = (V, \mathcal{E})$, there are alternative ways of stating this notion (known as Markov properties (Dawid & Lauritzen 1993)):

- *Pairwise Markov property (PMP)*: Associated to graph $g = (V, \mathcal{E})$, this property is satisfied if for any pair (i, j) of non-adjacent vertices

$$X_i \perp\!\!\!\perp X_j \mid \mathbf{x}_{V \setminus \{i, j\}}; \quad (3.6)$$

- *Local Markov property (LMP)*: Associated with $g = (V, \mathcal{E})$, this property is satisfied if for any vertex $i \in V$

$$X_i \perp\!\!\!\perp \mathbf{x}_{V \setminus cl(i)} \mid \mathbf{x}_{ne(i)} \text{ where } cl(i) = i \cup ne(i); \quad (3.7)$$

- *Global Markov property (GMP)*: Associated with $g = (V, \mathcal{E})$, for any triple (A, B, C) of disjoint subsets of the vertex set V , where B separates A from C , then $\mathbf{x}_A$ and $\mathbf{x}_C$ are independent given $\mathbf{x}_B$ and can be written as

$$\mathbf{x}_A \perp\!\!\!\perp \mathbf{x}_C \mid \mathbf{x}_B.$$

The following proposition gives the relationship between the Markov properties given above.

Proposition 3.2.2 *For any undirected graph $g = (V, \mathcal{E})$ and any probability distribution on $\mathbf{x}$,*

$$(GMP) \Rightarrow (LMP) \Rightarrow (PMP). \quad (3.8)$$

The above proposition states that the global Markov property (GMP) implies the local Markov property (LMP), which in sequence implies the pairwise Markov property (PMP). That is why the global Markov property (GMP) is conventionally considered as the Markov property (for proof of the proposition, see (Lauritzen 1996)). In the following theorem, it is shown that all the Markov properties stated above are equivalent.

Theorem 3.2.2 *(Pearl and Paz (Lauritzen 1996)). For any undirected graph $g = (V, \mathcal{E})$ and any probability distribution on $\mathbf{x}$ such that for all disjoint subsets A, B, C and D of V it holds that*

$$\text{if } A \perp\!\!\!\perp B \mid (C \cup D) \text{ and } A \perp\!\!\!\perp C \mid (B \cup D) \Rightarrow A \perp\!\!\!\perp (B \cup C) \mid D, \quad (3.9)$$

then

$$(GMP) \Leftrightarrow (LMP) \Leftrightarrow (PMP). \quad (3.10)$$

The above theorem means that under Condition (3.9) the global Markov property (GMP), the local Markov property (LMP) and the pairwise Markov property (PMP) are all equivalent. For proof of the proposition, see (Lauritzen 1996).

3.3 Gaussian graphical models

The Gaussian graphical model was originally introduced by Dempster (1972), who expressed a multivariate normal distribution in terms of the graphical interaction model. This model is also known as the covariance selection model (Dempster 1972) and concentration graph model (Drton & Perlman 2004).

Using undirected graphical models, we now characterise the multivariate Gaussian distribution, which is also known as the Gaussian random field, where an emphasis is given to the structure of the underlying graph.

Definition 3.3.1 (Multivariate Gaussian distribution (Whittaker 2009)) *A continuous v -dimensional random vector $\mathbf{X}$ has the multivariate Gaussian distribution if and only if its density has the following form*

$$f(\mathbf{x}) = (2\pi)^{-v/2} \det(\mathbf{\Sigma})^{-1/2} \exp\left(\frac{-1}{2}(\mathbf{x} - \boldsymbol{\mu})' \mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad (3.11)$$

where $\boldsymbol{\mu}$ is a $v \times 1$ mean vector and $\mathbf{\Sigma}$ is a $v \times v$ variance-covariance (positive definite) matrix with the entries $\sigma_{ij} = \text{Cov}(X_i, X_j)$, respectively, given by

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_v \end{pmatrix}, \quad \text{and} \quad \mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1v} \\ \vdots & & \vdots \\ \sigma_{v1} & \dots & \sigma_{vv} \end{pmatrix}.$$

This distribution is also denoted by $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$.

The interest here is in the inverse of the variance-covariance matrix $\mathbf{K} = \Sigma^{-1}$, which is referred to as the concentration/precision matrix, written as

$$\mathbf{K} = \begin{pmatrix} k_{11} & \dots & k_{1v} \\ \vdots & & \vdots \\ k_{v1} & \dots & k_{vv} \end{pmatrix}.$$

Now, rearranging terms inside the exponential brackets given in (3.11) can also be expressed in the form of the following exponential family:

$$f(\mathbf{x}) = \exp\left(\gamma + \Theta' \mathbf{x} - \frac{1}{2} \mathbf{x}' \mathbf{K} \mathbf{x}\right), \quad (3.12)$$

where $\Theta = \mathbf{K}\mu$ and γ is a normalising constant. Expression given in (3.12) can also be written as:

$$f(\mathbf{x}) = \exp\left(\gamma + \sum \Theta_i x_i - \frac{1}{2} \sum \sum k_{ij} x_i x_j\right). \quad (3.13)$$

From (3.13), using the factorising criterion given in (3.5), we can see that

$$(i, j) \notin \mathcal{E} \Leftrightarrow k_{ij} = 0 \Leftrightarrow X_i \perp\!\!\!\perp X_j \mid \mathbf{X}_{V \setminus \{i, j\}}.$$

Now we formally define Gaussian graphical models as follows:

Definition 3.3.2 *The v variate multivariate normal distribution $N(\mathbf{0}, \Sigma)$ is considered to be graphical with respect to an undirected graph $g = (V, \mathcal{E})$ if it specifies that the corresponding entries in the concentration matrix are zero i.e. $k_{ij} = 0$ whenever $\{i, j\} \notin \mathcal{E}$.*

In the next section, we explain the likelihood estimation for the Gaussian graphical models. For a complete description of these, see Lauritzen (1996),

Edwards (2000) and Højsgaard et al. (2012).

3.3.1 The likelihood

In this section, a review is given of the likelihood of decomposable Gaussian graphical models summarising the results of Lauritzen (1996) and Højsgaard et al. (2012).

Suppose that we have a sample of n observations $(X_1, \dots, X_n)$ of X following a multivariate normal distribution, i.e. $X \sim N(\mu, \Sigma)$. Also suppose that $\Omega = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$ represents the empirical covariance matrix and $\bar{X}$ is the sample mean. The log likelihood function based on the sample is expressed as (Højsgaard et al. 2012) :

$$L(K, \mu) = \frac{n}{2} \ln \det(K) - \frac{n}{2} \text{tr}(K\Omega) - \frac{n}{2} (\bar{X} - \mu)' K (\bar{X} - \mu).$$

The above expression is maximised by $\hat{\mu} = \bar{X}$ which makes $\frac{n}{2} (\bar{X} - \mu)' K (\bar{X} - \mu) = 0$, provided that K is fixed. Then the log-likelihood becomes

$$L(K) = \frac{n}{2} \ln \det(K) - \frac{n}{2} \text{tr}(K\Omega).$$

The above likelihood depends on the data only through the elements ω_{ij} of Ω for which $k_{ij} \neq 0$ in that $\text{tr}(K\Omega) = \sum_{i,j} k_{ij} \omega_{ij}$.

Now assume a graph $g = (V, \mathcal{E})$ can be decomposed as the triple (A, B, C) , where A, B and C are disjoint subsets of the vertex set V such that $A \cup B \cup C = V$. Also assume that B separates A from C , which implies that B has blocked all the paths that go from A to C . Let K_{AUB} , Σ_{AUB} and Ω_{AUB} be sub-matrices of K , Σ and Ω respectively, all having the dimension $|A| \times |B|$. A matrix $[K_{AUB}]^\Gamma$ corresponding to K_{AUB} , is a $v \times v$ symmetric matrix defined in (3.2) (the same applies to Σ_{AUB}

and $\Omega_{A \cup B}$).

let $\hat{K}_{A \cup B} = \Omega_{A \cup B}$, $\hat{K}_{B \cup C} = \Omega_{B \cup C}$ and $\hat{K}_B = \Omega_B$ refer to the maximum likelihood estimates (MLE) of K in the marginal model belonging to the given graph g . Then the maximum likelihood estimates of K can be written as:

$$\hat{K} = n \left[(\Omega_{A \cup B})^{-1} \right]^\Gamma + n \left[(\Omega_{B \cup C})^{-1} \right]^\Gamma - n \left[(\Omega_B)^{-1} \right]^\Gamma. \quad (3.14)$$

Under the following condition, the matrix inversions of the above estimator can be applied:

Condition 3.3.1 *The sample size $n > \max\{|A| + |B|, |B| + |C|\}$.*

In general, the related graph g can be decomposed into many cliques and separators. Suppose that the graph g is decomposable and let C denote the set of cliques and S the set of separators. Then, the general formula of (3.14) is (Lauritzen 1996):

$$\hat{K} = n \left\{ \sum_{C \in C} [(\Omega_C)^{-1}]^\Gamma - \sum_{S \in S} [(\Omega_S)^{-1}]^\Gamma \right\}. \quad (3.15)$$

Clearly, under the following condition, the above estimator exists:

Condition 3.3.2 *The sample size $n > \max_{C \in C} \{|C|\}$, i.e. the sample size should be greater than the length of any clique among all the C cliques.*

The maximum likelihood estimation is illustrated in the following example.

Example 3.3.1 *To describe the Gaussian graphical model in a real-world application, consider the example given in Whittaker (2009) where the dataset consists of examination marks of 88 students in five subjects: Mechanics, Vectors, Algebra, Analysis and Statistics.*

The empirical variance-covariance matrix for the marks data (in the same order of variables) is as follows:

$$\mathbf{\Omega} = \begin{pmatrix} 302.21 & 125.60 & 100.32 & 105.11 & 116.16 \\ 125.60 & 170.88 & 84.19 & 93.60 & 97.89 \\ 100.32 & 84.19 & 111.60 & 110.84 & 120.49 \\ 105.11 & 93.60 & 110.84 & 217.88 & 153.77 \\ 116.16 & 97.89 & 120.49 & 153.77 & 294.37 \end{pmatrix}.$$

Suppose that Mechanics and Vectors are independent of Analysis and Statistics given Algebra, i.e. $(\text{Mechanics}, \text{Vectors}) \perp\!\!\!\perp (\text{Analysis}, \text{Statistics}) \mid \text{Algebra}$ as shown in Figure 3.3.

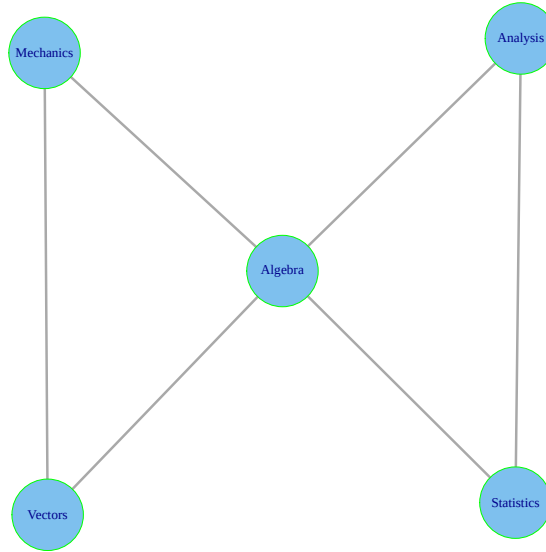


Figure 3.3: A graph based on the statement “Mechanics and Vectors are independent of Analysis and Statistics given Algebra” for the examination marks data.

The maximum likelihood estimate for the concentration matrix based on the graph

given in Figure 3.3 is obtained as follows:

$$\begin{aligned}
 \hat{\mathbf{K}} &= 88 \times \left\{ \begin{pmatrix} 0.005 & -0.002 & -0.003 & 0.000 & 0.000 \\ -0.002 & 0.010 & -0.006 & 0.000 & 0.000 \\ -0.003 & -0.006 & 0.016 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \end{pmatrix} + \begin{pmatrix} 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.022 & -0.008 & -0.005 \\ 0.000 & 0.000 & -0.008 & 0.010 & -0.002 \\ 0.000 & 0.000 & -0.005 & -0.002 & 0.007 \end{pmatrix} \right. \\
 &\quad \left. - \begin{pmatrix} 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.009 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \end{pmatrix} \right\} \\
 &= 88 \times \begin{pmatrix} 0.005 & -0.002 & -0.003 & 0.000 & 0.000 \\ -0.002 & 0.010 & -0.006 & 0.000 & 0.000 \\ -0.003 & -0.006 & 0.029 & -0.008 & -0.005 \\ 0.000 & 0.000 & -0.008 & 0.010 & -0.002 \\ 0.000 & 0.000 & -0.005 & -0.002 & 0.007 \end{pmatrix} = \begin{pmatrix} 0.440 & -0.176 & -0.264 & 0.000 & 0.000 \\ -0.176 & 0.880 & -0.528 & 0.000 & 0.000 \\ -0.264 & -0.528 & 2.552 & -0.704 & -0.440 \\ 0.000 & 0.000 & -0.704 & 0.880 & -0.176 \\ 0.000 & 0.000 & -0.440 & -0.176 & 0.616 \end{pmatrix}.
 \end{aligned}$$

3.3.2 Model selection for undirected graph

Gaussian graphical models selection is concerned with determining a graph g from a large number of well-fitting models in a given dataset. This involves dealing with issues, very similar to regression modelling framework, that arise in searching for an appropriate model. In the graphical context, model selection involves finding the proper layout of zeros in the concentration matrix that correspond to the conditional independence structure of the variables.

There are several methods available in the literature that are used for model selection in Gaussian graphical models. Among the common methods are the score-based search methods and the neighborhood selection method.

3.3.2.1 Score-based search

The score-based search methods involve model searching via a goodness-of-fit score in the entire space of the potential models. The examples of scores used are the Akaike information criterion (AIC) (Akaike 1974) and the Bayesian information criterion (BIC) (Schwarz 1978) that assess the extent of fitness between the data and the corresponding graph. The expressions for AIC and BIC are given as:

$$AIC = -2 \ln L_M(j) + \lambda \dim(M), \quad (3.16)$$

where $L_M(j)$ is the likelihood of the given model M , λ (which is usually equal 2 (Højsgaard et al. 2012)) is the rate of trade-off between the fit and parameters and $\dim(M)$ is the number of free parameters.

Changing the value of λ by $\ln(n)$ gives BIC written as

$$BIC = -2 \ln L_M(j) + \ln(n) \dim(M). \quad (3.17)$$

These scores help to find an improved fit, usually by one step change in the current graph. Stepwise procedures are used for this purpose and these consist of forward selection or backward elimination of an interaction term in the current model (Edwards 2000). These procedures usually adopt a greedy search by determining a neighbourhood structure. In forward selection procedures, an edge (interaction term) is added to the current model and the effect of this addition is considered for ignoring or adding the edge to the graph. The backward elimination procedures go in the reverse. Care should be taken while adding or removing an edge, which might result in a non-decomposable graph (Wermuth 1976). This means that the search should be restricted to the space of

decomposable models only, which results in a significant reduction in the search space. The process of adding/eliminating the edges is stopped with a graph that has the highest score among all the neighbouring graphs.

3.3.2.2 Neighborhood selection method

A simpler method for estimating the graph structure is proposed by Meinshausen & Bühlmann (2006), who estimate only the components k_{ij} of K that are nonzero. They implement this by fitting a Lasso regression using each variable as the response and the rest as predictors. However, in such a case, there is no guarantee to obtain a symmetric concentration matrix.

In order to address this problem, Meinshausen & Bühlmann (2006) suggest using an AND or OR rule. In the AND rule, an edge is selected if the coefficients of X_i on X_j and X_j on X_i are both nonzero; in the OR rule, an edge is selected if either of the coefficients is nonzero.

3.4 Conclusion

This chapter has given a detailed background about graphical models and their properties, with specific reference to Gaussian graphical models. The undirected graph theory and the concept of conditional independence and Markov properties have been discussed in detail and clarified with relevant examples. The associated notations used throughout the thesis have also been provided in this chapter. In addition, the concept of model selection in Gaussian graphical models has been introduced, with particular focus on the score-based search and neighbourhood selection methods.

CHAPTER 4

Graphical least square estimation

4.1 Introduction

As discussed in detail in Chapter 3, graphical models are an effective tool to graphically represent and investigate the conditional independencies among multivariate random variables, which is very helpful in extracting the hidden structures in the data. Inspired by the idea of graphical models, this chapter explores the possibility of using these models to estimate regression parameters in an unbiased way when the number of observations is less than the number of predictors ($n < v$) in the model.

When $n < v$, many methods, such as ridge (Hoerl & Kennard 1970), Lars (Efron et al. 2004) and Lasso (Tibshirani 2011) (discussed in Chapter 2), have been utilised to solve the linear regression model given in (2.1). These methods provide biased estimates, which, in some cases, may prevent consistent variable selection (Zhang 2010) and over-shrink the final coefficients (James & Radchenko 2009, Radchenko et al. 2011).

For regression analysis with $n > v$, OLS actually uses the saturated model assumption, i.e. any pair of covariates are correlated conditional on the other covariates. When $n < v$, the full saturated least squares cannot be used in that the sample covariance matrix cannot be inverted. A good alternative is the unsaturated model, which is achieved by adding constraints to the conditional correlation structure for the covariates. However, all existing methods do not

consider the conditional correlation and thus they may not be able to identify the true model when $n < v$. This might be very important in many areas such as genetic studies where the identification of gene networks (conditional correlations among variables) is needed.

This chapter develops an unbiased estimator called Graphical Least Squares Estimate (GLSE) which can provide a much better solution than all other existing methods. The proposed method has a strong connection with graphical models and takes into account the conditional correlation between covariates discussed in detail in Chapter 3. Therefore, it could possibly include the causal interpretation between response and covariates and identify the true regression model. Under certain mild assumptions, the estimator produced by the new method is unbiased.

Another advantage of GLSE is that it considers many more candidate models than existing methods do. For the regression model with covariates $(X_1, \dots, X_v)$, the estimates produced under the situation where all X_j are independent should be different from the regression estimates produced under the situation when some X_j are correlated. Considering the detailed conditional correlation structure among the predictors, this new method actually searches in a much larger space of potential regression models than all existing methods do. With v covariates, existing methods carry out model selection from 2^v different sets of models (at most). In contrast, if all covariates should be included in the model, the GLSE could search from $2^{v(v-1)/2}$ different models, since there will be many different models according to the different conditional correlation structures among the covariates X_j .

This chapter is structured as follows. The main methodology of GLSE and its properties are presented in Section 4.2, then the model selection for the underlying graph structure is provided in Section 4.3. The simulation studies

are given in Section 4.4 and the data analysis is presented in Section 4.5. The chapter ends with a conclusion given in Section 4.6.

4.2 The Graphical Least Squares estimates

4.2.1 The basic idea of Graphical Least Squares estimates

This section presents the basic idea of the GLSE for the regression model (2.1). It is assumed that all response values and covariate values have been centered and the covariates (predictors) can be both continuous and binary/categorical.

Suppose that the set V can be partitioned into disjoint sets A , B and C . Also, assume that the variables $(X_1, \dots, X_v)$ are factorising according to a given (known) decomposable graph g , which consists of two cliques $A \cup B$ and $B \cup C$ and a separator B . Then the proposed estimator for β associated with this graph g is given as

$$\hat{\beta}^s = \left[\left\{ (ssd_{A \cup B})^{-1} \right\}^\Gamma + \left\{ (ssd_{B \cup C})^{-1} \right\}^\Gamma - \left\{ (ssd_B)^{-1} \right\}^\Gamma \right] \mathbf{x}' \mathbf{y}. \quad (4.1)$$

For the existence of the matrix inversions in the above formula, Condition 3.3.1 ($n > \max\{|A| + |B|, |B| + |C|\}$) must hold. Note that Condition 3.3.1 is weaker than the standard OLS requirement $n > v = |A| + |B| + |C|$, which states that the given sample size n should be greater than the total number of covariates v .

As we only consider quantitative variables and qualitative variables, the above estimator can easily be calculated. On the other hand, if we have a categorical variable X_1 with $m > 2$ categories, this can still be practically carried out. In this case, we only need to recode X_1 into $m-1$ 0-1 dummy variables $X_{11}, \dots, X_{1,m-1}$ and the new graph vertex set becomes $\{X_{11}, \dots, X_{1,m-1}, X_2, \dots, X_v\}$.

To discuss the unbiasedness property of the above estimator, we first intro-

duce some definitions. For an estimator $\hat{\beta}$, we introduce three different types of unbiasedness regarding the regression model given in (2.1).

1. Unbiased estimate (UE): If $\mathbb{E}_{\mathbf{x},y}(\hat{\beta}) = \beta$, the estimator $\hat{\beta}$ is said to be unbiased.
2. Strong conditional unbiased estimate (SUE): If $\mathbb{E}_y(\hat{\beta}|\mathbf{x}) = \beta$ then the estimator $\hat{\beta}$ is said to be strongly conditionally unbiased.
3. Weak conditional unbiased estimate (WUE): If $\mathbb{E}_{\mathbf{x}_{V \setminus B},y}(\hat{\beta}|\mathbf{x}_B) = \beta$, then the estimator $\hat{\beta}$ is said to be weakly conditionally unbiased, where B denotes a proper nonempty subset of V .

It is easy to show that $\text{SUE} \Rightarrow \text{WUE} \Rightarrow \text{UE}$.

Write the covariate matrix as $\mathbf{x} = (\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C)$, where $\mathbf{x}_A$, $\mathbf{x}_B$ and $\mathbf{x}_C$ are the disjoint subsets of the variables set that constitute the covariate matrix. Then under the following condition, we can show that (4.1) is an unbiased estimate.

Condition 4.2.1 *The sets A , B and C form a decomposition of a given graph g , with B as the separator. The sets $\mathbf{x}_A$ and $\mathbf{x}_C$ are conditionally independent on $\mathbf{x}_B$, such that*

$$\begin{aligned}\mathbf{x}_C &= \mathbf{x}_B \cdot \mathbf{r}_{BC} + \xi_C, \quad \mathbb{E}(\xi_C) = \mathbf{0}, \\ \mathbf{x}_A &= \mathbf{x}_B \cdot \mathbf{r}_{BA} + \xi_A, \quad \mathbb{E}(\xi_A) = \mathbf{0},\end{aligned}\tag{4.2}$$

where $\mathbf{r}_{BC}$ and $\mathbf{r}_{AB}$ are an $b \times c$ and $a \times b$ dimensional constant matrix where $a = |A|$, $b = |B|$ and $c = |C|$.

Considering the above two conditions, the following theorem can be introduced to show that the proposed GLSE is unbiased.

Theorem 4.2.1 *Under Condition 3.3.1 and Condition 4.2.1, the estimator in (4.1) is unbiased and even weakly conditionally unbiased.*

Proof: We have

$$\begin{aligned}
 & E(\hat{\beta}^s | \mathbf{x}_B) \\
 &= E\left(\left\{[(ssd_{A \cup B})^{-1}]^\Gamma + [(ssd_{B \cup C})^{-1}]^\Gamma - [(ssd_B)^{-1}]^\Gamma\right\} \mathbf{x}' \boldsymbol{\beta} | \mathbf{x}_B\right) \\
 &= E\left[\left[\begin{pmatrix} \mathbf{I}_A & \mathbf{0} & (ssd_{A \cup B})^{-1} \mathbf{x}'_{A \cup B} \mathbf{x}_C \\ \mathbf{0} & \mathbf{I}_B & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ (ssd_{B \cup C})^{-1} \mathbf{x}'_{B \cup C} \mathbf{x}_A & \mathbf{I}_B & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_C \end{pmatrix} \right. \\
 &\quad \left. - \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ (ssd_B)^{-1} \mathbf{x}'_B \mathbf{x}_A & \mathbf{I}_B & (ssd_B)^{-1} \mathbf{x}'_B \mathbf{x}_C \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \right] \mathbf{x}_B \boldsymbol{\beta}.
 \end{aligned} \tag{4.3}$$

Condition 4.2.1 gives

$$E(\mathbf{x}_C | \mathbf{x}_B) = \mathbf{x}_B \mathbf{r}_{BC} = \mathbf{x}_{A \cup B} \begin{pmatrix} \mathbf{0} \\ \mathbf{r}_{BC} \end{pmatrix}$$

and we then have

$$\begin{aligned}
 & E\left[E\left[(ssd_{A \cup B})^{-1} \mathbf{x}'_{A \cup B} \mathbf{x}_C - \begin{pmatrix} \mathbf{0} \\ (ssd_B)^{-1} \mathbf{x}_B^T \mathbf{x}_C \end{pmatrix} \middle| \mathbf{x}_B, \mathbf{x}_A\right] \middle| \mathbf{x}_B\right] \\
 &= (ssd_{A \cup B})^{-1} \mathbf{x}'_{A \cup B} \mathbf{x}_{A \cup B} \begin{pmatrix} \mathbf{0} \\ \mathbf{r}_{BC} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ (ssd_B)^{-1} \mathbf{x}_B^T \mathbf{x}_B \mathbf{r}_{BC} \end{pmatrix} \\
 &= \begin{pmatrix} \mathbf{0} \\ \mathbf{r}_{BC} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \mathbf{r}_{BC} \end{pmatrix} = \mathbf{0}.
 \end{aligned} \tag{4.4}$$

Similarly we can prove

$$E\left[E\left[(ssd_{B \cup C})^{-1} \mathbf{x}'_{B \cup C} \mathbf{x}_A - \begin{pmatrix} (ssd_B)^{-1} \mathbf{x}'_B \mathbf{x}_A \\ \mathbf{0} \end{pmatrix} \middle| \mathbf{x}_B, \mathbf{x}_C\right] \middle| \mathbf{x}_B\right] = \mathbf{0}. \tag{4.5}$$

Substituting (5.3) and (4.5) back to (4.3), we have

$$E(\hat{\beta}^s | \mathbf{x}_B) = \beta$$

and therefore $E(\hat{\beta}^s) = \beta$. □

Remark 4.2.1 *The OLS estimate can be obtained only when $n > v$ is SUE, i.e. $E(\hat{\beta}^{ols} | \mathbf{x}) = \beta$. On the other hand, when $n < v$, all existing methods such as ridge regression, Lasso, or Lars are biased, i.e. $E(\hat{\beta}^{other}) \neq \beta$.*

Remark 4.2.2 *Note that Condition 4.2.1 will be satisfied in many cases. For example, if $\mathbf{x}_i = (X_{i1}, \dots, X_{iv})$ follows a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \Sigma)$ and the distribution of $\mathbf{x}_i$ factorises according to a decomposable graph $g = (V, \mathcal{E}_g)$, with cliques $A \cup B$ and $B \cup C$ and separator B , then Condition 4.2.1 implies the conditional independence of $\mathbf{x}_A$ and $\mathbf{x}_C$ given $\mathbf{x}_B$ and a linear relationship between $\mathbf{x}_A$ (and $\mathbf{x}_C$) and $\mathbf{x}_B$. Given a different underlying graph structure for the covariates, the GLSE will be different. It follows that the GLSE strongly depends on the graph structure that has been searched for. In practice, we can search for the best underlying graph structure, as discussed in Section 4.3. Moreover, even when $\mathbf{x}_i$ does not follow the multivariate normal distribution, the method still works, as shown in Scenario 2 in Section 4.4.*

Remark 4.2.3 *The linear assumption in Condition 4.2.1 will not limit the applicability of the method since variable transformation can always be used to achieve a linear relationship, if the covariates are quantitative variables. In addition, as any non-linear relation can be approximated via a polynomial, the nonlinear dependence on X can be viewed as linear dependence on $X, X^2, \dots$.*

4.2.2 The general form of Graphical Least Squares Estimates

In general, the associated graph g may be decomposed into many cliques or separators. Suppose that the graph g is decomposable and let C denote the set of cliques and S denote the set of separators. The general formula of the GLSE is

$$\hat{\beta}^g = \left[\sum_{C \in C} [(ssd_C)^{-1}]^T - \sum_{S \in S} [(ssd_S)^{-1}]^T \right] \mathbf{x}' \mathbf{y}. \quad (4.6)$$

Obviously, under Condition 3.3.2 given in Chapter 3, the GLSE exists.

Under the following condition, we can also show that $\hat{\beta}^g$ is unbiased.

Condition 4.2.2 Write the cliques and separators of g in the perfect ordering, as $C_1, \dots, C_q$ and $S_2, \dots, S_q$, such that

$$\begin{aligned} \mathbf{x}_{C_1 \setminus S_2} &= \mathbf{x}_{S_2} \cdot \mathbf{r}_{S_2, C_1 \setminus S_2} + \xi_1, \quad E(\xi_1) = \mathbf{0}, \\ \mathbf{x}_{C_k \setminus S_k} &= \mathbf{x}_{S_k} \cdot \mathbf{r}_{S_k, C_k \setminus S_k} + \xi_k, \quad E(\xi_k) = \mathbf{0}, \quad k = 2, \dots, q, \end{aligned}$$

where $\mathbf{r}_{S_k, C_k \setminus S_k}$ are constant matrices with dimensions $|S_k| \times (|C_k| - |S_k|)$;

Theorem 4.2.2 Under Condition 3.3.2 and Condition 4.2.2, the estimator in (4.6) is unbiased,

$$E(\hat{\beta}^g) = \beta$$

Proof: This theorem can be proved by mathematical induction, which is given below. The GLSE in (4.6) is associated with a graph g with cliques $C_1, \dots, C_q$ as a perfect sequence and separators $S_2, \dots, S_q$. For $\kappa = 2, \dots, q$, define the index

sets

$$\begin{aligned}
 H_\kappa &= \cup_{i=1}^\kappa C_i, \\
 H_\kappa^c &:= H_{\kappa+1} \setminus H_\kappa = C_{\kappa+1} \setminus S_{\kappa+1} \\
 J_\kappa &= H_\kappa \setminus S_{\kappa+1}
 \end{aligned} \tag{4.7}$$

and define the matrix

$$\mathbf{K}_\kappa = \sum_{i=1}^\kappa [(ssd_{C_i})^{-1}]^\Gamma - \sum_{i=2}^{\kappa-1} [(ssd_{S_i})^{-1}]^\Gamma. \tag{4.8}$$

Note that the elements in the i th rows and elements in the j th columns of $\mathbf{K}_\kappa$, for $i, j \in H_\kappa^c$, are all 0s. Therefore if we write

$$\mathbf{x}'_{H_{\kappa+1}} \mathbf{x}_{H_{\kappa+1}} = \begin{pmatrix} ssd_{H_\kappa} & \mathbf{x}'_{H_\kappa} \mathbf{x}_{H_\kappa^c} \\ \mathbf{x}'_{H_\kappa^c} \mathbf{x}_{H_\kappa} & ssd_{H_\kappa^c} \end{pmatrix} \tag{4.9}$$

we then have

$$\mathbf{K}_\kappa \cdot [\mathbf{x}'_{H_{\kappa+1}} \mathbf{x}_{H_{\kappa+1}}]^\Gamma = \mathbf{K}_\kappa \cdot [\mathbf{x}'_{H_\kappa} \mathbf{x}_{H_\kappa}]^\Gamma + \mathbf{K}_\kappa \cdot [\mathbf{x}'_{H_\kappa}]^\Gamma [\mathbf{x}_{H_\kappa^c}]^\Gamma. \tag{4.10}$$

From (4.6) and $E(\epsilon) = \mathbf{0}$ we know $E(\hat{\beta}^g) = E[\mathbf{K}_q \cdot \mathbf{x}' \mathbf{x} \cdot \beta]$. Therefore, we only need to show that

$$E[\mathbf{K}_q \cdot \mathbf{x}' \mathbf{x}] = \mathbf{I}. \tag{4.11}$$

Now we prove (4.11) via mathematical induction. Suppose, for $\kappa < q$, that

$$E[\mathbf{K}_\kappa \cdot [\mathbf{x}'_{H_\kappa} \mathbf{x}_{H_\kappa}]^\Gamma] = [\mathbf{I}_{H_\kappa}]^\Gamma. \tag{4.12}$$

Then from (4.8), (4.10) and the mathematical induction assumption (4.12) we have that, for $\kappa + 1$,

$$\begin{aligned}
& E \left[\mathbf{K}_{\kappa+1} \cdot [\mathbf{x}'_{H_{\kappa+1}} \mathbf{x}_{H_{\kappa+1}}]^\Gamma \right] \\
&= E \left[\left\{ \mathbf{K}_\kappa + \left[(ssd_{C_{\kappa+1}})^{-1} \right]^\Gamma - \left[(ssd_{S_{\kappa+1}})^{-1} \right]^\Gamma \right\} \cdot [\mathbf{x}'_{H_{\kappa+1}} \mathbf{x}_{H_{\kappa+1}}]^\Gamma \right] \\
&= E \left[\mathbf{K}_\kappa [\mathbf{x}'_{H_{\kappa+1}} \mathbf{x}_{H_{\kappa+1}}]^\Gamma \right] + E \left[\left\{ \left[(ssd_{C_{\kappa+1}})^{-1} \right]^\Gamma - \left[(ssd_{S_{\kappa+1}})^{-1} \right]^\Gamma \right\} \cdot [\mathbf{x}'_{H_{\kappa+1}} \mathbf{x}_{H_{\kappa+1}}]^\Gamma \right] \\
&= [\mathbf{I}_{H_\kappa}]^\Gamma + E \left[\mathbf{K}_\kappa \cdot [\mathbf{x}'_{H_\kappa}]^\Gamma [\mathbf{x}_{H_\kappa}^c]^\Gamma \right] + E \left[\left\{ \left[(ssd_{C_{\kappa+1}})^{-1} \right]^\Gamma - \left[(ssd_{S_{\kappa+1}})^{-1} \right]^\Gamma \right\} \cdot [\mathbf{x}'_{H_{\kappa+1}} \mathbf{x}_{H_{\kappa+1}}]^\Gamma \right].
\end{aligned} \tag{4.13}$$

Condition 4.2.2 gives $\mathbf{x}_{H_\kappa}^c = \mathbf{x}_{S_{\kappa+1}} \cdot \mathbf{r}_{S_{\kappa+1}, C_{\kappa+1} \setminus S_{\kappa+1}} + \boldsymbol{\xi}_{H_\kappa}^c$, for some $|S_{\kappa+1} \times (C_{\kappa+1} - S_{\kappa+1})|$ dimensional matrix $\mathbf{r}_{S_{\kappa+1}, C_{\kappa+1} \setminus S_{\kappa+1}}$ and some independent zero-mean residual matrix $\boldsymbol{\xi}_{H_\kappa}^c$. We can write $\mathbf{r}$ in terms of its elements via

$$\mathbf{r}_{S_{\kappa+1}, C_{\kappa+1} \setminus S_{\kappa+1}} = (r_{ij})_{i \in S_{\kappa+1}, j \in C_{\kappa+1} \setminus S_{\kappa+1}}. \tag{4.14}$$

If we define the expansion for this $\mathbf{r}_{S_{\kappa+1}, C_{\kappa+1} \setminus S_{\kappa+1}}$ to an $V \times V$ matrix $[\mathbf{r}_{S_{\kappa+1}, C_{\kappa+1} \setminus S_{\kappa+1}}]^\Gamma$ as

$$([\mathbf{r}_{S_{\kappa+1}, C_{\kappa+1} \setminus S_{\kappa+1}}]^\Gamma)_{ij} = \begin{cases} r_{ij} & \text{if } i \in S_{\kappa+1}, j \in C_{\kappa+1} \setminus S_{\kappa+1} \\ 0 & \text{otherwise,} \end{cases} \tag{4.15}$$

then Condition 4.2.2 further implies

$$[\mathbf{x}_{H_\kappa}^c]^\Gamma = [\mathbf{x}_{H_\kappa}]^\Gamma \cdot [\mathbf{r}_{S_{\kappa+1}, C_{\kappa+1} \setminus S_{\kappa+1}}]^\Gamma + [\boldsymbol{\xi}_{H_\kappa}^c]^\Gamma. \tag{4.16}$$

Therefore

$$E \left(\mathbf{K}_\kappa \cdot [\mathbf{x}'_{H_\kappa}]^\Gamma [\mathbf{x}_{H_\kappa}^c]^\Gamma \right) = [\mathbf{I}_{H_\kappa}]^\Gamma [\mathbf{r}_{S_{\kappa+1}, C_{\kappa+1} \setminus S_{\kappa+1}}]^\Gamma. \tag{4.17}$$

We also have the following terms,

$$\begin{aligned} & \mathbb{E} \left[\left[(ssd_{C_{k+1}})^{-1} \right]^\Gamma \cdot [\mathbf{x}'_{H_{k+1}} \mathbf{x}_{H_{k+1}}]^\Gamma \right] \\ &= [\mathbf{I}_{C_{k+1}}]^\Gamma + \mathbb{E} \left[\left[(ssd_{C_{k+1}})^{-1} \right]^\Gamma \cdot [\mathbf{x}'_{C_{k+1}}]^\Gamma [\mathbf{x}_{J_k}]^\Gamma \right] \end{aligned} \quad (4.18)$$

and

$$\begin{aligned} & -\mathbb{E} \left[\left[(ssd_{S_{k+1}})^{-1} \right]^\Gamma \cdot [\mathbf{x}'_{H_{k+1}} \mathbf{x}_{H_{k+1}}]^\Gamma \right] \\ &= -[\mathbf{I}_{S_{k+1}}]^\Gamma - \mathbb{E} \left[\left[(ssd_{S_{k+1}})^{-1} \right]^\Gamma \cdot [\mathbf{x}'_{S_{k+1}}]^\Gamma [\mathbf{x}_{J_k}]^\Gamma \right] - \mathbb{E} \left[\left[(ssd_{S_{k+1}})^{-1} \right]^\Gamma \cdot [\mathbf{x}'_{S_{k+1}}]^\Gamma [\mathbf{x}_{H_k^c}]^\Gamma \right] \\ &= -[\mathbf{I}_{S_{k+1}}]^\Gamma - \mathbb{E} \left[\left[(ssd_{S_{k+1}})^{-1} \right]^\Gamma \cdot [\mathbf{x}'_{S_{k+1}}]^\Gamma [\mathbf{x}_{J_k}]^\Gamma \right] - [\mathbf{I}_{S_{k+1}}]^\Gamma [\mathbf{r}_{S_{k+1}, C_{k+1} \setminus S_{k+1}}]^\Gamma. \end{aligned} \quad (4.19)$$

Therefore, substituting (4.17), (4.18) and (4.19) with (4.13) we know that

$$\begin{aligned} & \mathbb{E} \left[\mathbf{K}_{k+1} \cdot [\mathbf{x}'_{H_{k+1}} \mathbf{x}_{H_{k+1}}]^\Gamma \right] \\ &= [\mathbf{I}_{H_k}]^\Gamma + [\mathbf{I}_{H_k}]^\Gamma [\mathbf{r}_{S_{k+1}, C_{k+1} \setminus S_{k+1}}]^\Gamma + [\mathbf{I}_{C_{k+1}}]^\Gamma + \mathbb{E} \left[\left[(ssd_{C_{k+1}})^{-1} \right]^\Gamma \cdot [\mathbf{x}'_{C_{k+1}}]^\Gamma [\mathbf{x}_{J_k}]^\Gamma \right] \\ & \quad - [\mathbf{I}_{S_{k+1}}]^\Gamma - \mathbb{E} \left[\left[(ssd_{S_{k+1}})^{-1} \right]^\Gamma \cdot [\mathbf{x}'_{S_{k+1}}]^\Gamma [\mathbf{x}_{J_k}]^\Gamma \right] - [\mathbf{I}_{S_{k+1}}]^\Gamma [\mathbf{r}_{S_{k+1}, C_{k+1} \setminus S_{k+1}}]^\Gamma \\ &= [\mathbf{I}_{H_k}]^\Gamma + [\mathbf{I}_{C_{k+1}}]^\Gamma - [\mathbf{I}_{S_{k+1}}]^\Gamma = [\mathbf{I}_{H_{k+1}}]^\Gamma, \end{aligned} \quad (4.20)$$

where the result uses $[\mathbf{I}_{H_k}]^\Gamma [\mathbf{r}_{S_{k+1}, C_{k+1} \setminus S_{k+1}}]^\Gamma - [\mathbf{I}_{S_{k+1}}]^\Gamma [\mathbf{r}_{S_{k+1}, C_{k+1} \setminus S_{k+1}}]^\Gamma = \mathbf{0}$ and

$$\mathbb{E} \left[\left[(ssd_{C_{k+1}})^{-1} \right]^\Gamma \cdot [\mathbf{x}'_{C_{k+1}}]^\Gamma [\mathbf{x}_{J_k}]^\Gamma \right] - \mathbb{E} \left[\left[(ssd_{S_{k+1}})^{-1} \right]^\Gamma \cdot [\mathbf{x}'_{S_{k+1}}]^\Gamma [\mathbf{x}_{J_k}]^\Gamma \right] = \mathbf{0}. \quad (4.21)$$

Equation (4.21) is equivalent to the fact that from Condition 4.2.2, we also have

$\mathbf{x}_{J_k} = \mathbf{x}_{S_{k+1}} \cdot \mathbf{q}_{S_{k+1}, J_k} + \boldsymbol{\eta}_{J_k}$, for some $|S_{k+1}| \times |J_k|$ dimensional matrix $\mathbf{q}_{S_{k+1}, J_k}$ with $\mathbb{E}(\boldsymbol{\eta}_{J_k}) = \mathbf{0}$ and further

$$[\mathbf{x}_{J_k}]^\Gamma = [\mathbf{x}_{S_{k+1}}]^\Gamma \cdot [\mathbf{q}_{S_{k+1}, J_k}]^\Gamma + [\boldsymbol{\eta}_{J_k}]^\Gamma = [\mathbf{x}_{C_{k+1}}]^\Gamma \cdot [\mathbf{q}_{S_{k+1}, J_k}]^\Gamma + [\boldsymbol{\eta}_{J_k}]^\Gamma. \quad (4.22)$$

The theorem is then proved by mathematical induction with the fact that for $\kappa = 1$, $E \left[\mathbf{K}_2 \cdot [\mathbf{x}'_{H_2} \mathbf{x}_{H_2}]^\Gamma \right] = [\mathbf{I}_{H_2}]^\Gamma$ (from the results in 4.2.1 for a graph with only two cliques) and $[\mathbf{x}'_{H_q} \mathbf{x}_{H_q}]^\Gamma = \mathbf{x}' \mathbf{x}$.

□

4.2.3 Variance estimation for the GLSE

It has been shown that the proposed GLSE is unbiased in the given context. The next step is to estimate the exact or approximate variance of the GLSE. Now to get an expression for the variance of the GLSE, first we denote

$$\mathbf{K} = \sum_{C \in \mathcal{C}} [(\text{ssd}_C)^{-1}]^\Gamma - \sum_{S \in \mathcal{S}} [(\text{ssd}_S)^{-1}]^\Gamma,$$

then we can show that under Condition 4.2.2, the expression for the variance of the GLSE can be written as

$$\text{var}(\hat{\beta}^g) = E((\hat{\beta}^g - \beta)^2) = E \left[(\mathbf{K} \mathbf{x}' \mathbf{x} \beta - \beta) (\mathbf{K} \mathbf{x}' \mathbf{x} \beta - \beta)' + \sigma^2 (\mathbf{K} \mathbf{x}' \mathbf{x} \mathbf{K}') \right].$$

However, it is not easy to derive a simple estimator for the above variance of $\hat{\beta}^g$. Therefore, we need to consider other methods to obtain an approximate estimate for the variance of the estimator. The bootstrap estimate is a good alternative. In bootstrapping, we repeatedly sample from the given population the same number of observations as the population size using simple random sampling with replacement (detailed information about bootstrapping could be found in Efron & Tibshirani (1994)). The estimated graph is then used to estimate the regression coefficients for each bootstrap sample, and after the required number of bootstraps, the standard error of each estimate is obtained.

When we use bootstrap to estimate $\text{var}(\hat{\beta}^g)$, care should be taken in that a bootstrapped sample can be used only if the number of distinct observations in

the resampled data is greater than the number of variables within the maximum clique. This may happen only when the sample size is too small. To avoid this problem, one possibility is to restrict the cliques in order not to exceed a certain limit, which will lead to a highly sparse graph.

We can also use the following estimator for the variance,

$$\widehat{\text{var}}(\hat{\beta}^g) = (\mathbf{K}\mathbf{x}'\mathbf{x}\hat{\beta}^g - \hat{\beta}^g)(\mathbf{K}\mathbf{x}'\mathbf{x}\hat{\beta}^g - \hat{\beta}^g)' + \hat{\sigma}^2(\mathbf{K}\mathbf{x}'\mathbf{x}\mathbf{K}') \quad (4.23)$$

since any random variable is always an unbiased point estimate to its own mean. We will show that (4.23) does not work well for data with a small sample size, but when the sample size is large, the estimator works well.

4.2.4 Discussion on the unbiasedness and Condition 4.2.2

Condition 4.2.2 will be reasonable in many real applications. When there are many predictors, some of them are more likely to be conditionally independent given a subset of variables. For example, in bioinformatics studies, the covariates may be a pool of genes where some genes may be related due to an intermediate gene. There has been a vast pool of literature researching graphical models that can identify the conditional dependency structure for many random variables, even if the sample size is very small (see for examples Dobra et al. 2004, Krämer et al. 2009, Talluri & Shete 2014).

Condition 4.2.2 allows the possibility to study the detailed conditional independence structure under a regression framework, which helps solve the challenges when $n < v$. Traditional analyses did not consider the conditional correlation between covariates. It is acceptable to disregard this and apply the saturated model (any pair of covariates is conditionally correlated given the remaining) when the sample size is large enough. However, when $n < v$, certain

conditions have to be added to achieve a reasonable result. Traditional methods use other constraints or penalty terms and trade off the bias and variance. This distorts the real conditional correlation between covariates. That is why existing methods for $n < v$ may provide spurious correlations between covariates and predictors and cannot identify the real important causal predictors. For instance, Sun & Zhang (2010) showed that Lasso has a 50% chance of selecting a non-significant variable to predict y , thus leading to a spurious correlation.

4.3 Graphical model selection

In practice, the underlying graph structure g is unknown. We need to develop an algorithm to find which graph g is the best for the data. This is another advantage of GLSE since the underlying graph structure could lead to certain causal interpretations, which cannot be done via any existing methods.

Some covariance model selection methods were developed under the Bayesian framework, such as Jones et al. (2005) and Dai (2008). We here proposed a new one for graphical model selection under a regression framework. It is natural to consider the following criterion for model selection: the best graph is given by minimising a target function $\mathbb{T}(\boldsymbol{\beta}, g, \lambda_g)$:

$$(\hat{\boldsymbol{\beta}}, \hat{g}, \hat{\lambda}_g) = \arg \min_{\boldsymbol{\beta}, g \in \mathcal{G}, \lambda_g} \mathbb{T}(\boldsymbol{\beta}, g, \lambda_g) \quad (4.24)$$

$$\mathbb{T}(\boldsymbol{\beta}, g, \lambda_g) = \|\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\|^2 + \lambda_g |\mathcal{E}_g| \quad (4.25)$$

where $\mathcal{G}$ is the set of all possible graphs and $|\mathcal{E}_g|$ is the number of edges in graph g . This criterion aims to search for the estimate and graph structure, corresponding to the minimum of the sum of square errors ($\|\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\|^2$) plus a penalty term ($\lambda_g |\mathcal{E}_g|$).

The sum of square errors will decrease if more edges are added to the estimated graph. It should be noted, however, that the decrease in the sum of square errors will come to an end at a certain point even if more edges are added to the graph. In practice, we may fix the value of λ_g and only search for β and g . If one prefers a very sparse graph, then a large value of λ_g should be chosen; however, when preferring a more saturated graph, a small value of λ_g should be chosen.

We recommend that more moderate graphs be considered in practice because these will allow more information on the correlation structure. To show this, a graphical comparison is performed by generating a dataset similar to the one simulated in Scenario 1 in Section 4.4. For convenience, the graph generated is allowed to add only up to 30 edges shown on the y-axis in Figure 4.1. The figure shows that a smaller value of λ_g will give a smaller sum of square errors, i.e. more accurate prediction. Therefore, we may simply set $\lambda_g = 0$ and ignore the penalty term, except that in some situations a sparse graph is preferred. In the analysis of real data, $\lambda_g = 0$ is taken. Although one may argue that an overfitting problem might occur when $\lambda_g = 0$, the proposed method will not be affected by overfitting in that the cliques may grow to size n at most. This guarantees that the model does not become complicated to the extent of overfitting and will be much smaller than the entire space of the regression model. It also was not noticed in any of the simulation scenarios or data analyses. For instance, in the data analysis, while the total number of edges in the full graph should be 19,900, only 1,253 edges were obtained in the estimated graph using Algorithm 2.

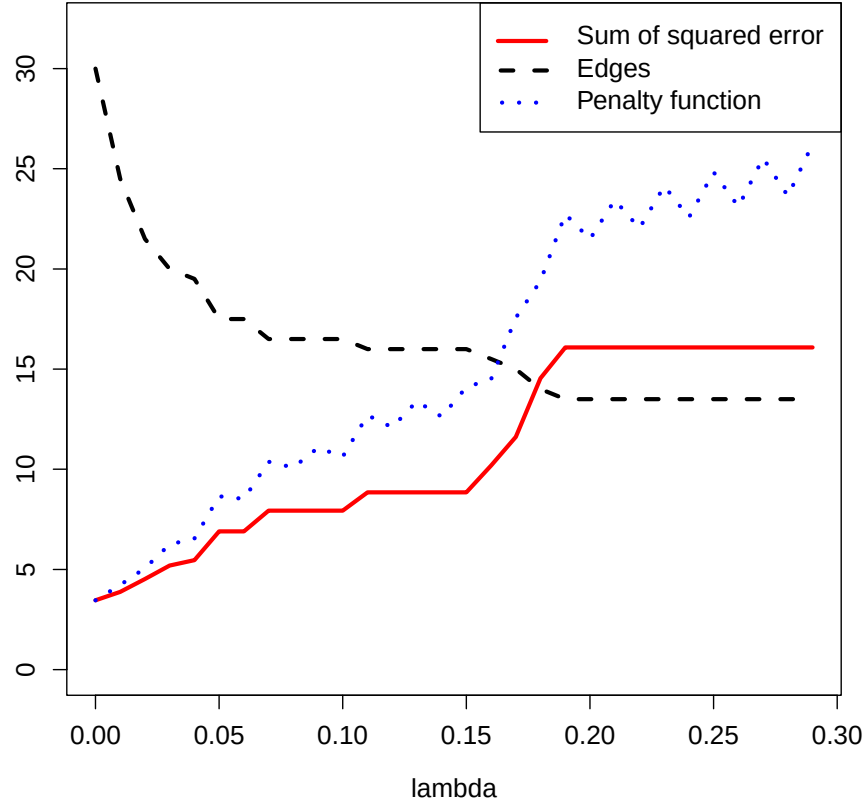


Figure 4.1: The effect of choosing different values of λ_g for the penalty term. Solid line: sum of squared errors; dotted line: penalty function; dashed line: the number of edges corresponding to the graph giving the minimum target function value.

When searching in the whole graph space $\mathcal{G}$, we can use a stepwise selection procedure. The method considers adding/deleting edges one by one to/from the current graph. When an edge under consideration is not in the current graph, it will be added if the addition makes an improvement in terms of the predetermined criteria; otherwise it will not be added. When an edge under consideration is in the current graph, it will be deleted if the deletion makes an improvement in terms of the predetermined criteria; otherwise it will not be deleted. Similar graphical model selection algorithms for multivariate Gaussian models have been proposed by Jones et al. (2005).

Algorithm 2 Pseudocode of the GLSE graph selection

-
- 1: Start graph $g = (V, \mathcal{E})$, which can be an empty (or a given decomposable) graph such that $n > \max_{C \in \mathcal{C}} |C|$.
 - 2: Generate all possible graphs, g_i , such that there is only one edge difference between g_i and the current graph g . All such g_i are decomposable and $n > \max_{C \in \mathcal{C}} |C|$.
 - 3: Find the graph g_i^* and the associated $\hat{\beta}^g$ such that g_i^* minimises the target function $\mathbb{T}(\cdot)$ (given in (4.25)).
 - 4: Go to step 2 with the selected graph g_i^* and iterate until the best one is found.
 - 5: Output g and $\hat{\beta}^g$.
-

Moreover, with the availability of modern parallel computational technology, the above algorithm can still handle large dimensions of variables in the datasets with less time (see, for example, the related discussion at the end of the Scenario 1). This means that more efficient computational tools can proportionally handle higher-dimensional variables. However, this was not the main focus of the current study and might be tackled in the future.

As an example, we used this algorithm to find the underlying graph structure for the dataset generated in Scenario 1 in Section 4.4. The true graph structure is given in Figure 4.3. The two selected graphs for different sample sizes are given in Figures 4.2.

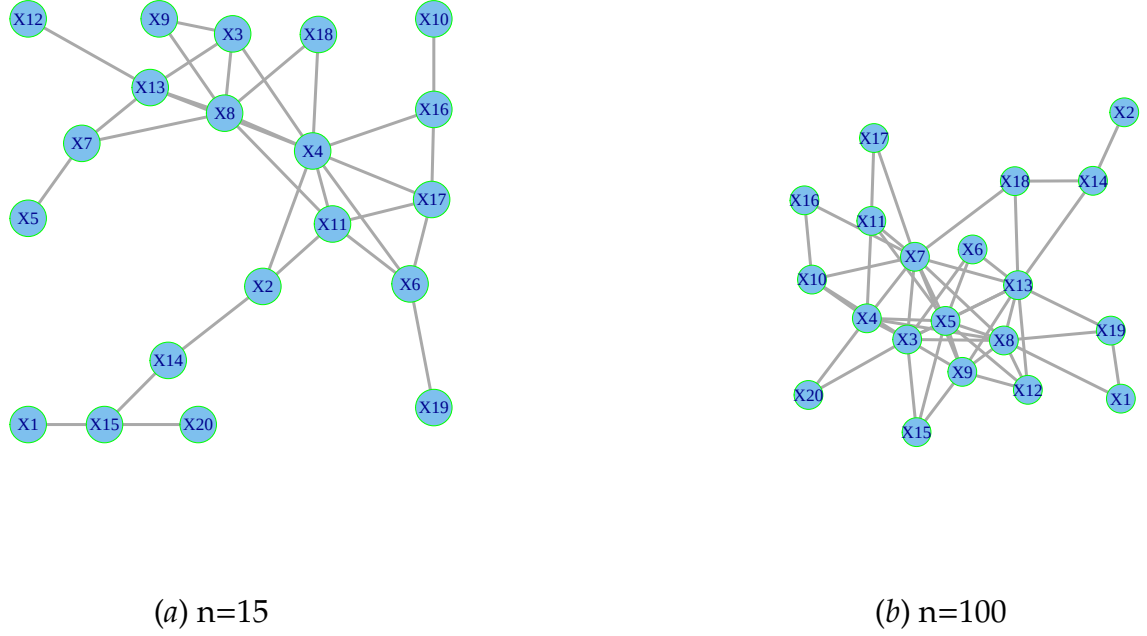


Figure 4.2: *The selected graphs.*

In the original graph there are 13 edges having conditional correlations that are greater than the absolute value 0.1. The graph chosen by the algorithm for $n = 15$, in Figure 4.2 (a), detected 9 of these edges in the original graph.

Also, the chosen graph detected 4 of the edges with conditional correlations less than or equal to the absolute value 0.1 in the original graph. The graph for $n = 100$ detected 11 edges having conditional correlations greater than absolute value 0.1 and 7 edges with conditional correlations less than or equal to the absolute value 0.1 of the original graph. It follows that by increasing the sample size, most of the original edges can be detected by using the algorithm.

The model selection step is the most challenging part for the GLSE method, since once a good graph structure is identified, the computation of the GLSE for regression parameters is straightforward. The computational cost is quite heavy to search the whole graph space, but step 2 of Algorithm 2 can be improved significantly via parallel computation.

One could stop the algorithm (starting from an empty graph) once it converges in terms of the sum of squared errors. This might significantly reduce the amount of computations required. However, this may produce a locally optimised graph.

An improved algorithm that can dramatically reduce the search space for finding the optimal graph is introduced in the next chapter.

4.4 Simulation studies

Detailed simulation studies are presented here to show that the GLSE can give unbiased estimation of regression coefficients compared to other existing methods such as Lasso, ridge and Lars. Six different simulation scenarios are considered here. The first simulation model considers the case of multivariate normally distributed covariates. The second simulation is similar to the first but has a larger number of variables to take the generalisation of the results into consideration, while the third simulation considers the case where the covariates are not multivariate normally distributed. The fourth is exactly similar to the third but, again, the number of variables is also larger. The fifth simulation is about the case of having a binary variable in the predictors set. In the final simulation study, we investigate the performance of the variance estimators obtained via the formula (4.23), Monte Carlo estimation and bootstrapping. The results of the GLSE on the simulated models are compared with those of Lasso, Lars and ridge regression, where cross-validation is used to obtain the penalty parameters for ridge and Lasso and find the number of steps required by Lars to select the optimal model. Note that the GLSE $\hat{\beta}^g$ depends on the graph g . Therefore, for Monte Carlo simulation studies, we use the same graph structure (known graph) and the algorithm (unknown graph) to find the GLSE for all replicated

simulations.

It is worth noting that the proposed method was not compared to grouped Lasso despite the close resemblance between them. This is due to the fact that the latter method either shrinks or selects the whole group of variables, unlike the GLSE method, which selects individual variables from the groups, becoming closer in function to Lasso. Furthermore, the grouped Lasso, contrary to the GLSE method, requires that the groups should be specified prior to implementing the method. For instance, in genes expression analysis, genes can be classified or grouped based on the biological pathway they belong to (Huang et al. 2012). In addition, even if we try to group the variables based on the cliques of any known graph, this will prove ineffective since the cliques will have shared variables (separators).

4.4.1 Scenario 1

The data in this scenario are simulated using model (2.1) with a total number of variables $v = 20$ following a multivariate normal distribution with mean $\mathbf{0}$ and variance covariance matrix Σ . The concentration matrix is associated with a graph that consists of 20 vertices, shown in Figure 4.3, where the conditional correlation is indicated by the number next to each line. Then the response $\mathbf{y}$ is generated from model (2.1), where the random errors ϵ are normally distributed with mean 0 and standard error $\sigma = 0.5$. We simulate 15 samples, on the 20 variables. The true parameter values of β are given in Table 4.1.

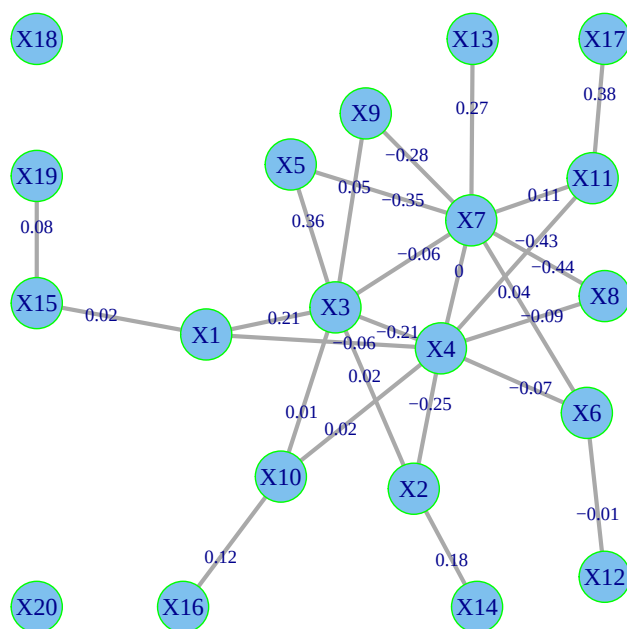


Figure 4.3: *Graph structure for covariates under Scenario 1.*

Table 4.1: Results from Scenario 1 (with known and unknown graphs), $n = 15, v = 20$.

	β	Ridge		Lasso		Lars		GLSE (with known graph)		$\hat{\beta}$	GLSE (with unknown graph)		$\hat{\beta}$	GLSE (with unknown graph)	
		$\hat{\beta}_{ridge}$	SD ^a	$\hat{\beta}_{lasso}$	SD ^a	$\hat{\beta}_{lars}$	SD ^a	SE (4.23) ^b	SD ^a	SE (4.23) ^b	SE (4.23) ^b	SD ^a	SE (4.23) ^b	SE (4.23) ^b	SE (Bootstrap) ^c
X ₁	0.5	0.256	0.594	0.129	0.472	0.050	0.300	0.456	3.571	1.605	1.612	0.279	0.326	1.081	1.293
X ₂	1	0.369	0.704	0.214	0.630	0.060	0.289	1.053	4.492	1.899	2.417	0.600	0.393	1.312	1.180
X ₃	0.6	0.092	0.578	0.063	0.470	0.026	0.253	0.472	4.208	1.839	1.594	0.218	0.390	1.489	1.280
X ₄	-1	-0.361	0.368	-0.391	0.498	-0.252	0.423	-1.009	2.734	1.178	0.919	-0.488	0.238	1.325	1.230
X ₅	0.7	0.393	0.337	0.423	0.510	0.323	0.471	0.698	2.380	0.924	0.816	0.538	0.220	0.884	0.779
X ₆	-1	-0.567	0.763	-0.317	0.658	-0.145	0.410	-0.975	4.541	1.863	2.042	-0.658	0.420	1.303	1.233
X ₇	1.1	0.399	0.332	0.526	0.582	0.439	0.562	1.098	2.911	1.130	1.174	0.565	0.196	0.599	0.959
X ₈	-0.6	-0.173	0.355	-0.219	0.448	-0.121	0.303	-0.602	2.358	0.965	1.011	-0.257	0.194	0.624	1.160
X ₉	1.5	0.885	0.627	0.720	0.844	0.445	0.718	1.585	3.965	1.479	1.361	1.159	0.348	1.081	1.226
X ₁₀	0.5	0.258	0.778	0.097	0.474	0.019	0.215	0.408	4.739	2.076	1.136	0.306	0.425	1.324	1.345
X ₁₁	0.8	0.243	0.287	0.274	0.426	0.214	0.364	0.808	2.233	0.888	1.016	0.363	0.194	0.802	0.950
X ₁₂	-0.8	-0.466	0.818	-0.218	0.588	-0.062	0.278	-0.807	4.623	1.912	1.625	-0.593	0.494	1.647	1.296
X ₁₃	1.8	0.822	0.475	1.051	0.813	0.849	0.771	1.809	2.873	0.796	0.818	1.102	0.212	0.734	1.032
X ₁₄	0.9	0.461	0.617	0.280	0.545	0.154	0.397	0.970	3.679	1.567	1.315	0.631	0.399	1.617	1.299
X ₁₅	-0.6	-0.295	0.571	-0.179	0.501	-0.102	0.321	-0.588	3.191	1.303	2.647	-0.452	0.316	0.969	1.307
X ₁₆	1.3	0.654	0.851	0.271	0.624	0.110	0.377	1.395	4.926	2.000	1.635	0.954	0.518	2.037	1.357
X ₁₇	0.1	-0.040	0.364	-0.008	0.381	-0.009	0.286	0.120	2.567	1.109	0.947	0.019	0.230	0.700	1.356
X ₁₈	1.7	0.882	0.934	0.467	0.825	0.218	0.611	1.599	5.135	2.094	1.985	1.241	0.509	1.754	1.258
X ₁₉	-2	-1.027	0.678	-0.890	0.930	-0.592	0.818	-1.935	3.914	1.441	1.084	-1.414	0.384	1.045	1.337
X ₂₀	0.8	0.426	0.814	0.167	0.559	0.079	0.305	0.662	4.429	1.868	1.873	0.465	0.484	1.457	1.296
Euclidean distance ^d		0.564		0.677		0.824		0.065				0.382			

^a SD: the Monte Carlo standard error for the 500 replicates.^b SE(4.23): the mean of the 500 standard error estimates using the error equation given in 4.23.^c SE(bootstrap): the mean of the 500 bootstrap standard error estimates.^d Euclidean distance : $\sqrt{\frac{\sum (\hat{\beta} - \hat{\beta})^2}{v}}$

Table 4.1 gives coefficient estimates, Monte Carlo estimates of standard errors and Euclidean distance for the ridge, Lasso, Lars and the proposed GLSE method. It also shows the standard error estimates using the bootstrap method for the GLSE method. It can be seen that GLSE outperforms all other three methods in the coefficients estimates and has the smallest Euclidean distance. Although the Monte Carlo estimates of standard errors for ridge, Lasso and Lars are smaller than those of the GLSE, the new method being an unbiased estimation may be preferred. We can also see from the results that the mean of the 500 bootstrap standard error estimates is very close to the Monte Carlo standard deviation for the 500 replicated estimates, which implies that bootstrap estimation for the standard error performs well. However, we also noted that the standard error estimate based on (4.23) does not work well for such a small sample size. We will show in Scenario 4.4.6 that (4.23) gives good estimates when n is large. In the case when the graph is estimated, the results given in the last 4 columns of Table 4.1 lead to the same conclusion. Here also, the GLSE outperforms the others in coefficient estimation. Another advantage of GLSE is its ability to identify the right sign ($\pm$) for the coefficient estimate, which is extremely important in portfolio optimisation, while the other methods lack accuracy in this regard.

The computational cost for the proposed algorithm is not heavy with modern parallel computing technology. Both serial and parallel computing time is considered when the true graph is unknown. It is noted that on a machine with 8GB of memory and a 2.7GHz processor, the time taken is approximately 10 hours. When parallel processing is used, with 60 cores, the computational time is reduced to approximately 9 minutes. For the case when the true graph structure is known, parallel computing is not needed as the time taken serially (using one processor) is just a few seconds.

4.4.2 Scenario 2

In order to confirm that the results from Scenario 4.4.1 generalise to other scenarios, a total of $n = 25$ observations are generated on $v = 30$ variables following a multivariate normal distribution with mean $\mathbf{0}$ and variance covariance matrix Σ . The graph representing the concentration matrix is shown in Figure 4.4, where the number next to each line refers to the conditional correlation. Then the response y is generated in a similar fashion to Scenario 4.4.1. The true parameter values of β are given in Table 4.2.

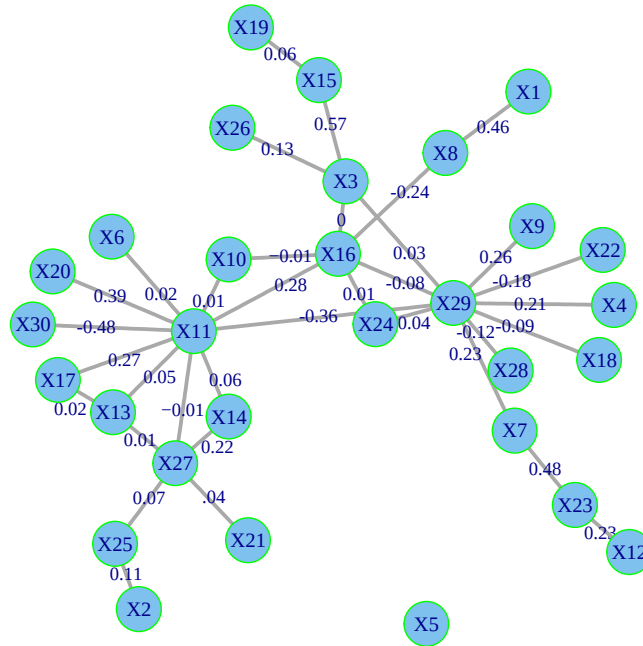


Figure 4.4: Graph structure for covariates under Scenario 2.

Table 4.2: Results from Scenario 2 (with known and unknown graphs), $n = 25$, $v = 30$.

	β	Ridge		Lasso		Lars		GLSE (with known graph)		$\hat{\beta}$	GLSE (with unknown graph)		$\hat{\beta}$	GLSE (bootstrap) ^c		GLSE (with unknown graph)		$\hat{\beta}$	GLSE (bootstrap) ^c	
		$\hat{\beta}_{\text{ridge}}$	SD ^a	$\hat{\beta}_{\text{lasso}}$	SD ^a	$\hat{\beta}_{\text{lars}}$	SD ^a	SE (4.23) ^b	SD ^a	SE (Bootstrap) ^c	SE (4.23) ^b	SD ^a	SE (Bootstrap) ^c	SE (4.23) ^b	SD ^a	SE (Bootstrap) ^c	SE (4.23) ^b	SD ^a	SE (Bootstrap) ^c	SE (Bootstrap) ^c
X ₁	0.5	0.3146	0.3687	0.2401	0.4236	0.1401	0.3302	0.538	2.4669	1.0664	1.0951	0.328	0.190	0.444	0.479	0.328	0.190	0.444	0.479	0.479
X ₂	1	0.6807	0.3809	0.644	0.5641	0.4711	0.5422	0.9319	2.4802	1.0562	1.0584	0.726	0.175	0.425	0.473	0.726	0.175	0.425	0.473	0.473
X ₃	0.6	0.4598	0.3077	0.3416	0.4355	0.2489	0.4177	0.55	2.3786	0.9713	0.9886	0.442	0.180	0.411	0.591	0.442	0.180	0.411	0.591	0.591
X ₄	-1	-0.5452	0.6216	-0.3999	0.6616	-0.2252	0.5286	-1.0141	3.762	1.5262	1.459	-0.545	0.310	0.800	0.936	-0.545	0.310	0.800	0.936	0.936
X ₅	0.7	0.4435	0.4274	0.3442	0.5056	0.2114	0.4295	0.7671	2.6991	1.1465	1.1048	0.518	0.211	0.498	0.607	0.518	0.211	0.498	0.607	0.607
X ₆	-1	-0.6027	0.5732	-0.4596	0.6637	-0.2709	0.5199	-1.0342	3.3225	1.3996	1.4863	-0.702	0.266	0.613	0.875	-0.702	0.266	0.613	0.875	0.875
X ₇	1.1	0.7501	0.6347	0.5885	0.8203	0.3792	0.7222	1.1342	3.7553	1.3757	1.5049	0.826	0.321	0.748	0.884	0.826	0.321	0.748	0.884	0.884
X ₈	-0.6	-0.271	0.4938	-0.1866	0.5068	-0.0856	0.3399	-0.7151	3.1674	1.381	1.3502	-0.275	0.239	0.595	0.832	-0.275	0.239	0.595	0.832	0.832
X ₉	1.5	1.0867	0.5934	0.9887	0.9109	0.7556	0.8656	1.4941	3.5637	1.3278	1.2949	1.246	0.303	0.674	1.127	1.246	0.303	0.674	1.127	1.127
X ₁₀	0.5	0.2388	0.6742	0.1624	0.7393	0.0603	0.5622	0.5373	4.2448	1.8296	1.8546	0.250	0.335	0.826	1.079	0.250	0.335	0.826	1.079	1.079
X ₁₁	0.8	0.5136	0.4485	0.3552	0.5579	0.2288	0.4744	0.8142	3.1418	1.3272	1.6204	0.511	0.208	0.484	0.779	0.511	0.208	0.484	0.779	0.779
X ₁₂	-0.8	-0.5496	0.5072	-0.4445	0.6063	-0.3401	0.5854	-0.7227	2.9628	1.2844	1.3092	-0.622	0.244	0.619	0.612	-0.622	0.244	0.619	0.612	0.612
X ₁₃	1.8	1.1713	0.695	1.1057	1.0315	0.8034	0.9913	1.8479	4.601	2.0063	1.9981	1.287	0.346	0.817	1.221	1.287	0.346	0.817	1.221	1.221
X ₁₄	0.9	0.645	0.5727	0.5394	0.7424	0.3518	0.6323	1.0165	3.8746	1.7223	1.7012	0.756	0.292	0.654	0.982	0.756	0.292	0.654	0.982	0.982
X ₁₅	-0.6	-0.2024	0.3302	-0.1272	0.3389	-0.0391	0.2165	-0.5895	2.3549	1.0214	0.9959	-0.234	0.179	0.429	0.566	-0.234	0.179	0.429	0.566	0.566
X ₁₆	1.3	0.8061	0.4311	0.7076	0.6699	0.4958	0.6024	1.4287	3.0836	1.1328	1.3171	0.874	0.219	0.487	0.608	0.874	0.219	0.487	0.608	0.608
X ₁₇	0.1	-0.035	0.7016	-0.0431	0.7243	-0.0468	0.5256	0.0237	4.6052	2.0616	2.0302	-0.046	0.352	0.883	1.220	-0.046	0.352	0.883	1.220	1.220
X ₁₈	1.7	1.1851	0.6443	1.125	0.9586	0.7747	0.8754	1.6647	3.6701	1.2822	1.3265	1.283	0.311	0.712	1.203	1.283	0.311	0.712	1.203	1.203
X ₁₉	-2	-1.2621	0.5698	-1.2762	0.8356	-0.9659	0.825	-2.0889	3.4923	1.2768	1.2422	-1.395	0.254	0.633	0.675	-1.395	0.254	0.633	0.675	0.675
X ₂₀	0.8	0.4746	0.5315	0.3239	0.5846	0.1737	0.4519	0.9002	3.4898	1.5765	1.4888	0.496	0.271	0.655	0.872	0.496	0.271	0.655	0.872	0.872
X ₂₁	0.8	0.5132	0.2875	0.4764	0.4207	0.3429	0.3882	0.8384	1.8636	0.8082	0.7399	0.546	0.132	0.319	0.359	0.546	0.132	0.319	0.359	0.359
X ₂₂	-0.3	-0.0722	0.6704	-0.1027	0.706	-0.041	0.5306	-0.3081	4.0943	1.654	1.6248	-0.108	0.337	0.810	1.326	-0.108	0.337	0.810	1.326	1.326
X ₂₃	2.5	1.6119	0.8364	1.5544	1.2147	1.1378	1.1728	2.6093	4.8785	1.81	1.9418	1.756	0.370	0.918	1.119	1.756	0.370	0.918	1.119	1.119
X ₂₄	0.2	0.1306	0.7137	0.0437	0.6762	0.0245	0.5254	0.2242	4.2386	1.6533	1.7673	0.166	0.343	0.871	1.261	0.166	0.343	0.871	1.261	1.261
X ₂₅	-0.4	-0.2614	0.2208	-0.1995	0.2924	-0.1209	0.25	-0.4344	1.8015	0.7904	0.7549	-0.288	0.109	0.250	0.296	-0.288	0.109	0.250	0.296	0.296
X ₂₆	1.9	1.1062	0.6942	0.9471	0.9454	0.6089	0.8267	1.9105	4.2637	1.7059	1.6543	1.159	0.308	0.731	1.170	1.159	0.308	0.731	1.170	1.170
X ₂₇	-0.05	-0.1587	0.3607	-0.1088	0.3789	-0.0583	0.2465	-0.023	3.2308	1.5223	1.5183	-0.128	0.192	0.472	0.499	-0.128	0.192	0.472	0.499	0.499
X ₂₈	1	0.7562	0.6461	0.5872	0.8342	0.4047	0.7218	0.9872	3.9346	1.4849	1.4844	0.862	0.326	0.757	1.172	0.862	0.326	0.757	1.172	1.172
X ₂₉	-2.2	-1.6236	0.6443	-1.9552	1.0673	-1.7665	1.0919	-2.2469	4.0603	1.5778	1.8321	-1.962	0.355	0.777	1.069	-1.962	0.355	0.777	1.069	1.069
X ₃₀	0.6	0.3233	0.6816	0.2082	0.709	0.1118	0.5568	0.5437	4.4135	1.9366	1.9761	0.327	0.342	0.840	1.237	0.327	0.342	0.840	1.237	1.237
Euclidean distance ^d		0.410		0.478		0.665		0.062												

^a SD: the Monte Carlo standard error for the 500 replicates.^b SE(4.23): the mean of the 500 standard error estimates using the error equation given in 4.23.^c SE(bootstrap): the mean of the 500 bootstrap standard error estimates.^d Euclidean distance: $\sqrt{\frac{\sum(\hat{\beta} - \hat{\beta}^2)}{v}}$

The results from this scenario are given in Table 4.2 and are obtained from a total of 500 realisations of the simulated data. They show that GLSE is better than ridge, Lasso and Lars in estimating the regression coefficients when the true graph structure is known and that it has the smallest Euclidean distance. In addition, the Monte Carlo estimates of the error show that the results are comparable to those of ridge, Lasso and Lars. The bootstrap estimates, obtained in the same fashion as in Scenario 1, are close to the Monte Carlo estimates of standard error. However, it can also be noted that the standard error estimate based on (4.23) is not effective for such a small sample size. In Scenario 4.4.6, we prove that (4.23) gives good estimates when n is large.

In the case of unknown graph structure, the GLSE is better than the rest in estimating regression coefficients and, consequently, has the smallest Euclidean distance, as per the results displayed in the last four columns of Table 4.2. For standard error estimates, the same conclusions could be drawn as those of the known graph structure.

4.4.3 Scenario 3

In this scenario, we also consider $n = 15$ and $v = 20$. The predictor variables set, $\mathbf{x}$, has the same graph structure as that in Scenario 1, but it does not follow a multivariate normal distribution. The idea of this simulation is to find out how the method performs in more general cases. The covariates are generated from a set of regression models with uniform random errors, which is given in detail in Appendix B.1. Then the response $\mathbf{y}$ is generated from model (2.1), where the random errors ϵ are normally distributed with mean 0 and standard error $\sigma = 0.5$. The true values of $\boldsymbol{\beta}$ are given in Table 4.3.

Table 4.3: Results from Scenario 2 (with known and unknown graphs), $n = 15, v = 20$.

	β	Ridge		Lasso		Lars		GLSE (with known graph)		$\hat{\beta}$	GLSE (with unknown graph)	
		$\hat{\beta}_{ridge}$	SD ^a	$\hat{\beta}_{lasso}$	SD ^a	$\hat{\beta}_{lars}$	SD ^a	SE (4.23) ^b	SD ^a	SE (4.23) ^b	SD ^a	SE (Bootstrap) ^c
X_1	0.5	-0.001	0.569	0.031	0.531	-0.032	0.341	4.124	1.697	1.717	0.028	0.083
X_2	1	0.714	0.575	0.625	0.834	0.450	0.744	4.232	1.694	1.989	0.911	0.084
X_3	0.6	0.339	0.376	0.287	0.515	0.228	0.448	4.177	1.692	2.511	0.434	0.065
X_4	-1	-0.240	0.283	-0.249	0.444	-0.126	0.331	3.533	1.483	1.881	-0.352	0.023
X_5	0.7	0.052	0.514	0.078	0.509	-0.033	0.331	4.050	1.547	1.423	0.140	0.079
X_6	-1	-0.165	0.541	-0.181	0.543	-0.049	0.261	4.303	1.894	1.988	-0.325	0.082
X_7	1.1	0.451	0.316	0.617	0.660	0.498	0.589	4.028	1.627	1.655	0.558	0.043
X_8	-0.6	-0.393	0.413	-0.272	0.575	-0.173	0.463	4.311	1.844	2.362	-0.438	0.079
X_9	1.5	0.439	0.563	0.466	0.758	0.190	0.489	4.139	1.520	1.992	0.541	0.069
X_{10}	0.5	0.124	0.483	0.162	0.579	0.108	0.344	4.292	1.863	2.453	0.216	0.072
X_{11}	0.8	0.334	0.495	0.248	0.568	0.125	0.375	4.556	1.994	1.930	0.465	0.081
X_{12}	-0.8	-0.416	0.674	-0.290	0.658	-0.123	0.411	4.131	1.817	2.249	-0.596	0.104
X_{13}	1.8	1.112	0.651	0.999	1.039	0.658	0.909	4.367	1.720	1.841	1.438	0.109
X_{14}	0.9	0.546	0.672	0.418	0.753	0.260	0.596	4.118	1.737	1.463	0.843	0.103
X_{15}	-0.6	-0.468	0.643	-0.317	0.677	-0.146	0.434	4.194	1.820	1.215	-0.595	0.112
X_{16}	1.3	0.752	0.698	0.592	0.845	0.283	0.596	4.181	1.799	1.354	0.967	0.097
X_{17}	0.1	0.145	0.621	0.091	0.541	0.058	0.318	4.074	1.826	1.823	0.126	0.093
X_{18}	1.7	0.976	0.662	0.851	0.930	0.508	0.764	3.865	1.603	0.843	1.311	0.089
X_{19}	-2	-1.123	0.818	-0.978	1.067	-0.580	0.840	4.377	1.769	1.498	-1.512	0.108
X_{20}	0.8	0.466	0.507	0.414	0.572	0.258	0.473	2.733	1.167	1.076	0.608	0.048
Euclidean distance ^d		0.571		0.614		0.805					0.424	

^a SD: the Monte Carlo standard error for the 500 replicates.^b SE(4.23): the mean of the 500 standard error estimates using the error equation given in 4.23.^c SE(bootstrap): the mean of the 500 bootstrap standard error estimates.^d Euclidean distance : $\sqrt{\frac{\sum(\beta - \hat{\beta})^2}{v}}$

The results based on Scenario 2 are given in Table 4.3 and obtained from the average of running 500 realisations of the dataset. It is obvious from the table that the GLSE is giving the closest estimates to the true coefficients compared to the other methods. This implies that even if the distribution of predictor variables is not a multivariate normal one, unbiased estimates for the regression coefficients can still be obtained by the GLSE. Similarly, as in Scenario 1, we also find that the bootstrap estimate for standard errors works well, but (4.23) does not provide good estimates. In addition, the Euclidean distance is smallest for the GLSE method.

In the case when the graph is estimated via Algorithm 1, the results reveal that GLSE still outperforms the rest of the methods and the Monte Carlo estimates of standard errors for GLSE are close to or sometimes smaller than those of ridge, Lasso and Lars.

4.4.4 Scenario 4

This scenario is similar to the previous one except in that the number of observations here is increased to $n = 25$ and the number of variables to $v = 30$. The predictor variables set $\mathbf{x}$ does not follow a multivariate normal distribution. The graph structure used to generate the data is given in Figure 4.5. An elaborate description of the data generation is given in Appendix B.2. The true values of $\boldsymbol{\beta}$ are given in Table 4.4.

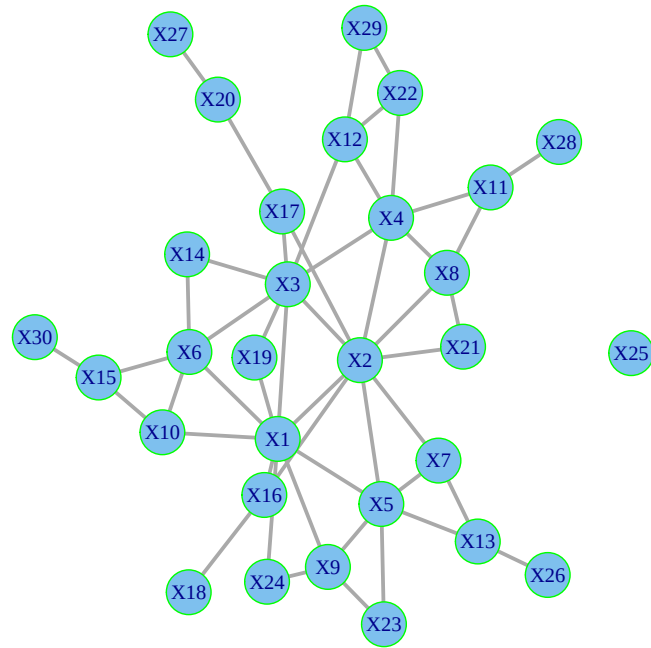


Figure 4.5: *Graph structure for covariates under Scenario 4.*

Table 4.4: Results from Scenario 4 (with known and unknown graphs), $n = 25, v = 30$.

	β	Ridge		Lasso		Lars		GLSE (with known graph)				GLSE (with unknown graph)			
		$\hat{\beta}_{ridge}$	SD ^a	$\hat{\beta}_{lasso}$	SD ^a	$\hat{\beta}_{lars}$	SD ^a	$\hat{\beta}$	SE (4.23) ^b	SD ^a	SE (Bootstrap) ^c	$\hat{\beta}$	SE (4.23) ^b	SD ^a	SE (Bootstrap) ^c
X_1	0.5	0.221	0.295	0.161	0.336	0.088	0.240	0.506	1.918	0.826	0.955	0.266	0.141	0.370	0.434
X_2	1	0.487	0.357	0.455	0.506	0.302	0.425	0.984	1.955	0.810	0.916	0.622	0.150	0.367	0.720
X_3	0.6	0.305	0.302	0.251	0.371	0.162	0.296	0.563	1.680	0.723	0.828	0.382	0.131	0.334	0.435
X_4	-1	-0.548	0.592	-0.425	0.700	-0.257	0.553	-1.018	3.561	1.583	1.639	-0.686	0.281	0.702	0.894
X_5	0.7	0.356	0.549	0.285	0.624	0.154	0.441	0.711	3.270	1.351	1.470	0.397	0.265	0.672	1.022
X_6	-1	-0.558	0.606	-0.407	0.682	-0.262	0.542	-1.030	3.382	1.513	1.575	-0.687	0.280	0.676	0.792
X_7	1.1	0.520	0.583	0.423	0.740	0.253	0.569	1.164	3.351	1.416	1.388	0.690	0.278	0.687	0.922
X_8	-0.6	-0.343	0.539	-0.275	0.636	-0.165	0.495	-0.661	3.220	1.415	1.482	-0.365	0.270	0.673	1.083
X_9	1.5	0.790	0.589	0.681	0.844	0.445	0.713	1.580	3.430	1.373	1.433	1.006	0.281	0.666	0.887
X_{10}	0.5	0.301	0.608	0.244	0.633	0.133	0.456	0.585	3.280	1.456	1.459	0.380	0.274	0.728	0.765
X_{11}	0.8	0.423	0.611	0.334	0.698	0.209	0.564	0.858	3.253	1.414	1.452	0.601	0.277	0.728	1.113
X_{12}	-0.8	-0.408	0.585	-0.321	0.681	-0.191	0.523	-0.820	3.395	1.443	1.476	-0.568	0.275	0.689	1.231
X_{13}	1.8	1.001	0.688	0.986	0.993	0.694	0.860	1.670	3.354	1.288	1.385	1.236	0.295	0.693	0.856
X_{14}	0.9	0.501	0.576	0.413	0.702	0.264	0.544	0.874	3.218	1.381	1.396	0.665	0.281	0.715	1.054
X_{15}	-0.6	-0.299	0.624	-0.207	0.607	-0.134	0.449	-0.626	3.163	1.378	1.465	-0.368	0.272	0.667	1.089
X_{16}	1.3	0.678	0.609	0.563	0.802	0.373	0.683	1.281	3.363	1.393	1.374	0.943	0.280	0.659	0.895
X_{17}	0.1	0.047	0.596	0.027	0.613	0.011	0.467	0.113	3.257	1.462	1.437	-0.015	0.278	0.664	0.963
X_{18}	1.7	0.965	0.678	0.895	0.962	0.667	0.885	1.731	3.430	1.346	1.331	1.151	0.290	0.711	0.913
X_{19}	-2	-1.070	0.701	-0.986	0.986	-0.711	0.861	-1.960	3.726	1.374	1.354	-1.410	0.291	0.705	1.097
X_{20}	0.8	0.434	0.581	0.316	0.646	0.206	0.519	0.760	3.273	1.414	1.431	0.538	0.272	0.634	0.975
X_{21}	0.8	0.437	0.567	0.339	0.629	0.195	0.471	0.916	3.217	1.390	1.388	0.530	0.281	0.697	0.845
X_{22}	-0.3	-0.174	0.583	-0.141	0.601	-0.086	0.418	-0.236	3.014	1.338	1.412	-0.280	0.260	0.676	0.998
X_{23}	2.5	1.417	0.765	1.558	1.203	1.235	1.113	2.450	3.774	1.368	1.305	1.767	0.304	0.725	0.873
X_{24}	0.2	0.114	0.556	0.037	0.539	0.044	0.417	0.303	3.136	1.393	1.385	0.178	0.273	0.700	1.079
X_{25}	-0.4	-0.219	0.251	-0.164	0.305	-0.103	0.242	-0.373	1.355	0.576	0.584	-0.284	0.120	0.298	0.373
X_{26}	1.9	1.027	0.651	0.943	0.981	0.667	0.807	1.966	3.646	1.368	1.273	1.305	0.291	0.699	0.826
X_{27}	-0.05	-0.022	0.585	0.008	0.551	0.004	0.448	-0.112	3.148	1.426	1.375	-0.003	0.276	0.668	0.917
X_{28}	1	0.556	0.604	0.453	0.735	0.287	0.591	0.968	3.247	1.373	1.370	0.729	0.276	0.702	0.821
X_{29}	-2.2	-1.189	0.702	-1.247	1.111	-0.927	1.030	-2.173	3.517	1.303	1.379	-1.574	0.303	0.748	0.924
X_{30}	0.6	0.317	0.621	0.242	0.657	0.151	0.495	0.633	3.158	1.429	1.390	0.361	0.279	0.720	0.931
Euclidean distance ^d		0.361		0.581		0.750		0.061				0.361			

^a SD: the Monte Carlo standard error for the 500 replicates.^b SE(4.23): the mean of the 500 standard error estimates using the error equation given in 4.23.^c SE(bootstrap): the mean of the 500 bootstrap standard error estimates.^d Euclidean distance : $\sqrt{\frac{\sum(\hat{\beta}-\tilde{\beta})^2}{v}}$

The results from this scenario are obtained from the average of running 500 realisations and are displayed in Table 4.4. It can be clearly seen that the GLSE is giving the closest estimates to the true coefficients compared to the other methods. This implies that even if the distribution of predictor variables is not a multivariate normal one, unbiased estimates for the regression coefficients can still be obtained by the GLSE, which makes it possible for these results to generalise to any dataset. It can also be noticed that the bootstrap estimate for standard errors works well, but the standard error estimate using formula (4.23) does not provide good estimates. In addition, the GLSE method gives the smallest Eucledian distance.

In the case where the graph is estimated (unknown graph structure), the results show that GLSE performs better than the rest of the methods, and the Monte Carlo estimates of standard errors for GLSE are close to or sometimes smaller than those of ridge, Lasso and Lars.

4.4.5 Scenario 5

In this scenario, we consider a model with one binary variable in the predictor set for the implementation of GLSE. We take $n = 15$ and $v = 20$. Suppose that X_{20} is binary, taking the value 0 or 1. We generate the data based on Figure 4.6, where the binary variable X_{20} belongs to the clique (X_3, X_4, X_7, X_{20}) . A detailed description of the simulation model is given in Appendix B.3. The response y is generated from model (2.1), where the random errors ϵ are normally distributed with mean 0 and standard error $\sigma = 0.5$. The true values of β are given in Table 4.5.

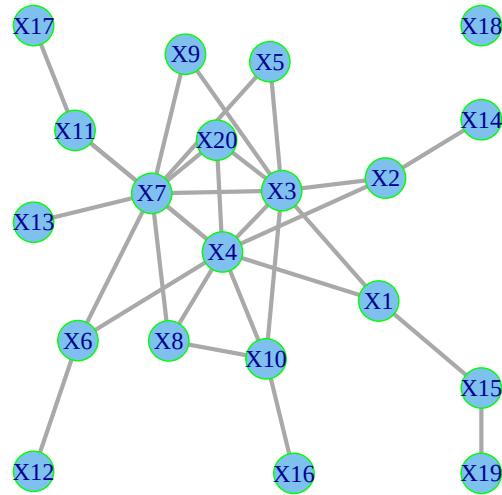


Figure 4.6: *True graph structure for covariates under Scenario 5.*

Table 4.5: Results from Scenario 3 (with known and unknown graphs), $n = 15, v = 20$.

	β	Ridge		Lasso		Lars		GLSE (with known graph)		$\hat{\beta}$	GLSE (with unknown graph)		$\hat{\beta}$	GLSE (with unknown graph)	
		$\hat{\beta}_{ridge}$	SD ^a	$\hat{\beta}_{lasso}$	SD ^a	$\hat{\beta}_{lars}$	SD ^a	SE (4.23) ^b	SD ^a		SE (4.23) ^b	SD ^a		SE (4.23) ^b	SD ^a
X ₁	0.5	0.411	0.651	0.074	0.353	0.207	0.555	0.363	4.246	1.995	2.101	0.482	2.826	1.374	1.592
X ₂	1	0.637	0.606	0.284	0.595	0.449	0.725	0.703	3.913	1.775	1.884	0.965	2.412	1.153	1.336
X ₃	0.6	0.182	0.319	0.137	0.349	0.215	0.399	0.625	2.479	1.032	1.441	0.195	1.394	0.677	0.716
X ₄	-1	-0.450	0.343	-0.474	0.539	-0.564	0.560	-0.938	2.689	1.085	1.533	-0.731	1.465	0.650	0.864
X ₅	0.7	0.409	0.679	0.084	0.326	0.223	0.613	0.736	4.128	1.683	1.812	0.902	2.929	1.344	1.449
X ₆	-1	-0.084	0.623	-0.025	0.283	-0.141	0.545	-1.023	4.245	1.650	1.807	0.579	2.894	1.179	1.253
X ₇	1.1	0.624	0.337	0.860	0.645	0.870	0.626	1.134	2.546	0.988	1.322	0.914	1.463	0.562	0.811
X ₈	-0.6	-0.423	0.590	-0.095	0.351	-0.213	0.590	-0.346	3.922	1.598	1.576	-0.717	2.639	1.315	1.643
X ₉	1.5	0.682	0.568	0.355	0.609	0.587	0.749	1.557	3.610	1.343	1.406	0.933	2.328	1.090	1.115
X ₁₀	0.5	0.255	0.735	0.050	0.329	0.143	0.588	0.435	4.270	1.924	2.130	0.408	3.046	1.485	1.684
X ₁₁	0.8	0.428	0.678	0.140	0.457	0.304	0.708	0.898	4.264	1.667	1.860	0.472	2.905	1.433	1.327
X ₁₂	-0.8	-0.422	0.695	-0.120	0.438	-0.241	0.633	-0.915	4.257	1.957	1.923	-0.847	3.069	1.547	1.752
X ₁₃	1.8	1.200	0.651	0.608	0.879	0.956	1.000	1.891	4.338	1.525	1.497	1.815	2.886	1.061	1.107
X ₁₄	0.9	0.433	0.719	0.118	0.407	0.293	0.662	0.626	4.240	1.946	1.822	0.498	3.024	1.535	1.877
X ₁₅	-0.6	-0.419	0.728	-0.099	0.366	-0.265	0.692	-0.620	4.106	1.900	2.024	-0.748	2.985	1.441	1.397
X ₁₆	1.3	0.677	0.705	0.188	0.481	0.389	0.695	1.348	4.238	1.928	1.822	1.130	2.922	1.521	1.645
X ₁₇	0.1	0.082	0.708	0.000	0.306	0.029	0.494	0.037	4.084	1.889	1.963	0.151	3.029	1.534	1.456
X ₁₈	1.7	0.954	0.664	0.559	0.777	0.851	0.876	1.754	3.739	1.504	1.413	1.380	2.470	1.176	1.142
X ₁₉	-2	-1.123	0.790	-0.496	0.811	-0.822	0.992	-1.996	4.566	1.973	1.827	-1.680	3.021	1.401	1.308
X ₂₀	0.8	0.442	0.401	0.400	0.464	0.505	0.479	0.808	1.798	0.645	0.725	0.700	1.219	0.543	0.655
Euclidean distance ^d		0.513		0.790		0.618		0.123				0.430			

^a SD: the Monte Carlo standard error for the 500 replicates.^b SE(4.23): the mean of the 500 standard error estimates using the error equation given in 4.23.^c SE(bootstrap): the mean of the 500 bootstrap standard error estimates.^d Euclidean distance : $\sqrt{\frac{\sum(\hat{\beta}-\beta)^2}{v}}$

The results based on this scenario are given in Table 4.5 and they are obtained from the average of running 500 realisations of the datasets. From the results, it is revealed that the estimates given by the proposed GLSE are the closest to the real coefficients as compared to ridge, Lars and Lasso. This means that the GLSE method still has higher efficacy in obtaining unbiased estimates of the regression coefficients when the dataset consists of both quantitative and binary variables. Besides, the Euclidean distance is smallest for the GLSE method. The Monte Carlo error estimates also reveal that the results are comparable to those of ridge, Lasso and Lars. As for the estimates of standard errors obtained by running 500 bootstraps from each of the 500 realisations, these prove to be close to the Monte Carlo estimates of the standard errors.

In addition, in the case of the unknown graph, the GLSE is outperforming the rest in coefficient estimates. The same conclusion could be drawn for the rest of the estimates.

4.4.6 Scenario 6

In this scenario we justify the performance of the standard error estimates for the GLSE, by the formula (4.23) and bootstrapping. We have seen in Sections 4.4.1, 4.4.3 and 4.4.5 that when $n < v$ the bootstrap standard errors perform well, but the standard error estimates obtained via (4.23) are not good enough.

If we increase n , the standard error estimates based on (4.23) become closer to the Monte Carlo estimates. Table 4.6 suggests that for a large value of n , both (4.23) and the bootstrap method can be used to calculate the standard errors for the estimates.

Table 4.6: *Results from Scenario 4 for $n > v$.*

Multivariate			Non-Multivariate		
SE(4.23) ^b	SD (Monte Carlo) ^a	SE (Bootstrap) ^c	SE(4.23) ^b	SD (Monte Carlo) ^a	SE (Bootstrap) ^c
0.353	0.362	0.364	0.427	0.434	0.387
0.410	0.421	0.498	0.469	0.483	0.428
0.346	0.345	0.356	0.199	0.191	0.191
0.211	0.211	0.244	0.121	0.111	0.103
0.216	0.214	0.243	0.426	0.420	0.395
0.423	0.420	0.520	0.406	0.379	0.337
0.227	0.210	0.221	0.110	0.093	0.080
0.224	0.224	0.243	0.372	0.354	0.350
0.372	0.364	0.366	0.429	0.413	0.406
0.463	0.477	0.608	0.439	0.440	0.400
0.201	0.196	0.227	0.377	0.366	0.337
0.480	0.492	0.522	0.465	0.476	0.455
0.228	0.177	0.158	0.437	0.402	0.397
0.363	0.362	0.427	0.462	0.471	0.514
0.310	0.319	0.333	0.462	0.468	0.433
0.504	0.513	0.502	0.473	0.488	0.525
0.248	0.250	0.306	0.443	0.459	0.522
0.537	0.527	0.578	0.399	0.400	0.415
0.364	0.347	0.371	0.485	0.458	0.432
0.456	0.466	0.565	0.345	0.341	0.351

^a SD: the Monte Carlo standard error for the 500 replicates.

^b SE(4.23): the mean of the 500 standard error estimates using the error equation given in 4.23.

^c SE(bootstrap): the mean of the 500 bootstrap standard error estimates.

4.5 Data analysis

We consider the data used in Scheetz et al. (2006), which consist of gene expressions from eye tissues of 12-week-old rats. The dataset, which has 120 observations on more than 31,000 genes, is available within the R library “flare” (Li et al. 2014), where the dimension of the dataset is reduced to 200 genes, which could have high possibilities to relate to TRIM32. The gene TRIM32 carries a significant biological information; for example, this gene might lead to the Bardet-Biedl syndrome, which is a heterogeneous genetic disease in several organs including eye retina (Huang et al. 2008).

In summary, the data analysis is a regression problem with $n = 120$ and $v = 200$ where identifying the correlation structure in the dataset is desirable in that most of the variables in the data are correlated. To show this, 30 variables are selected randomly from the data and their marginal and conditional correlation plots are given in Figure 4.7.

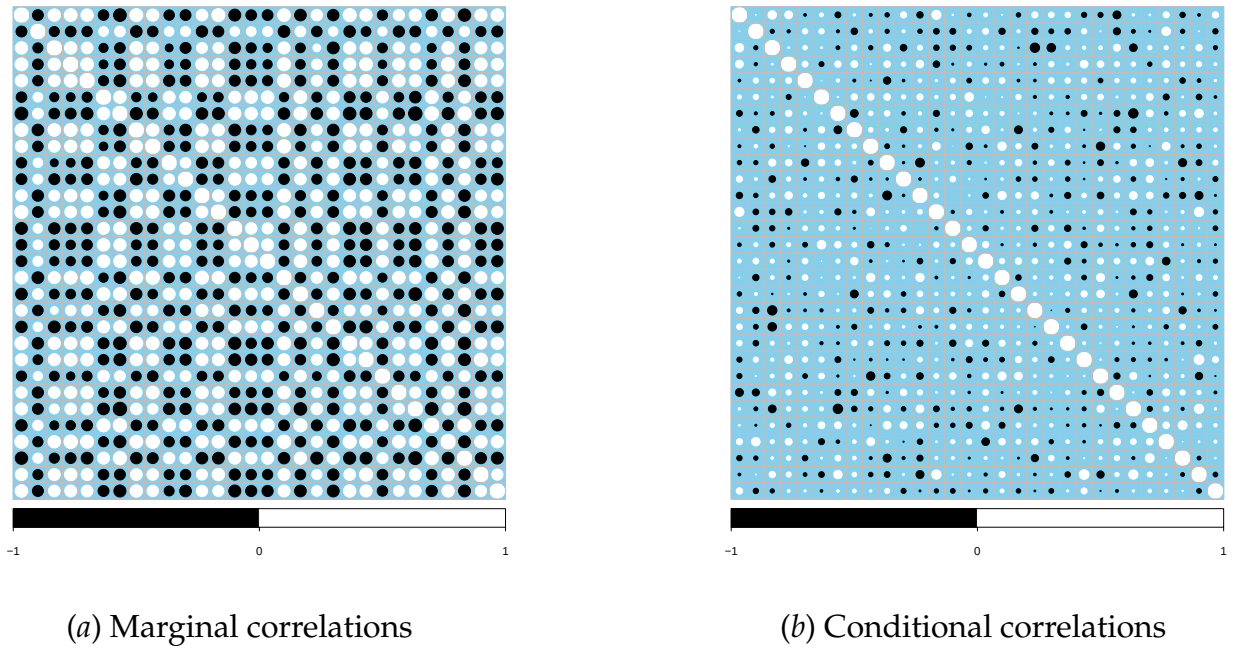


Figure 4.7: Marginal and conditional correlations plots for randomly chosen 30 variables of the data.

It is clear from Figure 4.7 that the most of the marginal correlations are strong, while the conditional correlations vary from weak to strong between the variables, which indicates that the data has a graph structure that has to be identified. Ridge, Lasso, Lars and our proposed GLSE method are applied on this dataset. Cross validation is used to obtain the penalty parameter for ridge and Lasso and to find the number of steps that Lars requires to find the optimal model. The Lasso method finds 24 non-zero coefficients while Lars finds 25 coefficients to be non-zero. Ridge regression selects all the variables and the results are not provided here as Ridge cannot give any information about significance

levels. The proposed method keeps all 200 variables in the model, among which 7 coefficients are significantly non-zero under a 10% significance level and 12 are significant under a 20% significance level. Note that bootstrap is used here to find the confidence interval for coefficient estimates and decide the significance level for the GLSE. The results are shown in Table 4.7. One advantage of GLSE is that it can provide the significance levels, while none of the other existing methods can do this properly. The reason for this is that reporting a standard error of a penalised estimator can give a mistaken impression of high precision. The penalised estimator completely ignores the inaccuracy introduced due to the bias. Therefore, confidence statements, such as those made based on bootstrap confidence intervals, might lead to incorrect decisions (Goeman et al. 2012, Van Houwelingen & Putter 2011).

Table 4.7: Variables selected and coefficient estimates by lars, lasso and GLSE. Significant at 5%: “***”, significant at 10%: “**” and significant at 20%: “ ”.

Lars		Lasso		GLSE	
Variable selected	$\hat{\beta}_{lars}$	Variable selected	$\hat{\beta}_{lasso}$	Variable selected	$\hat{\beta}_{GLSE}$
X_2	-0.009	X_2	-0.019	X_{48}	0.116
X_{11}	0.017	X_4	0.042	X_{54}	0.077 *
X_{13}	0.001	X_8	-0.031	X_{62}	-0.064 *
X_{42}	0.040	X_{11}	-0.011	X_{64}	0.087 **
X_{54}	0.041	X_{13}	0.021	X_{65}	0.068 *
X_{58}	0.008	X_{21}	-0.045	X_{87}	-0.169 **
X_{60}	0.013	X_{29}	-0.006	X_{102}	-0.076
X_{62}	-0.062	X_{33}	0.006	X_{146}	0.189 *
X_{76}	-0.012	X_{36}	0.039	X_{149}	-0.106
X_{87}	-0.116	X_{42}	0.100	X_{155}	0.078
X_{90}	-0.016	X_{54}	0.043	X_{170}	0.067
X_{106}	0.011	X_{55}	-0.007	X_{180}	0.118 **
X_{109}	-0.008	X_{58}	0.024		
X_{110}	-0.004	X_{60}	0.026		
X_{136}	-0.003	X_{62}	-0.079		
X_{146}	0.022	X_{65}	0.008		
X_{148}	0.008	X_{70}	-0.103		
X_{153}	0.079	X_{72}	-0.019		
X_{155}	0.025	X_{87}	-0.140		
X_{158}	-0.005	X_{106}	0.038		
X_{180}	0.021	X_{146}	0.031		
X_{185}	-0.027	X_{153}	0.152		
X_{187}	-0.015	X_{155}	0.062		
X_{188}	-0.018	X_{160}	0.047		
X_{200}	-0.043				

We compute the predictive mean square error (PMSE) for the selected model using cross-validation (Huang et al. 2008) through dividing the data randomly into a training set consisting of 90% of the observations and a test set consisting of the remaining 10% (Rosset & Zhu 2004). This process is repeated 300 times,

where a new random division is made each time. For GLSE, three different significance levels (5%, 10% and 20%) are considered when obtaining the predictive mean square error through cross-validation. In other words, at the 5% confidence level for instance, only those variables that are significant at this level are taken into consideration when running the cross-validation, while the other variables are discounted. The results for the dataset are given in Table 4.8. They reveal that the GLSE at 10% significance level gives the lowest predictive mean square error. In addition, the standard error of the predictive mean square error at 10% also has the lowest value. This indicates that the performance of GLSE at 10% is the best compared with the other methods in terms of the predictive mean square error.

Table 4.8: *Predictive mean square errors and their standard deviations.*

	Lars	Lasso	GLSE 5%	GLSE 10%	GLSE 20%
PMSE ^a	0.0090	0.0084	0.0091	0.0082	0.0111
SE-PMSE ^b	0.0136	0.0131	0.0167	0.0113	0.0170

^a PMSE: the predictive mean square error for the 500 replicates.

^b SE-PMSE: the mean of the 500 standard error estimates of the predictive mean square error.

The estimated graph showing the conditional correlations structure of genes is given in Figure 4.8. It was demonstrated by Scheetz et al. (2006) (using a graph representation) that some of the genes are highly correlated. This means that the genes have an independency structure that needs to be identified, which can provide very important biological information that can help in finding the undirected relation affecting the eye disease. Our graphical model has identified this really important structure, whereas none of the existing methods can.

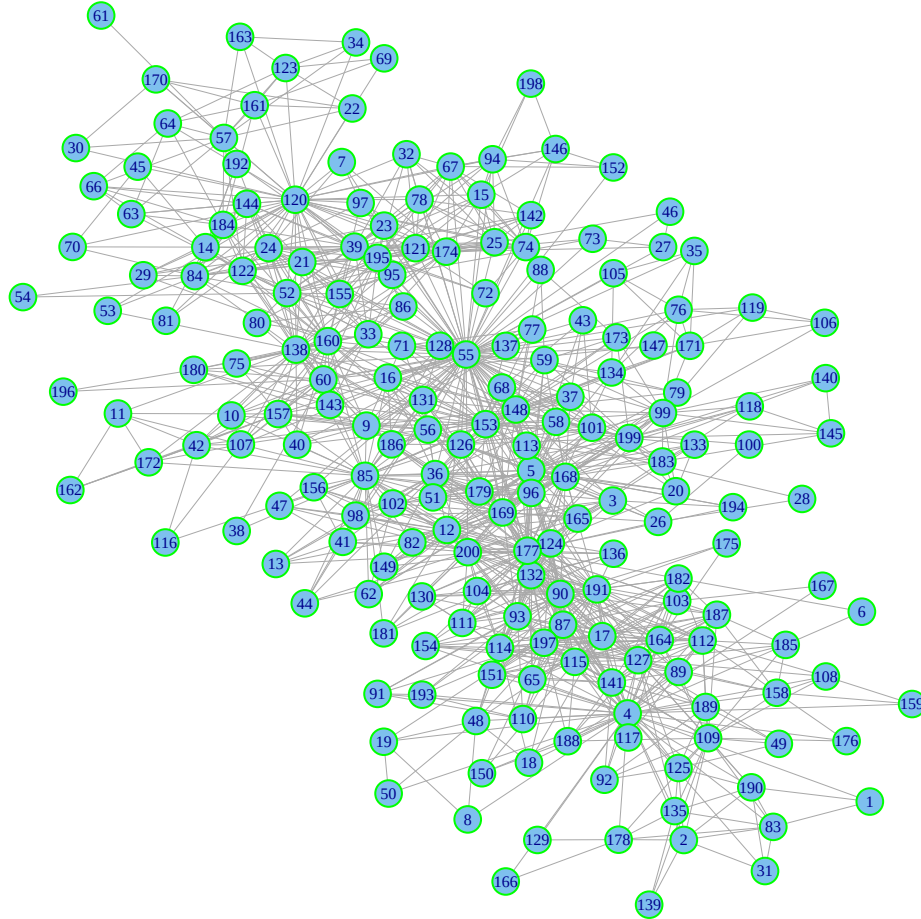


Figure 4.8: *Estimated gene network (conditional correlation structure) for the covariates in the real data.*

4.6 Conclusion

This chapter presented a novel method of estimating linear regression coefficients in an unbiased way under the conditional independence assumption. The method used the idea of Gaussian graphical models, which is very suitable for dealing with $n < p$ regression problems. The method is named the Graphical Least Square Estimation (GLSE) method. The unbiased estimator for

$n < p$ linear regression problems is supported by a theorem showing its unbiasedness. The results of the proposed method were compared with some well established methods (ridge, Lasso and Lars) using simulated and real datasets, revealing that the proposed GLSE outperforms the other established methods when required to estimate regression coefficients in an unbiased way.

CHAPTER 5

Hybridisation of graphical least squares and ridge regression

5.1 Introduction

The GLSE method proposed in Chapter 4 has good performance in terms of giving unbiased parameter estimates, considering that unbiased estimators are often preferable (Raol et al. 2004). However, in practice, the graph structure g is unknown and should be estimated based on the dataset. As a consequence, the whole graph space has to be searched in order to find the graph that is closest to the true graph, which can be a challenge in high-dimensional datasets with large numbers of covariates v (Fan et al. 2014, Fan & Lv 2010, Ferraty 2010, Johnstone & Titterton 2009). Examples of such datasets are genes expression and financial datasets. The potential weakness of the GLSE in executing this computation is that for a very large number of covariates, the involved computation is usually very heavy. Therefore, to overcome this computational issue, the ideas of ridge regression and GLSE are combined to propose a method named hybrid GLSE (HGLSE). This method imposes a ridge type penalty on some covariates while ensuring that the other covariate coefficient estimates are still unbiased. The HGLSE incorporates the advantages of both GLSE (unbiased) and ridge regression (computationally simple). In other words, HGLSE can give unbiased coefficient estimates for the most important variables, while penalising only the

less important variables. This idea could be achieved by predefining a yardstick that makes sure that the important variables are not missed out by using a pre-selection variable approach (Butler 1983, Guyon & Elisseeff 2003). As a consequence, a graphical model search will be implemented only in the several possible structures of these important variables and their possible relationships with less important ones. In addition, HGLSE helps to improve the performance of coefficients estimation and variable selection.

To empirically investigate the proposed method, the HGLSE is used here to construct non-sparse optimal portfolios using a financial dataset from London Stock Exchange (LSE). Four different sizes of portfolios are constructed and their in-and-out-of-sample performances are tracked. HGLSE is also applied to a higher-dimensional genes expression dataset (compared to that used in Chapter 4 with GLSE).

The rest of the chapter is organised as follows. Section 5.2 details the modified graphical least square estimation and briefly discusses the bias and variance of the proposed method. Section 5.3 provides the criteria for the graph model selection. Section 5.4 discusses the idea of portfolio optimisation in finance and the application of HGLSE to portfolio construction using a financial dataset. Section 5.5 describes the application of HGLSE to a genes expression dataset. Section 5.6 presents two sets of simulation scenarios. Finally, the conclusion of the chapter is given in Section 5.7.

5.2 The hybridised GLSE (HGLSE)

5.2.1 HGLSE with two cliques

Suppose that the vertex set V can be partitioned into disjoint sets A , B and C , where variables in set A are important variables but variables in set B and C are

less important. Also, assume that the variables $(X_1, \dots, X_v)$ are associated with a given decomposable graph g of two cliques $(A \cup B$ and $B \cup C)$ and a separator (B) . Then the proposed estimator for β can be as follows

$$\hat{\beta}^h = \left[\left\{ (ssd_{A \cup B})^{-1} \right\}^\Gamma + \left\{ (ssd_{B \cup C} + \lambda \mathbf{I}_{B \cup C})^{-1} \right\}^\Gamma - \left\{ (ssd_B)^{-1} \right\}^\Gamma \right] \mathbf{x}' \mathbf{y}, \quad (5.1)$$

where λ is the amount of penalty imposed on the variables with index set $B \cup C$. For the above estimator to exist, the following condition should be fulfilled,

Condition 5.2.1 *The sample size $n > \max\{|A| + |B|\}$.*

If we rewrite the covariate matrix $\mathbf{x}$ as $\mathbf{x} = (\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C)$ and the parameter β and its estimate $\hat{\beta}^h$ as $\beta = (\beta'_A, \beta'_B, \beta'_C)'$ and $\hat{\beta}^h = \left(\left(\hat{\beta}_A^h \right)', \left(\hat{\beta}_B^h \right)', \left(\hat{\beta}_C^h \right)' \right)'$, respectively, then based on the following Condition 5.2.2, the estimator $\hat{\beta}_A^h$ in (5.1) can be shown to be unbiased for β_A , while $\hat{\beta}_B^h$ and $\hat{\beta}_C^h$ are biased for β_B and β_C , respectively.

Condition 5.2.2 *The sets A , B and C form a decomposition of a given graph g where B is the separator. The sets $\mathbf{x}_A$ and $\mathbf{x}_C$ are conditionally independent given $\mathbf{x}_B$, such that*

$$\mathbf{x}_C = \mathbf{x}_B \cdot \mathbf{r}_{BC} + \xi_C, \quad \mathbb{E}(\xi_C) = \mathbf{0},$$

where $\mathbf{r}_{BC}$ is an $b \times c$ dimensional constant matrix and $b = |B|$, $c = |C|$.

To summarise the above arguments, the following theorem is introduced.

Theorem 5.2.1 *Under Condition 5.2.1 and Condition 5.2.2, the estimator in (5.1) is unbiased for the variables in set A while biased for those variables that are in set $B \cup C$,*

$$\mathbb{E}(\hat{\beta}_A^h) = \beta_A, \quad \mathbb{E}(\hat{\beta}_B^h) \neq \beta_B, \quad \mathbb{E}(\hat{\beta}_C^h) \neq \beta_C.$$

Proof: We have

$$\begin{aligned}
& E(\hat{\beta}^h | \mathbf{x}_B) \\
&= E \left(\left([(ssd_{A \cup B})^{-1}]^\Gamma + [(ssd_{B \cup C} + \lambda \mathbf{I}_{B \cup C})^{-1}]^\Gamma - [(ssd_B)^{-1}]^\Gamma \right) \mathbf{x}' \mathbf{x} \beta | \mathbf{x}_B \right) \\
&= E \left[\left(\begin{array}{ccc} \mathbf{I}_A & \mathbf{0} & (ssd_{A \cup B})^{-1} \mathbf{x}'_{A \cup B} \mathbf{x}_C \\ \mathbf{0} & \mathbf{I}_B & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right) \right. \\
&\quad + \left(\begin{array}{ccc} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ (ssd_{B \cup C} + \lambda \mathbf{I}_{B \cup C})^{-1} \mathbf{x}'_{B \cup C} \mathbf{x}_A & (ssd_{B \cup C} + \lambda \mathbf{I}_{B \cup C})^{-1} \mathbf{x}'_{B \cup C} \mathbf{x}_B & (ssd_{B \cup C} + \lambda \mathbf{I}_{B \cup C})^{-1} \mathbf{x}'_{B \cup C} \mathbf{x}_C \end{array} \right) \\
&\quad \left. - \left(\begin{array}{ccc} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ (ssd_B)^{-1} \mathbf{x}'_B \mathbf{x}_A & \mathbf{I}_B & (ssd_B)^{-1} \mathbf{x}'_B \mathbf{x}_C \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right) \right] \mathbf{x}_B \beta. \tag{5.2}
\end{aligned}$$

Using Condition 5.2.2, we have

$$E(\mathbf{x}_C | \mathbf{x}_B) = \mathbf{x}_B \mathbf{r}_{B,C} = \mathbf{x}_{A \cup B} \begin{pmatrix} \mathbf{0} \\ \mathbf{r}_{B,C} \end{pmatrix},$$

therefore,

$$\begin{aligned}
& E \left[E \left[(ssd_{A \cup B})^{-1} \mathbf{x}'_{A \cup B} \mathbf{x}_C - \begin{pmatrix} \mathbf{0} \\ (ssd_B)^{-1} \mathbf{x}_B^T \mathbf{x}_C \end{pmatrix} \middle| \mathbf{x}_B, \mathbf{x}_A \right] \middle| \mathbf{x}_B \right] \\
&= (ssd_{A \cup B})^{-1} \mathbf{x}'_{A \cup B} \mathbf{x}_{A \cup B} \begin{pmatrix} \mathbf{0} \\ \mathbf{r}_{BC} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ (ssd_B)^{-1} \mathbf{x}_B^T \mathbf{x}_B \mathbf{r}_{BC} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{0} \\ \mathbf{r}_{BC} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \mathbf{r}_{BC} \end{pmatrix} = \mathbf{0}. \tag{5.3}
\end{aligned}$$

Then Equation (5.2) becomes

$$E(\hat{\beta}^h | \mathbf{x}_B) = \begin{pmatrix} \mathbf{I}_A & \mathbf{0} & \mathbf{0} \\ * & * & * \\ * & * & * \end{pmatrix} \beta \quad (5.4)$$

and further

$$E(\hat{\beta}_A^h) = \beta_A, \text{ but } E(\hat{\beta}_B^h) \neq \beta_B, E(\hat{\beta}_C^h) \neq \beta_C.$$

□

5.2.2 The general form of HGLSE

We now assume that the graph g has a set of cliques $\mathcal{C} = \{C_1, \dots, C_q\}$ (a perfect ordering) with separators $\mathcal{S} = \{S_2, \dots, S_q\}$. Consider that $C^* \in \mathcal{C}$, which can be a very large clique but the variables in it are less important than those not belonging to it. Denote $\mathcal{C}' = \mathcal{C} \setminus \{C^*\}$ as the set of cliques not having C^* . The general HGLSE is given by

$$\hat{\mathbf{b}}^g = \left[\sum_{C \in \mathcal{C}'} [(\text{ssd}_C)^{-1}]^\Gamma + [(\text{ssd}_{C^*} + \lambda \mathbf{I}_{C^*})^{-1}]^\Gamma - \sum_{S \in \mathcal{S}} [(\text{ssd}_S)^{-1}]^\Gamma \right] \mathbf{x}' \mathbf{y}. \quad (5.5)$$

For the HGLSE to exist, the following condition must be met.

Condition 5.2.3 *The sample size $n > \max_{C \in \mathcal{C}'} \{|C|\}$.*

Then we have the following theorem:

Proposition 5.2.1 *Given that Condition 4.2.2 and Condition 5.2.3 hold true, the estimator in (5.5) is unbiased for all variables $V \setminus C^*$ and biased for those in C^* .*

Proof: Following Theorem 4.2.2 , Proposition 5.2.1 is easily established.

The above Proposition 5.2.1 can be proved by mathematical induction similar to Theorem 4.2.2.

5.2.3 Bias and variance of HGLSE

In order to estimate the variance of the unbiased estimates, the bootstrap method is used where samples of size n are repeatedly drawn from the observed sample using simple random sampling with replacement (Efron & Tibshirani 1994). However, as mentioned in Chapter 4, care should be taken to avoid the problem of multi-collinearity within a clique (except the clique C^*).

For the less important covariates, as mentioned previously, the variance estimates are not desirable in that the variance is not very meaningful for biased estimates (Goeman et al. 2012). Estimation via penalised methods is employed to reduce the variance of estimators by making the estimates substantially biased.

5.3 Graph structure selection

Similar to the application of the GLSE method in Chapter 4, the unknown graph g needs to be estimated in order to apply the HGLSE method. A new iteration algorithm is proposed to search the space and select the graph that best fits the data under a regression framework, and to find HGLSE.

The proposed algorithm has two key stages in each iteration: 1. finding the best graph g^* , its associated estimate $\hat{b}^g$ and the variances of $\hat{b}^g$ based on given C^* , the set of less important variables; 2. selecting the less important variables C^* based on $\hat{b}^g$ and its variance. The algorithm can start with a guess of C^* . A ridge penalty is then applied on this clique to make the inversion of the aforementioned clique possible.

In the first stage, the algorithm searches for the best graph structure g^* with the condition that the clique C^* is fixed, by minimising the target function $\mathbb{T}(\beta, g)$ (sum of square errors) given as follows:

$$\begin{aligned} (\hat{\mathbf{b}}^g, g^*) &= \arg \min_{g \in \mathcal{G}_{C^*}} \mathbb{T}(\beta, g), \\ \mathbb{T}(\beta, g) &= \|\mathbf{y} - \mathbf{x}\beta\|^2 \end{aligned} \quad (5.6)$$

where $\hat{\mathbf{b}}^g$ is given in (5.5) and $\mathcal{G}_{C^*}$ denotes the space of all decomposable graphs with the clique C^* and satisfying Condition 5.2.3. It should be noted that the target function above is without the penalty term $(\lambda_g |\mathcal{E}_g|)$ since the latter is not required, due to the conclusion reached in the example given in Section 4.3. Bootstrapping is then performed to calculate the variance and the significance level of the unbiased estimates for the important variables (those not in C^*).

In the second stage, we update C^* . Denote the variables (for those unbiased estimates only) above a certain significance threshold level as D . We then apply Lasso to the saturated regression model with variables $\mathbf{x}_D$ fixed. The variables with coefficient estimates which are shrunk to 0 by Lasso are treated as less important variables and form the new clique C^* .

This process is repeated until there is convergence where the method selects the same sets of important variables twice. The algorithm is given as follows:

Algorithm 3 Pseudo-code of the hybridised iteration algorithm

1: **Initialisation:** Select covariates C^* that are less important and $V \setminus C^*$ as the set for the important ones.

Iteration – repeated until the convergence with respect to the estimate $\hat{b}^g$

-
- 2: Search the graph space (similar to step 2 of Algorithm 2) and find the graph g^* and $\hat{b}^g$ such that they minimise the target function $\mathbb{T}(\cdot)$ (given in (5.6)), on the condition that C^* is a clique of g^* .
 - 3: Based on g^* , do bootstrapping to find the variance and significance levels for the variables in $V \setminus C^*$.
 - 4: Select variables in $V \setminus C^*$, which are above a certain significance threshold level. Denote the set of the selected variables as D .
 - 5: Consider the regression model with all covariates V and use Lasso, with the variables in D fixed, to do variable selection for all other variables in $V \setminus D$. The non-selected variables by Lasso form the new C^* in the next iteration.
 - 6: Go back to step 2.
-

In Algorithm 3, the second step can be parallelised to work if several central processing units are available in the computational environment. The computational cost can therefore be improved significantly via parallel computation.

5.4 Portfolio optimisation

This section introduces the idea of portfolio optimisation and includes a review of how it has been addressed in the literature, then moves to discuss the application of HGLSE to portfolio construction and finally applies the HGLSE method on financial dataset.

In finance, a portfolio is considered as a collection of two or more risk (or risk-free) assets such as shares, government bonds and derivative securities which

are held directly by investors or managed by a financial institution (Ennew et al. 2014). Investors or financial institution managers seek to efficiently allocate and diversify their capital over a number of available assets by creating a portfolio that leads to maximising expected returns on the investment and minimising the associated risk. Such an efficient allocation of capital among different assets could be made by portfolio optimisation.

The issue of efficient capital allocation through portfolio optimisation was solved by Markowitz (1952) using a model known as the Mean-Variance model or Markowitz theory. For a portfolio consisting of v assets, with expected return vector μ and covariance matrix Σ , the Markowitz model selects an efficient (optimal) portfolio weight vector w (column vector), which minimises the risk (volatility) of the portfolio for a predetermined targeted expected return. In other words, it is a trade-off between two factors, the risk and the return of the portfolio. Here w can be interpreted as the amount of capital invested in each asset in the portfolio. The construction of such an efficient (optimal) portfolio for a given targeted expected return R is done through finding the optimal portfolio weights w by solving the following quadratic optimisation problem (Ledoit & Wolf 2003):

$$\underset{w}{\text{minimise}} \quad w' \Sigma w \quad \text{subject to} \quad \mu' w = R, \quad w' \mathbf{1}_v = 1 \quad (5.7)$$

where $\mathbf{1}_v$ denotes a $v \times 1$ vector of 1s. The set of all efficient portfolios lies on a risk-return relationship curve called efficient frontier (see Figure 5.1) where every point on the curve offers the highest expected return for a given level of risk or the lowest risk for a given level of expected return (Lee & Lee 2010).

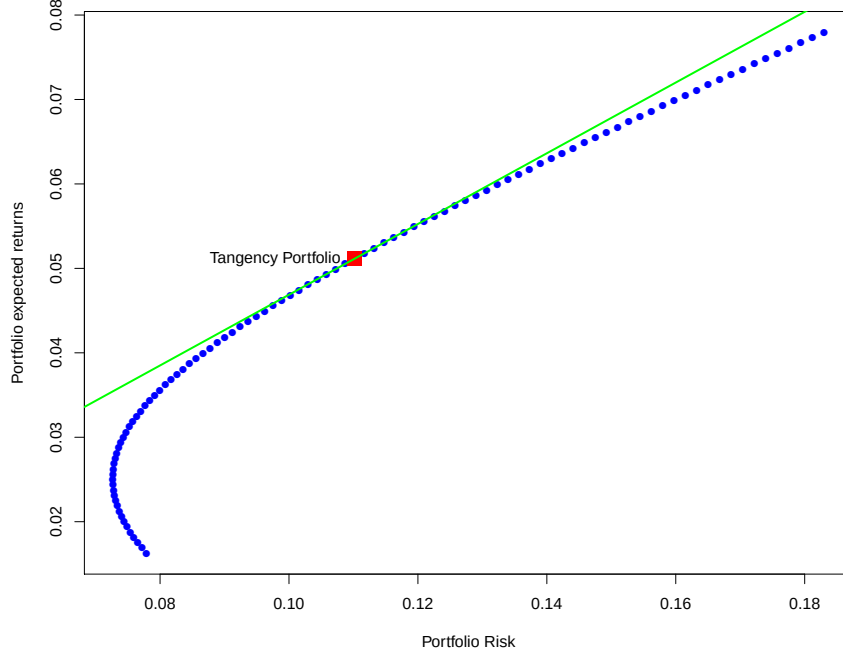


Figure 5.1: Graph of efficient portfolios curve and tangency portfolio.

Investors, however, usually prefer to use the Sharpe ratio (the ratio of return to the standard deviation of the return) to evaluate a portfolio performance (Lai et al. 2011). The combination of the v risk assets which gives the maximum Sharpe ratio among portfolios on the efficient frontier is called tangency portfolio w^* (Gilli et al. 2011) and is graphically represented as a point where a line through the origin (or any other point in the vertical axis if the portfolio includes risk-free assets) is tangent to the curve of efficient frontier (see Figure 5.1).

In line with this, Markowitz theory requires that the population mean μ and covariance matrix Σ of the assets' returns be known. However, in practice, these two parameters are unknown and should be estimated using a historical dataset. Britten-Jones (1999) shows that the tangency portfolio (the optimal solution) based on the historical data is given as:

$$\hat{w}^* = \frac{\hat{\beta}}{\hat{\beta}' \mathbf{1}_v} \quad (5.8)$$

where $\hat{\beta}$ is the OLS estimate of the coefficient parameter β for the linear regression model in (2.1). However, y here is different from that in the general regression and, hence, it is a column vector of 1s with length n , while x is the assets' return. In the regression model, the total number of observations n is usually much less than the total number of assets v , i.e. $\hat{\beta}$ is not available based on standard regression methods (as mentioned earlier), which is the main challenge for such a modern portfolio optimisation problem.

Therefore, in order to overcome the issue for model (2.1) when $n < v$, many recent studies (e.g. Carrasco & Noumon 2011, DeMiguel et al. 2009, Fastrich et al. 2015, Still & Kondor 2010) focused on regularisation methods such as ridge regression (Hoerl & Kennard 1970), Lasso (Tibshirani 2011), Lars (Efron et al. 2004), adaptive Lasso (Zou 2006) and the Dantzig selector (Candes & Tao 2007). However, as mentioned before, all of their estimates are biased giving emphasis to the famous notion of "bias-variance tradeoff", which might over-shrink the coefficients (James & Radchenko 2009, Radchenko et al. 2011) and potentially produce inaccurate portfolio weights. As mentioned in Chapter 2, some of these methods (notably Lasso and Lars) also suffer from the problem of not selecting more than n covariates (Hastie et al. 2009, Zou & Hastie 2005) and giving a sparse portfolio (Fastrich et al. 2015), which is less preferable for risk management and diversification.

5.4.1 Applying HGLSE to portfolio optimisation

Unlike the aforementioned regularisation methods, HGLSE can give unbiased coefficient estimates for the most important assets (assets that may have high return and low risk) and maintain them in the portfolio with large weights, while penalising only the weights of the other less important assets. Such an advantage will lead to increasing the Sharpe ratio and the expected rate of

returns, and decreasing the risk of the portfolio for both in-and-out-of-sample periods. It can also generate diversified portfolios (like the ridge method) across a large number of stocks, as it produces a non-sparse portfolio (including all stocks in the market). The diversified portfolio risk is very important (Malkiel & Xu 1997), because if one or more sectors of the economy decline, the other sectors could help in reducing the significant loss due to market fluctuations.

The HGLSE method also has the properties of graphical models, as the latter can depict the conditional correlation of assets. Our method can thus find this structure when there are many predictors/assets, some of which are more likely to be conditionally independent given a subset of assets. For instance, the return of a portfolio may depend on a large number of assets. If we simply look at the correlation of these assets, they may all be correlated marginally. However, some of them may be independent given a certain subset of other assets (for a related example, see Carvalho & West 2007). This advantage is a distinguishing feature of the HGLSE method against all other existing methods.

5.4.2 Financial data analysis

As the focus here is on non-sparse portfolios, the result of our method is only compared with that of the ridge method.

Monthly returns of more than 850 stocks are used in this study (which have a complete return history over the last 10 years on London Stock Exchange) starting from 31/07/2005 and ending on 31/07/2015 (120 months). The year 2008, which is the peak period of the financial crisis (Ivashina & Scharfstein 2010), has been removed from the data in order to avoid its influence on the analysis. The marginal and conditional correlations of the stocks are shown in Figure 5.2 via a small sample of 30 stocks selected randomly. Figure 5.2 displays that most of the marginal correlations range from weak to moderate, while the conditional ones

vary between very weak and weak with a few moderate and strong ones. This means that there is an underlying graph structure that needs to be identified. In the world of finance, this may imply that the portfolio's risk can be lowered by taking into consideration the correlations between assets (Levi 2009).

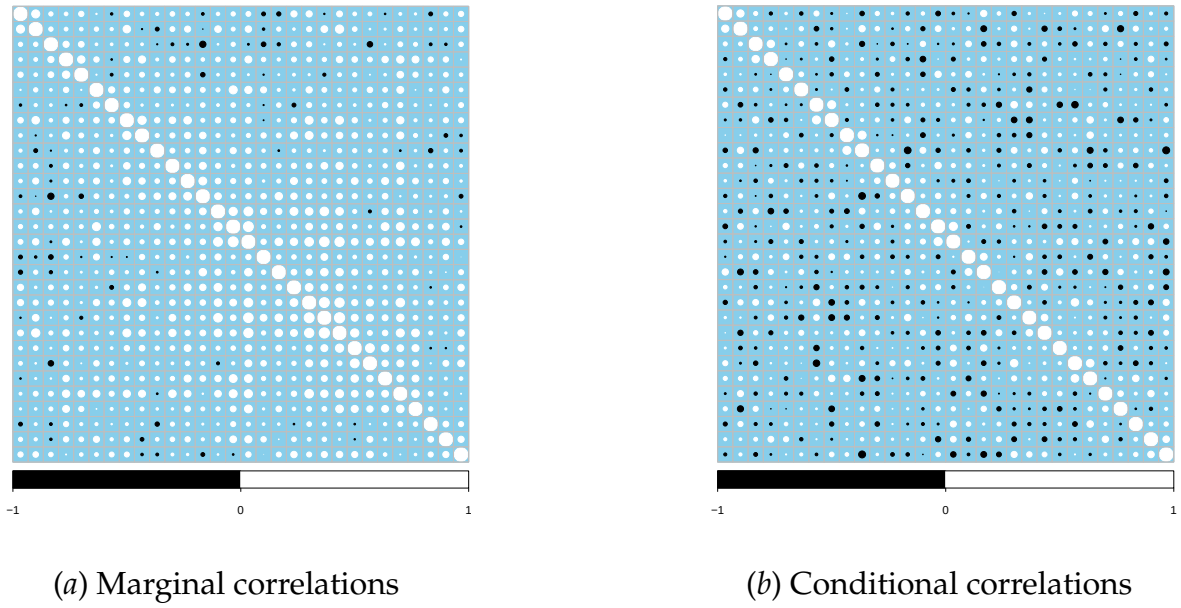


Figure 5.2: Marginal and conditional correlations plots for randomly chosen 30 assets of the data.

Out of the 850 stocks, 100, 200, 300 and 400 have been selected at random to construct non-sparse portfolios using ridge and HGLSE methods. Those portfolios are constructed using the past 96 months (8 years) of stock return. Subsequently, the portfolios are kept for one year, and their monthly out-of-sample returns are observed. The in-sample period of portfolios is from 31/07/2005 to 31/12/2007 and from 01/01/2009 to 31/07/2014, and the out-of-sample period lasts from 01/08/2014 to 31/07/2015.

When using the ridge method to construct the portfolio, cross-validation is utilised to obtain the penalty parameter that gives the smallest prediction mean square error. The Sharpe ratios, expected returns and risk for both in-and-out-of-sample periods are computed in Table 5.1 and used to evaluate the performance

of different sizes of the obtained portfolios. The results in Table 5.1 reveal that the HGLSE method outperforms the ridge method in terms of Sharpe ratio, expected rate of returns and risk of portfolios for both in-and-out-of-sample periods. Moreover, it can also be noticed that the HGLSE method reduces the risk of portfolios for both in-and-out of sample periods more than ridge, when the size of the portfolio increases; except for the 200 sample portfolio, the risk is almost the same. Thus, it can be argued that the HGLSE method is a good strategy in large data set as it could construct a diversified portfolio with smaller the risk. The 200 sample portfolio has a high number of small companies which have more risk investments than large companies (Vassalou & Xing 2004). However, the Sharp ratios of the 200 sample portfolio constructed by the HGLSE is significantly much higher compared to the ones produced by the ridge method for both in-and-out-sample periods.

In the first 100-assets portfolio, HGLSE selects 29 out of 100 as important assets based on the significance of their estimated coefficients. These assets have a significant impact on the returns of the portfolio, as they contribute by more than 50% of the total portfolio returns for the in-sample period and approximately 85% of the total portfolio returns for the out-of-sample period. In the rest of portfolios (200, 300, and 400 stocks), the HGLSE method selects 40, 45 and 60 assets, respectively, as important assets. These assets yield a considerable effect on the total portfolios' return, as they account for in-sample periods by nearly 44%, 53% and 53%, and for out-of-sample periods by approximately 47%, 85% and 88%, respectively. Although the contribution of 47% of the 200-assets portfolio's return by 40 important assets for the out-of-sample period appears substantial, this value is still relatively small if compared with the 85% and 88% accounted for by 45 and 60 assets, respectively, in other portfolios. This is due to the same reason mentioned above regarding the fact that the 200 sample mainly

comprises smaller firms which are deemed risky for investment.

Table 5.1: *Portfolios' sizes and in-and-out-of-sample portfolio's Sharpe ratios, expected returns and risk found by ridge and HGLSE*

Portfolio size	Methods	Sharpe ratio	Expected returns	Portfolio's risk
100 stocks (in sample)	Ridge	1.044	0.050	0.048
	HGLSE	1.164	0.054	0.046
100 stocks (out of sample)	Ridge	0.376	0.020	0.053
	HGLSE	0.672	0.034	0.050
200 stocks (in sample)	Ridge	1.195	0.051	0.043
	HGLSE	1.721	0.073	0.042
200 stocks (out of sample)	Ridge	0.302	0.017	0.057
	HGLSE	0.447	0.025	0.056
300 stocks (in sample)	Ridge	1.579	0.060	0.038
	HGLSE	1.954	0.064	0.033
300 stocks (out of sample)	Ridge	0.394	0.023	0.058
	HGLSE	0.754	0.035	0.046
400 stocks (in sample)	Ridge	1.537	0.051	0.033
	HGLSE	2.002	0.062	0.031
400 stocks (out of sample)	Ridge	0.538	0.020	0.037
	HGLSE	0.715	0.024	0.033

5.5 Genes expression data analysis

The aim of this dataset analysis is to study how the HGLSE performs in selecting the most significant probes compare to Lars and Lasso.

A preprocessed form of the dataset used in Scheetz et al. (2006) is considered in this chapter (similar to the dataset used in Chapter 4). The reason for using

this gene expression dataset is that it is believed that there are only a few genes that are important to the response variable (Huang et al. 2008). This dataset consists of gene expressions from eye tissues of 12-week-old rats with 120 observations on more than 31,000 genes. The robust multi-chip averaging (RMA) method (Bolstad et al. 2003) along with a logarithmic scale was used to normalise and analyse the intensity levels and to obtain representative expression values for all the probe sets. Not fully expressed genes were removed on the basis of two criteria: firstly, the empirical distribution of RMA normalised values where the maximum expression value was not greater than the 25th percentile of the entire set of RMA expression values, and, secondly, a gene not exhibiting at least two-fold variation in expression level among the 120 F_2 animals. About 18,975 probes fulfilled these criteria and could have high possibilities to relate to TRIM32. As mentioned in Chapter 4, this gene (TRIM32) has significant biological information. For instance, it might lead to a heterogeneous genetic disease in several organs including the eye retina called the “Bardet-Biedl syndrome” (Huang et al. 2008).

The 18,975 probes were reduced by Huang et al. (2008) to 200 and by Wang et al. (2012) to 300 genes because they believed, as stated before, that there are only a few genes that are important to the response variable. In the current study, 400 probes were selected based on the same criteria used by Huang et al. (2008) to assess the proposed method when the number of variables is much greater than the number of observations. Before analysing the data, 30 variables of the 400 probes were selected randomly to examine the marginal and conditional correlations, plots of which are shown in Figure 5.3. It can be seen that most of the marginal correlations between the variables are strong, while the conditional ones range between weak and strong. This indicates that there is a graph structure in the data that needs to be explored.

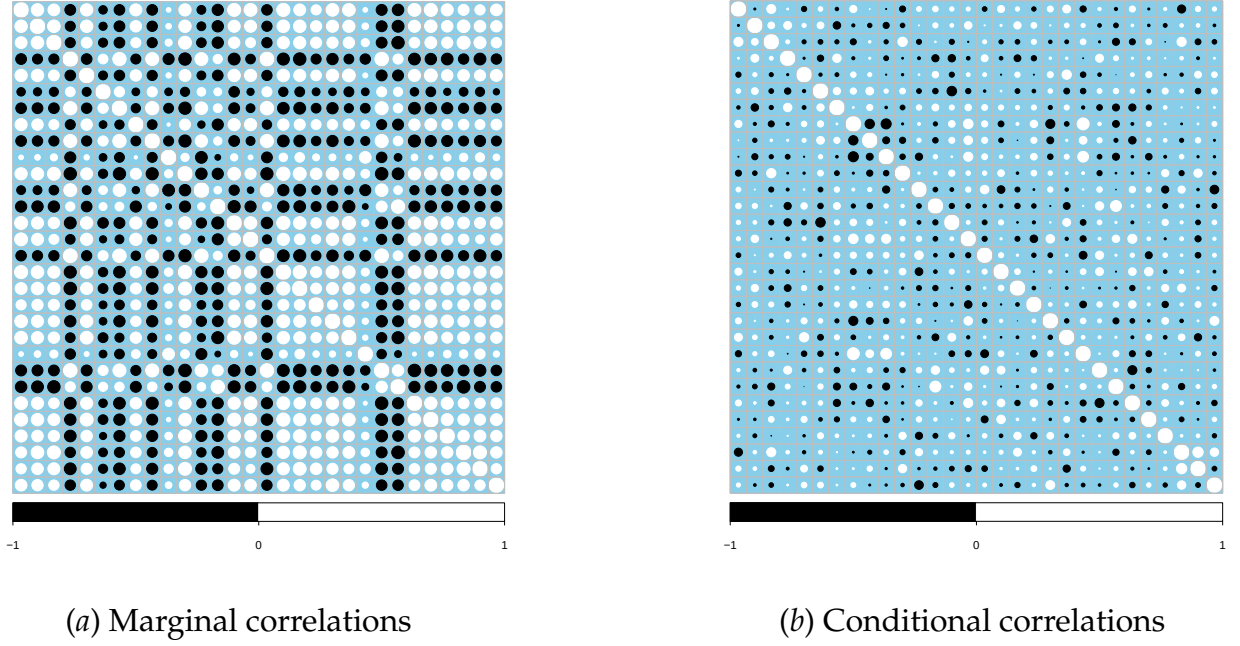


Figure 5.3: *Marginal and conditional correlations plots for randomly chosen 30 probes of the data.*

Ridge, Lasso, Lars and HGLSE methods are applied to the dataset. Cross-validation is used to obtain the penalty parameters for ridge and Lasso and find the number of steps required by Lars to select the optimal model. The results for the dataset are given in Table 5.2. They indicate that a total of 29 variables are chosen by HGLSE while 22 are chosen by Lasso and only 16 by Lars. It can also be seen from Table 5.2 that some variables selected by HGLSE are not present in either Lasso or Lars selected sets. Wang et al. (2012) found that probes such as 1383996_at (X_{206}) have the highest frequency of selection based on their study. This probe is also found by Lars, Lasso and HGLSE in the current study. Moreover, Wang et al. (2012) found that 1393382_at (X_{343}) is selected by all methods under comparison in their study, whereas in this study this probe is only selected by our proposed method (HGLSE).

Table 5.2: Variables selected and coefficient estimates by Lars, Lasso and HGLSE.

Lars		Lasso		HGLSE	
Variable selected	$\hat{\beta}_{lars}$	Variable selected	$\hat{\beta}_{lasso}$	Variable selected	$\hat{\beta}_{hglse}$
X_{11}	-0.013	X_{11}	-0.023	X_1	-0.053
X_{19}	-0.024	X_{19}	-0.004	X_{11}	0.012
X_{37}	0.029	X_{27}	-0.001	X_{21}	-0.009
X_{118}	0.003	X_{37}	0.047	X_{27}	-0.038
X_{129}	0.022	X_{87}	0.001	X_{37}	0.049
X_{130}	0.005	X_{115}	0.007	X_{84}	0.045
X_{158}	-0.006	X_{129}	0.021	X_{87}	0.016
X_{182}	0.048	X_{130}	0.006	X_{115}	0.007
X_{194}	0.019	X_{158}	-0.007	X_{126}	0.04
X_{198}	-0.029	X_{179}	0.018	X_{129}	-0.039
X_{206}	0.044	X_{182}	0.042	X_{130}	0.035
X_{210}	0	X_{190}	0.003	X_{131}	0.014
X_{251}	-0.002	X_{194}	0.022	X_{149}	-0.017
X_{268}	0.059	X_{198}	-0.03	X_{177}	-0.05
X_{353}	-0.003	X_{202}	0.002	X_{179}	0.087
X_{366}	0.007	X_{206}	0.065	X_{182}	0.015
		X_{256}	0	X_{194}	0.023
		X_{268}	0.063	X_{198}	-0.013
		X_{349}	0.019	X_{206}	0.069
		X_{353}	-0.001	X_{256}	0.023
		X_{366}	0.025	X_{268}	0.059
		X_{375}	0.003	X_{296}	-0.027
				X_{343}	0.023
				X_{348}	0.014
				X_{349}	0.043
				X_{353}	0.02
				X_{366}	0.071
				X_{371}	-0.022
				X_{375}	0.036

Table 5.3: Predictive mean square errors and their standard deviations.

	Lars	Lasso	HGLSE
PMSE ^a	0.0293	0.0306	0.0255
SE-PMSE ^b	0.0798	0.0830	0.0511

^a PMSE: the predictive mean square error for the 500 replicates.^b SE-PMSE: the mean of the 500 standard error estimates of the predictive mean square error.

Similar to the data analysis in Chapter 4, the predictive mean square error (PMSE) for the selected model is also calculated using cross-validation. The results for the dataset are given in Table 5.3. They indicate that HGLSE yields the lowest PMSE and that its standard error also has the lowest value. This means that HGLSE outperforms the other methods in terms of the predictive mean square error.

The estimated graph showing the conditional correlations structure for the informative and non-informative genes is given in Figure 5.4. As mentioned in the previous chapter, some of the genes are highly correlated with each other. As a consequence of this, they regulate the genes responsible for the eye disease. This assures the presence of an independency structure among genes, which is identified by the proposed HGLSE and can yield important biological information for finding the undirected association behind the eye disease. This also remains a distinguishing feature for HGLSE.

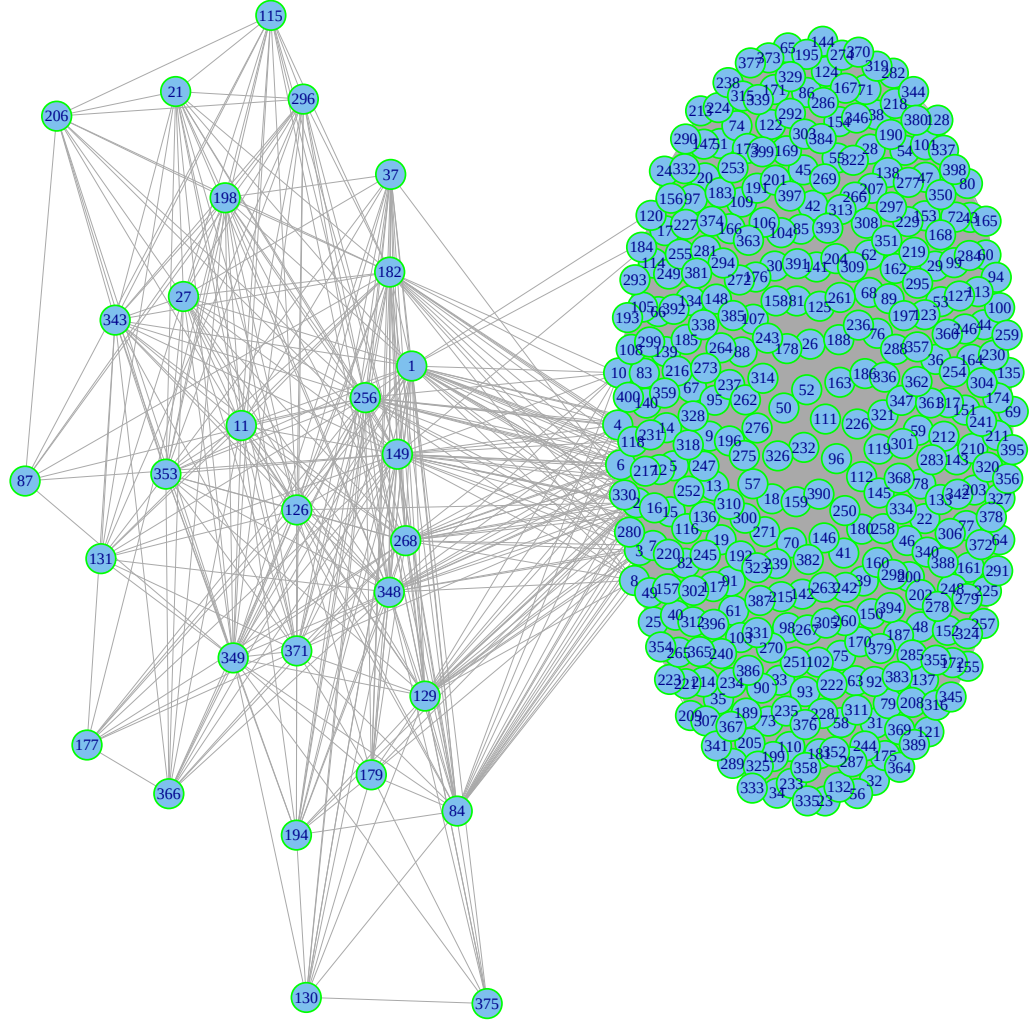


Figure 5.4: Estimated gene network (conditional correlation structure) for the covariates in the real data when some variables are considered as informative. The dense clique is C^* .

5.6 Simulation study

This section provides two different sets of simulation scenarios in order to further evaluate the HGLSE method in selecting significant variables and constructing a non-spares optimal portfolio. As far as selecting significant variables is con-

cerned, the method is compared with Lars and Lasso, while it is compared with ridge in terms of portfolio construction. In the first set of scenarios, the response is not a vector of 1s, while in the second set of scenarios the response is a vector of 1s to mimic portfolio optimisation.

5.6.1 First simulation set

Four simulation scenarios are presented here for the assessment of the proposed HGLSE in producing unbiased estimates for the important variables (in all scenarios) and testing the accuracy of variable selection (in Scenarios 2, 3 and 4). The accuracy of variable selection is assessed by finding the frequency of correctly identified zero and nonzero coefficients in the 500 replicated simulations based on three measurement criteria: sensitivity, specificity and model size (discussed in detail in 2.3.1), in addition to a fourth criterion, which is the number of times important variables are selected.

Scenario 1 is based on a case of normally distributed covariates, provided that the C^* is known. Since we suppose C^* is known, we do not need the full Algorithm 3, but step 2 (finding the estimate and graph) and step 3 (finding the significance level and standard deviation) are required. This simulation scenario is used to justify how appropriate the standard deviation (or confidence intervals) of the estimate given by the bootstrap method is and how the important variables are estimated. Scenario 2 is based on multivariate normally distributed predictor variables, where the important variables are supposed to be unknown. Algorithm 3 is therefore applied here. Scenario 3 is similar to Scenario 2; however, the covariates do not follow multivariate normal distribution. Finally, in Scenario 4, unlike in the previous three scenarios, there exists a case of a binary variable in predictor variables. The results of the four scenarios are compared with those of ridge, Lars and Lasso.

5.6.1.1 Scenario 1

In this scenario, we use the model (2.1) as given in Chapter 2 to generate the response y where the random errors are normally distributed with mean 0 and standard error $\sigma = 1$. A total of $v = 20$ predictor variables are considered, which follow a multivariate normal distribution with mean vector $\mathbf{0}$ and variance covariance matrix Σ , and a total of $n = 15$ samples are generated. The true β values are given in Table 5.4, where five predictor variables are important (X_1, X_5, X_{10}, X_{15} and X_{20}) as they have non-zero regression parameter values. The true graph used in this scenario to generate the data is given in Figure 5.5, where the conditional correlations for the important variables are shown on the corresponding edges.

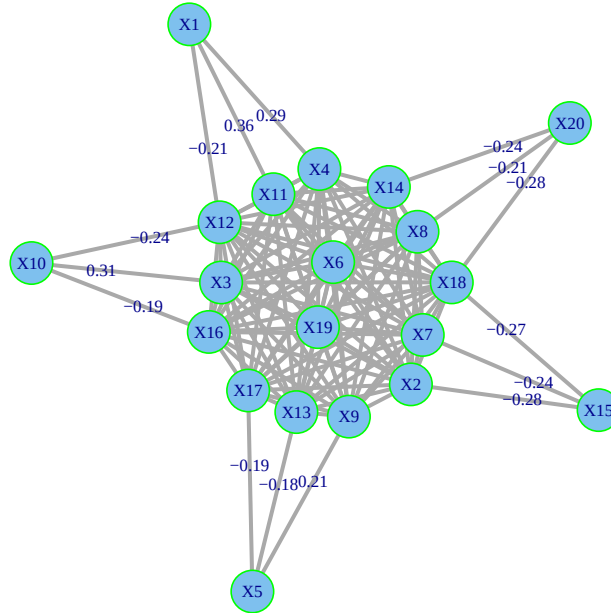


Figure 5.5: Graph structure for covariates under Scenario 1.

Table 5.4: Results for Scenario 1, $n = 15, v = 20$.

	β	Ridge			Lasso			LARS			HGLSE			CP ^d	
		$\hat{\beta}_{Ridge}$	SD ^a	MSE ^b	$\hat{\beta}_{Lasso}$	SD ^a	MSE ^b	$\hat{\beta}_{Lars}$	SD ^a	MSE ^b	$\hat{\beta}_{HGLSE}$	SD ^a	MSE ^b		SE (Bootstrap) ^c
X_1	-1	-0.492	0.535	0.544	-0.51	0.613	0.616	-0.327	0.489	0.692	-0.989	0.502	0.252	0.824	0.976
X_2	0	-0.166	0.604	0.392	-0.048	0.482	0.235	-0.023	0.207	0.043	0.045	0.389	0.153	1.144	
X_3	0	0.125	0.533	0.300	0.064	0.488	0.242	0.047	0.282	0.082	-0.021	0.28	0.079	0.562	
X_4	0	-0.06	0.499	0.253	-0.04	0.404	0.165	-0.039	0.24	0.059	-0.008	0.207	0.043	0.363	
X_5	1.8	0.896	0.499	1.066	1.213	0.746	0.901	1.007	0.747	1.187	1.797	0.443	0.196	0.66	0.974
X_6	0	0.039	0.491	0.243	0.014	0.407	0.166	0.001	0.259	0.067	-0.003	0.138	0.019	0.184	
X_7	0	-0.14	0.585	0.362	-0.013	0.449	0.202	-0.009	0.194	0.038	-0.012	0.048	0.002	0.053	
X_8	0	-0.111	0.526	0.289	-0.022	0.465	0.217	-0.021	0.217	0.048	-0.012	0.062	0.004	0.052	
X_9	0	0.106	0.493	0.254	0.053	0.441	0.197	0.061	0.256	0.069	0.027	0.067	0.005	0.049	
X_{10}	1.3	0.651	0.492	0.663	0.796	0.661	0.691	0.614	0.585	0.813	1.315	0.460	0.212	0.788	0.970
X_{11}	0	-0.089	0.435	0.197	-0.077	0.427	0.188	-0.051	0.23	0.056	-0.016	0.065	0.004	0.056	
X_{12}	0	0.011	0.487	0.237	0.043	0.422	0.180	0.021	0.21	0.045	0.003	0.065	0.004	0.049	
X_{13}	0	-0.013	0.415	0.172	-0.007	0.389	0.151	-0.01	0.252	0.064	-0.002	0.066	0.004	0.053	
X_{14}	0	-0.046	0.429	0.186	0.008	0.433	0.188	-0.01	0.253	0.064	-0.011	0.072	0.005	0.057	
X_{15}	2	1.021	0.495	1.203	1.497	0.746	0.810	1.345	0.751	0.993	1.998	0.436	0.190	0.622	0.964
X_{16}	0	0.023	0.538	0.290	0.017	0.458	0.210	-0.006	0.244	0.060	0.004	0.05	0.003	0.042	
X_{17}	0	-0.056	0.523	0.277	-0.018	0.464	0.216	-0.02	0.228	0.052	-0.006	0.069	0.005	0.055	
X_{18}	0	-0.339	0.514	0.379	-0.051	0.48	0.233	-0.053	0.24	0.060	-0.053	0.054	0.006	0.047	
X_{19}	0	0.077	0.553	0.312	0.059	0.438	0.195	0.041	0.233	0.056	0.011	0.049	0.003	0.046	
X_{20}	1.6	0.8	0.390	0.792	1.232	0.645	0.551	1.091	0.614	0.636	1.595	0.360	0.130	0.563	0.964
Euclidean distance ^e		0.787			0.495			0.669			0.009				

^a SD: The Monte Carlo standard error for the 500 replicates.^b MSE: Mean square error $(\|\hat{\beta} - \beta\|^2 + (SD)^2)/n$.^c SE: The mean of the 500 standard error estimates using the bootstrap.^d CP: Coverage probability for 95% CI.^e Euclidean distance: $\sqrt{\frac{\sum(\hat{\beta} - \beta)^2}{V/C^*}}$.

The results from this scenario are given in Table 5.4, which shows that HGLSE gives better results than ridge, Lasso and Lars in estimating the coefficients for both important and less important variables and yields the smallest Euclidean distance for important variables. The Monte Carlo estimates of standard errors and the mean square error (MSE) for HGLSE are smaller than those of ridge, Lasso and Lars, therefore, the performance of HGLSE is superior. Additionally, for the HGLSE method, the mean of the 500 bootstrap standard error estimates (SE^c) is in general slightly larger than the Monte Carlo standard error (SD^b) of the 500 replicated estimates, which is because of the small sample size.

In terms of coverage probability of the 95% confidence interval, the results are reasonable for the 5 important variables, implying that bootstrap estimation for the confidence intervals of the important predictor coefficients is acceptable. We did not present the coverage probability for the biased estimates (less important variables), since confidence interval is of no use for biased estimates.

5.6.1.2 Scenario 2

In this scenario, a total of $n = 25$ observations are generated on $v = 40$ variables from multivariate normal distribution with mean vector $\mathbf{0}$ and variance covariance matrix Σ . The values for the response variable y are generated using Model (2.1), where the random errors ϵ are normally distributed with mean 0 and standard error $\sigma = 1$. The true regression coefficients for these 10 important variables are given in the second column of Table 5.5. The aim of this scenario is to see how HGLSE performs in assessing the coefficient estimates compared to ridge, Lars and Lasso. Additionally, it aims to examine the performance of HGLSE in selecting the important variables compared to Lars and Lasso based on the aforementioned four criteria (sensitivity, specificity, model size and frequency of important variables selection). A total of 500 independent realisations are made

in this simulation. The graph structure for covariates under Scenario 2 is given in Figure 5.6, where the conditional correlations of the important variables are shown on the corresponding edges.

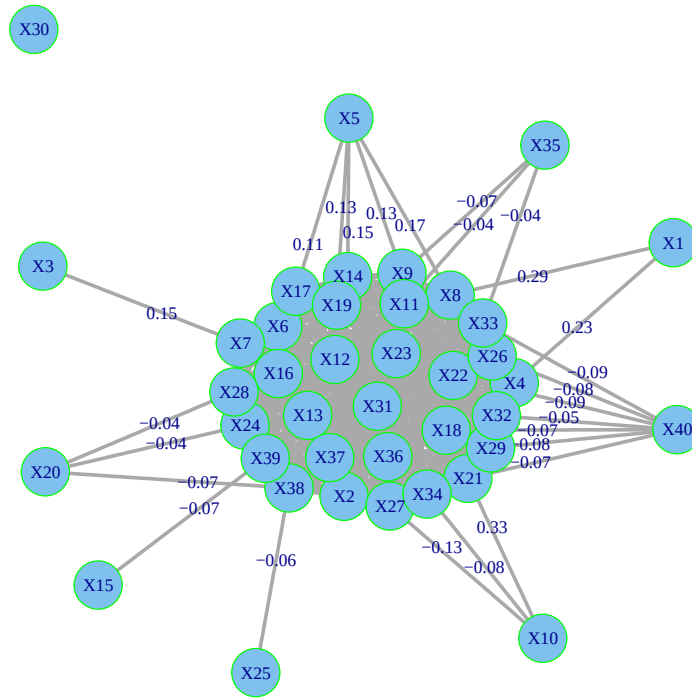


Figure 5.6: *Graph structure for covariates under Scenarios 2.*

The results from the average of running 500 realisations are given in Table 5.5. The table consists of the estimated coefficients by ridge, Lasso, Lars and the proposed HGLSE, in addition to the Euclidean distance for important variables. The table also includes the means of the measurement criteria (sensitivity, specificity and model size), in addition to the number of times the important variables are selected by Lasso, Lars and HGLSE. The simulation standard deviations for the three aforementioned criteria are displayed in parentheses in the table.

Table 5.5: *Results for Scenario 2, $n = 25, v = 40$.*

Variables	β	Ridge		Lasso		Lars		HGLSE		Number of times variables selected		
		$\hat{\beta}_{ridge}$	MSE ^a	$\hat{\beta}_{Lasso}$	MSE ^a	$\hat{\beta}_{Lars}$	MSE ^a	$\hat{\beta}_{HGLSE}$	MSE ^a	Lasso	Lars	HGLSE
X_1	2.6	1.274	2.309	1.396	2.675	1.155	3.273	1.553	1.946	396	352	402
X_3	3	1.552	2.781	1.791	3.083	1.526	3.784	1.803	2.514	422	389	428
X_5	2.5	1.200	2.103	1.417	2.243	1.159	2.856	1.572	1.509	418	374	433
X_{10}	3.3	1.586	3.668	1.846	3.966	1.489	5.105	2.050	2.637	412	359	446
X_{15}	2.6	1.259	2.514	1.281	3.114	1.007	3.813	1.515	2.231	365	303	370
X_{20}	2.4	1.155	2.143	1.187	2.621	0.964	3.194	1.528	1.734	367	308	395
X_{25}	2.5	1.286	2.034	1.380	2.414	1.156	2.970	1.582	1.521	403	348	424
X_{30}	3.5	1.846	3.539	2.191	3.777	1.898	4.603	2.206	2.854	423	389	438
X_{35}	-2.7	-1.293	2.453	-1.606	2.359	-1.336	2.992	-1.706	1.654	427	391	449
X_{40}	2.8	1.317	2.646	1.615	2.624	1.345	3.362	1.712	1.792	425	370	448
Euclidean distance ^b		1.422		1.224		1.493		1.076				
Sensitivity		-		81.1% (0.243)		71.7% (0.263)		84.6% (0.148)				
Specificity		-		67.7% (0.152)		78.6% (0.100)		75.6% (0.105)				
Size		-		17.846 (6.370)		13.59 (5.085)		15.798 (3.166)				

^a MSE: Mean square error ($(\hat{\beta} - \beta)^2 + (SD : \text{The Monte Carlo standard error})^2$).

^b Euclidean distance : $\sqrt{\frac{\sum (\hat{\beta} - \beta)^2}{V_{IC^*}}}$.

The results given in Table 5.5 indicate that HGLSE is the best in estimating coefficients of important variables among all the other methods as it gives the smallest Euclidean distance and mean square error. For variable selection, the results suggest that HGLSE outperforms Lasso and Lars in terms of the selected most important covariates as well as the highest levels of sensitivity (with the lowest standard deviation), although the mean of the model size in Lasso is greater than that in HGLSE. As for specificity, although HGLSE yields a higher value than that produced by Lasso, its value is slightly lower than that of Lars, which is due to the fact that Lars selects sparse models. It should be noted that the results of HGLSE with regards to the model size are rather consistent with those of the other two methods (Lars and Lasso), as can be seen from the

standard deviations for model size.

5.6.1.3 Scenario 3

This scenario is exactly similar to Scenario 2 in all aspects except in terms of predictor variables, $\mathbf{x}$, which do not follow a multivariate normal distribution. The idea of this simulation is to find out how the method performs in more general cases. The covariates are generated from a set of regression models with uniform random errors given in detail in Appendix C.1. The graph structure used for generating the covariates in Scenario 3 is given in Figure 5.7.

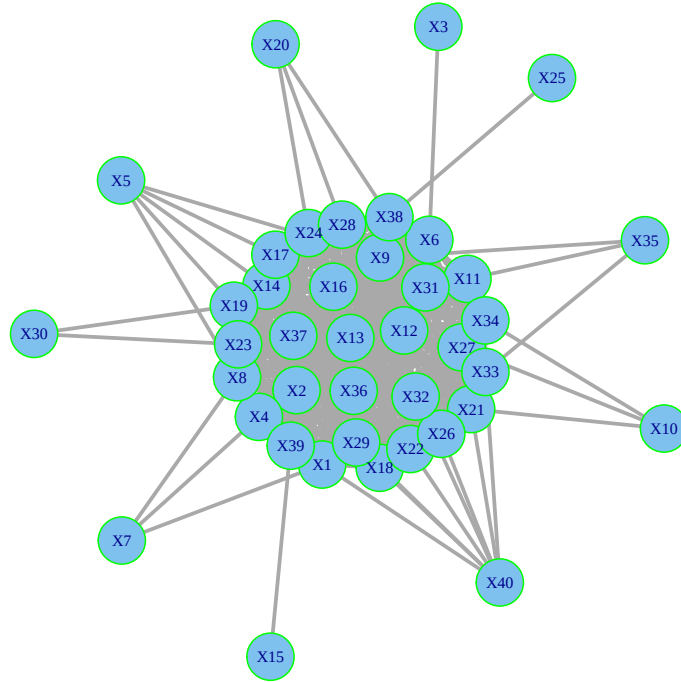


Figure 5.7: Graph structure for covariates under Scenario 3.

Similar to Scenario 2, in the case of coefficient estimation, the results are compared with those of ridge, Lasso and Lars. As for variables selection, the same criteria used in Scenario 2 are utilised and the results are compared with

those of Lasso and Lars. In this scenario, the penalty of Lasso is adjusted in order to give approximately the same number of selected variables the HGLSE does. Similarly, the number of steps is adjusted for Lars to give approximately the same number of variables selected by HGLSE. This is performed in order to have the same complexity models for Lasso, Lars and HGLSE and, consequently, achieve a more accurate comparison between them.

Table 5.6: Results for Scenario 3, $n = 25, v = 40$.

Variables	β	Ridge		Lasso		Lars		HGLSE		Number of times variables selected		
		$\hat{\beta}_{ridge}$	MSE ^a	$\hat{\beta}_{Lasso}$	MSE ^a	$\hat{\beta}_{Lars}$	MSE ^a	$\hat{\beta}_{HGLSE}$	MSE ^a	Lasso	Lars	HGLSE
X_3	3	1.335	3.455	1.413	4.454	1.369	4.647	1.710	3.303	363	349	352
X_5	2.5	1.111	2.372	1.137	3.259	1.116	3.379	1.360	2.537	342	343	340
X_7	2.6	1.188	2.508	1.322	3.296	1.330	3.273	1.424	2.758	360	351	370
X_{10}	3.3	1.261	4.556	1.636	4.528	1.596	4.747	1.678	3.829	404	389	411
X_{15}	2.6	1.091	2.857	1.099	3.644	1.092	3.774	1.391	2.633	333	322	352
X_{20}	2.4	1.124	2.241	1.093	3.299	1.066	3.379	1.366	2.574	316	306	338
X_{25}	2.5	1.197	2.353	1.208	3.296	1.105	3.585	1.387	2.645	346	324	324
X_{30}	3.5	1.609	4.342	1.975	4.728	1.988	4.845	2.217	3.338	408	402	429
X_{35}	-2.7	-0.987	3.293	-1.134	3.755	-1.126	3.903	-1.222	3.098	346	342	348
X_{40}	2.8	0.961	3.658	1.129	3.976	1.122	4.042	1.225	3.180	361	346	386
Euclidean distance ^b		1.623		1.483		1.506		1.306				
Sensitivity		-		71.6% (0.248)		69.5% (0.252)		73.0% (0.158)				
Specificity		-		66.3% (0.156)		70.5% (0.133)		68.1% (0.100)				
Size		-		17.258 (6.524)		15.794 (5.921)		16.872 (3.213)				

^a MSE: Mean square error $([(\hat{\beta} - \beta)^2 + (SD : \text{The Monte Carlo standard error})^2])$.

^b Euclidean distance : $\sqrt{\frac{\sum(\hat{\beta} - \beta)^2}{V/C^2}}$.

The results reveal that HGLSE has better performance than ridge, Lasso and Lars in terms of coefficient estimation of the important variables with the smallest mean square error and Euclidean distance. Likewise, the results also show that HGLSE is superior to Lasso and Lars as it achieves the highest sensitivity levels, with the lowest standard deviation. As for specificity, although HGLSE

yields the middle value among other methods, it is more consistent as it has the lowest standard deviation. In terms of model size, all the methods produce models with approximately equal sizes. However, as indicated by the values of standard deviations for the three methods, HGLSE is much more consistent in model size than Lasso and Lars, evidenced by its considerably small value.

5.6.1.4 Scenario 4

In this scenario HGLSE is applied to the case of having one binary predictor variable in the data. The data is generated from model (2.1), and their generation mechanism is shown in Appendix C.2. A total of $n = 25$ observations are generated on $v = 40$ features. The same graph in Scenario 3 is used here in generating this dataset. The variables that are assumed to be significant in predicting the response y are given in Table 5.7.

Table 5.7: Results for Scenario 4, $n = 25, v = 40$.

Variables	β	Ridge		Lasso		Lars		HGLSE		Number of times variables selected		
		$\hat{\beta}_{ridge}$	MSE ^a	$\hat{\beta}_{Lasso}$	MSE ^a	$\hat{\beta}_{Lars}$	MSE ^a	$\hat{\beta}_{HGLSE}$	MSE ^a	Lasso	Lars	HGLSE
X_3	3	1.503	3.148	1.595	4.216	1.162	5.619	1.927	2.854	376	302	371
X_5	2.5	1.116	2.470	1.193	3.141	0.880	4.056	1.481	2.490	349	291	378
X_7	2.6	1.281	2.327	1.326	3.245	0.991	4.211	1.596	2.430	360	303	398
X_{10}	3.3	1.412	4.145	1.730	4.421	1.357	5.731	1.853	3.445	402	362	416
X_{15}	2.6	1.180	2.568	1.206	3.435	0.899	4.385	1.376	2.663	358	283	356
X_{20}	2.4	1.120	2.350	1.010	3.307	0.735	4.149	1.449	2.317	317	255	350
X_{25}	2.5	1.229	2.366	1.196	3.391	0.851	4.409	1.534	2.465	332	268	364
X_{30}	3.5	1.706	4.075	1.957	4.839	1.582	6.137	2.268	3.175	401	359	430
X_{35}	-2.7	-1.134	2.944	-1.210	3.699	-0.875	4.809	-1.484	2.641	352	293	375
X_{40}	2.8	1.070	3.502	1.275	3.860	0.906	5.123	1.377	2.922	358	294	392
Euclidean distance ^b		1.530		1.424		1.771		1.168				
Sensitivity		-		72.1% (0.274)		60.2% (0.283)		76.6% (0.145)				
Specificity		-		68.5% (0.158)		80.1% (0.105)		67.7% (0.100)				
Size		-		16.65 (6.915)		11.728 (5.472)		17.338 (3.085)				

^a MSE: Mean square error ($(\hat{\beta} - \beta)^2 + (SD : \text{The Monte Carlo standard error})^2$).

^b Euclidean distance : $\sqrt{\frac{\sum (\hat{\beta} - \beta)^2}{V/C^*}}$.

A total of 500 realisations are made on the data and the results are given in Table 5.7. Similar to Scenarios 2 and 3, the results of HGLSE are compared with those of ridge, Lasso and Lars in terms of coefficient estimation. As for variables selection, the same set of criteria (2.3.1) is used and the results are compared with those of Lasso and Lars.

The results show that HGLSE outperforms ridge, Lasso and Lars in estimating the coefficients of the important variables with the smallest values of mean square error in most of these variables and the smallest Euclidean distance among them. Similarly, the results indicate that HGLSE performs better than Lasso and Lars in terms of variable selection and yields the highest levels of sensitivity (with the lowest standard deviation). As for specificity, although

HGLSE yields the lowest value among other methods, it is more consistent as it has the lowest standard deviation. In terms of model size, HGLSE produces models with an approximate size to that produced by Lasso but a larger size than that of Lars; however, HGLSE is much more consistent in model size than Lasso and Lars, as indicated by its very small value of standard deviation compared to those in Lasso and Lars.

5.6.2 Second simulation set

Two simulation studies are carried out here for the assessment of HGLSE in constructing an optimal portfolio. Scenario 5 is based on a case of normally distributed covariates, where the response y is a vector of 1s. Scenario 6 is similar to Scenario 5; however, the covariates do not follow multivariate normal distribution. In the two scenarios, we compare HGLSE with ridge regression. We do not compare HGLSE with Lasso or Lars because these methods only provide sparse portfolios (not keeping all covariates in the model), which is not preferable in practice (the diversified portfolio risk via a saturated model is very important (Malkiel & Xu 1997)).

5.6.2.1 Scenario 5

In this scenario, a total of $n = 48$ observations are generated on $v = 40$ variables from multivariate normal distribution with mean 0.01 and variance covariance matrix Σ (similar to Gilli et al. (2011)), where 36 observations (to mimic three-year monthly financial data) are used for estimating the regression coefficients and the portfolio performance (Sharpe ratios, expected returns and risk) for the in-sample period, and the remaining 12 observations (to mimic one-year monthly financial data) for computing the portfolio's performance for the out-of-sample period. The true portfolio weight w is derived based on the true covariance

matrix Σ (Britten-Jones 1999). Then the regression parameters β are estimated on the basis that the response y is a vector of 1s and the estimated weights $\hat{w}^*$ are derived from $\hat{\beta}$ via equation (5.8).

The aim of this scenario is to study how HGLSE performs in weight estimates and optimal portfolio construction. A total of 500 independent realisations are made in this simulation. The graph structure for covariates which are used in generating the dataset under Scenario 5 is given in Figure 5.6.

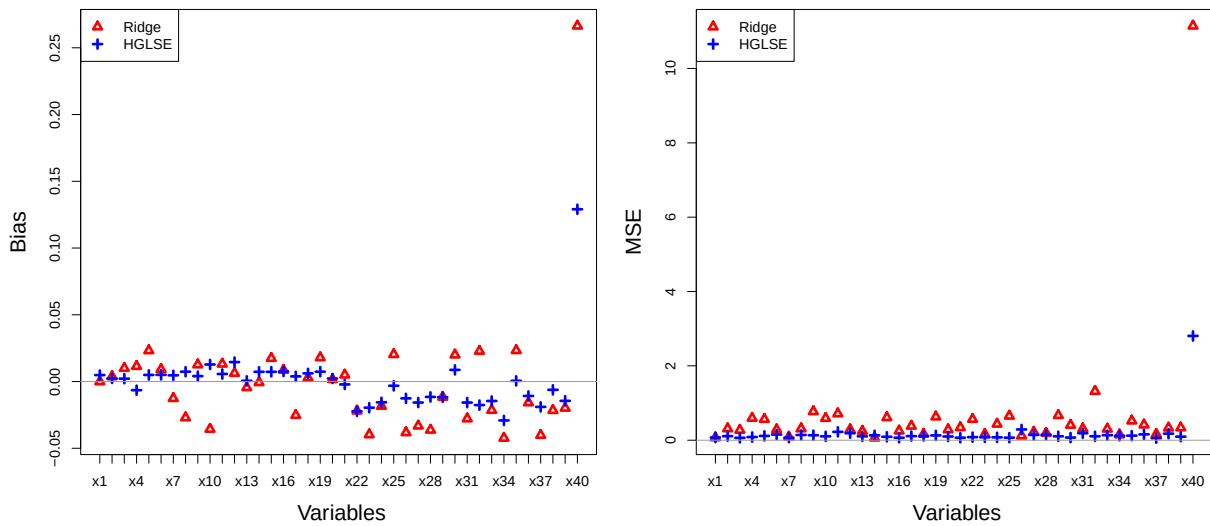
Since there are too many covariates to display, we simply summarise the results in Table 5.8 and Figure 5.8. Table 5.8 shows the means of in-and-out-of-sample portfolio's Sharpe ratios, expected returns and portfolio's risk from the 500 simulated data (in practice the portfolio's weights are of more interest than the actual coefficient estimate). Figure 5.8 displays the bias and the mean square errors of the tangent portfolio weight estimates.

The results given in Table 5.8 reveal that HGLSE gives higher means of in-and-out-of-sample portfolio's Sharpe ratios and lower risk from the 500 simulated data. In spite of the fact that the returns of the ridge portfolio for the in-sample period are higher than the HGLSE ones, the standard deviations of the returns for ridge are much higher than the ones for HGLSE. This indicates that the HGLSE method is likely to yield more stable returns than ridge. In addition, the out-of-sample expected returns are positive for HGLSE but negative for ridge, which is essential in the world of finance. On the other hand, as far as the Sharpe ratio is concerned, it can be seen from Table 5.8 that the standard deviations for the HGLSE method are slightly higher than those for ridge, and this difference is still relatively small when compared to the differences in standard deviations for expected returns and error. As for the weight estimates, Figure 5.8 indicates that HGLSE outperforms ridge in estimating the weights, as the bias in most of the weights found by HGLSE is closer to zero than that in ridge.

In addition, Figure 5.8 also shows that the mean square errors of HGLSE are much lower than those of ridge, which means that HGLSE is more consistent than ridge in estimating the weights of the portfolio. Moreover, the Euclidean distance for HGLSE (0.023) is smaller than that for ridge (0.047).

Table 5.8: The means and standard deviations of in-and-out-of-sample portfolio's Sharpe ratios, expected returns and risk from the 500 simulated data (partially consistent with Ledoit & Wolf (2014)).

	Ridge		HGLSE	
	In-sample	Out of sample	In-sample	Out of sample
Sharpe ratio	0.257	0.218	0.567	0.456
Expected returns	0.176	-0.075	0.158	0.055
Portfolio's risk	1.556	1.486	0.744	0.778
	Standard deviation		Standard deviation	
	In-sample	Out of sample	In-sample	Out of sample
Sharpe ratio	0.297	0.349	0.438	0.460
Expected returns	2.500	4.139	0.571	1.005
Portfolio's risk	9.492	8.367	2.336	2.819



(a) The bias in estimating weights

(b) Mean square errors

Figure 5.8: Comparison of the bias in estimating weights and mean square errors by ridge and HGLSE methods from datasets generated in Scenario 5.

5.6.2.2 Scenario 6

This scenario is similar to Scenario 5 in all aspects except in terms of predictor variables, $\mathbf{x}$, which do not follow a multivariate normal distribution. The idea of this simulation is to find out how the method performs in more general cases. The covariates are generated from a set of regression models with uniform random errors, which is given in detail in Appendix C.3. Similar to Scenario 5, the true portfolio weight $\mathbf{w}$ is also derived based on the true covariance matrix Σ , whose concentration matrix Σ^{-1} is given in Figure 5.9.

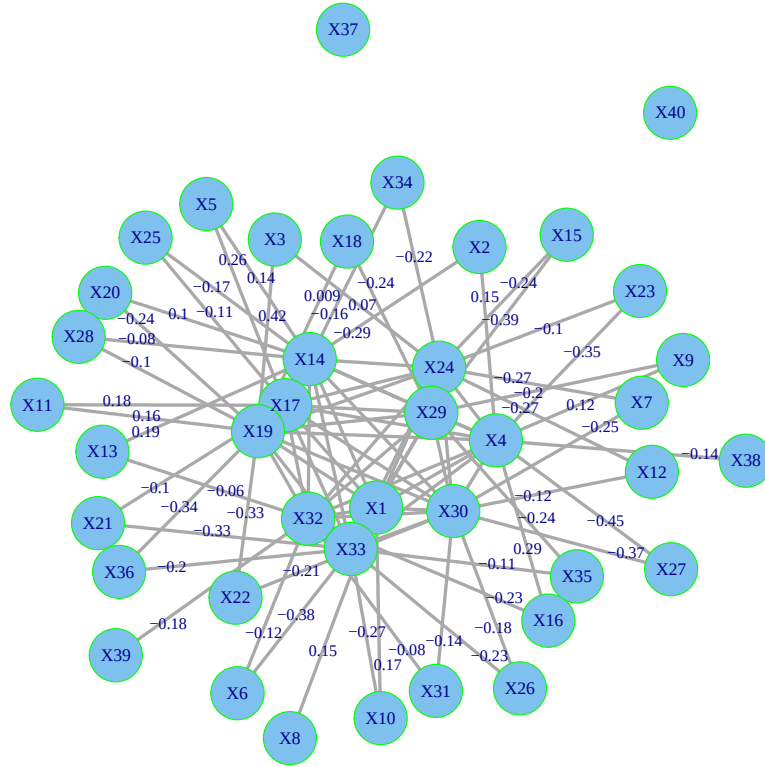


Figure 5.9: Graph structure for covariates under Scenario 6.

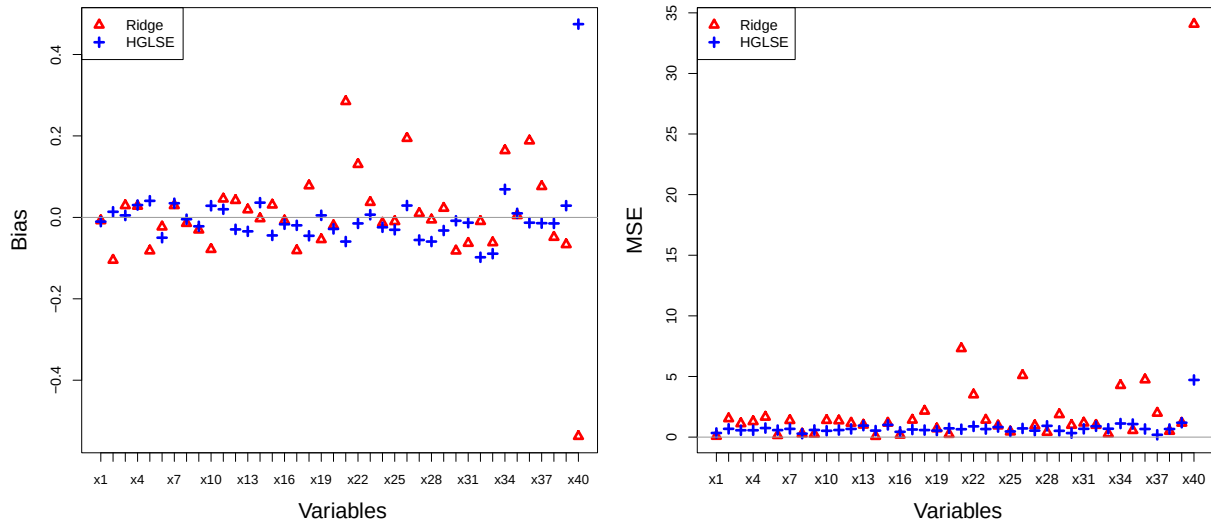
The results based on Scenario 6 are given in Table 5.9 and Figure 5.10. It can be seen from Table 5.9 that, out of the 500 simulated data, HGLSE yields higher means of the portfolio's Sharpe ratio and lower risk than ridge does for the in-

sample period. However, for the out-of-sample period, ridge gives higher means of the portfolio's Sharpe ratio than HGLSE does. It should be noted, though, that this may not be very accurate due to the fact that the mean standard deviations of the portfolio's risk and expected returns are very high compared to those under HGLSE. In addition, the out-of-sample expected returns are negative for ridge but positive for HGLSE, which is desirable in finance. With regards to the weight estimates, Figure 5.10 shows that HGLSE is better able to estimate the weights than ridge can do, since the bias detected by HGLSE in most of the weights is closer to zero than that detected by ridge. Moreover, the mean square error of HGLSE is much lower than that of ridge, which indicates that the performance of HGLSE is superior to that of ridge. Similarly, the Euclidean distance for HGLSE (0.084) is smaller than that for ridge (0.119).

The results of this scenario imply that even if the distribution of predictor variables does not follow multivariate normal distribution, unbiased estimates for the portfolio weights can still be obtained by HGLSE.

Table 5.9: The means and standard deviations of in-and-out-of-sample portfolio's Sharpe ratios, expected returns and risk from the 500 simulated data (partially consistent with Ledoit & Wolf (2014)).

	Ridge		HGLSE	
	In-sample	Out of sample	In-sample	Out of sample
Sharpe ratio	0.024	0.026	0.047	0.019
Expected returns	0.430	-1.200	0.092	0.128
Portfolio's risk	4.418	4.723	2.753	2.713
	Standard deviation		Standard deviation	
	In-sample	Out of sample	In-sample	Out of sample
Sharpe ratio	0.160	0.312	0.164	0.332
Expected returns	9.270	26.972	1.326	1.315
Portfolio's risk	59.607	67.214	7.068	7.254



(a) The bias in estimating weights

(b) Mean square errors

Figure 5.10: Comparison of mean square errors and the bias in estimating weights by ridge and HGLSE methods from data sets generated in Scenario 6.

5.7 Conclusion

A modified form of GLSE (given in Chapter 4) was proposed in this chapter. In the proposed method, GLSE is hybridised with a ridge type penalty imposed on some of the variables to become Hybridised Graphical Least Square Estimation (HGLSE). The ridge type penalty is imposed on those less important variables in a dataset. In this chapter, Lasso has been used as the variable pre-selection method. The chapter also gave a theorem related to the unbiasedness of the important variables while showing that biased coefficient estimates are obtained for less important variables due to the ridge type penalty.

In this chapter, the HGLSE method was also utilised to construct portfolio weights. For this purpose, a financial data analysis and two simulations were carried out to examine the performance of HGLSE in optimising portfolios. The results from the data analysis and the two simulation scenarios indicate that HGLSE can always outperform the ridge method, which also produces non-sparse portfolios, in maximising the portfolio's Sharpe ratios and expected rates of returns and minimising their risk.

For further assessing the proposed method, genes expression data analysis and four simulations were carried out in the chapter and the results were compared with those from Lasso and Lars in terms of data analysis and those from ridge, Lasso and Lars in terms of the simulation scenarios.

CHAPTER 6

Conclusion and future directions

This thesis has investigated issues related to finding regression models for high-dimensional regression datasets (with $n < v$), which continue to be a topical subject of interest in modern research (Hastie et al. 2009). It is well-known that the ordinary least squares (OLS) regression method is incapable of providing regression estimates for such datasets. To solve this problem, several alternative methods have been utilised, such as ridge regression, Lasso and Lars. Although these methods can deal with these problems when $n < v$, they yield biased estimates for regression coefficients. The bias in these methods is towards zero, as they shrink the coefficients in this direction. Some of these methods (notably Lasso and Lars) also have a model selection problem; they cannot include all the variables in the model if $n < v$. They can choose a maximum number of n variables (Hastie et al. 2009, Zou & Hastie 2005) in $n < v$ regression problems. Despite the fact that ridge regression does not face this problem and can include all the variables in the model, interpreting the model can be difficult when there are a great number of variables.

To solve the above mentioned $n < v$ regression problems, this thesis proposed the idea of graphical least square estimation (GLSE). The proposed procedure is based on the theory of graphical models, where least square estimation is used with a conditional independency structure existing in the dataset. One of the advantages of the proposed method is that coefficient estimates are unbiased under certain conditions. It also enjoys some properties of graphical models. For

example, as graphical models can depict the conditional correlation of variables, our method can thus find this structure. This advantage is a distinguishing feature of the GLSE against all the other existing methods. This feature makes it possible to consider the causal interpretation within predictors themselves and their relationship with the response, which helps to identify the true regression model. In addition to these advantages, the main results obtained in the thesis are summarised in the remainder of this chapter.

Initially, the thesis presented the core idea of graphical least square estimation (see Chapter 4), with specific and general forms of regression estimators. It was demonstrated by two supporting theorems (one for each of the two forms) that the proposed graphical least square estimator is unbiased. This might be desirable when unbiased estimation of regression coefficients is required, such as in financial applications. To further support the unbiasedness property of the proposed estimator, 4 different simulation scenarios were considered. The results of GLSE from the first three simulations were compared with those of ridge regression, Lasso and Lars in terms of coefficients estimates and Monte Carlo error estimates; bootstrap standard errors estimated for GLSE were also obtained from the generated data. The results indicated that the proposed GLSE outperforms all the other methods in terms of coefficients. In the fourth scenario, which was implemented to check the effect of increasing sample size n on the error estimates obtained via Formula (4.23), it is observed that the error estimates via the formula approach the Monte Carlo estimates if the sample is increased.

In addition to the above 4 simulation studies for the core GLSE estimator, analysis was conducted for a gene expression dataset. The results reveal that GLSE, unlike the other methods, is able to identify the conditional dependency structure of variables (graph) present in the data. Moreover, the GLSE gives the

smallest prediction mean square error obtained via the cross-validation method.

The proposed GLSE is extended to take a hybridised form of the graphical least square estimator (HGLSE) by imposing a ridge type penalty on those variables that are considered less important. This idea might be very helpful in reducing the search space of potential models. A specific form of the proposed modified estimator is given and is grounded in a theorem that shows its unbiasedness for the important variables. In the current version of using the estimator, Lasso is used at the initial stage to select important covariates.

Portfolio optimisation is a key concept in the world of finance as it efficiently allocates the amount of capital (weight) invested in each asset. As discussed in Section 5.4, optimising a portfolio can be achieved using a regression framework. Based on this, the proposed estimator (HGLSE) was applied to construct portfolios using a financial dataset from London Stock Exchange. Four different portfolios were constructed and the results indicate that HGLSE can always outperform ridge in maximising the Sharpe ratio and expected returns of the portfolio and minimising its risk. The proposed method also gives a better diversified portfolio risk in small and large datasets.

Additionally, an analysis of a gene expression dataset was carried out. The results show that HGLSE finds different probes from those found by Lasso and Lars and is able to detect the same highly frequent probes found by Wang et al. (2012). In addition, the prediction mean square error of HGLSE, obtained by the cross-validation method, proves to be smaller than those of Lasso and Lars.

Two sets of simulation scenarios were performed to assess the efficiency of HGLSE in terms of coefficient estimation and variable selection and to mimic portfolio optimisation. In the first set, 4 simulation scenarios were carried out. It is observed from the first scenario that the proposed method outperforms the others in terms of coefficient estimation for the important variables. The

Monte Carlo estimates of standard errors and the mean square error (MSE) for HGLSE are smaller than those of ridge, Lasso and Lars. As for the rest of the scenarios, the experiments implemented show that non-zero coefficients are better estimated by HGLSE than the other methods. In addition, in terms of variable selection, HGLSE outperforms Lasso and Lars.

In the second set, two simulations were conducted in order to mimic portfolio optimisation. The results were compared only with those of ridge (since only non-spare portfolios were considered). They indicated that HGLSE can always outperform ridge in maximising the Sharpe ratio and expected returns of the portfolio, and minimising its risk.

Considering both GLSE and HGLSE, it can be stated that the main difference between them and the traditional graphical models is that unlike the latter, the proposed methods do not consider the multivariate normality assumption for the predictors, but use a linear relationship assumption among them. This increases the scope of the application of the proposed methods in that variables in a dataset are not always normally distributed. Moreover, for high-dimensional ($n < v$) regression problems where unbiased estimates or saturated regression models are more important, the new methods provide a good alternative to existing regularisation methods.

6.1 Limitations and future direction

Further research is needed to improve the algorithms in the two proposed methods (GLSE and HGLSE) in order to reduce the computational cost. In addition, as suggested in the following sections, future research may be interested in exploring the possibility of creating new conditions that can assist the proposed methods in dealing with discrete variables, and in finding out the feasibility of

applying HGLSE to sparse portfolios. Moreover, one area that lends itself for further investigation is the likely extension of the methods to a larger family of distribution, such as elliptical distributions.

6.1.1 Discrete variables and robustness of the methods

With regard to discrete variables in the predictors datasets, both GLSE and HGLSE were applied to cases where only one binary variable was involved. To illustrate the case of having two binary variables in the predictor set, Condition 4.2.1 requires that the sets A , B and C form a decomposition of a given graph g with B as the separator. Then if either $A \subset \Psi$ or $C \subset \Psi$ and $B \subset \Psi$ (either $\mathbf{x}_A$ or $\mathbf{x}_C$ and $\mathbf{x}_B$ are quantitative). If $A \subset \Psi$ and $C \subset \Delta$, then the requirement in (4.2.1), $\mathbf{x}_A = \mathbf{x}_B \cdot \mathbf{r}_{BA} + \xi_A, \mathbb{E}(\xi_A) = \mathbf{0}$, is reasonable since $\mathbf{x}_A$ are quantitative variables. On the other hand, $\mathbf{x}_C = \mathbf{x}_B \cdot \mathbf{r}_{BC} + \xi_C, \mathbb{E}(\xi_C) = \mathbf{0}$ is also reasonable since it can be viewed as $\mathbf{x}_B = \mathbf{x}_C \cdot \mathbf{r}_{CB}^{-1} - \xi_C \cdot \mathbf{r}_{CB}^{-1}$ where $\mathbf{x}_B$ is quantitative and $\mathbf{r}_{CB}^{-1}$ is the generalised matrix inversion. Therefore, even if $\mathbf{x}_A$ or $\mathbf{x}_C$ are binary variables, the method still works, which indicates the robustness of the method. Moreover, in this case, the graph will be weakly decomposed (for illustration, see the example in 3.1.2).

However, in the case of having more than one categorical variable, Condition 4.2.1 may fail if $A \subset \Delta$ and $B \subset \Delta$ or $C \subset \Delta$ and $B \subset \Delta$, which is further illustrated in Simulation D.1 in Appendix D. In this simulation, there are two binary variables in the predictor set, one of which is a clique and the other is a separator. It can be seen that GLSE still works but gives a biased estimate for the binary variable in the clique. Nevertheless, this estimate of the coefficient is still much less biased than that given by the ridge, Lasso and Lars methods, indicating the robustness of GLSE. To obtain an unbiased estimate in such a situation using Condition 4.2.1, a new condition can be created where the separator B should

always belong to Ψ . This condition can guarantee that Condition 4.2.1 does not fail. However, how the two methods perform under the previously mentioned condition is an area for further research.

6.1.2 Sparse portfolio applications

While this thesis has addressed non-sparse (saturated) portfolios as far as HGLSE is concerned, looking into sparse portfolios remains a topic for extensive research. It would be interesting to see how HGLSE performs in terms of selecting assets that increase the portfolio's returns and reduce its risks in comparison with the performance of Lars and Lasso. This is due to the fact that HGLSE can have better selection than the other two methods, as was shown in the simulation scenarios in Chapter 5, and is efficient in dealing with highly correlated data, as was found in the real financial data analysis in Chapter 5.

6.1.3 Elliptical distributions

As both GLSE and HGLSE can be applied to multivariate normally distributed data, extending the proposed methods to elliptical distributions seems to form a natural generalisation of our methods to a larger family of distributions, since multivariate normal distribution is also a special case of this family (Christensen 1991). However, there might be some difficulties while doing so; for example, in the case of student t-distribution (one of the most commonly used elliptical distributions), the density function does not factor according to the corresponding graph (Finegold & Drton 2009) and thus there is no conditional independence. To see this, consider a v -dimensional random vector $\mathbf{X}$. If $\mathbf{X}$ is elliptically distributed, then its density function $f(\cdot)$ has the following form:

$$f(\mathbf{x}) = \delta \cdot \varphi((\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})) \quad (6.1)$$

where δ is the scale parameter (Lausen et al. 2014). As mentioned above, student t-distribution is a well known example of elliptical distributions and needs to be briefly defined.

The random vector X has a centered multivariate t-distribution with parameters Σ, μ, ν if its density has the following form:

$$f(x) = \frac{\Lambda[(\nu + v)/2]}{\Lambda(\nu/2)(\pi\nu)^{v/2}|\Sigma|^{1/2}} \left[1 + \frac{1}{\nu} x' \Sigma^{-1} x \right]^{-(\nu+v)/2} \quad (6.2)$$

where ν is the degree of freedom. As explained earlier, the above density cannot factorise according to the graph; hence, the density factorisation are investigated in the following lines.

Suppose that the set V can be partitioned into disjoint sets A, B and C . Also, assume that the random vector X is factorising according to a given (known) decomposable graph g , which consists of two cliques $A \cup B$ and $B \cup C$ and a separator B . Then the conditional density of the multivariate t-distribution has to satisfy the following:

$$f(x_{A \cup B}, x_{C \cup B} | x_B) = \frac{\left[1 + \frac{1}{\nu} (x_{A \cup B})' \Sigma_{A \cup B}^{-1} (x_{A \cup B}) \right]^{-(\nu+v_{A \cup B})/2} \left[1 + \frac{1}{\nu} (x_{B \cup C})' \Sigma^{-1} (x_{B \cup C}) \right]^{-(\nu+v_{B \cup C})/2}}{\left[1 + \frac{1}{\nu} (x_B)' \Sigma_B^{-1} (x_B) \right]^{-(\nu+v_B)/2}} \quad (6.3)$$

As can be seen in the above equation, the terms given in the exponent can be decomposed, whereas the base terms given in the square brackets cannot. As a result, it follows that Condition 4.2.1 cannot be fulfilled by the general elliptical distributions. Moreover, It should be noted that the components of X (follow a multivariate t-distribution) are not correlated does not mean that they are independent unless $\nu \rightarrow \infty$, where the distribution becomes a multivariate normal distribution (Kibria & Joarder 2006, Spanos 1999).

Moreover, the t-distribution can be used to generate the data in a similar way

to Scenarios 2 and 3 in Chapter 4 and Scenario 3 in Chapter 5. However, the sum of two t-distribution variables will not yield t-distribution, unlike the case in normal distribution, where the sum of two normally distributed variables gives normal distribution. If t-distribution is used to generate the data, as in the previously mentioned scenarios, then the resulting random variables will have unknown distribution. For this reason, extending the proposed methods to other elliptical distributions might be an appealing future direction.

Bibliography

- Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.
- Bai, J. & Shi, S. (2011), 'Estimating high dimensional covariance matrices and its applications.', *Annals of Economics & Finance* **12**(2), 199–215.
- Blalock, H. M. (1971), *Causal Models in the Social Sciences*, Chicago, Aldine.
- Bolstad, B. M., Irizarry, R. A., Åstrand, M. & Speed, T. P. (2003), 'A comparison of normalization methods for high density oligonucleotide array data based on variance and bias', *Bioinformatics* **19**(2), 185–193.
- Breheny, P. & Huang, J. (2015), 'Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors', *Statistics and Computing* **25**(2), 173–187.
- Britten-Jones, M. (1999), 'The sampling error in estimates of mean-variance efficient portfolio weights', *The Journal of Finance* **54**(2), 655–671.
- Butler, R. W. (1983), 'Optimal properties of one-step variable selection in regression', *The Annals of Statistics* **11**(1), 219–224.
- Candes, E. & Tao, T. (2007), 'The dantzig selector: Statistical estimation when p is much larger than n', *The Annals of Statistics* **35**(6), 2313–2351.
- Carrasco, M. & Noumon, N. (2011), Optimal portfolio selection using regularization, Technical report, University of Montreal. Retrieved from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.716.6710&rep=rep1&type=pdf>.

- Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q. & West, M. (2012), 'High-dimensional sparse factor modeling: Applications in gene expression genomics', *Journal of the American Statistical Association* **103**(484), 1438–1456.
- Carvalho, C. M. & West, M. (2007), 'Dynamic matrix-variate graphical models', *Bayesian Analysis* **2**(1), 69–97.
- Christensen, R. (1991), *Linear models for multivariate, time series, and spatial data*, Springer Science & Business Media.
- Croiseau, P. & Cordell, H. J. (2009), 'Analysis of North American rheumatoid arthritis consortium data using a penalized logistic regression approach', *BMC Proceedings* **3**(7), 1–5.
- Dai, H. (2008), 'Perfect sampling methods for random forests', *Advances in Applied Probability* **40**(3), 897–917.
- Dawid, A. P. & Lauritzen, S. L. (1993), 'Hyper markov laws in the statistical analysis of decomposable graphical models', *The Annals of Statistics* **21**(3), 1272–1317.
- DeMiguel, V., Garlappi, L., Nogales, F. J. & Uppal, R. (2009), 'A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms', *Management Science* **55**(5), 798–812.
- Dempster, A. P. (1972), 'Covariance selection', *Biometrics* **28**(1), 157–175.
- Dobra, A., Hans, C., Jones, B., Nevins, J. R., Yao, G. & West, M. (2004), 'Sparse graphical models for exploring gene expression data', *Journal of Multivariate Analysis* **90**(1), 196–212.
- Drton, M. & Perlman, M. D. (2004), 'Model selection for Gaussian concentration graphs', *Biometrika* **91**(3), 591–602.

- Edwards, D. (2000), *Introduction to Graphical Modelling*, Springer.
- Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. (2004), 'Least angle regression', *The Annals of statistics* **32**(2), 407–499.
- Efron, B., Hastie, T. & Tibshirani, R. (2007), 'Discussion of "Dantzig selector" by Candes and Tao', *Annals of Statistics* **35**(6), 2358–2364.
- Efron, B. & Tibshirani, R. J. (1994), *An Introduction to the Bootstrap*, CRC press.
- Ennew, C., Watkins, T. & Wright, M. (2014), *Cases in Marketing Financial Services*, Oxford, Butterworth-Heinemann Limited.
- Fan, J., Han, F. & Liu, H. (2014), 'Challenges of big data analysis', *National Science Review* **1**(2), 293–314.
- Fan, J. & Li, R. (2001), 'Variable selection via nonconcave penalized likelihood and its oracle properties', *Journal of the American Statistical Association* **96**(456), 1348–1360.
- Fan, J. & Lv, J. (2010), 'A selective overview of variable selection in high dimensional feature space', *Statistica Sinica* **20**(1), 101.
- Fastrich, B., Paterlini, S. & Winker, P. (2015), 'Constructing optimal sparse portfolios using regularization methods', *Computational Management Science* **12**(3), 417–434.
- Ferraty, F. (2010), 'High-dimensional data: A fascinating statistical challenge', *Journal of Multivariate Analysis* **101**(2), 305–306.
- Finegold, M. A. & Drton, M. (2009), Robust graphical modeling with t-distributions, in 'Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence', AUAI Press, pp. 169–176.

- Friedman, J., Hastie, T. & Tibshirani, R. (2010), *A note on the group lasso and a sparse group lasso*. Old Dominion University, Department of Computer Science. Retrieved from: <http://www.cs.odu.edu/~sji/classes/dm-2013s/papers/sparse-grlasso.pdf>.
- Gilli, M., Maringer, D. & Schumann, E. (2011), *Numerical Methods and Optimization in Finance*, Academic Press.
- Goeman, J., Meijer, R. & Chaturvedi, N. (2012), L1 and l2 penalized regression models, Technical report, Documentation for the *penalized* package for the statistical environment R. Retrieved from: <https://cran.r-project.org/web/packages/penalized/vignettes/penalized.pdf>.
- Golumbic, M. C. (2004), *Algorithmic Graph Theory and Perfect Graphs*, Vol. 57, Elsevier.
- Guo, P., Zeng, F., Hu, X., Zhang, D., Zhu, S., Deng, Y. & Hao, Y. (n.d.), 'Improved variable selection algorithm using a lasso-type penalty, with an application to assessing hepatitis b infection relevant factors in community residents', *PloS one* **10**(7).
- Guyon, I. & Elisseeff, A. (2003), 'An introduction to variable and feature selection', *The Journal of Machine Learning Research* **3**, 1157–1182.
- Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J. & Tibshirani, R. (2009), *The Elements of Statistical Learning*, Vol. 2, Springer.
- Hoerl, A. E. & Kennard, R. W. (1970), 'Ridge regression: Biased estimation for nonorthogonal problems', *Technometrics* **12**(1), 55–67.
- Højsgaard, S., Edwards, D. & Lauritzen, S. (2012), *Graphical models with R*, Springer.

- Huang, J., Breheny, P. & Ma, S. (2012), 'A selective review of group selection in high-dimensional models', *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* **27**(4).
- Huang, J., Ma, S. & Zhang, C.-H. (2008), 'Adaptive lasso for sparse high-dimensional regression models', *Statistica Sinica* **18**, 1603–1618.
- Ivashina, V. & Scharfstein, D. (2010), 'Bank lending during the financial crisis of 2008', *Journal of Financial economics* **97**(3), 319–338.
- James, G. M. & Radchenko, P. (2009), 'A generalized dantzig selector with shrinkage tuning', *Biometrika* **96**(2), 323–337.
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013), *An Introduction to Statistical Learning*, Vol. 112, Springer.
- Johnstone, I. M. & Titterton, D. M. (2009), 'Statistical challenges of high-dimensional data', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **367**(1906), 4237–4253.
- Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C. & West, M. (2005), 'Experiments in stochastic computation for high-dimensional graphical models', *Statistical Science* **20**(4), 388–400.
- Jordan, M. I. (1998), *Learning in Graphical Models*, Vol. 89, Springer.
- Kibria, B. G. & Joarder, A. H. (2006), 'A short review of multivariate t-distribution', *Journal of Statistical research* **40**(1), 59–72.
- Krämer, N., Schäfer, J. & Boulesteix, A.-L. (2009), 'Regularized estimation of large-scale gene association networks using graphical Gaussian models', *BMC Bioinformatics* p. 384.

- Lai, T. L., Xing, H. & Chen, Z. (2011), 'Mean-variance portfolio optimization when means and covariances are unknown', *The Annals of Applied Statistics* 5(2A), 798–823.
- Lauritzen, S. L. (1996), *Graphical Models*, Oxford, Oxford University Press.
- Lausen, B., Poel, D. V. d. & Ultsch, A. (2014), *Algorithms from and for nature and life*, Springer.
- Ledoit, O. & Wolf, M. (2003), 'Improved estimation of the covariance matrix of stock returns with an application to portfolio selection', *Journal of Empirical Finance* 10(5), 603–621.
- Ledoit, O. & Wolf, M. (2014), Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks, Technical report. Retrieved from: <http://ssrn.com/abstract=2383361>.
- Lee, C.-F. & Lee, J. (2010), *Handbook of Quantitative Finance and Risk Management*, Springer Science & Business Media.
- Levi, M. D. (2009), *International Finance (5th edition)*, London: Routledge.
- Li, Q. & Shao, J. (2015), 'Regularizing lasso: A consistent variable selection method', *Statistica Sinica* 25, 975–992.
- Li, X., Zhao, T., Wang, L., Yuan, X. & Liu, H. (2014), *flare: Family of Lasso Regression*. R package version 1.1.0. Retrieved from: <http://CRAN.R-project.org/package=flare>.
- Maimon, O. & Rokach, L. (2005), *Data Mining and Knowledge Discovery Handbook*, Vol. 2, Springer.

- Malkiel, B. G. & Xu, Y. (1997), 'Risk and return revisited', *The Journal of Portfolio Management* **23**(3), 9–14.
- Markowitz, H. (1952), 'Portfolio selection', *The Journal of Finance* **7**(1), 77–91.
- Meinshausen, N. & Bühlmann, P. (2006), 'High-dimensional graphs and variable selection with the lasso', *The Annals of Statistics* **34**(3), 1436–1462.
- Mordelet, F., Horton, J., Hartemink, A. J., Engelhardt, B. E. & Gordân, R. (2013), 'Stability selection for regression-based models of transcription factor–DNA binding specificity', *Bioinformatics* **29**(13), i117–i125.
- Radchenko, P., James, G. M. et al. (2011), 'Improved variable selection with forward-lasso adaptive shrinkage', *The Annals of Applied Statistics* **5**(1), 427–448.
- Raol, J. R., Girija, G. & Singh, J. (2004), *Modelling and Parameter Estimation of Dynamic Systems*, Vol. 65, Iet.
- Rosset, S. & Zhu, J. (2004), 'Discussion of "least angle regression" by Efron et al', *Annals of Statistics* **32**(2), 469–475.
- Scheetz, T. E., Kim, K.-Y. A., Swiderski, R. E., Philp, A. R., Braun, T. A., Knudtson, K. L., Dorrance, A. M., DiBona, G. F., Huang, J. & Casavant, T. L. (2006), 'Regulation of gene expression in the mammalian eye and its relevance to eye disease', *Proceedings of the National Academy of Sciences* **103**(39), 14429–14434.
- Schwarz, G. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**(2), 461–464.
- Seetharaman, I. (2013), Consistent bi-level variable selection via composite group bridge penalized regression, PhD thesis, Kansas State University.

- Spanos, A. (1999), *Probability theory and statistical inference: econometric modeling with observational data*, Cambridge University Press.
- Still, S. & Kondor, I. (2010), 'Regularizing portfolio optimization', *New Journal of Physics* **12**(7).
- Sun, T. & Zhang, C.-H. (2010), 'Comments on: ℓ_1 -penalization for mixture regression models', *Test* **19**(2), 270–275.
- Talluri, R. & Shete, S. (2014), 'Gaussian graphical models for phenotypes using pedigree data and exploratory analysis using networks with genetic and nongenetic factors based on genetic analysis workshop 18 data', **8**, S99.
- Tibshirani, R. (2011), 'Regression shrinkage and selection via the lasso: A retrospective', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **58**(1), 267–288.
- Uslu, V. R., Egrioglu, E. & Bas, E. (2014), 'Finding optimal value for the shrinkage parameter in ridge regression via particle swarm optimization', *American Journal of Intelligent Systems* **4**(4), 142–147.
- Van Houwelingen, H. & Putter, H. (2011), *Dynamic Prediction in Clinical Survival Analysis*, CRC Press.
- Vassalou, M. & Xing, Y. (2004), 'Default risk in equity returns', *The Journal of Finance* **59**(2), 831–868.
- Wang, L., Wu, Y. & Li, R. (2012), 'Quantile regression for analyzing heterogeneity in ultra-high dimension', *Journal of the American Statistical Association* **107**(497), 214–222.
- Wermuth, N. (1976), 'Model search among multiplicative models', *Biometrics* **32**, 253–263.

- Whittaker, J. (2009), *Graphical Models in Applied Multivariate Statistics*, Wiley Publishing.
- Wold, H. (1954), 'Causality and econometrics', *Econometrica: Journal of the Econometric Society* **22**(2), 162–177.
- Wright, S. (1923), 'The theory of path coefficients: A reply to Niles's criticism', *Genetics* **8**(3), 239–255.
- Yuan, M., Ekici, A., Lu, Z. & Monteiro, R. (2007), 'Dimension reduction and coefficient estimation in multivariate linear regression', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**(3), 329–346.
- Yuan, M. & Lin, Y. (2006), 'Model selection and estimation in regression with grouped variables', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(1), 49–67.
- Zhang, C.-H. (2010), 'Nearly unbiased variable selection under minimax concave penalty', *The Annals of Statistics* **38**(2), 894–942.
- Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American statistical association* **101**(476), 1418–1429.
- Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2), 301–320.
- Zou, H. & Zhang, H. H. (2009), 'On the adaptive elastic-net with a diverging number of parameters', *Annals of Statistics* **37**(4), 1733–1751.

APPENDIX A

Concentration matrix for Chapter 2 simulations

Below is the full concentration matrix for the the simulation performed in Chapter 2. Due to the size of the concentration matrix K , it has been divided into two parts, K_1 and K_2 i.e. $K = [K_1 \ K_2]$.

$$K_1 = \begin{pmatrix} 3.155 & 0 & 0 & -0.181 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.25 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4.602 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.494 & 0 & 0 & -0.137 & 0 & 0 \\ 0 & 0 & 3.446 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.86 & 0 \\ -0.181 & 0 & 0 & 4.372 & 0 & -0.372 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.513 & 0 & 0 & -0.166 & 0 & -0.632 \\ 0 & 0 & 0 & 0 & 2.737 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.127 & 0 & 0 \\ 0 & 0 & 0 & -0.372 & 0 & 2.082 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3.112 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.274 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3.421 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4.987 & -0.272 & 0 & 0 & 0 & -0.525 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.272 & 4.019 & 0 & 0 & 0 & -0.487 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3.828 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4.194 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4.274 & 0 & 0 & 0 & 0 & 0 & -0.498 \\ -0.25 & -0.494 & 0 & -0.513 & 0 & 0 & 0 & 0 & -0.525 & -0.487 & 0 & 0 & 0 & 4.259 & 0 & -0.341 & -0.112 & 0 & -0.591 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.198 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.341 & 0 & 2.239 & 0 & 0 & 0 \\ 0 & -0.137 & 0 & -0.166 & -0.127 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.112 & 0 & 0 & 3.211 & 0 & -0.622 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.865 & 0 & 0 \\ 0 & 0 & -0.86 & -0.632 & 0 & 0 & -0.274 & 0 & 0 & 0 & 0 & -0.498 & -0.591 & 0 & 0 & 0 & -0.622 & 0 & 4.859 \\ 0 & 0 & 0 & -0.646 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3.977 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.242 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.175 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.581 & 0 & 0 & 0 & 0 & -0.596 \\ 0 & 0 & 0 & -0.396 & 0 & 0 & 0.729 & 0 & 0 & 0 & 0.258 & 0.332 & 0.241 & 0.692 & 0.277 & 0 & 0.612 & 0 & 0.779 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.423 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.172 & 0.778 & 0.196 & 0 & 0 & 0.402 & 0 & 0 & 0 & 0.706 & 0 & 0.244 & 0 & 0.734 & 0.137 & 0.573 & 0.207 & 0 \\ 0 & 0 & 0.713 & 0.785 & 0 & 0 & 0 & 0 & 0 & 0 & 0.73 & 0 & 0 & 0.752 & 0 & 0.731 & 0.39 & 0 & 0.261 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.79 & 0 & 0 & 0.164 & 0.129 & 0 & 0 & 0 & 0 & 0 & 0.49 & 0 & 0 & 0.759 & 0.309 & 0.658 & 0 \\ 0 & 0 & 0.768 & 0.182 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.448 & 0.495 & 0 & 0.411 & 0.202 & 0 & 0.545 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.624 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.733 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.453 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.209 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.781 & 0.736 & 0 & 0 & 0 & 0.799 & 0 & 0 & 0.377 & 0 & 0 & 0.547 & 0 & 0 & 0.374 & 0 & 0.116 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.58 & 0 & 0 & 0.483 & 0 & 0 \\ 0 & 0 & 0 & 0.483 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.15 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

$K_2 =$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	0	0	0	0	0	0	0	0	0.172	0.713	0	0	0.768	0	0	0	0	0.781477	0	0	
	0	0	0	-0.396	0	0.423	0	0	0.778	0.785	0	0.79	0.182	0	0	0.733	0.453	0.736	0	0.483	
	0	0	0	0	0	0	0	0	0.196	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0.729	0	0	0	0	0	0	0	0.164	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0.402	0	0	0.129	0	0	0	0	0	0	0.799	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0.258	0	0	0	0	0	0	0.73	0	0	0	0	0	0	0	0.377	0	0
	0	0	0	0.332	0	0	0	0	0.706	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0.241	0	0	0	0	0	0	0	0	0.448	0	0	0	0	0	0	0	0
	-0.242	0	-0.581	0.692	0	0	0	0	0.244	0.752	0	0.49	0.495	0.624	0	0	0.209	0.547	0.58	0.15	
	0	0	0	0.277	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0.734	0.731	0	0	0.411	0	0	0	0	0	0	0	0
	0	-0.175	0	0.612	0	0	0	0	0.137	0.39	0	0.759	0.202	0	0	0	0	0.374	0.483	0	
	0	0	0	0	0	0	0	0	0.573	0	0	0.309	0	0	0	0	0	0	0	0	0
	0	0	-0.596	0.779	0	0	0	0	0.207	0.261	0	0.658	0.545	0	0	0	0	0.116	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2.016	0	0	0	0	0	0	0	0	0	0	0.406	0.239	0	0	0	0	0	0	0	0
	0	4.691	0	0	0	0	0	0	0	0	0	0.339	0	0	0	0	0	0.548	0	0	0
	0	0	3.927	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	4.125	0	0.138	0.106	0	0.741	0.722	0.612	0.694	0.61	0	0	0.309	0	0.722	0	0	0
	0	0	0	0	4.721	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0.138	0	4.762	0	0	0	0.611	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0.106	0	0	4.055	0	0	0	0	0	0.308	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	4.458	0	0	0	0	0	0	0	0	0	0.196	0	0	0
	0	0	0																		

APPENDIX B

Data generation procedure in Chapter 4

This appendix presents the process of data generation for Scenario 3, 4 and 5 of Chapter 4.

B.1 Scenario 3

Using the same graph structure given in Scenario 1 (Chapter 4), the perfect ordering of cliques (excluding X_{18} and X_{20}) is $C_1 = \{X_3, X_4, X_7\}$, $C_2 = \{X_1, X_3, X_4\}$, $C_3 = \{X_2, X_3, X_4\}$, $C_4 = \{X_3, X_5, X_7\}$, $C_5 = \{X_4, X_6, X_7\}$, $C_6 = \{X_4, X_7, X_8\}$, $C_7 = \{X_4, X_7, X_{11}\}$, $C_8 = \{X_3, X_7, X_9\}$, $C_9 = \{X_3, X_4, X_{10}\}$, $C_{10} = \{X_6, X_{12}\}$, $C_{11} = \{X_7, X_{13}\}$, $C_{12} = \{X_2, X_{14}\}$, $C_{13} = \{X_1, X_{15}\}$, $C_{14} = \{X_{10}, X_{16}\}$, $C_{15} = \{X_{11}, X_{17}\}$, $C_{16} = \{X_{15}, X_{19}\}$.

We generated 3 variables X_3, X_4, X_7 from $MVN(\mathbf{0}, \Sigma^{-1} = L_1)$ where

$$K_1 = \begin{pmatrix} 0.230 & 0.080 & 0.019 \\ 0.080 & 0.602 & 0.007 \\ 0.019 & 0.007 & 0.472 \end{pmatrix}.$$

Variable X_{18} is generated from $N(0, 0.352)$ and variable X_{20} is generated from $N(0, 0.453)$.

Then we use the linear condition

$$\mathbf{x}_{C_k \setminus S_k} = \mathbf{x}_{S_k} \cdot \mathbf{r}_{S_k, C_k \setminus S_k} + \boldsymbol{\xi}_k, \quad (\text{B.1.1})$$

to generate all the rest variables, as follows:

1. We have $C_2 \setminus C_1 = \{X_1\}$, $C_2 \cap C_1 = \{X_3, X_4\}$ and generate

$$X_1 = (X_3, X_4) \cdot \mathbf{r}_{S_2, C_2 \setminus S_2} + \xi_2, \quad (\text{B.1.2})$$

$$\text{where } \mathbf{r}_{S_2, C_2 \setminus S_2} = \begin{pmatrix} -0.3 \\ 0.25 \end{pmatrix}, \text{ and } \xi_2 \sim U(-0.5, 0.5);$$

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_2\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_3, X_4\}$ and generate

$$X_2 = (X_3, X_4) \cdot \mathbf{r}_{S_3, C_3 \setminus S_3} + \xi_3, \quad (\text{B.1.3})$$

$$\text{where } \mathbf{r}_{S_3, C_3 \setminus S_3} = \begin{pmatrix} 0.45 \\ -0.35 \end{pmatrix}, \text{ and } \xi_3 \sim U(-0.5, 0.5);$$

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_5\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_3, X_7\}$ and generate

$$X_5 = (X_3, X_7) \cdot \mathbf{r}_{S_4, C_4 \setminus S_4} + \xi_4, \quad (\text{B.1.4})$$

$$\text{where } \mathbf{r}_{S_4, C_4 \setminus S_4} = \begin{pmatrix} 0.4 \\ -0.3 \end{pmatrix}, \text{ and } \xi_4 \sim U(-0.5, 0.5);$$

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_6\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_4, X_7\}$ and generate

$$X_6 = (X_4, X_7) \cdot \mathbf{r}_{S_5, C_5 \setminus S_5} + \xi_5, \quad (\text{B.1.5})$$

$$\text{where } \mathbf{r}_{S_5, C_5 \setminus S_5} = \begin{pmatrix} -0.2 \\ 0.3 \end{pmatrix}, \text{ and } \xi_5 \sim U(-0.5, 0.5);$$

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_8\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_4, X_7\}$ and generate

$$X_8 = (X_4, X_7) \cdot \mathbf{r}_{S_6, C_6 \setminus S_6} + \xi_6, \quad (\text{B.1.6})$$

$$\text{where } \mathbf{r}_{S_6, C_6 \setminus S_6} = \begin{pmatrix} -0.23 \\ -0.5 \end{pmatrix}, \text{ and } \xi_6 \sim U(-0.5, 0.5);$$

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_{11}\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_4, X_7\}$ and generate

$$X_{11} = (X_4, X_7) \cdot \mathbf{r}_{S_7, C_7 \setminus S_7} + \xi_7, \quad (\text{B.1.7})$$

$$\text{where } \mathbf{r}_{S_7, C_7 \setminus S_7} = \begin{pmatrix} 0.3 \\ 0.2 \end{pmatrix}, \text{ and } \xi_7 \sim U(-0.5, 0.5);$$

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_9\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_3, X_7\}$ and generate

$$X_9 = (X_3, X_7) \cdot \mathbf{r}_{S_8, C_8 \setminus S_8} + \xi_8, \quad (\text{B.1.8})$$

$$\text{where } \mathbf{r}_{S_8, C_8 \setminus S_8} = \begin{pmatrix} 0.26 \\ -0.35 \end{pmatrix}, \text{ and } \xi_8 \sim U(-0.5, 0.5);$$

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_{10}\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_3, X_4\}$ and generate

$$X_{10} = (X_3, X_4) \cdot \mathbf{r}_{S_9, C_9 \setminus S_9} + \xi_9, \quad (\text{B.1.9})$$

$$\text{where } \mathbf{r}_{S_9, C_9 \setminus S_9} = \begin{pmatrix} 0.2 \\ 0.35 \end{pmatrix}, \text{ and } \xi_9 \sim U(-0.5, 0.5);$$

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_{12}\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_6\}$ and generate

$$X_{12} = X_6 \cdot \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} + \xi_{10}, \quad (\text{B.1.10})$$

$$\text{where } \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} = (0.25), \text{ and } \xi_{10} \sim U(-0.5, 0.5);$$

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{13}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_7\}$ and generate

$$X_{13} = X_7 \cdot \mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} + \xi_{11}, \quad (\text{B.1.11})$$

where $\mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} = (0.3)$, and $\xi_{11} \sim U(-0.5, 0.5)$;

11. We have $C_{12} \setminus \{\cup_{k=1}^{11} C_k\} = \{X_{14}\}$, $C_{12} \cap \{\cup_{k=1}^{11} C_k\} = \{X_2\}$ and generate

$$X_{14} = X_2 \cdot \mathbf{r}_{s_{12}, C_{12} \setminus s_{12}} + \xi_{12}, \quad (\text{B.1.12})$$

where $\mathbf{r}_{s_{12}, C_{12} \setminus s_{12}} = (0.32)$, and $\xi_{12} \sim U(-0.5, 0.5)$;

12. We have $C_{13} \setminus \{\cup_{k=1}^{12} C_k\} = \{X_{15}\}$, $C_{13} \cap \{\cup_{k=1}^{12} C_k\} = \{X_1\}$ and generate

$$X_{15} = X_1 \cdot \mathbf{r}_{s_{13}, C_{13} \setminus s_{13}} + \xi_{13}, \quad (\text{B.1.13})$$

where $\mathbf{r}_{s_{13}, C_{13} \setminus s_{13}} = (0.2)$, and $\xi_{13} \sim U(-0.5, 0.5)$;

13. We have $C_{14} \setminus \{\cup_{k=1}^{13} C_k\} = \{X_{16}\}$, $C_{14} \cap \{\cup_{k=1}^{13} C_k\} = \{X_{10}\}$ and generate

$$X_{16} = X_{10} \cdot \mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} + \xi_{14}, \quad (\text{B.1.14})$$

where $\mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} = (0.35)$, and $\xi_{14} \sim U(-0.5, 0.5)$;

14. We have $C_{15} \setminus \{\cup_{k=1}^{14} C_k\} = \{X_{17}\}$, $C_{15} \cap \{\cup_{k=1}^{14} C_k\} = \{X_{11}\}$ and generate

$$X_{17} = X_{11} \cdot \mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} + \xi_{15}, \quad (\text{B.1.15})$$

where $\mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} = (0.4)$, and $\xi_{15} \sim U(-0.5, 0.5)$;

15. We have $C_{16} \setminus \{\cup_{k=1}^{15} C_k\} = \{X_{19}\}$, $C_{16} \cap \{\cup_{k=1}^{15} C_k\} = \{X_{15}\}$ and generate

$$X_{19} = X_{15} \cdot \mathbf{r}_{S_{16}, C_{16} \setminus S_{16}} + \xi_{16}, \quad (\text{B.1.16})$$

where $\mathbf{r}_{S_{16}, C_{16} \setminus S_{16}} = (0.3)$, and $\xi_{16} \sim U(-0.5, 0.5)$;

B.2 Scenario 4

Using the same graph structure given in Scenario 4 (Chapter 4), the perfect ordering of cliques (excluding X_{25}) is

$$\begin{aligned} C_1 &= \{X_1, X_2, X_3\}, C_2 = \{X_2, X_3, X_4\}, C_3 = \{X_1, X_2, X_5\}, C_4 = \{X_1, X_3, X_6\}, \\ C_5 &= \{X_2, X_5, X_7\}, C_6 = \{X_2, X_4, X_8\}, C_7 = \{X_1, X_5, X_9\}, C_8 = \{X_1 X_6, X_{10}\}, \\ C_9 &= \{X_4, X_8 X_{11}\}, C_{10} = \{X_3, X_4, X_{12}\}, C_{11} = \{X_5, X_7, X_{13}\}, C_{12} = \{X_3, X_6, X_{14}\}, \\ C_{13} &= \{X_6, X_{10}, X_{15}\}, C_{14} = \{X_1, X_2, X_{16}\}, C_{15} = \{X_2, X_3, X_{17}\}, C_{16} = \{X_1, X_3, X_{19}\}, \\ C_{17} &= \{X_2, X_8, X_{21}\}, C_{18} = \{X_4, X_{12}, X_{22}\}, C_{19} = \{X_5, X_9, X_{23}\}, C_{20} = \{X_1, X_9, X_{24}\}, \\ C_{21} &= \{X_{12}, X_{22}, X_{29}\}, C_{22} = \{X_{18}, X_{16}\}, C_{23} = \{X_{17}, X_{20}\}, C_{24} = \{X_{13}, X_{26}\}, \\ C_{25} &= \{X_{20}, X_{27}\}, C_{26} = \{X_{11}, X_{28}\}, C_{27} = \{X_{15}, X_{30}\}. \end{aligned}$$

We generated 3 variables X_3, X_4, X_7 from $MVN(\mathbf{0}, \Sigma^{-1} = L_1)$ where

$$K_1 = \begin{pmatrix} 0.321 & 0.009 & 0.001 \\ 0.009 & 0.346 & 0.005 \\ 0.001 & 0.005 & 0.388 \end{pmatrix}.$$

Variable X_{25} is generated from $N(0, 0.472)$.

Then we use the linear condition

$$\mathbf{x}_{C_k \setminus S_k} = \mathbf{x}_{S_k} \cdot \mathbf{r}_{S_k, C_k \setminus S_k} + \xi_k, \quad (\text{B.2.17})$$

to generate all the rest variables, as follows:

1. We have $C_2 \setminus C_1 = \{X_4\}$, $C_2 \cap C_1 = \{X_2, X_3\}$ and generate

$$X_4 = (X_2, X_3) \cdot \mathbf{r}_{S_2, C_2 \setminus S_2} + \xi_2, \quad (\text{B.2.18})$$

$$\text{where } \mathbf{r}_{S_2, C_2 \setminus S_2} = \begin{pmatrix} 0.01 \\ 0.05 \end{pmatrix}, \text{ and } \xi_2 \sim U(-0.5, 0.5);$$

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_5\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_1, X_2\}$ and generate

$$X_5 = (X_1, X_2) \cdot \mathbf{r}_{S_3, C_3 \setminus S_3} + \xi_3, \quad (\text{B.2.19})$$

$$\text{where } \mathbf{r}_{S_3, C_3 \setminus S_3} = \begin{pmatrix} -0.12 \\ -0.01 \end{pmatrix}, \text{ and } \xi_3 \sim U(-0.5, 0.5);$$

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_6\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_1, X_3\}$ and generate

$$X_6 = (X_1, X_3) \cdot \mathbf{r}_{S_4, C_4 \setminus S_4} + \xi_4, \quad (\text{B.2.20})$$

$$\text{where } \mathbf{r}_{S_4, C_4 \setminus S_4} = \begin{pmatrix} 0.08 \\ -0.05 \end{pmatrix}, \text{ and } \xi_4 \sim U(-0.5, 0.5);$$

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_7\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_2, X_5\}$ and generate

$$X_7 = (X_2, X_5) \cdot \mathbf{r}_{S_5, C_5 \setminus S_5} + \xi_5, \quad (\text{B.2.21})$$

$$\text{where } \mathbf{r}_{S_5, C_5 \setminus S_5} = \begin{pmatrix} -0.1 \\ 0.08 \end{pmatrix}, \text{ and } \xi_5 \sim U(-0.5, 0.5);$$

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_8\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_2, X_4\}$ and generate

$$X_8 = (X_2, X_4) \cdot \mathbf{r}_{S_6, C_6 \setminus S_6} + \xi_6, \quad (\text{B.2.22})$$

where $\mathbf{r}_{S_6, C_6 \setminus S_6} = \begin{pmatrix} 0.05 \\ 0.1 \end{pmatrix}$, and $\xi_6 \sim U(-0.5, 0.5)$;

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_9\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_1, X_5\}$ and generate

$$X_9 = (X_1, X_5) \cdot \mathbf{r}_{S_7, C_7 \setminus S_7} + \xi_7, \quad (\text{B.2.23})$$

where $\mathbf{r}_{S_7, C_7 \setminus S_7} = \begin{pmatrix} -0.1 \\ 0.1 \end{pmatrix}$, and $\xi_7 \sim U(-0.5, 0.5)$;

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_{10}\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_1, X_6\}$ and generate

$$X_{10} = (X_1, X_6) \cdot \mathbf{r}_{S_8, C_8 \setminus S_8} + \xi_8, \quad (\text{B.2.24})$$

where $\mathbf{r}_{S_8, C_8 \setminus S_8} = \begin{pmatrix} -0.001 \\ -0.01 \end{pmatrix}$, and $\xi_8 \sim U(-0.5, 0.5)$;

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_{11}\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_4, X_8\}$ and generate

$$X_{11} = (X_4, X_8) \cdot \mathbf{r}_{S_9, C_9 \setminus S_9} + \xi_9, \quad (\text{B.2.25})$$

where $\mathbf{r}_{S_9, C_9 \setminus S_9} = \begin{pmatrix} -0.02 \\ 0.05 \end{pmatrix}$, and $\xi_9 \sim U(-0.5, 0.5)$;

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_{12}\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_3, X_4\}$ and generate

$$X_{12} = (X_3, X_4) \cdot \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} + \xi_{10}, \quad (\text{B.2.26})$$

where $\mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} = \begin{pmatrix} -0.03 \\ -0.01 \end{pmatrix}$, and $\xi_{10} \sim U(-0.5, 0.5)$;

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{13}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_5, X_7\}$ and generate

$$X_{13} = (X_5, X_7) \cdot \mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} + \xi_{11}, \quad (\text{B.2.27})$$

$$\text{where } \mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} = \begin{pmatrix} -0.012 \\ 0.02 \end{pmatrix}, \text{ and } \xi_{11} \sim U(-0.5, 0.5);$$

11. We have $C_{12} \setminus \{\cup_{k=1}^{11} C_k\} = \{X_{14}\}$, $C_{12} \cap \{\cup_{k=1}^{11} C_k\} = \{X_3, X_6\}$ and generate

$$X_{14} = (X_3, X_6) \cdot \mathbf{r}_{s_{12}, C_{12} \setminus s_{12}} + \xi_{12}, \quad (\text{B.2.28})$$

$$\text{where } \mathbf{r}_{s_{12}, C_{12} \setminus s_{12}} = \begin{pmatrix} -0.015 \\ -0.021 \end{pmatrix}, \text{ and } \xi_{12} \sim U(-0.5, 0.5);$$

12. We have $C_{13} \setminus \{\cup_{k=1}^{12} C_k\} = \{X_{15}\}$, $C_{13} \cap \{\cup_{k=1}^{12} C_k\} = \{X_6, X_{10}\}$ and generate

$$X_{15} = (X_6, X_{10}) \cdot \mathbf{r}_{s_{13}, C_{13} \setminus s_{13}} + \xi_{13}, \quad (\text{B.2.29})$$

$$\text{where } \mathbf{r}_{s_{13}, C_{13} \setminus s_{13}} = \begin{pmatrix} -0.07 \\ 0.018 \end{pmatrix}, \text{ and } \xi_{13} \sim U(-0.5, 0.5);$$

13. We have $C_{14} \setminus \{\cup_{k=1}^{13} C_k\} = \{X_{16}\}$, $C_{14} \cap \{\cup_{k=1}^{13} C_k\} = \{X_1, X_2\}$ and generate

$$X_{16} = (X_1, X_2) \cdot \mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} + \xi_{14}, \quad (\text{B.2.30})$$

$$\text{where } \mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} = \begin{pmatrix} 0.02 \\ 0.01 \end{pmatrix}, \text{ and } \xi_{14} \sim U(-0.5, 0.5);$$

14. We have $C_{15} \setminus \{\cup_{k=1}^{14} C_k\} = \{X_{17}\}$, $C_{15} \cap \{\cup_{k=1}^{14} C_k\} = \{X_2, X_3\}$ and generate

$$X_{17} = (X_2, X_3) \cdot \mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} + \xi_{15}, \quad (\text{B.2.31})$$

where $\mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} = \begin{pmatrix} -0.051 \\ 0.058 \end{pmatrix}$, and $\xi_{15} \sim U(-0.5, 0.5)$;

15. We have $C_{16} \setminus \{\cup_{k=1}^{15} C_k\} = \{X_{19}\}$, $C_{16} \cap \{\cup_{k=1}^{15} C_k\} = \{X_1, X_3\}$ and generate

$$X_{19} = (X_1, X_3) \cdot \mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} + \xi_{16}, \quad (\text{B.2.32})$$

where $\mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} = \begin{pmatrix} -0.05 \\ 0.03 \end{pmatrix}$, and $\xi_{16} \sim U(-0.5, 0.5)$;

16. We have $C_{17} \setminus \{\cup_{k=1}^{16} C_k\} = \{X_{21}\}$, $C_{17} \cap \{\cup_{k=1}^{16} C_k\} = \{X_2, X_8\}$ and generate

$$X_{21} = (X_2, X_8) \cdot \mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} + \xi_{17}, \quad (\text{B.2.33})$$

where $\mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} = \begin{pmatrix} -0.05 \\ -0.03 \end{pmatrix}$, and $\xi_{17} \sim U(-0.5, 0.5)$;

17. We have $C_{18} \setminus \{\cup_{k=1}^{17} C_k\} = \{X_{22}\}$, $C_{18} \cap \{\cup_{k=1}^{17} C_k\} = \{X_4, X_{12}\}$ and generate

$$X_{22} = (X_4, X_{12}) \cdot \mathbf{r}_{s_{18}, C_{18} \setminus s_{18}} + \xi_{18}, \quad (\text{B.2.34})$$

where $\mathbf{r}_{s_{18}, C_{18} \setminus s_{18}} = \begin{pmatrix} 0.2 \\ -0.2 \end{pmatrix}$, and $\xi_{18} \sim U(-0.5, 0.5)$;

18. We have $C_{19} \setminus \{\cup_{k=1}^{18} C_k\} = \{X_{23}\}$, $C_{19} \cap \{\cup_{k=1}^{18} C_k\} = \{X_5, X_9\}$ and generate

$$X_{23} = (X_5, X_9) \cdot \mathbf{r}_{s_{19}, C_{19} \setminus s_{19}} + \xi_{19}, \quad (\text{B.2.35})$$

where $\mathbf{r}_{s_{19}, C_{19} \setminus s_{19}} = \begin{pmatrix} -0.026 \\ 0.022 \end{pmatrix}$, and $\xi_{19} \sim U(-0.5, 0.5)$;

19. We have $C_{20} \setminus \{\cup_{k=1}^{19} C_k\} = \{X_{24}\}$, $C_{20} \cap \{\cup_{k=1}^{19} C_k\} = \{X_1, X_9\}$ and generate

$$X_{24} = (X_1, X_9) \cdot \mathbf{r}_{S_{20}, C_{20} \setminus S_{20}} + \xi_{20}, \quad (\text{B.2.36})$$

$$\text{where } \mathbf{r}_{S_{20}, C_{20} \setminus S_{20}} = \begin{pmatrix} 0.02 \\ -0.05 \end{pmatrix}, \text{ and } \xi_{20} \sim U(-0.5, 0.5);$$

20. We have $C_{21} \setminus \{\cup_{k=1}^{20} C_k\} = \{X_{29}\}$, $C_{21} \cap \{\cup_{k=1}^{20} C_k\} = \{X_{12}, X_{22}\}$ and generate

$$X_{29} = (X_{12}, X_{22}) \cdot \mathbf{r}_{S_{21}, C_{21} \setminus S_{21}} + \xi_{21}, \quad (\text{B.2.37})$$

$$\text{where } \mathbf{r}_{S_{21}, C_{21} \setminus S_{21}} = \begin{pmatrix} -0.02 \\ -0.01 \end{pmatrix}, \text{ and } \xi_{21} \sim U(-0.5, 0.5);$$

21. We have $C_{22} \setminus \{\cup_{k=1}^{21} C_k\} = \{X_{18}\}$, $C_{22} \cap \{\cup_{k=1}^{21} C_k\} = \{X_{16}\}$ and generate

$$X_{18} = (X_{16}) \cdot \mathbf{r}_{S_{22}, C_{22} \setminus S_{22}} + \xi_{22}, \quad (\text{B.2.38})$$

$$\text{where } \mathbf{r}_{S_{22}, C_{22} \setminus S_{22}} = \begin{pmatrix} -0.01 \end{pmatrix}, \text{ and } \xi_{22} \sim U(-0.5, 0.5);$$

22. We have $C_{23} \setminus \{\cup_{k=1}^{22} C_k\} = \{X_{20}\}$, $C_{23} \cap \{\cup_{k=1}^{22} C_k\} = \{X_{17}\}$ and generate

$$X_{20} = (X_{17}) \cdot \mathbf{r}_{S_{23}, C_{23} \setminus S_{23}} + \xi_{23}, \quad (\text{B.2.39})$$

$$\text{where } \mathbf{r}_{S_{23}, C_{23} \setminus S_{23}} = \begin{pmatrix} -0.05 \end{pmatrix}, \text{ and } \xi_{23} \sim U(-0.5, 0.5);$$

23. We have $C_{24} \setminus \{\cup_{k=1}^{23} C_k\} = \{X_{26}\}$, $C_{24} \cap \{\cup_{k=1}^{23} C_k\} = \{X_{13}\}$ and generate

$$X_{26} = (X_{13}) \cdot \mathbf{r}_{S_{24}, C_{24} \setminus S_{24}} + \xi_{24}, \quad (\text{B.2.40})$$

$$\text{where } \mathbf{r}_{S_{24}, C_{24} \setminus S_{24}} = \begin{pmatrix} -0.03 \end{pmatrix}, \text{ and } \xi_{24} \sim U(-0.5, 0.5);$$

24. We have $C_{25} \setminus \{\cup_{k=1}^{24} C_k\} = \{X_{27}\}$, $C_{25} \cap \{\cup_{k=1}^{24} C_k\} = \{X_{20}\}$ and generate

$$X_{27} = (X_{20}) \cdot \mathbf{r}_{s_{25}, C_{25} \setminus s_{25}} + \xi_{25}, \quad (\text{B.2.41})$$

where $\mathbf{r}_{s_{25}, C_{25} \setminus s_{25}} = (0.125)$, and $\xi_{25} \sim U(-0.5, 0.5)$;

25. We have $C_{26} \setminus \{\cup_{k=1}^{25} C_k\} = \{X_{28}\}$, $C_{26} \cap \{\cup_{k=1}^{25} C_k\} = \{X_{11}\}$ and generate

$$X_{28} = (X_{11}) \cdot \mathbf{r}_{s_{26}, C_{26} \setminus s_{26}} + \xi_{26}, \quad (\text{B.2.42})$$

where $\mathbf{r}_{s_{26}, C_{26} \setminus s_{26}} = (0.05)$, and $\xi_{26} \sim U(-0.5, 0.5)$;

26. We have $C_{27} \setminus \{\cup_{k=1}^{26} C_k\} = \{X_{30}\}$, $C_{27} \cap \{\cup_{k=1}^{26} C_k\} = \{X_{15}\}$ and generate

$$X_{30} = (X_{15}) \cdot \mathbf{r}_{s_{27}, C_{27} \setminus s_{27}} + \xi_{27}, \quad (\text{B.2.43})$$

where $\mathbf{r}_{s_{27}, C_{27} \setminus s_{27}} = (0.05)$, and $\xi_{27} \sim U(-0.5, 0.5)$;

B.3 Scenario 5

This simulation includes a binary variable X_{20} among the predictors which take two values 0 and 1.

From the graph structure given in Scenario 3, the perfect ordering of cliques (excluding X_{18}) is $C_1 = \{X_3, X_4, X_7, X_{20}\}$, $C_2 = \{X_1, X_3, X_4\}$, $C_3 = \{X_2, X_3, X_4\}$, $C_4 = \{X_3, X_5, X_7\}$, $C_5 = \{X_4, X_6, X_7\}$, $C_6 = \{X_4, X_7, X_8\}$, $C_7 = \{X_4, X_7, X_{11}\}$, $C_8 = \{X_3, X_7, X_9\}$, $C_9 = \{X_3, X_4, X_{10}\}$, $C_{10} = \{X_6, X_{12}\}$, $C_{11} = \{X_7, X_{13}\}$, $C_{12} = \{X_2, X_{14}\}$, $C_{13} = \{X_1, X_{15}\}$, $C_{14} = \{X_{10}, X_{16}\}$, $C_{15} = \{X_{11}, X_{17}\}$, $C_{16} = \{X_{15}, X_{19}\}$. The variables X_3, X_4, X_7 are generated from a multi-variate normal distribution with a mean

vector $\mathbf{0}$ and $\Sigma = L_1$ if $X_{20} = 1$ or L_2 if $X_{20} = 0$, where

$$L_1 = \begin{pmatrix} 0.230 & 0.080 & 0.019 \\ 0.080 & 0.602 & 0.007 \\ 0.0192 & 0.007 & 0.472 \end{pmatrix},$$

and

$$L_2 = \begin{pmatrix} 0.377 & -0.093 & 0.092 \\ -0.093 & 0.387 & -0.061 \\ 0.092 & -0.061 & 0.246 \end{pmatrix}.$$

Variable X_{18} is generated from $N(0, 0.433)$

Then we use the linear condition

$$\mathbf{x}_{C_k \setminus S_k} = \mathbf{x}_{S_k} \cdot \mathbf{r}_{S_k, C_k \setminus S_k} + \boldsymbol{\xi}_k, \quad (\text{B.3.44})$$

In case when $X_{20} = 1$ the variables are generated as:

1. We have $C_2 \setminus C_1 = \{X_1\}$, $C_2 \cap C_1 = \{X_3, X_4\}$ and generate

$$X_1 = (X_3, X_4) \cdot \mathbf{r}_{S_2, C_2 \setminus S_2} + \xi_2, \quad (\text{B.3.45})$$

where $\mathbf{r}_{S_2, C_2 \setminus S_2} = \begin{pmatrix} -0.3 \\ 0.25 \end{pmatrix}$, and $\xi_2 \sim U(-0.5, 0.5)$;

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_2\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_3, X_4\}$ and generate

$$X_2 = (X_3, X_4) \cdot \mathbf{r}_{S_3, C_3 \setminus S_3} + \xi_3, \quad (\text{B.3.46})$$

where $\mathbf{r}_{S_3, C_3 \setminus S_3} = \begin{pmatrix} 0.45 \\ -0.35 \end{pmatrix}$, and $\xi_3 \sim U(-0.5, 0.5)$;

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_5\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_3, X_7\}$ and generate

$$X_5 = (X_3, X_7) \cdot \mathbf{r}_{S_4, C_4 \setminus S_4} + \xi_4, \quad (\text{B.3.47})$$

$$\text{where } \mathbf{r}_{S_4, C_4 \setminus S_4} = \begin{pmatrix} 0.4 \\ -0.3 \end{pmatrix}, \text{ and } \xi_4 \sim U(-0.5, 0.5);$$

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_6\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_4, X_7\}$ and generate

$$X_6 = (X_4, X_7) \cdot \mathbf{r}_{S_5, C_5 \setminus S_5} + \xi_5, \quad (\text{B.3.48})$$

$$\text{where } \mathbf{r}_{S_5, C_5 \setminus S_5} = \begin{pmatrix} -0.2 \\ 0.3 \end{pmatrix}, \text{ and } \xi_5 \sim U(-0.5, 0.5);$$

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_8\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_4, X_7\}$ and generate

$$X_8 = (X_4, X_7) \cdot \mathbf{r}_{S_6, C_6 \setminus S_6} + \xi_6, \quad (\text{B.3.49})$$

$$\text{where } \mathbf{r}_{S_6, C_6 \setminus S_6} = \begin{pmatrix} -0.23 \\ -0.5 \end{pmatrix}, \text{ and } \xi_6 \sim U(-0.5, 0.5);$$

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_{11}\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_4, X_7\}$ and generate

$$X_{11} = (X_4, X_7) \cdot \mathbf{r}_{S_7, C_7 \setminus S_7} + \xi_7, \quad (\text{B.3.50})$$

$$\text{where } \mathbf{r}_{S_7, C_7 \setminus S_7} = \begin{pmatrix} 0.3 \\ 0.2 \end{pmatrix}, \text{ and } \xi_7 \sim U(-0.5, 0.5);$$

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_9\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_3, X_7\}$ and generate

$$X_9 = (X_3, X_7) \cdot \mathbf{r}_{S_8, C_8 \setminus S_8} + \xi_8, \quad (\text{B.3.51})$$

where $\mathbf{r}_{s_8, C_8 \setminus s_8} = \begin{pmatrix} 0.26 \\ -0.35 \end{pmatrix}$, and $\xi_8 \sim U(-0.5, 0.5)$;

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_{10}\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_3, X_4\}$ and generate

$$X_{10} = (X_3, X_4) \cdot \mathbf{r}_{s_9, C_9 \setminus s_9} + \xi_9, \quad (\text{B.3.52})$$

where $\mathbf{r}_{s_9, C_9 \setminus s_9} = \begin{pmatrix} 0.2 \\ 0.35 \end{pmatrix}$, and $\xi_9 \sim U(-0.5, 0.5)$;

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_{12}\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_6\}$ and generate

$$X_{12} = X_6 \cdot \mathbf{r}_{s_{10}, C_{10} \setminus s_{10}} + \xi_{10}, \quad (\text{B.3.53})$$

where $\mathbf{r}_{s_{10}, C_{10} \setminus s_{10}} = (0.25)$, and $\xi_{10} \sim U(-0.5, 0.5)$;

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{13}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_7\}$ and generate

$$X_{13} = X_7 \cdot \mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} + \xi_{11}, \quad (\text{B.3.54})$$

where $\mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} = (0.3)$, and $\xi_{11} \sim U(-0.5, 0.5)$;

11. We have $C_{12} \setminus \{\cup_{k=1}^{11} C_k\} = \{X_{14}\}$, $C_{12} \cap \{\cup_{k=1}^{11} C_k\} = \{X_2\}$ and generate

$$X_{14} = X_2 \cdot \mathbf{r}_{s_{12}, C_{12} \setminus s_{12}} + \xi_{12}, \quad (\text{B.3.55})$$

where $\mathbf{r}_{s_{12}, C_{12} \setminus s_{12}} = (0.32)$, and $\xi_{12} \sim U(-0.5, 0.5)$;

12. We have $C_{13} \setminus \{\cup_{k=1}^{12} C_k\} = \{X_{15}\}$, $C_{13} \cap \{\cup_{k=1}^{12} C_k\} = \{X_1\}$ and generate

$$X_{15} = X_1 \cdot \mathbf{r}_{s_{13}, C_{13} \setminus s_{13}} + \xi_{13}, \quad (\text{B.3.56})$$

where $\mathbf{r}_{s_{13}, C_{13} \setminus s_{13}} = (0.2)$, and $\xi_{13} \sim U(-0.5, 0.5)$;

13. We have $C_{14} \setminus \{\cup_{k=1}^{13} C_k\} = \{X_{16}\}$, $C_{14} \cap \{\cup_{k=1}^{13} C_k\} = \{X_{10}\}$ and generate

$$X_{16} = X_{10} \cdot \mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} + \xi_{14}, \quad (\text{B.3.57})$$

where $\mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} = (0.35)$, and $\xi_{14} \sim U(-0.5, 0.5)$;

14. We have $C_{15} \setminus \{\cup_{k=1}^{14} C_k\} = \{X_{17}\}$, $C_{15} \cap \{\cup_{k=1}^{14} C_k\} = \{X_{11}\}$ and generate

$$X_{17} = X_{11} \cdot \mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} + \xi_{15}, \quad (\text{B.3.58})$$

where $\mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} = (0.4)$, and $\xi_{15} \sim U(-0.5, 0.5)$;

15. We have $C_{16} \setminus \{\cup_{k=1}^{15} C_k\} = \{X_{19}\}$, $C_{16} \cap \{\cup_{k=1}^{15} C_k\} = \{X_{15}\}$ and generate

$$X_{19} = X_{15} \cdot \mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} + \xi_{16}, \quad (\text{B.3.59})$$

where $\mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} = (0.3)$, and $\xi_{16} \sim U(-0.5, 0.5)$;

In case when $X_{20} = 0$ the variables are generated as:

1. We have $C_2 \setminus C_1 = \{X_1\}$, $C_2 \cap C_1 = \{X_3, X_4\}$ and generate

$$X_1 = (X_3, X_4) \cdot \mathbf{r}_{s_2, C_2 \setminus s_2} + \xi_2, \quad (\text{B.3.60})$$

where $\mathbf{r}_{s_2, C_2 \setminus s_2} = \begin{pmatrix} -0.25 \\ 0.12 \end{pmatrix}$, and $\xi_2 \sim U(-0.5, 0.5)$;

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_2\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_3, X_4\}$ and generate

$$X_2 = (X_3, X_4) \cdot \mathbf{r}_{S_3, C_3 \setminus S_3} + \xi_3, \quad (\text{B.3.61})$$

$$\text{where } \mathbf{r}_{S_3, C_3 \setminus S_3} = \begin{pmatrix} -0.5 \\ -0.15 \end{pmatrix}, \text{ and } \xi_2 \sim U(-0.5, 0.5);$$

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_5\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_3, X_7\}$ and generate

$$X_5 = (X_3, X_7) \cdot \mathbf{r}_{S_4, C_4 \setminus S_4} + \xi_4, \quad (\text{B.3.62})$$

$$\text{where } \mathbf{r}_{S_4, C_4 \setminus S_4} = \begin{pmatrix} 0.4 \\ 0.2 \end{pmatrix}, \text{ and } \xi_4 \sim U(-0.5, 0.5);$$

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_6\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_4, X_7\}$ and generate

$$X_6 = (X_4, X_7) \cdot \mathbf{r}_{S_5, C_5 \setminus S_5} + \xi_5, \quad (\text{B.3.63})$$

$$\text{where } \mathbf{r}_{S_5, C_5 \setminus S_5} = \begin{pmatrix} -0.2 \\ 0.5 \end{pmatrix}, \text{ and } \xi_5 \sim U(-0.5, 0.5);$$

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_8\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_4, X_7\}$ and generate

$$X_8 = (X_4, X_7) \cdot \mathbf{r}_{S_6, C_6 \setminus S_6} + \xi_6, \quad (\text{B.3.64})$$

$$\text{where } \mathbf{r}_{S_6, C_6 \setminus S_6} = \begin{pmatrix} 0.12 \\ -0.15 \end{pmatrix}, \text{ and } \xi_6 \sim U(-0.5, 0.5);$$

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_{11}\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_4, X_7\}$ and generate

$$X_{11} = (X_4, X_7) \cdot \mathbf{r}_{S_7, C_7 \setminus S_7} + \xi_7, \quad (\text{B.3.65})$$

where $\mathbf{r}_{s_7, C_7 \setminus s_7} = \begin{pmatrix} 0.15 \\ -0.13 \end{pmatrix}$, and $\xi_7 \sim U(-0.5, 0.5)$;

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_9\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_3, X_7\}$ and generate

$$X_9 = (X_3, X_7) \cdot \mathbf{r}_{s_8, C_8 \setminus s_8} + \xi_8, \quad (\text{B.3.66})$$

where $\mathbf{r}_{s_8, C_8 \setminus s_8} = \begin{pmatrix} -0.3 \\ 0.18 \end{pmatrix}$, and $\xi_8 \sim U(-0.5, 0.5)$;

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_{10}\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_3, X_4\}$ and generate

$$X_{10} = (X_3, X_4) \cdot \mathbf{r}_{s_9, C_9 \setminus s_9} + \xi_9, \quad (\text{B.3.67})$$

where $\mathbf{r}_{s_9, C_9 \setminus s_9} = \begin{pmatrix} 0.3 \\ -0.2 \end{pmatrix}$, and $\xi_9 \sim U(-0.5, 0.5)$;

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_{12}\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_6\}$ and generate

$$X_{12} = X_6 \cdot \mathbf{r}_{s_{10}, C_{10} \setminus s_{10}} + \xi_{10}, \quad (\text{B.3.68})$$

where $\mathbf{r}_{s_{10}, C_{10} \setminus s_{10}} = \begin{pmatrix} 0.45 \end{pmatrix}$, and $\xi_{10} \sim U(-0.5, 0.5)$;

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{13}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_7\}$ and generate

$$X_{13} = X_7 \cdot \mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} + \xi_{11}, \quad (\text{B.3.69})$$

where $\mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} = \begin{pmatrix} 0.35 \end{pmatrix}$, and $\xi_{11} \sim U(-0.5, 0.5)$;

11. We have $C_{12} \setminus \{\cup_{k=1}^{11} C_k\} = \{X_{14}\}$, $C_{12} \cap \{\cup_{k=1}^{11} C_k\} = \{X_2\}$ and generate

$$X_{14} = X_2 \cdot \mathbf{r}_{s_{12}, C_{12} \setminus s_{12}} + \xi_{12}, \quad (\text{B.3.70})$$

where $\mathbf{r}_{s_{12}, C_{12} \setminus s_{12}} = (0.15)$, and $\xi_{12} \sim U(-0.5, 0.5)$;

12. We have $C_{13} \setminus \{\cup_{k=1}^{12} C_k\} = \{X_{15}\}$, $C_{13} \cap \{\cup_{k=1}^{12} C_k\} = \{X_1\}$ and generate

$$X_{15} = X_1 \cdot \mathbf{r}_{s_{13}, C_{13} \setminus s_{13}} + \xi_{13}, \quad (\text{B.3.71})$$

where $\mathbf{r}_{s_{13}, C_{13} \setminus s_{13}} = (0.3)$, and $\xi_{13} \sim U(-0.5, 0.5)$;

13. We have $C_{14} \setminus \{\cup_{k=1}^{13} C_k\} = \{X_{16}\}$, $C_{14} \cap \{\cup_{k=1}^{13} C_k\} = \{X_{10}\}$ and generate

$$X_{16} = X_{10} \cdot \mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} + \xi_{14}, \quad (\text{B.3.72})$$

where $\mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} = (-0.28)$, and $\xi_{14} \sim U(-0.5, 0.5)$;

14. We have $C_{15} \setminus \{\cup_{k=1}^{14} C_k\} = \{X_{17}\}$, $C_{15} \cap \{\cup_{k=1}^{14} C_k\} = \{X_{11}\}$ and generate

$$X_{17} = X_{11} \cdot \mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} + \xi_{15}, \quad (\text{B.3.73})$$

where $\mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} = (0.19)$, and $\xi_{15} \sim U(-0.5, 0.5)$;

15. We have $C_{16} \setminus \{\cup_{k=1}^{15} C_k\} = \{X_{19}\}$, $C_{16} \cap \{\cup_{k=1}^{15} C_k\} = \{X_{15}\}$ and generate

$$X_{19} = X_{15} \cdot \mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} + \xi_{16}, \quad (\text{B.3.74})$$

where $\mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} = (0.2)$, and $\xi_{16} \sim U(-0.5, 0.5)$;

APPENDIX C

Data generated procedure in Chapter 5

This appendix shows the process of data generation for the first simulation set (Scenario 3 and 4) and the second simulation set (Scenario 6) of Chapter 5.

C.1 Scenario 3

Below are the details about how the data are generated for Scenario 3.

Using the graph structure given in Figure 5.7, the perfect ordering of cliques is

$$C_1 = \{X_1, X_2, X_4, X_6, X_8, X_9, X_{11}, X_{12}, X_{13}, X_{14}, X_{16}, X_{17}, X_{18}, X_{19}, X_{21}, X_{22}, X_{23}, X_{24}, X_{26}, X_{27}, X_{28}, X_{29}, X_{31}, X_{32}, X_{33}, X_{34}, X_{36}, X_{37}, X_{38}, X_{39}\},$$

$$C_2 = \{X_1, X_{18}, X_{21}, X_{22}, X_{26}, X_{29}, X_{32}, X_{33}, X_{40}\},$$

$$C_3 = \{X_5, X_8, X_9, X_{14}, X_{17}, X_{19}\}, C_4 = \{X_1, X_4, X_7, X_8\}, C_5 = \{X_{10}, X_{21}, X_{27}, X_{34}\},$$

$$C_6 = \{X_{20}, X_{24}, X_{28}, X_{38}\}, C_7 = \{X_9, X_{11}, X_{33}, X_{35}\}, C_8 = \{X_{19}, X_{23}, X_{30}\}, C_9 = \{X_3, X_6\},$$

$$C_{10} = \{X_{15}, X_{39}\}, C_{11} = \{X_{25}, X_{38}\}.$$

All variables in C_1 ($C_1 = C^*$) are generated from $MVN(\mathbf{0}, \Sigma^{-1} = \mathbf{K}_1)$ where

$$\mathbf{K}_1 = \begin{pmatrix} 2.649 & -0.363 & -0.366 & -0.271 & -0.293 & -0.355 & -0.238 & -0.144 & -0.379 & -0.356 & -0.199 & -0.263 & -0.273 & -0.277 & -0.362 & 0.220 & 0.212 & 0.132 & 0.103 & 0.236 & 0.114 & 0.162 & 0.206 & 0.160 & 0.237 & 0.183 & 0.105 & 0.282 & 0.136 & 0.145 \\ -0.363 & 2.186 & -0.247 & -0.110 & -0.368 & -0.172 & -0.168 & -0.241 & -0.380 & -0.146 & -0.102 & -0.133 & -0.161 & -0.387 & -0.216 & 0.269 & 0.241 & 0.104 & 0.222 & 0.181 & 0.199 & 0.214 & 0.275 & 0.102 & 0.232 & 0.143 & 0.172 & 0.136 & 0.145 & 0.195 \\ -0.366 & -0.247 & 2.548 & -0.383 & -0.390 & -0.389 & -0.318 & -0.357 & -0.201 & -0.170 & -0.115 & -0.193 & -0.301 & -0.109 & -0.275 & 0.273 & 0.115 & 0.205 & 0.150 & 0.242 & 0.259 & 0.243 & 0.146 & 0.106 & 0.119 & 0.202 & 0.231 & 0.199 & 0.269 & 0.283 \\ -0.271 & -0.110 & -0.383 & 2.351 & -0.355 & -0.287 & -0.301 & -0.353 & -0.373 & -0.120 & -0.336 & -0.322 & -0.322 & -0.220 & -0.154 & 0.177 & 0.153 & 0.113 & 0.269 & 0.140 & 0.151 & 0.160 & 0.249 & 0.205 & 0.168 & 0.108 & 0.287 & 0.250 & 0.157 & 0.229 \\ -0.293 & -0.368 & -0.390 & -0.355 & 1.942 & -0.270 & -0.274 & -0.164 & -0.241 & -0.275 & -0.312 & -0.396 & -0.325 & -0.103 & -0.350 & 0.208 & 0.195 & 0.145 & 0.263 & 0.278 & 0.186 & 0.230 & 0.196 & 0.168 & 0.140 & 0.246 & 0.291 & 0.289 & 0.267 & 0.281 \\ -0.355 & -0.172 & -0.389 & -0.287 & -0.270 & 1.944 & -0.273 & -0.148 & -0.351 & -0.110 & -0.376 & -0.298 & -0.155 & -0.139 & -0.229 & 0.178 & 0.237 & 0.242 & 0.270 & 0.293 & 0.182 & 0.115 & 0.223 & 0.242 & 0.153 & 0.112 & 0.124 & 0.213 & 0.125 & 0.246 \\ -0.238 & -0.168 & -0.318 & -0.301 & -0.274 & -0.273 & 2.418 & -0.308 & -0.356 & -0.278 & -0.186 & -0.126 & -0.180 & -0.138 & -0.145 & 0.197 & 0.162 & 0.189 & 0.164 & 0.199 & 0.176 & 0.128 & 0.266 & 0.201 & 0.226 & 0.102 & 0.241 & 0.185 & 0.218 & 0.150 \\ -0.144 & -0.241 & -0.357 & -0.353 & -0.164 & -0.148 & -0.308 & 2.391 & -0.260 & -0.236 & -0.346 & -0.120 & -0.304 & -0.319 & -0.313 & 0.183 & 0.161 & 0.146 & 0.240 & 0.186 & 0.181 & 0.122 & 0.115 & 0.171 & 0.208 & 0.179 & 0.224 & 0.145 & 0.234 & 0.208 \\ -0.379 & -0.380 & -0.201 & -0.373 & -0.241 & -0.351 & -0.356 & -0.260 & 2.422 & -0.167 & -0.267 & -0.304 & -0.333 & -0.377 & -0.190 & 0.200 & 0.205 & 0.177 & 0.161 & 0.259 & 0.103 & 0.241 & 0.180 & 0.153 & 0.238 & 0.159 & 0.117 & 0.222 & 0.279 & 0.293 \\ -0.356 & -0.146 & -0.170 & -0.120 & -0.275 & -0.110 & -0.278 & -0.236 & -0.167 & 2.689 & -0.285 & -0.110 & -0.142 & -0.110 & -0.246 & 0.155 & 0.295 & 0.295 & 0.212 & 0.213 & 0.125 & 0.220 & 0.261 & 0.212 & 0.128 & 0.207 & 0.188 & 0.143 & 0.236 & 0.214 \\ -0.199 & -0.102 & -0.115 & -0.336 & -0.312 & -0.376 & -0.186 & -0.346 & -0.267 & -0.285 & 2.242 & -0.179 & -0.239 & -0.324 & -0.391 & 0.150 & 0.233 & 0.126 & 0.235 & 0.230 & 0.262 & 0.214 & 0.249 & 0.295 & 0.271 & 0.281 & 0.264 & 0.184 & 0.252 & 0.131 \\ -0.263 & -0.133 & -0.193 & -0.322 & -0.396 & -0.298 & -0.126 & -0.120 & -0.304 & -0.110 & -0.179 & 2.668 & -0.267 & -0.295 & -0.362 & 0.206 & 0.171 & 0.117 & 0.103 & 0.283 & 0.166 & 0.286 & 0.144 & 0.167 & 0.106 & 0.272 & 0.166 & 0.298 & 0.261 & 0.193 \\ -0.273 & -0.161 & -0.301 & -0.322 & -0.325 & -0.155 & -0.180 & -0.304 & -0.333 & -0.142 & -0.239 & -0.267 & 1.838 & -0.148 & -0.157 & 0.251 & 0.144 & 0.136 & 0.151 & 0.215 & 0.112 & 0.297 & 0.168 & 0.153 & 0.299 & 0.228 & 0.224 & 0.280 & 0.228 & 0.103 \\ -0.277 & -0.387 & -0.109 & -0.220 & -0.103 & -0.139 & -0.138 & -0.319 & -0.377 & -0.110 & -0.324 & -0.295 & -0.148 & 1.558 & -0.188 & 0.101 & 0.239 & 0.139 & 0.200 & 0.195 & 0.176 & 0.143 & 0.250 & 0.246 & 0.116 & 0.101 & 0.211 & 0.217 & 0.196 & 0.182 \\ -0.362 & -0.216 & -0.275 & -0.154 & -0.350 & -0.229 & -0.145 & -0.313 & -0.190 & -0.246 & -0.391 & -0.362 & -0.157 & -0.188 & 2.317 & 0.198 & 0.297 & 0.237 & 0.128 & 0.244 & 0.116 & 0.227 & 0.220 & 0.227 & 0.139 & 0.171 & 0.130 & 0.261 & 0.198 & 0.279 \\ 0.220 & 0.269 & 0.273 & 0.177 & 0.208 & 0.178 & 0.197 & 0.183 & 0.200 & 0.155 & 0.150 & 0.206 & 0.251 & 0.101 & 0.198 & 2.044 & 0.172 & 0.264 & 0.234 & 0.266 & 0.276 & 0.144 & 0.184 & 0.152 & 0.220 & 0.267 & 0.294 & 0.135 & 0.263 & 0.137 \\ 0.212 & 0.241 & 0.115 & 0.153 & 0.195 & 0.237 & 0.162 & 0.161 & 0.205 & 0.295 & 0.233 & 0.171 & 0.144 & 0.239 & 0.297 & 0.172 & 1.037 & 0.269 & 0.244 & 0.299 & 0.287 & 0.201 & 0.204 & 0.205 & 0.291 & 0.149 & 0.216 & 0.109 & 0.241 & 0.143 \\ 0.132 & 0.104 & 0.205 & 0.113 & 0.145 & 0.242 & 0.189 & 0.146 & 0.177 & 0.295 & 0.126 & 0.117 & 0.136 & 0.139 & 0.237 & 0.264 & 0.269 & 2.210 & 0.236 & 0.237 & 0.293 & 0.167 & 0.132 & 0.163 & 0.217 & 0.258 & 0.103 & 0.271 & 0.199 & 0.119 \\ 0.103 & 0.222 & 0.150 & 0.269 & 0.263 & 0.270 & 0.164 & 0.240 & 0.161 & 0.212 & 0.235 & 0.103 & 0.151 & 0.200 & 0.128 & 0.234 & 0.244 & 0.236 & 1.676 & 0.181 & 0.105 & 0.300 & 0.166 & 0.126 & 0.199 & 0.216 & 0.130 & 0.270 & 0.126 & 0.144 \\ 0.236 & 0.181 & 0.242 & 0.140 & 0.278 & 0.293 & 0.199 & 0.186 & 0.259 & 0.213 & 0.230 & 0.283 & 0.215 & 0.195 & 0.244 & 0.266 & 0.299 & 0.237 & 0.181 & 1.315 & 0.199 & 0.294 & 0.246 & 0.253 & 0.203 & 0.202 & 0.173 & 0.166 & 0.145 & 0.202 \\ 0.114 & 0.199 & 0.259 & 0.151 & 0.186 & 0.182 & 0.176 & 0.181 & 0.103 & 0.125 & 0.262 & 0.166 & 0.112 & 0.176 & 0.116 & 0.276 & 0.287 & 0.293 & 0.105 & 0.199 & 2.815 & 0.209 & 0.104 & 0.176 & 0.112 & 0.130 & 0.274 & 0.203 & 0.150 & 0.279 \\ 0.162 & 0.214 & 0.243 & 0.160 & 0.230 & 0.115 & 0.128 & 0.122 & 0.241 & 0.220 & 0.214 & 0.286 & 0.297 & 0.143 & 0.227 & 0.144 & 0.201 & 0.167 & 0.300 & 0.294 & 0.209 & 1.160 & 0.206 & 0.297 & 0.161 & 0.148 & 0.181 & 0.179 & 0.264 & 0.234 \\ 0.206 & 0.275 & 0.146 & 0.249 & 0.196 & 0.223 & 0.266 & 0.115 & 0.180 & 0.261 & 0.249 & 0.144 & 0.168 & 0.250 & 0.220 & 0.184 & 0.204 & 0.132 & 0.166 & 0.246 & 0.104 & 0.206 & 1.910 & 0.177 & 0.264 & 0.137 & 0.260 & 0.289 & 0.250 & 0.212 \\ 0.160 & 0.102 & 0.106 & 0.205 & 0.168 & 0.242 & 0.201 & 0.171 & 0.153 & 0.212 & 0.295 & 0.167 & 0.153 & 0.246 & 0.227 & 0.152 & 0.205 & 0.163 & 0.126 & 0.253 & 0.176 & 0.297 & 0.177 & 1.146 & 0.158 & 0.121 & 0.120 & 0.186 & 0.281 & 0.115 \\ 0.237 & 0.232 & 0.119 & 0.168 & 0.140 & 0.153 & 0.226 & 0.208 & 0.238 & 0.128 & 0.271 & 0.106 & 0.299 & 0.116 & 0.139 & 0.220 & 0.291 & 0.217 & 0.199 & 0.203 & 0.112 & 0.161 & 0.264 & 0.158 & 1.178 & 0.199 & 0.187 & 0.294 & 0.211 & 0.286 \\ 0.183 & 0.143 & 0.202 & 0.108 & 0.246 & 0.112 & 0.102 & 0.179 & 0.159 & 0.207 & 0.281 & 0.272 & 0.228 & 0.101 & 0.171 & 0.267 & 0.149 & 0.258 & 0.216 & 0.202 & 0.130 & 0.148 & 0.137 & 0.121 & 0.199 & 2.194 & 0.148 & 0.185 & 0.257 & 0.220 \\ 0.105 & 0.172 & 0.231 & 0.287 & 0.291 & 0.124 & 0.241 & 0.224 & 0.117 & 0.188 & 0.264 & 0.166 & 0.224 & 0.211 & 0.130 & 0.294 & 0.216 & 0.103 & 0.130 & 0.173 & 0.274 & 0.181 & 0.260 & 0.120 & 0.187 & 0.148 & 1.229 & 0.121 & 0.147 & 0.143 \\ 0.282 & 0.136 & 0.199 & 0.250 & 0.289 & 0.213 & 0.185 & 0.145 & 0.222 & 0.143 & 0.184 & 0.298 & 0.280 & 0.217 & 0.261 & 0.135 & 0.109 & 0.271 & 0.270 & 0.166 & 0.203 & 0.179 & 0.289 & 0.186 & 0.294 & 0.185 & 0.121 & 1.793 & 0.239 & 0.257 \\ 0.136 & 0.145 & 0.269 & 0.157 & 0.267 & 0.125 & 0.218 & 0.234 & 0.279 & 0.236 & 0.252 & 0.261 & 0.228 & 0.196 & 0.198 & 0.263 & 0.241 & 0.199 & 0.126 & 0.145 & 0.150 & 0.264 & 0.250 & 0.281 & 0.211 & 0.257 & 0.147 & 0.239 & 2.425 & 0.273 \\ 0.145 & 0.195 & 0.283 & 0.229 & 0.281 & 0.246 & 0.150 & 0.208 & 0.293 & 0.214 & 0.131 & 0.193 & 0.103 & 0.182 & 0.279 & 0.137 & 0.143 & 0.119 & 0.144 & 0.202 & 0.279 & 0.234 & 0.212 & 0.115 & 0.286 & 0.220 & 0.143 & 0.257 & 0.273 & 2.033 \end{pmatrix}.$$

Then the following linear condition is used

$$\mathbf{x}_{C_k \setminus S_k} = \mathbf{x}_{S_k} \cdot \mathbf{r}_{S_k, C_k \setminus S_k} + \boldsymbol{\xi}_k, \quad (\text{C.1.1})$$

in order to generate the rest of the variables, as follows:

1. We have $C_2 \setminus C_1 = \{X_{40}\}$, $C_2 \cap C_1 = \{X_1, X_{18}, X_{21}, X_{22}, X_{26}, X_{29}, X_{32}, X_{33}\}$ and generate

$$X_{40} = (X_1, X_{18}, X_{21}, X_{22}, X_{26}, X_{29}, X_{32}, X_{33}) \cdot \mathbf{r}_{S_2, C_2 \setminus S_2} + \boldsymbol{\xi}_2, \quad (\text{C.1.2})$$

where $\mathbf{r}_{S_2, C_2 \setminus S_2} = (0.15, 0.25, 0.1, 0.25, 0.27, 0.2, 0.1, 0.2)'$, and $\boldsymbol{\xi}_2 \sim U(-0.5, 0.5)$;

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_5\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_8, X_9, X_{14}, X_{17}, X_{19}\}$ and generate

$$X_5 = (X_8, X_9, X_{14}, X_{17}, X_{19}) \cdot \mathbf{r}_{S_3, C_3 \setminus S_3} + \boldsymbol{\xi}_3, \quad (\text{C.1.3})$$

where $\mathbf{r}_{S_3, C_3 \setminus S_3} = (0.2, 0.1, -0.2, .031, .01)'$, and $\xi_3 \sim U(-0.5, 0.5)$;

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_7\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_1, X_4, X_8\}$ and generate

$$X_7 = (X_1, X_4, X_8) \cdot \mathbf{r}_{S_4, C_4 \setminus S_4} + \xi_4, \quad (\text{C.1.4})$$

where $\mathbf{r}_{S_4, C_4 \setminus S_4} = (0.1, -0.2, 0.3)'$, and $\xi_4 \sim U(-0.5, 0.5)$;

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_{10}\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_{21}, X_{27}, X_{34}\}$ and generate

$$X_{10} = (X_{21}, X_{27}, X_{34}) \cdot \mathbf{r}_{S_5, C_5 \setminus S_5} + \xi_5, \quad (\text{C.1.5})$$

where $\mathbf{r}_{S_5, C_5 \setminus S_5} = (-0.2, 0.3, 0.2)'$, and $\xi_5 \sim U(-0.5, 0.5)$;

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_{20}\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_{24}, X_{28}, X_{38}\}$ and generate

$$X_{20} = (X_{24}, X_{28}, X_{38}) \cdot \mathbf{r}_{S_6, C_6 \setminus S_6} + \xi_6, \quad (\text{C.1.6})$$

where $\mathbf{r}_{S_6, C_6 \setminus S_6} = (0.15, 0.1, 0.2)'$, and $\xi_6 \sim U(-0.5, 0.5)$;

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_{35}\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_9, X_{11}, X_{33}\}$ and generate

$$X_{35} = (X_9, X_{11}, X_{33}) \cdot \mathbf{r}_{S_7, C_7 \setminus S_7} + \xi_7, \quad (\text{C.1.7})$$

where $\mathbf{r}_{S_7, C_7 \setminus S_7} = (0.2, 0.15, 0.3)'$, and $\xi_7 \sim U(-0.5, 0.5)$;

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_{30}\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_{19}, X_{23}\}$ and generate

$$X_{30} = (X_{19}, X_{23}) \cdot \mathbf{r}_{S_8, C_8 \setminus S_8} + \xi_8, \quad (\text{C.1.8})$$

where $\mathbf{r}_{S_8, C_8 \setminus S_8} = \begin{pmatrix} 0.2 \\ -0.1 \end{pmatrix}$, and $\xi_8 \sim U(-0.5, 0.5)$;

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_3\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_6\}$ and generate

$$X_3 = (X_6) \cdot \mathbf{r}_{S_9, C_9 \setminus S_9} + \xi_9, \quad (\text{C.1.9})$$

where $\mathbf{r}_{S_9, C_9 \setminus S_9} = (-0.47)$, and $\xi_9 \sim U(-0.5, 0.5)$;

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_{15}\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_{39}\}$ and generate

$$X_{15} = (X_{39}) \cdot \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} + \xi_{10}, \quad (\text{C.1.10})$$

where $\mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} = (0.6)$, and $\xi_{10} \sim U(-0.5, 0.5)$;

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{25}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_{38}\}$ and generate

$$X_{25} = (X_{38}) \cdot \mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} + \xi_{11}, \quad (\text{C.1.11})$$

where $\mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} = (0.5)$, and $\xi_{11} \sim U(-0.5, 0.5)$.

C.2 Scenario 4

Below are the details about how the data are generated for Scenario 4.

Using the graph structure given in Figure 5.7, the perfect ordering of cliques is

$C_1 = \{X_1, X_2, X_4, X_6, X_8, X_9, X_{11}, X_{12}, X_{13}, X_{14}, X_{16}, X_{17}, X_{18}, X_{19}, X_{21}, X_{22}, X_{23}, X_{24}, X_{26}, X_{27}, X_{28}, X_{29}, X_{31}, X_{32}, X_{33}, X_{34}, X_{36}, X_{37}, X_{38}, X_{39}\}$,
 $C_2 = \{X_1, X_{18}, X_{21}, X_{22}, X_{26}, X_{29}, X_{32}, X_{33}, X_{40}\}$, $C_3 = \{X_5, X_8, X_9, X_{14}, X_{17}, X_{19}\}$, $C_4 = \{X_1, X_4, X_7, X_8\}$, $C_5 = \{X_{10}, X_{21}, X_{27}, X_{34}\}$, $C_6 = \{X_{20}, X_{24}, X_{28}, X_{38}\}$, $C_7 = \{X_9, X_{11}, X_{33}, X_{35}\}$,
 $C_8 = \{X_{19}, X_{23}, X_{30}\}$, $C_9 = \{X_3, X_6\}$, $C_{10} = \{X_{15}, X_{39}\}$, $C_{11} = \{X_{25}, X_{38}\}$. All variables in C_1 ($C_1 = C^*$) are generated from a multi-variate normal distribution with a mean vector $\mathbf{0}$ and $\Sigma^{-1} = L_1$ if $X_{12} = 1$ or L_2 if $X_{12} = 0$, where

$$L_1 = \begin{pmatrix} 9.948 & -0.639 & -0.582 & -0.444 & -0.401 & -0.807 & -0.656 & -0.630 & -0.331 & -0.874 & -0.330 & -0.738 & -0.477 & -0.646 & -0.732 & 0.574 & 0.657 & 0.584 & 0.500 & 0.594 & 0.792 & 0.585 & 0.631 & 0.764 & 0.896 & 0.826 & 0.851 & 0.799 & 0.599 & 0.529 \\ -0.639 & 6.498 & -0.819 & -0.577 & -0.773 & -0.674 & -0.659 & -0.401 & -0.455 & -0.568 & -0.482 & -0.338 & -0.760 & -0.861 & -0.863 & 0.843 & 0.841 & 0.770 & 0.856 & 0.504 & 0.851 & 0.627 & 0.793 & 0.700 & 0.755 & 0.798 & 0.731 & 0.825 & 0.812 & 0.618 \\ -0.582 & -0.819 & 4.546 & -0.401 & -0.624 & -0.448 & -0.676 & -0.337 & -0.475 & -0.612 & -0.565 & -0.404 & -0.406 & -0.316 & -0.818 & 0.648 & 0.839 & 0.593 & 0.536 & 0.514 & 0.598 & 0.651 & 0.823 & 0.728 & 0.766 & 0.693 & 0.574 & 0.744 & 0.560 & 0.558 \\ -0.444 & -0.577 & -0.401 & 6.507 & -0.853 & -0.306 & -0.443 & -0.366 & -0.896 & -0.737 & -0.653 & -0.534 & -0.793 & -0.703 & -0.651 & 0.647 & 0.876 & 0.831 & 0.510 & 0.792 & 0.605 & 0.722 & 0.509 & 0.852 & 0.540 & 0.700 & 0.856 & 0.880 & 0.879 & 0.771 \\ -0.401 & -0.773 & -0.624 & -0.853 & 8.536 & -0.427 & -0.673 & -0.362 & -0.788 & -0.329 & -0.707 & -0.333 & -0.334 & -0.429 & -0.357 & 0.692 & 0.524 & 0.633 & 0.758 & 0.753 & 0.876 & 0.825 & 0.574 & 0.882 & 0.838 & 0.756 & 0.508 & 0.728 & 0.729 & 0.590 \\ -0.807 & -0.674 & -0.448 & -0.306 & -0.427 & 9.461 & -0.889 & -0.722 & -0.381 & -0.878 & -0.832 & -0.803 & -0.595 & -0.765 & -0.840 & 0.839 & 0.577 & 0.568 & 0.643 & 0.896 & 0.871 & 0.520 & 0.644 & 0.802 & 0.709 & 0.763 & 0.754 & 0.679 & 0.560 & 0.567 \\ -0.656 & -0.659 & -0.676 & -0.443 & -0.673 & -0.889 & 8.649 & -0.733 & -0.863 & -0.514 & -0.781 & -0.715 & -0.662 & -0.578 & -0.662 & 0.514 & 0.869 & 0.780 & 0.572 & 0.783 & 0.727 & 0.790 & 0.779 & 0.678 & 0.779 & 0.679 & 0.600 & 0.633 & 0.576 & 0.870 \\ -0.630 & -0.401 & -0.337 & -0.366 & -0.362 & -0.722 & -0.733 & 2.668 & -0.595 & -0.729 & -0.415 & -0.891 & -0.805 & -0.554 & -0.451 & 0.798 & 0.881 & 0.575 & 0.541 & 0.536 & 0.557 & 0.858 & 0.741 & 0.666 & 0.789 & 0.818 & 0.626 & 0.641 & 0.554 & 0.500 \\ -0.331 & -0.455 & -0.475 & -0.896 & -0.788 & -0.381 & -0.863 & -0.595 & 6.837 & -0.801 & -0.548 & -0.316 & -0.327 & -0.529 & -0.351 & 0.791 & 0.595 & 0.886 & 0.572 & 0.743 & 0.805 & 0.507 & 0.868 & 0.889 & 0.841 & 0.873 & 0.759 & 0.551 & 0.647 & 0.700 \\ -0.874 & -0.568 & -0.612 & -0.737 & -0.329 & -0.878 & -0.514 & -0.729 & -0.801 & 1.141 & -0.670 & -0.882 & -0.506 & -0.703 & -0.706 & 0.528 & 0.818 & 0.530 & 0.765 & 0.807 & 0.710 & 0.884 & 0.827 & 0.797 & 0.591 & 0.839 & 0.779 & 0.782 & 0.551 & 0.683 \\ -0.330 & -0.482 & -0.565 & -0.653 & -0.707 & -0.832 & -0.781 & -0.415 & -0.548 & -0.670 & 1.568 & -0.379 & -0.792 & -0.731 & -0.522 & 0.699 & 0.827 & 0.595 & 0.507 & 0.770 & 0.714 & 0.501 & 0.695 & 0.809 & 0.782 & 0.877 & 0.843 & 0.854 & 0.571 & 0.843 \\ -0.738 & -0.338 & -0.404 & -0.534 & -0.333 & -0.803 & -0.715 & -0.891 & -0.316 & -0.882 & -0.379 & 6.873 & -0.869 & -0.367 & -0.404 & 0.744 & 0.501 & 0.552 & 0.570 & 0.841 & 0.538 & 0.747 & 0.639 & 0.552 & 0.589 & 0.621 & 0.517 & 0.524 & 0.774 & 0.738 \\ -0.477 & -0.760 & -0.406 & -0.793 & -0.334 & -0.595 & -0.662 & -0.805 & -0.327 & -0.506 & -0.792 & -0.869 & 4.622 & -0.424 & -0.773 & 0.634 & 0.658 & 0.571 & 0.647 & 0.584 & 0.516 & 0.649 & 0.878 & 0.898 & 0.613 & 0.675 & 0.732 & 0.705 & 0.680 & 0.530 \\ -0.646 & -0.861 & -0.316 & -0.703 & -0.429 & -0.765 & -0.578 & -0.554 & -0.529 & -0.703 & -0.731 & -0.367 & -0.424 & 7.699 & -0.824 & 0.713 & 0.697 & 0.645 & 0.854 & 0.881 & 0.769 & 0.889 & 0.558 & 0.683 & 0.818 & 0.604 & 0.808 & 0.839 & 0.828 & 0.615 \\ -0.732 & -0.863 & -0.818 & -0.651 & -0.357 & -0.840 & -0.662 & -0.451 & -0.351 & -0.706 & -0.522 & -0.404 & -0.773 & -0.824 & 8.296 & 0.723 & 0.844 & 0.815 & 0.855 & 0.590 & 0.880 & 0.811 & 0.732 & 0.561 & 0.716 & 0.505 & 0.547 & 0.882 & 0.847 & 0.772 \\ 0.574 & 0.843 & 0.648 & 0.647 & 0.692 & 0.839 & 0.514 & 0.798 & 0.791 & 0.528 & 0.699 & 0.744 & 0.634 & 0.713 & 0.723 & 8.717 & 0.596 & 0.541 & 0.797 & 0.865 & 0.563 & 0.733 & 0.769 & 0.549 & 0.877 & 0.857 & 0.764 & 0.652 & 0.839 & 0.506 \\ 0.657 & 0.841 & 0.839 & 0.876 & 0.524 & 0.577 & 0.869 & 0.881 & 0.595 & 0.818 & 0.827 & 0.501 & 0.658 & 0.697 & 0.844 & 0.596 & 0.032 & 0.850 & 0.612 & 0.614 & 0.790 & 0.680 & 0.701 & 0.707 & 0.696 & 0.845 & 0.720 & 0.799 & 0.553 & 0.896 \\ 0.584 & 0.770 & 0.593 & 0.831 & 0.633 & 0.568 & 0.780 & 0.575 & 0.886 & 0.530 & 0.595 & 0.552 & 0.571 & 0.645 & 0.815 & 0.541 & 0.850 & 8.743 & 0.733 & 0.604 & 0.744 & 0.762 & 0.737 & 0.668 & 0.666 & 0.544 & 0.549 & 0.869 & 0.634 & 0.586 \\ 0.500 & 0.856 & 0.536 & 0.510 & 0.758 & 0.643 & 0.572 & 0.541 & 0.572 & 0.765 & 0.507 & 0.570 & 0.647 & 0.854 & 0.855 & 0.797 & 0.612 & 0.733 & 7.684 & 0.712 & 0.599 & 0.771 & 0.894 & 0.517 & 0.864 & 0.604 & 0.676 & 0.642 & 0.843 & 0.618 \\ 0.594 & 0.504 & 0.514 & 0.792 & 0.753 & 0.896 & 0.783 & 0.536 & 0.743 & 0.807 & 0.770 & 0.841 & 0.584 & 0.881 & 0.590 & 0.865 & 0.614 & 0.604 & 0.712 & 8.172 & 0.657 & 0.578 & 0.615 & 0.727 & 0.666 & 0.673 & 0.555 & 0.664 & 0.723 & 0.690 \\ 0.792 & 0.851 & 0.598 & 0.605 & 0.876 & 0.871 & 0.727 & 0.557 & 0.805 & 0.710 & 0.714 & 0.538 & 0.516 & 0.769 & 0.880 & 0.563 & 0.790 & 0.744 & 0.599 & 0.657 & 8.400 & 0.533 & 0.626 & 0.780 & 0.600 & 0.811 & 0.754 & 0.527 & 0.787 & 0.802 \\ 0.585 & 0.627 & 0.651 & 0.722 & 0.825 & 0.520 & 0.790 & 0.858 & 0.507 & 0.884 & 0.501 & 0.747 & 0.649 & 0.889 & 0.811 & 0.733 & 0.680 & 0.762 & 0.771 & 0.578 & 0.533 & 8.846 & 0.583 & 0.895 & 0.752 & 0.816 & 0.757 & 0.681 & 0.854 & 0.761 \\ 0.631 & 0.793 & 0.823 & 0.509 & 0.574 & 0.644 & 0.779 & 0.741 & 0.868 & 0.827 & 0.695 & 0.639 & 0.878 & 0.558 & 0.732 & 0.769 & 0.701 & 0.737 & 0.894 & 0.615 & 0.626 & 0.583 & 7.347 & 0.795 & 0.820 & 0.637 & 0.803 & 0.567 & 0.550 & 0.681 \\ 0.764 & 0.700 & 0.728 & 0.852 & 0.882 & 0.802 & 0.678 & 0.666 & 0.889 & 0.797 & 0.809 & 0.552 & 0.898 & 0.683 & 0.561 & 0.549 & 0.707 & 0.668 & 0.517 & 0.727 & 0.780 & 0.895 & 0.795 & 9.060 & 0.698 & 0.607 & 0.863 & 0.586 & 0.764 & 0.643 \\ 0.896 & 0.755 & 0.766 & 0.540 & 0.838 & 0.709 & 0.779 & 0.789 & 0.841 & 0.591 & 0.782 & 0.589 & 0.613 & 0.818 & 0.716 & 0.877 & 0.696 & 0.666 & 0.864 & 0.666 & 0.600 & 0.752 & 0.820 & 0.698 & 9.203 & 0.766 & 0.756 & 0.678 & 0.719 & 0.839 \\ 0.826 & 0.798 & 0.693 & 0.700 & 0.756 & 0.763 & 0.679 & 0.818 & 0.873 & 0.839 & 0.877 & 0.621 & 0.675 & 0.604 & 0.505 & 0.857 & 0.845 & 0.544 & 0.604 & 0.673 & 0.811 & 0.816 & 0.637 & 0.607 & 0.766 & 4.871 & 0.555 & 0.634 & 0.652 & 0.699 \\ 0.851 & 0.731 & 0.574 & 0.856 & 0.508 & 0.754 & 0.600 & 0.626 & 0.759 & 0.779 & 0.843 & 0.517 & 0.732 & 0.808 & 0.547 & 0.764 & 0.720 & 0.549 & 0.676 & 0.555 & 0.754 & 0.757 & 0.803 & 0.863 & 0.756 & 0.555 & 4.893 & 0.822 & 0.638 & 0.505 \\ 0.799 & 0.825 & 0.744 & 0.880 & 0.728 & 0.679 & 0.633 & 0.641 & 0.551 & 0.782 & 0.854 & 0.524 & 0.705 & 0.839 & 0.882 & 0.652 & 0.799 & 0.869 & 0.642 & 0.664 & 0.527 & 0.681 & 0.567 & 0.586 & 0.678 & 0.634 & 0.822 & 1.668 & 0.744 & 0.856 \\ 0.599 & 0.812 & 0.560 & 0.879 & 0.729 & 0.560 & 0.576 & 0.554 & 0.647 & 0.551 & 0.571 & 0.774 & 0.680 & 0.828 & 0.847 & 0.839 & 0.553 & 0.634 & 0.843 & 0.723 & 0.787 & 0.854 & 0.550 & 0.764 & 0.719 & 0.652 & 0.638 & 0.744 & 8.418 & 0.800 \\ 0.529 & 0.618 & 0.558 & 0.771 & 0.590 & 0.567 & 0.870 & 0.500 & 0.700 & 0.683 & 0.843 & 0.738 & 0.530 & 0.615 & 0.772 & 0.506 & 0.896 & 0.586 & 0.618 & 0.690 & 0.802 & 0.761 & 0.681 & 0.643 & 0.839 & 0.699 & 0.505 & 0.856 & 0.800 & 3.231 \end{pmatrix}.$$

$$L_2 = \begin{pmatrix} 2.649 & -0.363 & -0.366 & -0.271 & -0.293 & -0.355 & -0.238 & -0.144 & -0.379 & -0.356 & -0.199 & -0.263 & -0.273 & -0.277 & -0.362 & 0.220 & 0.212 & 0.132 & 0.103 & 0.236 & 0.114 & 0.162 & 0.206 & 0.160 & 0.237 & 0.183 & 0.105 & 0.282 & 0.136 & 0.145 \\ -0.363 & 2.186 & -0.247 & -0.110 & -0.368 & -0.172 & -0.168 & -0.241 & -0.380 & -0.146 & -0.102 & -0.133 & -0.161 & -0.387 & -0.216 & 0.269 & 0.241 & 0.104 & 0.222 & 0.181 & 0.199 & 0.214 & 0.275 & 0.102 & 0.232 & 0.143 & 0.172 & 0.136 & 0.145 & 0.195 \\ -0.366 & -0.247 & 2.548 & -0.383 & -0.390 & -0.389 & -0.318 & -0.357 & -0.201 & -0.170 & -0.115 & -0.193 & -0.301 & -0.109 & -0.275 & 0.273 & 0.115 & 0.205 & 0.150 & 0.242 & 0.259 & 0.243 & 0.146 & 0.106 & 0.119 & 0.202 & 0.231 & 0.199 & 0.269 & 0.283 \\ -0.271 & -0.110 & -0.383 & 2.351 & -0.355 & -0.287 & -0.301 & -0.353 & -0.373 & -0.120 & -0.336 & -0.322 & -0.322 & -0.220 & -0.154 & 0.177 & 0.153 & 0.113 & 0.269 & 0.140 & 0.151 & 0.160 & 0.249 & 0.205 & 0.168 & 0.108 & 0.287 & 0.250 & 0.157 & 0.229 \\ -0.293 & -0.368 & -0.390 & -0.355 & 1.942 & -0.270 & -0.274 & -0.164 & -0.241 & -0.275 & -0.312 & -0.396 & -0.325 & -0.103 & -0.350 & 0.208 & 0.195 & 0.145 & 0.263 & 0.278 & 0.186 & 0.230 & 0.196 & 0.168 & 0.140 & 0.246 & 0.291 & 0.289 & 0.267 & 0.281 \\ -0.355 & -0.172 & -0.389 & -0.287 & -0.270 & 1.944 & -0.273 & -0.148 & -0.351 & -0.110 & -0.376 & -0.298 & -0.155 & -0.139 & -0.229 & 0.178 & 0.237 & 0.242 & 0.230 & 0.242 & 0.153 & 0.112 & 0.124 & 0.213 & 0.121 & 0.214 & 0.123 & 0.125 & 0.246 \\ -0.238 & -0.168 & -0.318 & -0.301 & -0.274 & -0.273 & 2.418 & -0.308 & -0.356 & -0.278 & -0.186 & -0.126 & -0.180 & -0.138 & -0.145 & 0.197 & 0.162 & 0.189 & 0.164 & 0.199 & 0.176 & 0.128 & 0.266 & 0.201 & 0.226 & 0.102 & 0.241 & 0.185 & 0.218 & 0.150 \\ -0.144 & -0.241 & -0.357 & -0.353 & -0.164 & -0.148 & -0.308 & 2.391 & -0.260 & -0.236 & -0.346 & -0.120 & -0.304 & -0.319 & -0.313 & 0.183 & 0.161 & 0.146 & 0.240 & 0.186 & 0.181 & 0.122 & 0.115 & 0.171 & 0.208 & 0.179 & 0.224 & 0.145 & 0.234 & 0.208 \\ -0.379 & -0.380 & -0.201 & -0.373 & -0.241 & -0.351 & -0.356 & -0.260 & 2.422 & -0.167 & -0.267 & -0.304 & -0.333 & -0.377 & -0.190 & 0.200 & 0.205 & 0.177 & 0.161 & 0.259 & 0.103 & 0.241 & 0.180 & 0.153 & 0.238 & 0.159 & 0.117 & 0.222 & 0.279 & 0.293 \\ -0.356 & -0.146 & -0.170 & -0.120 & -0.275 & -0.110 & -0.278 & -0.236 & -0.167 & 2.689 & -0.285 & -0.110 & -0.142 & -0.110 & -0.246 & 0.155 & 0.295 & 0.295 & 0.212 & 0.213 & 0.125 & 0.220 & 0.261 & 0.212 & 0.128 & 0.207 & 0.188 & 0.143 & 0.236 & 0.214 \\ -0.199 & -0.102 & -0.115 & -0.336 & -0.312 & -0.376 & -0.186 & -0.346 & -0.267 & -0.285 & 2.242 & -0.179 & -0.239 & -0.324 & -0.391 & 0.150 & 0.233 & 0.126 & 0.235 & 0.230 & 0.262 & 0.214 & 0.249 & 0.295 & 0.271 & 0.281 & 0.264 & 0.184 & 0.252 & 0.131 \\ -0.263 & -0.133 & -0.193 & -0.322 & -0.396 & -0.298 & -0.126 & -0.120 & -0.304 & -0.110 & -0.179 & 2.668 & -0.267 & -0.295 & -0.362 & 0.206 & 0.171 & 0.117 & 0.103 & 0.283 & 0.166 & 0.286 & 0.144 & 0.167 & 0.106 & 0.272 & 0.166 & 0.298 & 0.261 & 0.193 \\ -0.273 & -0.161 & -0.301 & -0.322 & -0.325 & -0.155 & -0.180 & -0.304 & -0.333 & -0.142 & -0.239 & -0.267 & 1.838 & -0.148 & -0.157 & 0.251 & 0.144 & 0.136 & 0.151 & 0.215 & 0.112 & 0.297 & 0.168 & 0.153 & 0.299 & 0.228 & 0.224 & 0.280 & 0.228 & 0.103 \\ -0.277 & -0.387 & -0.109 & -0.220 & -0.103 & -0.139 & -0.138 & -0.319 & -0.377 & -0.110 & -0.324 & -0.295 & -0.148 & 1.558 & -0.188 & 0.101 & 0.239 & 0.139 & 0.200 & 0.195 & 0.176 & 0.143 & 0.250 & 0.246 & 0.116 & 0.101 & 0.211 & 0.217 & 0.196 & 0.182 \\ -0.362 & -0.216 & -0.275 & -0.154 & -0.350 & -0.229 & -0.145 & -0.313 & -0.190 & -0.246 & -0.391 & -0.362 & -0.157 & -0.188 & 2.317 & 0.198 & 0.297 & 0.237 & 0.128 & 0.244 & 0.116 & 0.227 & 0.220 & 0.227 & 0.139 & 0.171 & 0.130 & 0.261 & 0.198 & 0.279 \\ 0.220 & 0.269 & 0.273 & 0.177 & 0.208 & 0.178 & 0.197 & 0.183 & 0.200 & 0.155 & 0.150 & 0.206 & 0.251 & 0.101 & 0.198 & 2.044 & 0.172 &$$

in case when $X_{12} = 1$ the variables are generated as:

1. We have $C_2 \setminus C_1 = \{X_{40}\}$, $C_2 \cap C_1 = \{X_1, X_{18}, X_{21}, X_{22}, X_{26}, X_{29}, X_{32}, X_{33}\}$ and generate

$$X_{40} = (X_1, X_{18}, X_{21}, X_{22}, X_{26}, X_{29}, X_{32}, X_{33}) \cdot \mathbf{r}_{s_2, C_2 \setminus s_2} + \xi_2, \quad (\text{C.2.13})$$

where $\mathbf{r}_{s_2, C_2 \setminus s_2} = (0.15, 0.25, 0.1, 0.25, 0.27, 0.2, 0.1, 0.2)'$, and the elements in ξ_2 are generated from $U(-0.5, 0.5)$;

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_5\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_8, X_9, X_{14}, X_{17}, X_{19}\}$ and generate

$$X_5 = (X_8, X_9, X_{14}, X_{17}, X_{19}) \cdot \mathbf{r}_{s_3, C_3 \setminus s_3} + \xi_3, \quad (\text{C.2.14})$$

where $\mathbf{r}_{s_3, C_3 \setminus s_3} = (0.2, 0.1, -0.2, .031, .01)'$, and the elements in ξ_3 are generated from $U(-0.5, 0.5)$;

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_7\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_1, X_4, X_8\}$ and generate

$$X_7 = (X_1, X_4, X_8) \cdot \mathbf{r}_{s_4, C_4 \setminus s_4} + \xi_4, \quad (\text{C.2.15})$$

where $\mathbf{r}_{s_4, C_4 \setminus s_4} = (0.1, -0.2, 0.3)'$, and $\xi_4 \sim U(-0.5, 0.5)$;

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_{10}\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_{21}, X_{27}, X_{34}\}$ and generate

$$X_{10} = (X_{21}, X_{27}, X_{34}) \cdot \mathbf{r}_{s_5, C_5 \setminus s_5} + \xi_5, \quad (\text{C.2.16})$$

where $\mathbf{r}_{s_5, C_5 \setminus s_5} = (-0.2, 0.3, 0.2)'$, and $\xi_5 \sim U(-0.5, 0.5)$;

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_{20}\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_{24}, X_{28}, X_{38}\}$ and generate

$$X_{20} = (X_{24}, X_{28}, X_{38}) \cdot \mathbf{r}_{S_6, C_6 \setminus S_6} + \xi_6, \quad (\text{C.2.17})$$

where $\mathbf{r}_{S_6, C_6 \setminus S_6} = (0.15, 0.1, 0.2)'$, and $\xi_6 \sim U(-0.5, 0.5)$;

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_{35}\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_9, X_{11}, X_{33}\}$ and generate

$$X_{35} = (X_9, X_{11}, X_{33}) \cdot \mathbf{r}_{S_7, C_7 \setminus S_7} + \xi_7, \quad (\text{C.2.18})$$

where $\mathbf{r}_{S_7, C_7 \setminus S_7} = (0.2, 0.15, 0.3)'$, and $\xi_7 \sim U(-0.5, 0.5)$;

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_{30}\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_{19}, X_{23}\}$ and generate

$$X_{30} = (X_{19}, X_{23}) \cdot \mathbf{r}_{S_8, C_8 \setminus S_8} + \xi_8, \quad (\text{C.2.19})$$

where $\mathbf{r}_{S_8, C_8 \setminus S_8} = \begin{pmatrix} 0.2 \\ -0.1 \end{pmatrix}$, and $\xi_8 \sim U(-0.5, 0.5)$;

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_3\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_6\}$ and generate

$$X_3 = (X_6) \cdot \mathbf{r}_{S_9, C_9 \setminus S_9} + \xi_9, \quad (\text{C.2.20})$$

where $\mathbf{r}_{S_9, C_9 \setminus S_9} = (-0.47)$, and $\xi_9 \sim U(-0.5, 0.5)$;

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_{15}\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_{39}\}$ and generate

$$X_{15} = (X_{39}) \cdot \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} + \xi_{10}, \quad (\text{C.2.21})$$

where $\mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} = (0.6)$, and $\xi_{10} \sim U(-0.5, 0.5)$;

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{25}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_{38}\}$ and generate

$$X_{25} = (X_{38}) \cdot \mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} + \xi_{11}, \quad (\text{C.2.22})$$

where $\mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} = (0.5)$, and $\xi_{11} \sim U(-0.5, 0.5)$.

In case when $X_{12} = 0$ the variables are generated as:

1. We have $C_2 \setminus C_1 = \{X_{40}\}$, $C_2 \cap C_1 = \{X_1, X_{18}, X_{21}, X_{22}, X_{26}, X_{29}, X_{32}, X_{33}\}$ and generate

$$X_{40} = (X_1, X_{18}, X_{21}, X_{22}, X_{26}, X_{29}, X_{32}, X_{33}) \cdot \mathbf{r}_{s_2, C_2 \setminus s_2} + \xi_2, \quad (\text{C.2.23})$$

where $\mathbf{r}_{s_2, C_2 \setminus s_2} = (0.15, 0.25, 0.1, 0.25, 0.27, 0.2, 0.1, 0.2)'$, and the elements in ξ_2 are generated from $U(-0.5, 0.5)$;

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_5\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_8, X_9, X_{14}, X_{17}, X_{19}\}$ and generate

$$X_5 = (X_8, X_9, X_{14}, X_{17}, X_{19}) \cdot \mathbf{r}_{s_3, C_3 \setminus s_3} + \xi_3, \quad (\text{C.2.24})$$

where $\mathbf{r}_{s_3, C_3 \setminus s_3} = (0.2, 0.1, -0.2, .031, .01)'$, and $\xi_3 \sim U(-0.5, 0.5)$;

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_7\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_1, X_4, X_8\}$ and generate

$$X_7 = (X_1, X_4, X_8) \cdot \mathbf{r}_{s_4, C_4 \setminus s_4} + \xi_4, \quad (\text{C.2.25})$$

where $\mathbf{r}_{s_4, C_4 \setminus s_4} = (0.1, -0.2, 0.3)'$, and $\xi_4 \sim U(-0.5, 0.5)$;

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_{10}\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_{21}, X_{27}, X_{34}\}$ and generate

$$X_{10} = (X_{21}, X_{27}, X_{34}) \cdot \mathbf{r}_{S_5, C_5 \setminus S_5} + \xi_5, \quad (\text{C.2.26})$$

where $\mathbf{r}_{S_5, C_5 \setminus S_5} = (-0.2, 0.3, 0.2)'$, and $\xi_5 \sim U(-0.5, 0.5)$;

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_{20}\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_{24}, X_{28}, X_{38}\}$ and generate

$$X_{20} = (X_{24}, X_{28}, X_{38}) \cdot \mathbf{r}_{S_6, C_6 \setminus S_6} + \xi_6, \quad (\text{C.2.27})$$

where $\mathbf{r}_{S_6, C_6 \setminus S_6} = (0.15, 0.1, 0.2)'$, and $\xi_6 \sim U(-0.5, 0.5)$;

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_{35}\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_9, X_{11}, X_{33}\}$ and generate

$$X_{35} = (X_9, X_{11}, X_{33}) \cdot \mathbf{r}_{S_7, C_7 \setminus S_7} + \xi_7, \quad (\text{C.2.28})$$

where $\mathbf{r}_{S_7, C_7 \setminus S_7} = (0.2, 0.15, 0.3)'$, and $\xi_7 \sim U(-0.5, 0.5)$;

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_{30}\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_{19}, X_{23}\}$ and generate

$$X_{30} = (X_{19}, X_{23}) \cdot \mathbf{r}_{S_8, C_8 \setminus S_8} + \xi_8, \quad (\text{C.2.29})$$

where $\mathbf{r}_{S_8, C_8 \setminus S_8} = \begin{pmatrix} 0.2 \\ -0.1 \end{pmatrix}$, and $\xi_8 \sim U(-0.5, 0.5)$;

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_3\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_6\}$ and generate

$$X_3 = (X_6) \cdot \mathbf{r}_{S_9, C_9 \setminus S_9} + \xi_9, \quad (\text{C.2.30})$$

where $\mathbf{r}_{S_9, C_9 \setminus S_9} = (-0.47)$, and $\xi_9 \sim U(-0.5, 0.5)$;

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_{15}\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_{39}\}$ and generate

$$X_{15} = (X_{39}) \cdot \mathbf{r}_{s_{10}, C_{10} \setminus s_{10}} + \xi_{10}, \quad (\text{C.2.31})$$

where $\mathbf{r}_{s_{10}, C_{10} \setminus s_{10}} = (0.6)$, and $\xi_{10} \sim U(-0.5, 0.5)$;

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{25}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_{38}\}$ and generate

$$X_{25} = (X_{38}) \cdot \mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} + \xi_{11}, \quad (\text{C.2.32})$$

where $\mathbf{r}_{s_{11}, C_{11} \setminus s_{11}} = (0.5)$, and $\xi_{11} \sim U(-0.5, 0.5)$.

C.3 Scenario 6

Below are the details about how the data are generated for Scenario 6.

Using the graph structure given in Figure 5.9, the perfect ordering of cliques is $C_1 = \{X_1, X_4, X_{14}, X_{17}, X_{19}, X_{24}, X_{29}, X_{30}, X_{32}, X_{33}\}$, $C_2 = \{X_{11}, X_{17}, X_{19}\}$, $C_3 = \{X_{12}, X_{24}, X_{30}\}$, $C_4 = \{X_{13}, X_{14}, X_{32}\}$, $C_5 = \{X_2, X_4, X_{14}\}$, $C_6 = \{X_{15}, X_{24}, X_{29}\}$, $C_7 = \{X_{16}, X_4, X_{32}\}$, $C_8 = \{X_3, X_{19}, X_{24}\}$, $C_9 = \{X_{18}, X_{19}, X_{29}\}$, $C_{10} = \{X_5, X_{14}, X_{17}\}$, $C_{11} = \{X_{14}, X_{19}, X_{20}\}$, $C_{12} = \{X_{17}, X_{21}, X_{33}\}$, $C_{13} = \{X_{19}, X_{22}, X_{30}\}$, $C_{14} = \{X_4, X_{23}, X_{24}\}$, $C_{15} = \{X_6, X_{32}, X_{33}\}$, $C_{16} = \{X_{14}, X_{17}, X_{25}\}$, $C_{17} = \{X_{26}, X_{30}, X_{33}\}$, $C_{18} = \{X_4, X_{27}, X_{30}\}$, $C_{19} = \{X_{14}, X_{19}, X_{28}\}$, $C_{20} = \{X_7, X_{24}, X_{30}\}$, $C_{21} = \{X_4, X_9, X_{29}\}$, $C_{22} = \{X_{30}, X_{31}, X_{32}\}$, $C_{23} = \{X_1, X_{10}, X_{33}\}$, $C_{24} = \{X_{14}, X_{24}, X_{34}\}$, $C_{25} = \{X_{29}, X_{33}, X_{35}\}$, $C_{26} = \{X_{19}, X_{33}, X_{36}\}$, $C_{27} = \{X_1, X_8\}$, $C_{28} = \{X_4, X_{38}\}$, $C_{29} = \{X_{32}, X_{39}\}$, $C_{30} = \{X_{37}\}$, $C_{31} = \{X_{40}\}$.

All variables in C_1 ($C_1 = C^*$) are generated from $MVN(\mathbf{0}, \Sigma^{-1} = \mathbf{K}_1)$ where

$$\mathbf{K}_1 = \begin{pmatrix} 1.882 & -0.587 & -0.602 & -0.474 & -0.573 & 0.499 & 0.494 & 0.517 & 0.941 & 0.558 \\ -0.587 & 6.498 & -0.712 & -0.676 & -0.774 & 0.735 & 0.212 & 0.609 & 0.779 & 0.310 \\ -0.602 & -0.712 & 1.141 & -0.891 & -0.356 & 0.875 & 0.405 & 0.216 & 0.696 & 0.907 \\ -0.474 & -0.676 & -0.891 & 6.873 & -0.596 & 0.705 & 0.428 & 0.722 & 0.796 & 0.835 \\ -0.573 & -0.774 & -0.356 & -0.596 & 7.699 & 0.499 & 0.629 & 0.726 & 0.885 & 0.764 \\ 0.499 & 0.735 & 0.875 & 0.705 & 0.499 & 8.743 & 0.455 & 0.402 & 0.905 & 0.923 \\ 0.494 & 0.212 & 0.405 & 0.428 & 0.629 & 0.455 & 8.846 & 0.893 & 0.311 & 0.579 \\ 0.517 & 0.609 & 0.216 & 0.722 & 0.726 & 0.402 & 0.893 & 2.244 & 0.474 & 0.427 \\ 0.941 & 0.779 & 0.696 & 0.796 & 0.885 & 0.905 & 0.311 & 0.474 & 9.060 & 0.367 \\ 0.558 & 0.310 & 0.907 & 0.835 & 0.764 & 0.923 & 0.579 & 0.427 & 0.367 & 9.203 \end{pmatrix}.$$

Variable X_{37} is generated from $N(0.01, 0.973276)$ and variable X_{40} is generated from

$N(0.01, 1.531571)$.

Then the following linear condition is used

$$\mathbf{x}_{C_k \setminus S_k} = \mathbf{x}_{S_k} \cdot \mathbf{r}_{S_k, C_k \setminus S_k} + \boldsymbol{\xi}_k, \quad (\text{C.3.33})$$

in order to generate the rest of the variables, as follows:

1. We have $C_2 \setminus C_1 = \{X_{11}\}$, $C_2 \cap C_1 = \{X_{17}, X_{19}\}$ and generate

$$X_{11} = (X_{17}, X_{19}) \cdot \mathbf{r}_{S_2, C_2 \setminus S_2} + \xi_2, \quad (\text{C.3.34})$$

where $\mathbf{r}_{S_2, C_2 \setminus S_2} = \begin{pmatrix} -0.30 \\ 0.40 \end{pmatrix}$, and $\xi_2 \sim U(-0.5, 0.5)$;

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_{12}\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_{24}, X_{30}\}$ and generate

$$X_{12} = (X_{24}, X_{30}) \cdot \mathbf{r}_{S_3, C_3 \setminus S_3} + \xi_3, \quad (\text{C.3.35})$$

$$\text{where } \mathbf{r}_{S_3, C_3 \setminus S_3} = \begin{pmatrix} 0.20 \\ 0.30 \end{pmatrix}, \text{ and } \xi_3 \sim U(-0.5, 0.5);$$

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_{13}\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_{14}, X_{32}\}$ and generate

$$X_{13} = (X_{14}, X_{32}) \cdot \mathbf{r}_{S_4, C_4 \setminus S_4} + \xi_4, \quad (\text{C.3.36})$$

$$\text{where } \mathbf{r}_{S_4, C_4 \setminus S_4} = \begin{pmatrix} -0.10 \\ 0.20 \end{pmatrix}, \text{ and } \xi_4 \sim U(-0.5, 0.5);$$

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_2\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_4, X_{14}\}$ and generate

$$X_2 = (X_4, X_{14}) \cdot \mathbf{r}_{S_5, C_5 \setminus S_5} + \xi_5, \quad (\text{C.3.37})$$

$$\text{where } \mathbf{r}_{S_5, C_5 \setminus S_5} = \begin{pmatrix} -0.20 \\ -0.30 \end{pmatrix}, \text{ and } \xi_5 \sim U(-0.5, 0.5);$$

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_{15}\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_{24}, X_{29}\}$ and generate

$$X_{15} = (X_{24}, X_{29}) \cdot \mathbf{r}_{S_6, C_6 \setminus S_6} + \xi_6, \quad (\text{C.3.38})$$

$$\text{where } \mathbf{r}_{S_6, C_6 \setminus S_6} = \begin{pmatrix} 0.20 \\ 0.30 \end{pmatrix}, \text{ and } \xi_6 \sim U(-0.5, 0.5);$$

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_{16}\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_4, X_{32}\}$ and generate

$$X_{16} = (X_4, X_{32}) \cdot \mathbf{r}_{S_7, C_7 \setminus S_7} + \xi_7, \quad (\text{C.3.39})$$

where $\mathbf{r}_{S_7, C_7 \setminus S_7} = \begin{pmatrix} -0.10 \\ 0.40 \end{pmatrix}$, and $\xi_7 \sim U(-0.5, 0.5)$;

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_3\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_{19}, X_{24}\}$ and generate

$$X_3 = (X_{19}, X_{24}) \cdot \mathbf{r}_{S_8, C_8 \setminus S_8} + \xi_8, \quad (\text{C.3.40})$$

where $\mathbf{r}_{S_8, C_8 \setminus S_8} = \begin{pmatrix} 0.40 \\ 0.10 \end{pmatrix}$, and $\xi_8 \sim U(-0.5, 0.5)$;

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_{18}\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_{19}, X_{29}\}$ and generate

$$X_{18} = (X_{19}, X_{29}) \cdot \mathbf{r}_{S_9, C_9 \setminus S_9} + \xi_9, \quad (\text{C.3.41})$$

where $\mathbf{r}_{S_9, C_9 \setminus S_9} = \begin{pmatrix} -0.30 \\ 0.35 \end{pmatrix}$, and $\xi_9 \sim U(-0.5, 0.5)$;

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_5\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_{14}, X_{17}\}$ and generate

$$X_5 = (X_{14}, X_{17}) \cdot \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} + \xi_{10}, \quad (\text{C.3.42})$$

where $\mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} = \begin{pmatrix} -0.20 \\ -0.10 \end{pmatrix}$, and $\xi_{10} \sim U(-0.5, 0.5)$;

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{20}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_{14}, X_{19}\}$ and generate

$$X_{20} = (X_{14}, X_{19}) \cdot \mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} + \xi_{11}, \quad (\text{C.3.43})$$

where $\mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} = \begin{pmatrix} -0.30 \\ 0.10 \end{pmatrix}$, and $\xi_{11} \sim U(-0.5, 0.5)$;

11. We have $C_{12} \setminus \{\cup_{k=1}^{11} C_k\} = \{X_{21}\}$, $C_{12} \cap \{\cup_{k=1}^{11} C_k\} = \{X_{17}, X_{33}\}$ and generate

$$X_{21} = (X_{17}, X_{33}) \cdot \mathbf{r}_{S_{12}, C_{12} \setminus S_{12}} + \xi_{12}, \quad (\text{C.3.44})$$

$$\text{where } \mathbf{r}_{S_{12}, C_{12} \setminus S_{12}} = \begin{pmatrix} 0.28 \\ -0.30 \end{pmatrix}, \text{ and } \xi_{12} \sim U(-0.5, 0.5);$$

12. We have $C_{13} \setminus \{\cup_{k=1}^{12} C_k\} = \{X_{22}\}$, $C_{13} \cap \{\cup_{k=1}^{12} C_k\} = \{X_{19}, X_{30}\}$ and generate

$$X_{22} = (X_{19}, X_{30}) \cdot \mathbf{r}_{S_{13}, C_{13} \setminus S_{13}} + \xi_{13}, \quad (\text{C.3.45})$$

$$\text{where } \mathbf{r}_{S_{13}, C_{13} \setminus S_{13}} = \begin{pmatrix} 0.20 \\ -0.10 \end{pmatrix}, \text{ and } \xi_{13} \sim U(-0.5, 0.5);$$

13. We have $C_{14} \setminus \{\cup_{k=1}^{13} C_k\} = \{X_{23}\}$, $C_{14} \cap \{\cup_{k=1}^{13} C_k\} = \{X_4, X_{24}\}$ and generate

$$X_{23} = (X_4, X_{24}) \cdot \mathbf{r}_{S_{14}, C_{14} \setminus S_{14}} + \xi_{14}, \quad (\text{C.3.46})$$

$$\text{where } \mathbf{r}_{S_{14}, C_{14} \setminus S_{14}} = \begin{pmatrix} -0.10 \\ 0.40 \end{pmatrix}, \text{ and } \xi_{14} \sim U(-0.5, 0.5);$$

14. We have $C_{15} \setminus \{\cup_{k=1}^{14} C_k\} = \{X_6\}$, $C_{15} \cap \{\cup_{k=1}^{14} C_k\} = \{X_{32}, X_{33}\}$ and generate

$$X_6 = (X_{32}, X_{33}) \cdot \mathbf{r}_{S_{15}, C_{15} \setminus S_{15}} + \xi_{15}, \quad (\text{C.3.47})$$

$$\text{where } \mathbf{r}_{S_{15}, C_{15} \setminus S_{15}} = \begin{pmatrix} 0.30 \\ 0.20 \end{pmatrix}, \text{ and } \xi_{15} \sim U(-0.5, 0.5);$$

15. We have $C_{16} \setminus \{\cup_{k=1}^{15} C_k\} = \{X_{25}\}$, $C_{16} \cap \{\cup_{k=1}^{15} C_k\} = \{X_{14}, X_{17}\}$ and generate

$$X_{25} = (X_{14}, X_{17}) \cdot \mathbf{r}_{S_{16}, C_{16} \setminus S_{16}} + \xi_{16}, \quad (\text{C.3.48})$$

where $\mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} = \begin{pmatrix} 0.80 \\ 0.20 \end{pmatrix}$, and $\xi_{16} \sim U(-0.5, 0.5)$;

16. We have $C_{17} \setminus \{\cup_{k=1}^{16} C_k\} = \{X_{26}\}$, $C_{17} \cap \{\cup_{k=1}^{16} C_k\} = \{X_{30}, X_{33}\}$ and generate

$$X_{26} = (X_{30}, X_{33}) \cdot \mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} + \xi_{17}, \quad (\text{C.3.49})$$

where $\mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} = \begin{pmatrix} 0.33 \\ -0.25 \end{pmatrix}$, and $\xi_{17} \sim U(-0.5, 0.5)$;

17. We have $C_{18} \setminus \{\cup_{k=1}^{17} C_k\} = \{X_{27}\}$, $C_{18} \cap \{\cup_{k=1}^{17} C_k\} = \{X_4, X_{30}\}$ and generate

$$X_{27} = (X_4, X_{30}) \cdot \mathbf{r}_{s_{18}, C_{18} \setminus s_{18}} + \xi_{18}, \quad (\text{C.3.50})$$

where $\mathbf{r}_{s_{18}, C_{18} \setminus s_{18}} = \begin{pmatrix} 0.60 \\ -0.40 \end{pmatrix}$, and $\xi_{18} \sim U(-0.5, 0.5)$;

18. We have $C_{19} \setminus \{\cup_{k=1}^{18} C_k\} = \{X_{28}\}$, $C_{19} \cap \{\cup_{k=1}^{18} C_k\} = \{X_{14}, X_{19}\}$ and generate

$$X_{28} = (X_{14}, X_{19}) \cdot \mathbf{r}_{s_{19}, C_{19} \setminus s_{19}} + \xi_{19}, \quad (\text{C.3.51})$$

where $\mathbf{r}_{s_{19}, C_{19} \setminus s_{19}} = \begin{pmatrix} 0.30 \\ -0.20 \end{pmatrix}$, and $\xi_{19} \sim U(-0.5, 0.5)$;

19. We have $C_{20} \setminus \{\cup_{k=1}^{19} C_k\} = \{X_7\}$, $C_{20} \cap \{\cup_{k=1}^{19} C_k\} = \{X_{24}, X_{30}\}$ and generate

$$X_7 = (X_{24}, X_{30}) \cdot \mathbf{r}_{s_{20}, C_{20} \setminus s_{20}} + \xi_{20}, \quad (\text{C.3.52})$$

where $\mathbf{r}_{s_{20}, C_{20} \setminus s_{20}} = \begin{pmatrix} -0.20 \\ 0.40 \end{pmatrix}$, and $\xi_{20} \sim U(-0.5, 0.5)$;

20. We have $C_{21} \setminus \{\cup_{k=1}^{20} C_k\} = \{X_9\}$, $C_{21} \cap \{\cup_{k=1}^{20} C_k\} = \{X_4, X_{29}\}$ and generate

$$X_9 = (X_4, X_{29}) \cdot \mathbf{r}_{S_{21}, C_{21} \setminus S_{21}} + \xi_{21}, \quad (\text{C.3.53})$$

$$\text{where } \mathbf{r}_{S_{21}, C_{21} \setminus S_{21}} = \begin{pmatrix} 0.10 \\ 0.20 \end{pmatrix}, \text{ and } \xi_{21} \sim U(-0.5, 0.5);$$

21. We have $C_{22} \setminus \{\cup_{k=1}^{21} C_k\} = \{X_{31}\}$, $C_{22} \cap \{\cup_{k=1}^{21} C_k\} = \{X_{30}, X_{32}\}$ and generate

$$X_{31} = (X_{30}, X_{32}) \cdot \mathbf{r}_{S_{22}, C_{22} \setminus S_{22}} + \xi_{22}, \quad (\text{C.3.54})$$

$$\text{where } \mathbf{r}_{S_{22}, C_{22} \setminus S_{22}} = \begin{pmatrix} 0.20 \\ -0.30 \end{pmatrix}, \text{ and } \xi_{22} \sim U(-0.5, 0.5);$$

22. We have $C_{23} \setminus \{\cup_{k=1}^{22} C_k\} = \{X_{10}\}$, $C_{23} \cap \{\cup_{k=1}^{22} C_k\} = \{X_1, X_{33}\}$ and generate

$$X_{10} = (X_1, X_{33}) \cdot \mathbf{r}_{S_{23}, C_{23} \setminus S_{23}} + \xi_{23}, \quad (\text{C.3.55})$$

$$\text{where } \mathbf{r}_{S_{23}, C_{23} \setminus S_{23}} = \begin{pmatrix} -0.35 \\ 0.17 \end{pmatrix}, \text{ and } \xi_{23} \sim U(-0.5, 0.5);$$

23. We have $C_{24} \setminus \{\cup_{k=1}^{23} C_k\} = \{X_{34}\}$, $C_{24} \cap \{\cup_{k=1}^{23} C_k\} = \{X_{14}, X_{24}\}$ and generate

$$X_{34} = (X_{14}, X_{24}) \cdot \mathbf{r}_{S_{24}, C_{24} \setminus S_{24}} + \xi_{24}, \quad (\text{C.3.56})$$

$$\text{where } \mathbf{r}_{S_{24}, C_{24} \setminus S_{24}} = \begin{pmatrix} -0.20 \\ 0.30 \end{pmatrix}, \text{ and } \xi_{24} \sim U(-0.5, 0.5);$$

24. We have $C_{25} \setminus \{\cup_{k=1}^{24} C_k\} = \{X_{35}\}$, $C_{25} \cap \{\cup_{k=1}^{24} C_k\} = \{X_{29}, X_{33}\}$ and generate

$$X_{35} = (X_{29}, X_{33}) \cdot \mathbf{r}_{S_{25}, C_{25} \setminus S_{25}} + \xi_{25}, \quad (\text{C.3.57})$$

where $\mathbf{r}_{s_{25}, C_{25} \setminus s_{25}} = \begin{pmatrix} 0.60 \\ 0.10 \end{pmatrix}$, and $\xi_{25} \sim U(-0.5, 0.5)$;

25. We have $C_{26} \setminus \{\cup_{k=1}^{25} C_k\} = \{X_{36}\}$, $C_{26} \cap \{\cup_{k=1}^{25} C_k\} = \{X_{19}, X_{33}\}$ and generate

$$X_{36} = (X_{19}, X_{33}) \cdot \mathbf{r}_{s_{26}, C_{26} \setminus s_{26}} + \xi_{26}, \quad (\text{C.3.58})$$

where $\mathbf{r}_{s_{26}, C_{26} \setminus s_{26}} = \begin{pmatrix} -0.20 \\ -0.30 \end{pmatrix}$, and $\xi_{26} \sim U(-0.5, 0.5)$;

26. We have $C_{27} \setminus \{\cup_{k=1}^{26} C_k\} = \{X_8\}$, $C_{27} \cap \{\cup_{k=1}^{26} C_k\} = \{X_1\}$ and generate

$$X_8 = X_1 \cdot \mathbf{r}_{s_{27}, C_{27} \setminus s_{27}} + \xi_{27}, \quad (\text{C.3.59})$$

where $\mathbf{r}_{s_{27}, C_{27} \setminus s_{27}} = (-0.85)$, and $\xi_{27} \sim U(-0.5, 0.5)$;

27. We have $C_{28} \setminus \{\cup_{k=1}^{27} C_k\} = \{X_{38}\}$, $C_{28} \cap \{\cup_{k=1}^{27} C_k\} = \{X_4\}$ and generate

$$X_{38} = X_4 \cdot \mathbf{r}_{s_{28}, C_{28} \setminus s_{28}} + \xi_{28}, \quad (\text{C.3.60})$$

where $\mathbf{r}_{s_{28}, C_{28} \setminus s_{28}} = (0.47)$, and $\xi_{28} \sim U(-0.5, 0.5)$;

28. We have $C_{29} \setminus \{\cup_{k=1}^{28} C_k\} = \{X_{39}\}$, $C_{29} \cap \{\cup_{k=1}^{28} C_k\} = \{X_{32}\}$ and generate

$$X_{39} = X_{32} \cdot \mathbf{r}_{s_{29}, C_{29} \setminus s_{29}} + \xi_{29}, \quad (\text{C.3.61})$$

where $\mathbf{r}_{s_{29}, C_{29} \setminus s_{29}} = (-0.65)$, and $\xi_{29} \sim U(-0.5, 0.5)$.

APPENDIX D

GLSE simulation for more than one binary predictor variable in the data

D.1 Scenario 1

In this scenario, we have two binary predictor variables in the dataset to prove the failure of Condition 4.2.1 in giving unbiased coefficient estimates for GLSE. A total of $n = 15$ observations are generated on $v = 21$ variables based on Figure D.1, where variables X_{21} and X_{18} form the binary predictor set. The variable X_{18} separates X_{21} from the rest. The details of data generation are provided in Section D.2. Then the response y is generated from model (2.1), where the random errors ϵ are normally distributed with mean 0 and standard error $\sigma = 1$. The true values of β are given in Table D.1.

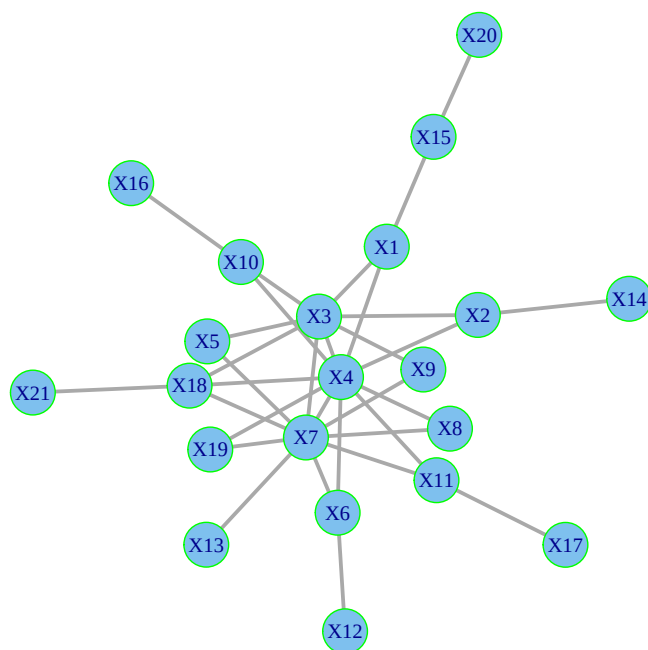


Figure D.1: *Graph structure for covariates under Scenario 1.*

Table D.1: *Results from Scenario 1, $n = 15, v = 21$.*

	β	Ridge $\hat{\beta}$	Lasso $\hat{\beta}$	Lars $\hat{\beta}$	GLSE $\hat{\beta}$ (with known graph)	GLSE $\hat{\beta}$ (with unknown graph)
X_1	0.5	0.104	0.081	0.019	0.584	0.058
X_2	1	0.556	0.367	0.174	1.179	0.812
X_3	0.6	0.295	0.168	0.227	0.642	0.378
X_4	-1	-0.221	-0.112	-0.083	-0.793	-0.413
X_5	0.7	0.272	0.176	0.090	0.64	0.414
X_6	-1	-0.193	-0.147	-0.045	-1.096	-0.284
X_7	1.1	0.443	0.261	0.382	1.122	0.636
X_8	-0.6	-0.462	-0.304	-0.183	-0.592	-0.549
X_9	1.5	0.475	0.280	0.191	1.385	0.839
X_{10}	0.5	0.277	0.187	0.128	0.581	0.326
X_{11}	0.8	0.257	0.129	0.054	0.737	0.500
X_{12}	-0.8	-0.326	-0.145	-0.056	-0.996	-0.651
X_{13}	1.8	0.995	0.741	0.339	2.109	1.414
X_{14}	0.9	0.508	0.334	0.116	0.944	0.801
X_{15}	-0.6	-0.188	-0.077	-0.039	-0.781	-0.517
X_{16}	1.3	0.623	0.398	0.195	1.362	1.013
X_{17}	0.1	0.122	0.035	0.033	0.099	0.088
X_{18}	1.7	0.822	0.966	1.038	1.779	1.606
X_{19}	-2	-0.662	-0.418	-0.167	-2.055	-1.237
X_{20}	0.8	0.385	0.197	0.043	0.681	0.602
X_{21}	2	0.959	1.274	1.311	2.066	1.762

The results of this scenario are obtained from the average of running 500 realisations and are shown in Table D.1. It is clear that GLSE fails to give unbiased estimates for X_{21} due to the fact that Condition 4.2.1 cannot be utilised. However, the estimate of the coefficient of X_{21} is still noticeably less biased than that given by the ridge, Lasso and Lars methods, which indicates the robustness of the method.

D.2 Data generation procedure in Scenario 1 Appendix D

This simulation includes binary variables X_{18} and X_{21} among the predictors which take two values 0 and 1 as follows:

$$X_{21} = (1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1) \text{ and } X_{18} = (1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1).$$

From the graph structure given in the above scenario , the perfect ordering of cliques is

$$\begin{aligned} C_1 &= \{X_4, X_7, X_3, X_{18}\}, C_2 = \{X_4, X_1, X_3\}, C_3 = \{X_4, X_2, X_3\}, C_4 = \{X_5, X_3, X_7\}, \\ C_5 &= \{X_4, X_7, X_8\}, C_6 = \{X_9, X_7, X_3\}, C_7 = \{X_4, X_{10}, X_3\}, C_8 = \{X_4, X_7, X_{11}\}, \\ C_9 &= \{X_4, X_7, X_6\}, C_{10} = \{X_4, X_7, X_{19}\}, C_{11} = \{X_{14}, X_2\}, C_{12} = \{X_{15}, X_1\}, \\ C_{13} &= \{X_{16}, X_{10}\}, C_{14} = \{X_{17}, X_{11}\}, C_{15} = \{X_{12}, X_6\}, C_{16} = \{X_{13}, X_7\}, \\ C_{17} &= \{X_{15}, X_{20}\}, C_{18} = \{X_{21}, X_{18}\}. \end{aligned}$$

The variables X_3, X_4, X_7 are generated from a multivariate normal distribution with a mean vector $\mathbf{0}$ and $\Sigma = L_1$ if $X_{21} = 0$ and $X_{18} = 0$ or L_2 if $X_{21} = 0$ and $X_{18} = 1$ or L_3 if $X_{21} = 1$ and $X_{18} = 0$ or L_4 if $X_{21} = 1$ and $X_{18} = 1$ where

$$L_1 = \begin{pmatrix} 4.575 & -0.605 & -0.177 \\ -0.605 & 1.740 & -0.001 \\ -0.177 & -0.001 & 2.128 \end{pmatrix},$$

$$L_2 = \begin{pmatrix} 1.295 & -0.313 & 0.005 \\ -0.313 & 2.394 & 0.001 \\ 0.005 & 0.001 & 1.849 \end{pmatrix},$$

$$L_3 = \begin{pmatrix} 3.397 & -0.712 & 0.100 \\ -0.712 & 1.627 & 0.001 \\ 0.100 & 0.001 & 1.771 \end{pmatrix},$$

and

$$L_4 = \begin{pmatrix} 2.889 & 0.615 & 0.790 \\ 0.615 & 3.141 & 0.820 \\ 0.790 & 0.820 & 3.988 \end{pmatrix}.$$

Then we use the linear condition

$$\mathbf{x}_{C_k \setminus S_k} = \mathbf{x}_{S_k} \cdot \mathbf{r}_{S_k, C_k \setminus S_k} + \xi_k, \quad (\text{D.2.1})$$

In the case when $X_{21} = 0$ and $X_{18} = 0$ the variables are generated as:

1. We have $C_2 \setminus C_1 = \{X_1\}$, $C_2 \cap C_1 = \{X_3, X_4\}$ and generate

$$X_1 = (X_3, X_4) \cdot \mathbf{r}_{S_2, C_2 \setminus S_2} + \xi_2, \quad (\text{D.2.2})$$

$$\text{where } \mathbf{r}_{S_2, C_2 \setminus S_2} = \begin{pmatrix} -0.1 \\ 0.2 \end{pmatrix}, \text{ and } \xi_2 \sim U(-0.5, 0.5);$$

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_2\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_3, X_4\}$ and generate

$$X_2 = (X_3, X_4) \cdot \mathbf{r}_{S_3, C_3 \setminus S_3} + \xi_3, \quad (\text{D.2.3})$$

$$\text{where } \mathbf{r}_{S_3, C_3 \setminus S_3} = \begin{pmatrix} 0.36 \\ -0.3 \end{pmatrix}, \text{ and } \xi_3 \sim U(-0.5, 0.5);$$

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_5\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_3, X_7\}$ and generate

$$X_5 = (X_3, X_7) \cdot \mathbf{r}_{S_4, C_4 \setminus S_4} + \xi_4, \quad (\text{D.2.4})$$

where $\mathbf{r}_{s_4, c_4 \setminus s_4} = \begin{pmatrix} 0.3 \\ -0.25 \end{pmatrix}$, and $\xi_4 \sim U(-0.5, 0.5)$;

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_8\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_4, X_7\}$ and generate

$$X_8 = (X_4, X_7) \cdot \mathbf{r}_{s_5, c_5 \setminus s_5} + \xi_5, \quad (\text{D.2.5})$$

where $\mathbf{r}_{s_5, c_5 \setminus s_5} = \begin{pmatrix} -0.23 \\ -0.5 \end{pmatrix}$, and $\xi_5 \sim U(-0.5, 0.5)$;

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_9\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_3, X_7\}$ and generate

$$X_9 = (X_3, X_7) \cdot \mathbf{r}_{s_6, c_6 \setminus s_6} + \xi_6, \quad (\text{D.2.6})$$

where $\mathbf{r}_{s_6, c_6 \setminus s_6} = \begin{pmatrix} 0.2 \\ 0.35 \end{pmatrix}$, and $\xi_6 \sim U(-0.5, 0.5)$;

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_{10}\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_3, X_4\}$ and generate

$$X_{10} = (X_3, X_4) \cdot \mathbf{r}_{s_7, c_7 \setminus s_7} + \xi_7, \quad (\text{D.2.7})$$

where $\mathbf{r}_{s_7, c_7 \setminus s_7} = \begin{pmatrix} 0.15 \\ 0.35 \end{pmatrix}$, and $\xi_7 \sim U(-0.5, 0.5)$;

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_{11}\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_4, X_7\}$ and generate

$$X_{11} = (X_4, X_7) \cdot \mathbf{r}_{s_8, c_8 \setminus s_8} + \xi_8, \quad (\text{D.2.8})$$

where $\mathbf{r}_{s_8, c_8 \setminus s_8} = \begin{pmatrix} 0.28 \\ 0.2 \end{pmatrix}$, and $\xi_8 \sim U(-0.5, 0.5)$;

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_6\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_4, X_7\}$ and generate

$$X_6 = (X_4, X_7) \cdot \mathbf{r}_{S_9, C_9 \setminus S_9} + \xi_9, \quad (\text{D.2.9})$$

$$\text{where } \mathbf{r}_{S_9, C_9 \setminus S_9} = \begin{pmatrix} -0.1 \\ 0.2 \end{pmatrix}, \text{ and } \xi_9 \sim U(-0.5, 0.5);$$

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_{19}\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_4, X_7\}$ and generate

$$X_{19} = (X_4, X_7) \cdot \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} + \xi_{10}, \quad (\text{D.2.10})$$

$$\text{where } \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} = \begin{pmatrix} -0.2 \\ 0.09 \end{pmatrix}, \text{ and } \xi_{10} \sim U(-0.5, 0.5);$$

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{14}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_2\}$ and generate

$$X_{14} = X_2 \cdot \mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} + \xi_{11}, \quad (\text{D.2.11})$$

$$\text{where } \mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} = \begin{pmatrix} 0.32 \end{pmatrix}, \text{ and } \xi_{11} \sim U(-0.5, 0.5);$$

11. We have $C_{12} \setminus \{\cup_{k=1}^{11} C_k\} = \{X_{15}\}$, $C_{12} \cap \{\cup_{k=1}^{11} C_k\} = \{X_1\}$ and generate

$$X_{15} = X_1 \cdot \mathbf{r}_{S_{12}, C_{12} \setminus S_{12}} + \xi_{12}, \quad (\text{D.2.12})$$

$$\text{where } \mathbf{r}_{S_{12}, C_{12} \setminus S_{12}} = \begin{pmatrix} -0.1 \end{pmatrix}, \text{ and } \xi_{12} \sim U(-0.5, 0.5);$$

12. We have $C_{13} \setminus \{\cup_{k=1}^{12} C_k\} = \{X_{16}\}$, $C_{13} \cap \{\cup_{k=1}^{12} C_k\} = \{X_{10}\}$ and generate

$$X_{16} = X_{10} \cdot \mathbf{r}_{S_{13}, C_{13} \setminus S_{13}} + \xi_{13}, \quad (\text{D.2.13})$$

$$\text{where } \mathbf{r}_{S_{13}, C_{13} \setminus S_{13}} = \begin{pmatrix} 0.35 \end{pmatrix}, \text{ and } \xi_{13} \sim U(-0.5, 0.5);$$

13. We have $C_{14} \setminus \{\cup_{k=1}^{13} C_k\} = \{X_{17}\}$, $C_{14} \cap \{\cup_{k=1}^{13} C_k\} = \{X_{11}\}$ and generate

$$X_{17} = X_{11} \cdot \mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} + \xi_{14}, \quad (\text{D.2.14})$$

where $\mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} = (0.45)$, and $\xi_{14} \sim U(-0.5, 0.5)$;

14. We have $C_{15} \setminus \{\cup_{k=1}^{14} C_k\} = \{X_{12}\}$, $C_{15} \cap \{\cup_{k=1}^{14} C_k\} = \{X_6\}$ and generate

$$X_{12} = X_6 \cdot \mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} + \xi_{15}, \quad (\text{D.2.15})$$

where $\mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} = (0.25)$, and $\xi_{15} \sim U(-0.5, 0.5)$;

15. We have $C_{16} \setminus \{\cup_{k=1}^{15} C_k\} = \{X_{13}\}$, $C_{16} \cap \{\cup_{k=1}^{15} C_k\} = \{X_7\}$ and generate

$$X_{13} = X_7 \cdot \mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} + \xi_{16}, \quad (\text{D.2.16})$$

where $\mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} = (0.25)$, and $\xi_{16} \sim U(-0.5, 0.5)$;

16. We have $C_{17} \setminus \{\cup_{k=1}^{16} C_k\} = \{X_{20}\}$, $C_{17} \cap \{\cup_{k=1}^{16} C_k\} = \{X_{15}\}$ and generate

$$X_{20} = X_{15} \cdot \mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} + \xi_{17}, \quad (\text{D.2.17})$$

where $\mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} = (0.4)$, and $\xi_{17} \sim U(-0.5, 0.5)$;

In case when $X_{21} = 0$ and $X_{18} = 1$ the variables are generated as:

1. We have $C_2 \setminus C_1 = \{X_1\}$, $C_2 \cap C_1 = \{X_3, X_4\}$ and generate

$$X_1 = (X_3, X_4) \cdot \mathbf{r}_{s_2, C_2 \setminus s_2} + \xi_2, \quad (\text{D.2.18})$$

where $\mathbf{r}_{s_2, C_2 \setminus s_2} = \begin{pmatrix} -0.15 \\ 0.09 \end{pmatrix}$, and $\xi_2 \sim U(-0.5, 0.5)$;

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_2\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_3, X_4\}$ and generate

$$X_2 = (X_3, X_4) \cdot \mathbf{r}_{s_3, C_3 \setminus s_3} + \xi_3, \quad (\text{D.2.19})$$

where $\mathbf{r}_{s_3, C_3 \setminus s_3} = \begin{pmatrix} -0.1 \\ -0.13 \end{pmatrix}$, and $\xi_3 \sim U(-0.5, 0.5)$;

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_5\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_3, X_7\}$ and generate

$$X_5 = (X_3, X_7) \cdot \mathbf{r}_{s_4, C_4 \setminus s_4} + \xi_4, \quad (\text{D.2.20})$$

where $\mathbf{r}_{s_4, C_4 \setminus s_4} = \begin{pmatrix} 0.5 \\ -0.1 \end{pmatrix}$, and $\xi_4 \sim U(-0.5, 0.5)$;

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_8\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_4, X_7\}$ and generate

$$X_8 = (X_4, X_7) \cdot \mathbf{r}_{s_5, C_5 \setminus s_5} + \xi_5, \quad (\text{D.2.21})$$

where $\mathbf{r}_{s_5, C_5 \setminus s_5} = \begin{pmatrix} 0.12 \\ -0.2 \end{pmatrix}$, and $\xi_5 \sim U(-0.5, 0.5)$;

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_9\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_3, X_7\}$ and generate

$$X_9 = (X_3, X_7) \cdot \mathbf{r}_{s_6, C_6 \setminus s_6} + \xi_6, \quad (\text{D.2.22})$$

where $\mathbf{r}_{s_6, C_6 \setminus s_6} = \begin{pmatrix} -0.3 \\ 0.18 \end{pmatrix}$, and $\xi_6 \sim U(-0.5, 0.5)$;

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_{10}\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_3, X_4\}$ and generate

$$X_{10} = (X_3, X_4) \cdot \mathbf{r}_{S_7, C_7 \setminus S_7} + \xi_7, \quad (\text{D.2.23})$$

$$\text{where } \mathbf{r}_{S_7, C_7 \setminus S_7} = \begin{pmatrix} 0.45 \\ -0.1 \end{pmatrix}, \text{ and } \xi_7 \sim U(-0.5, 0.5);$$

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_{11}\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_4, X_7\}$ and generate

$$X_{11} = (X_4, X_7) \cdot \mathbf{r}_{S_8, C_8 \setminus S_8} + \xi_8, \quad (\text{D.2.24})$$

$$\text{where } \mathbf{r}_{S_8, C_8 \setminus S_8} = \begin{pmatrix} 0.15 \\ -0.13 \end{pmatrix}, \text{ and } \xi_8 \sim U(-0.5, 0.5);$$

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_6\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_4, X_7\}$ and generate

$$X_6 = (X_4, X_7) \cdot \mathbf{r}_{S_9, C_9 \setminus S_9} + \xi_9, \quad (\text{D.2.25})$$

$$\text{where } \mathbf{r}_{S_9, C_9 \setminus S_9} = \begin{pmatrix} -0.15 \\ 0.2 \end{pmatrix}, \text{ and } \xi_9 \sim U(-0.5, 0.5);$$

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_{19}\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_4, X_7\}$ and generate

$$X_{19} = (X_4, X_7) \cdot \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} + \xi_{10}, \quad (\text{D.2.26})$$

$$\text{where } \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} = \begin{pmatrix} -0.15 \\ 0.09 \end{pmatrix}, \text{ and } \xi_{10} \sim U(-0.5, 0.5);$$

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{14}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_2\}$ and generate

$$X_{14} = X_2 \cdot \mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} + \xi_{11}, \quad (\text{D.2.27})$$

$$\text{where } \mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} = \begin{pmatrix} 0.15 \end{pmatrix}, \text{ and } \xi_{11} \sim U(-0.5, 0.5);$$

11. We have $C_{12} \setminus \{\cup_{k=1}^{11} C_k\} = \{X_{15}\}$, $C_{12} \cap \{\cup_{k=1}^{11} C_k\} = \{X_1\}$ and generate

$$X_{15} = X_1 \cdot \mathbf{r}_{S_{12}, C_{12} \setminus S_{12}} + \xi_{12}, \quad (\text{D.2.28})$$

where $\mathbf{r}_{S_{12}, C_{12} \setminus S_{12}} = (0.05)$, and $\xi_{12} \sim U(-0.5, 0.5)$;

12. We have $C_{13} \setminus \{\cup_{k=1}^{12} C_k\} = \{X_{16}\}$, $C_{13} \cap \{\cup_{k=1}^{12} C_k\} = \{X_{10}\}$ and generate

$$X_{16} = X_{10} \cdot \mathbf{r}_{S_{13}, C_{13} \setminus S_{13}} + \xi_{13}, \quad (\text{D.2.29})$$

where $\mathbf{r}_{S_{13}, C_{13} \setminus S_{13}} = (-0.15)$, and $\xi_{13} \sim U(-0.5, 0.5)$;

13. We have $C_{14} \setminus \{\cup_{k=1}^{13} C_k\} = \{X_{17}\}$, $C_{14} \cap \{\cup_{k=1}^{13} C_k\} = \{X_{11}\}$ and generate

$$X_{17} = X_{11} \cdot \mathbf{r}_{S_{14}, C_{14} \setminus S_{14}} + \xi_{14}, \quad (\text{D.2.30})$$

where $\mathbf{r}_{S_{14}, C_{14} \setminus S_{14}} = (0.3)$, and $\xi_{14} \sim U(-0.5, 0.5)$;

14. We have $C_{15} \setminus \{\cup_{k=1}^{14} C_k\} = \{X_{12}\}$, $C_{15} \cap \{\cup_{k=1}^{14} C_k\} = \{X_6\}$ and generate

$$X_{12} = X_6 \cdot \mathbf{r}_{S_{15}, C_{15} \setminus S_{15}} + \xi_{15}, \quad (\text{D.2.31})$$

where $\mathbf{r}_{S_{15}, C_{15} \setminus S_{15}} = (0.45)$, and $\xi_{15} \sim U(-0.5, 0.5)$;

15. We have $C_{16} \setminus \{\cup_{k=1}^{15} C_k\} = \{X_{13}\}$, $C_{16} \cap \{\cup_{k=1}^{15} C_k\} = \{X_7\}$ and generate

$$X_{13} = X_7 \cdot \mathbf{r}_{S_{16}, C_{16} \setminus S_{16}} + \xi_{16}, \quad (\text{D.2.32})$$

where $\mathbf{r}_{S_{16}, C_{16} \setminus S_{16}} = (0.3)$, and $\xi_{16} \sim U(-0.5, 0.5)$;

16. We have $C_{17} \setminus \{\cup_{k=1}^{16} C_k\} = \{X_{20}\}$, $C_{17} \cap \{\cup_{k=1}^{16} C_k\} = \{X_{15}\}$ and generate

$$X_{20} = X_{15} \cdot \mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} + \xi_{17}, \quad (\text{D.2.33})$$

where $\mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} = \begin{pmatrix} 0.2 \end{pmatrix}$, and $\xi_{17} \sim U(-0.5, 0.5)$;

In case when $X_{21} = 1$ and $X_{18} = 0$ the variables are generated as:

1. We have $C_2 \setminus C_1 = \{X_1\}$, $C_2 \cap C_1 = \{X_3, X_4\}$ and generate

$$X_1 = (X_3, X_4) \cdot \mathbf{r}_{s_2, C_2 \setminus s_2} + \xi_2, \quad (\text{D.2.34})$$

where $\mathbf{r}_{s_2, C_2 \setminus s_2} = \begin{pmatrix} -0.1 \\ 0.2 \end{pmatrix}$, and $\xi_2 \sim U(-0.5, 0.5)$;

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_2\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_3, X_4\}$ and generate

$$X_2 = (X_3, X_4) \cdot \mathbf{r}_{s_3, C_3 \setminus s_3} + \xi_3, \quad (\text{D.2.35})$$

where $\mathbf{r}_{s_3, C_3 \setminus s_3} = \begin{pmatrix} 0.36 \\ -0.3 \end{pmatrix}$, and $\xi_3 \sim U(-0.5, 0.5)$;

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_5\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_3, X_7\}$ and generate

$$X_5 = (X_3, X_7) \cdot \mathbf{r}_{s_4, C_4 \setminus s_4} + \xi_4, \quad (\text{D.2.36})$$

where $\mathbf{r}_{s_4, C_4 \setminus s_4} = \begin{pmatrix} 0.3 \\ -0.25 \end{pmatrix}$, and $\xi_4 \sim U(-0.5, 0.5)$;

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_8\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_4, X_7\}$ and generate

$$X_8 = (X_4, X_7) \cdot \mathbf{r}_{s_5, C_5 \setminus s_5} + \xi_5, \quad (\text{D.2.37})$$

where $\mathbf{r}_{S_5, C_5 \setminus S_5} = \begin{pmatrix} -0.23 \\ -0.5 \end{pmatrix}$, and $\xi_5 \sim U(-0.5, 0.5)$;

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_9\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_3, X_7\}$ and generate

$$X_9 = (X_3, X_7) \cdot \mathbf{r}_{S_6, C_6 \setminus S_6} + \xi_6, \quad (\text{D.2.38})$$

where $\mathbf{r}_{S_6, C_6 \setminus S_6} = \begin{pmatrix} 0.2 \\ 0.35 \end{pmatrix}$, and $\xi_6 \sim U(-0.5, 0.5)$;

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_{10}\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_3, X_4\}$ and generate

$$X_{10} = (X_3, X_4) \cdot \mathbf{r}_{S_7, C_7 \setminus S_7} + \xi_7, \quad (\text{D.2.39})$$

where $\mathbf{r}_{S_7, C_7 \setminus S_7} = \begin{pmatrix} 0.15 \\ 0.35 \end{pmatrix}$, and $\xi_7 \sim U(-0.5, 0.5)$;

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_{11}\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_4, X_7\}$ and generate

$$X_{11} = (X_4, X_7) \cdot \mathbf{r}_{S_8, C_8 \setminus S_8} + \xi_8, \quad (\text{D.2.40})$$

where $\mathbf{r}_{S_8, C_8 \setminus S_8} = \begin{pmatrix} 0.28 \\ 0.2 \end{pmatrix}$, and $\xi_8 \sim U(-0.5, 0.5)$;

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_6\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_4, X_7\}$ and generate

$$X_6 = (X_4, X_7) \cdot \mathbf{r}_{S_9, C_9 \setminus S_9} + \xi_9, \quad (\text{D.2.41})$$

where $\mathbf{r}_{S_9, C_9 \setminus S_9} = \begin{pmatrix} -0.1 \\ 0.2 \end{pmatrix}$, and $\xi_9 \sim U(-0.5, 0.5)$;

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_{19}\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_4, X_7\}$ and generate

$$X_{19} = (X_4, X_7) \cdot \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} + \xi_{10}, \quad (\text{D.2.42})$$

$$\text{where } \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} = \begin{pmatrix} -0.2 \\ 0.09 \end{pmatrix}, \text{ and } \xi_{10} \sim U(-0.5, 0.5);$$

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{14}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_2\}$ and generate

$$X_{14} = X_2 \cdot \mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} + \xi_{11}, \quad (\text{D.2.43})$$

$$\text{where } \mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} = \begin{pmatrix} 0.32 \end{pmatrix}, \text{ and } \xi_{11} \sim U(-0.5, 0.5);$$

11. We have $C_{12} \setminus \{\cup_{k=1}^{11} C_k\} = \{X_{15}\}$, $C_{12} \cap \{\cup_{k=1}^{11} C_k\} = \{X_1\}$ and generate

$$X_{15} = X_1 \cdot \mathbf{r}_{S_{12}, C_{12} \setminus S_{12}} + \xi_{12}, \quad (\text{D.2.44})$$

$$\text{where } \mathbf{r}_{S_{12}, C_{12} \setminus S_{12}} = \begin{pmatrix} -0.1 \end{pmatrix}, \text{ and } \xi_{12} \sim U(-0.5, 0.5);$$

12. We have $C_{13} \setminus \{\cup_{k=1}^{12} C_k\} = \{X_{16}\}$, $C_{13} \cap \{\cup_{k=1}^{12} C_k\} = \{X_{10}\}$ and generate

$$X_{16} = X_{10} \cdot \mathbf{r}_{S_{13}, C_{13} \setminus S_{13}} + \xi_{13}, \quad (\text{D.2.45})$$

$$\text{where } \mathbf{r}_{S_{13}, C_{13} \setminus S_{13}} = \begin{pmatrix} 0.35 \end{pmatrix}, \text{ and } \xi_{13} \sim U(-0.5, 0.5);$$

13. We have $C_{14} \setminus \{\cup_{k=1}^{13} C_k\} = \{X_{17}\}$, $C_{14} \cap \{\cup_{k=1}^{13} C_k\} = \{X_{11}\}$ and generate

$$X_{17} = X_{11} \cdot \mathbf{r}_{S_{14}, C_{14} \setminus S_{14}} + \xi_{14}, \quad (\text{D.2.46})$$

$$\text{where } \mathbf{r}_{S_{14}, C_{14} \setminus S_{14}} = \begin{pmatrix} 0.45 \end{pmatrix}, \text{ and } \xi_{14} \sim U(-0.5, 0.5);$$

14. We have $C_{15} \setminus \{\cup_{k=1}^{14} C_k\} = \{X_{12}\}$, $C_{15} \cap \{\cup_{k=1}^{14} C_k\} = \{X_6\}$ and generate

$$X_{12} = X_6 \cdot \mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} + \xi_{15}, \quad (\text{D.2.47})$$

where $\mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} = (0.25)$, and $\xi_{15} \sim U(-0.5, 0.5)$;

15. We have $C_{16} \setminus \{\cup_{k=1}^{15} C_k\} = \{X_{13}\}$, $C_{16} \cap \{\cup_{k=1}^{15} C_k\} = \{X_7\}$ and generate

$$X_{13} = X_7 \cdot \mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} + \xi_{16}, \quad (\text{D.2.48})$$

where $\mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} = (0.25)$, and $\xi_{16} \sim U(-0.5, 0.5)$;

16. We have $C_{17} \setminus \{\cup_{k=1}^{16} C_k\} = \{X_{20}\}$, $C_{17} \cap \{\cup_{k=1}^{16} C_k\} = \{X_{15}\}$ and generate

$$X_{20} = X_{15} \cdot \mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} + \xi_{17}, \quad (\text{D.2.49})$$

where $\mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} = (0.4)$, and $\xi_{17} \sim U(-0.5, 0.5)$;

In case when $X_{21} = 1$ and $X_{18} = 1$ the variables are generated as:

1. We have $C_2 \setminus C_1 = \{X_1\}$, $C_2 \cap C_1 = \{X_3, X_4\}$ and generate

$$X_1 = (X_3, X_4) \cdot \mathbf{r}_{s_2, C_2 \setminus s_2} + \xi_2, \quad (\text{D.2.50})$$

where $\mathbf{r}_{s_2, C_2 \setminus s_2} = \begin{pmatrix} -0.15 \\ 0.09 \end{pmatrix}$, and $\xi_2 \sim U(-0.5, 0.5)$;

2. We have $C_3 \setminus \{\cup_{k=1}^2 C_k\} = \{X_2\}$, $C_3 \cap \{\cup_{k=1}^2 C_k\} = \{X_3, X_4\}$ and generate

$$X_2 = (X_3, X_4) \cdot \mathbf{r}_{s_3, C_3 \setminus s_3} + \xi_3, \quad (\text{D.2.51})$$

where $\mathbf{r}_{s_3, C_3 \setminus s_3} = \begin{pmatrix} -0.1 \\ -0.13 \end{pmatrix}$, and $\xi_3 \sim U(-0.5, 0.5)$;

3. We have $C_4 \setminus \{\cup_{k=1}^3 C_k\} = \{X_5\}$, $C_4 \cap \{\cup_{k=1}^3 C_k\} = \{X_3, X_7\}$ and generate

$$X_5 = (X_3, X_7) \cdot \mathbf{r}_{s_4, C_4 \setminus s_4} + \xi_4, \quad (\text{D.2.52})$$

where $\mathbf{r}_{s_4, C_4 \setminus s_4} = \begin{pmatrix} 0.5 \\ -0.1 \end{pmatrix}$, and $\xi_4 \sim U(-0.5, 0.5)$;

4. We have $C_5 \setminus \{\cup_{k=1}^4 C_k\} = \{X_8\}$, $C_5 \cap \{\cup_{k=1}^4 C_k\} = \{X_4, X_7\}$ and generate

$$X_8 = (X_4, X_7) \cdot \mathbf{r}_{s_5, C_5 \setminus s_5} + \xi_5, \quad (\text{D.2.53})$$

where $\mathbf{r}_{s_5, C_5 \setminus s_5} = \begin{pmatrix} 0.12 \\ -0.2 \end{pmatrix}$, and $\xi_5 \sim U(-0.5, 0.5)$;

5. We have $C_6 \setminus \{\cup_{k=1}^5 C_k\} = \{X_9\}$, $C_6 \cap \{\cup_{k=1}^5 C_k\} = \{X_3, X_7\}$ and generate

$$X_9 = (X_3, X_7) \cdot \mathbf{r}_{s_6, C_6 \setminus s_6} + \xi_6, \quad (\text{D.2.54})$$

where $\mathbf{r}_{s_6, C_6 \setminus s_6} = \begin{pmatrix} -0.3 \\ 0.18 \end{pmatrix}$, and $\xi_6 \sim U(-0.5, 0.5)$;

6. We have $C_7 \setminus \{\cup_{k=1}^6 C_k\} = \{X_{10}\}$, $C_7 \cap \{\cup_{k=1}^6 C_k\} = \{X_3, X_4\}$ and generate

$$X_{10} = (X_3, X_4) \cdot \mathbf{r}_{s_7, C_7 \setminus s_7} + \xi_7, \quad (\text{D.2.55})$$

where $\mathbf{r}_{s_7, C_7 \setminus s_7} = \begin{pmatrix} 0.45 \\ -0.1 \end{pmatrix}$, and $\xi_7 \sim U(-0.5, 0.5)$;

7. We have $C_8 \setminus \{\cup_{k=1}^7 C_k\} = \{X_{11}\}$, $C_8 \cap \{\cup_{k=1}^7 C_k\} = \{X_4, X_7\}$ and generate

$$X_{11} = (X_4, X_7) \cdot \mathbf{r}_{S_8, C_8 \setminus S_8} + \xi_8, \quad (\text{D.2.56})$$

$$\text{where } \mathbf{r}_{S_8, C_8 \setminus S_8} = \begin{pmatrix} 0.15 \\ -0.13 \end{pmatrix}, \text{ and } \xi_8 \sim U(-0.5, 0.5);$$

8. We have $C_9 \setminus \{\cup_{k=1}^8 C_k\} = \{X_6\}$, $C_9 \cap \{\cup_{k=1}^8 C_k\} = \{X_4, X_7\}$ and generate

$$X_6 = (X_4, X_7) \cdot \mathbf{r}_{S_9, C_9 \setminus S_9} + \xi_9, \quad (\text{D.2.57})$$

$$\text{where } \mathbf{r}_{S_9, C_9 \setminus S_9} = \begin{pmatrix} -0.15 \\ 0.2 \end{pmatrix}, \text{ and } \xi_9 \sim U(-0.5, 0.5);$$

9. We have $C_{10} \setminus \{\cup_{k=1}^9 C_k\} = \{X_{19}\}$, $C_{10} \cap \{\cup_{k=1}^9 C_k\} = \{X_4, X_7\}$ and generate

$$X_{19} = (X_4, X_7) \cdot \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} + \xi_{10}, \quad (\text{D.2.58})$$

$$\text{where } \mathbf{r}_{S_{10}, C_{10} \setminus S_{10}} = \begin{pmatrix} -0.15 \\ 0.09 \end{pmatrix}, \text{ and } \xi_{10} \sim U(-0.5, 0.5);$$

10. We have $C_{11} \setminus \{\cup_{k=1}^{10} C_k\} = \{X_{14}\}$, $C_{11} \cap \{\cup_{k=1}^{10} C_k\} = \{X_2\}$ and generate

$$X_{14} = X_2 \cdot \mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} + \xi_{11}, \quad (\text{D.2.59})$$

$$\text{where } \mathbf{r}_{S_{11}, C_{11} \setminus S_{11}} = \begin{pmatrix} 0.15 \end{pmatrix}, \text{ and } \xi_{11} \sim U(-0.5, 0.5);$$

11. We have $C_{12} \setminus \{\cup_{k=1}^{11} C_k\} = \{X_{15}\}$, $C_{12} \cap \{\cup_{k=1}^{11} C_k\} = \{X_1\}$ and generate

$$X_{15} = X_1 \cdot \mathbf{r}_{S_{12}, C_{12} \setminus S_{12}} + \xi_{12}, \quad (\text{D.2.60})$$

$$\text{where } \mathbf{r}_{S_{12}, C_{12} \setminus S_{12}} = \begin{pmatrix} 0.05 \end{pmatrix}, \text{ and } \xi_{12} \sim U(-0.5, 0.5);$$

12. We have $C_{13} \setminus \{\cup_{k=1}^{12} C_k\} = \{X_{16}\}$, $C_{13} \cap \{\cup_{k=1}^{12} C_k\} = \{X_{10}\}$ and generate

$$X_{16} = X_{10} \cdot \mathbf{r}_{s_{13}, C_{13} \setminus s_{13}} + \xi_{13}, \quad (\text{D.2.61})$$

where $\mathbf{r}_{s_{13}, C_{13} \setminus s_{13}} = (-0.15)$, and $\xi_{13} \sim U(-0.5, 0.5)$;

13. We have $C_{14} \setminus \{\cup_{k=1}^{13} C_k\} = \{X_{17}\}$, $C_{14} \cap \{\cup_{k=1}^{13} C_k\} = \{X_{11}\}$ and generate

$$X_{17} = X_{11} \cdot \mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} + \xi_{14}, \quad (\text{D.2.62})$$

where $\mathbf{r}_{s_{14}, C_{14} \setminus s_{14}} = (0.3)$, and $\xi_{14} \sim U(-0.5, 0.5)$;

14. We have $C_{15} \setminus \{\cup_{k=1}^{14} C_k\} = \{X_{12}\}$, $C_{15} \cap \{\cup_{k=1}^{14} C_k\} = \{X_6\}$ and generate

$$X_{12} = X_6 \cdot \mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} + \xi_{15}, \quad (\text{D.2.63})$$

where $\mathbf{r}_{s_{15}, C_{15} \setminus s_{15}} = (0.45)$, and $\xi_{15} \sim U(-0.5, 0.5)$;

15. We have $C_{16} \setminus \{\cup_{k=1}^{15} C_k\} = \{X_{13}\}$, $C_{16} \cap \{\cup_{k=1}^{15} C_k\} = \{X_7\}$ and generate

$$X_{13} = X_7 \cdot \mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} + \xi_{16}, \quad (\text{D.2.64})$$

where $\mathbf{r}_{s_{16}, C_{16} \setminus s_{16}} = (0.3)$, and $\xi_{16} \sim U(-0.5, 0.5)$;

16. We have $C_{17} \setminus \{\cup_{k=1}^{16} C_k\} = \{X_{20}\}$, $C_{17} \cap \{\cup_{k=1}^{16} C_k\} = \{X_{15}\}$ and generate

$$X_{20} = X_{15} \cdot \mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} + \xi_{17}, \quad (\text{D.2.65})$$

where $\mathbf{r}_{s_{17}, C_{17} \setminus s_{17}} = (0.2)$, and $\xi_{17} \sim U(-0.5, 0.5)$;

APPENDIX E

R package

The package consists of functions for estimating regression parameters and bootstrap errors using graphical models for high-dimensional data ($n < v$). Both serial and parallel computations are allowed. The functions given in the package are illustrated with real data examples.

Package ‘HDLRGM’

September 28, 2016

Type Package

Title High-dimensional linear regression via graphical models

Version 1.0

Date 2016-08-02

Author Saeed Aldahmani and Hongsheng Dai

Maintainer Maintainer: Saeed Aldahmani <skaald@essex.ac.uk>

Description The package consists of functions for estimating linear regression via graphical models especially when the number of observations is smaller than the number of variables. The first function (called GLSE) is giving unbiased estimated regression models under some conditions on the predictors. The second function (called HGLSE) is producing unbiased estimation for some variables. Finally, the package also provides a function for estimating bootstrap errors of the GLSE parameters. The package supports both serial and parallel computations for estimating the graph.

License GPL-2

Imports gRbase,glmnet

Lazyload yes

R topics documented:

High-dimensional linear regression via graphical models-package	2
Boot.Error	3
GLSE	4
HGLSE	6
qsbralks	9
Index	11

Description

The package consists of functions for estimating linear regression via graphical models especially when the number of observations is smaller than the number of variables. The first function (called GLSE) is giving unbiased estimated regression models under some conditions on the predictors. The second function (called HGLSE) is producing unbiased estimation for some variables. Finally, the package also provides a function for estimating bootstrap errors of the GLSE parameters.

Details

Package: High-dimensional linear regression via graphical models
Type: Package
Version: 1.0
Date: 2016-08-20
License: GPL-2

The package supports both serial and parallel computations for estimating the graph.

Author(s)

Saeed Aldahmani and Hongsheng Dai

Maintainer: Saeed Aldahmani <skaald@essex.ac.uk>

References

Aldahmani, S. and Dai, H. (2015). Unbiased Estimation for Linear Regression When $n < v$. International Journal of Statistics and Probability, 4(3), p61.

Aldahmani, S. Dai, H. and Zhang, Q., Hybrid Graphical Least Square Estimation and its application in Portfolio Selection, submitte.

Dethlefsen, C., and Højsgaard, S. (2005). A common platform for graphical models in R: The gR-base package. Journal of Statistical Software, 14(17), 1-12. URL <http://www.jstatsoft.org/v14/i17/>.

Jerome F, Trevor H, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22. URL <http://www.jstatsoft.org/v33/i01/>.

See Also

GLSE, Boot.Error, qsbralks

Description

The function provides bootstrap error estimation and significance for the estimates obtained via the GLSE.

Usage

```
Boot.Error(Graph, x, y, n.boot, sig.level = 0.05, centering = TRUE)
```

Arguments

Graph	An object of class GLSE.
x	A matrix of $n \times v$ dimension consisting of the predictors values.
y	A vector of length n of the response variable.
n.boot	Number of bootstrap samples.
sig.level	Significance level. Must be pre-specified.
centering	Logical. If equal to TRUE then all the variables in x and the response y are centered. The default is TRUE.

Value

A data frame consisting of:

Beta	The estimated regression coefficients based on the provided graph.
Sttandard.Error	Bootstrap estimates of the standard errors.
Significance	Significance of the estimates based on the specified quantile points. "***" means significance at the specified level while "+" means no significance.

Author(s)

Saeed Aldahmani and Hongsheng Dai

References

- Aldahmani, S. and Dai, H. (2015). Unbiased Estimation for Linear Regression When $n < v$. International Journal of Statistics and Probability, 4(3), p61.
- Aldahmani, S. Dai, H. and Zhang, Q., Hybrid Graphical Least Square Estimation and its application in Portfolio Selection, submitte.
- Dethlefsen, C., and Højsgaard, S. (2005). A common platform for graphical models in R: The gR-base package. Journal of Statistical Software, 14(17), 1-12. URL <http://www.jstatsoft.org/v14/i17/>.
- Jerome F, Trevor H, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22. URL <http://www.jstatsoft.org/v33/i01/>.

See Also

[GLSE](#)

Examples

```
library(gRbase)
#load the data
data(qsbralks)
x <- as.matrix(qsbralks[,1:21])
y <- as.matrix(qsbralks[,22])

# Get the graph

Graph <- GLSE(x,y,print.info=TRUE,parallel=FALSE)$Graph

# Call Boot.Error(.) .

results <- Boot.Error(Graph,x,y,n.boot=100)

# Results

results
```

GLSE

Graphical least sqauare estimation

Description

The function estimates linear regression via graphical models in an unbiased way. This function supports both serial and parallel computations. However, the available computational resources should be considered before choosing one of the two options. For any given data, an estimated graph, regression parameters and errors are returned.

Usage

```
GLSE(x, y, centering = TRUE, lambda = 0, parallel = FALSE,
      clusters = NULL, print.info = TRUE,c.size=NULL)
```

Arguments

x	A matrix of n x v dimension consisting of the predictors values.
y	A vector of length n of the response variable.
centering	Logical. If equal to TRUE then all the variables in x and the response y are centered. The default is TRUE.
lambda	The amount of regularisation for shrinking the graph. The default is 0.
parallel	Logical. Whether or not to do parallel computation. If set to TRUE, then clusters must also be provided. The default is FALSE.
clusters	The number of child nodes that must be created in the cluster if the parallel option is set to TRUE.
print.info	Logical. If set equal to TRUE, the number of searches done for finding the best graph are printed. The default is FALSE.
c.size	Size of the clique to be allowed. The default is n.

Details

The parallel options also work for Windows. Care should be taken in creating the cluster as all types of clusters do not work for Windows, e.g. FORK.

Value

A list of three values:

Graph	The estimated graph for the given data.
Beta	The vector of estimated regression parameters based on the estimated graph.
SSE	Sum of squares of errors for the parameters.
Error	Estimated errors for the parameters.

Author(s)

Saeed Aldahmani and Hongsheng Dai

References

- Aldahmani, S. and Dai, H. (2015). Unbiased Estimation for Linear Regression When $n < v$. International Journal of Statistics and Probability, 4(3), p61.
- Aldahmani, S. Dai, H. and Zhang, Q., Hybrid Graphical Least Square Estimation and its application in Portfolio Selection, submitte.
- Dethlefsen, C., and Højsgaard, S. (2005). A common platform for graphical models in R: The gR-base package. Journal of Statistical Software, 14(17), 1-12. URL <http://www.jstatsoft.org/v14/i17/>.
- Jerome F, Trevor H, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22. URL <http://www.jstatsoft.org/v33/i01/>.

See Also

[Boot.Error](#)

Examples

```
#####

# Example with serail precessing.

#####
library(gRbase)
library(parallel)
#load the data
data(qsbralks)
x <- as.matrix(qsbralks[,1:21])
y <- as.matrix(qsbralks[,22])

# Get the graph

# Call GLSE(.) .

results <- GLSE(x,y,print.info=TRUE,parallel=FALSE)
```

```

# The returned objects are

results$Graph
results$Beta
results$SSE
results$Error

#####

# Example with parallel precessing.

#####

#load the data
data(qsbralks)
x <- qsbralks[,1:21]
y <- qsbralks[,21]

# Create cluster

cl <- makeCluster(detectCores())

# Pass the gRbase package to all the nodes

clusterEvalQ(cl, library(gRbase))

# Call GLSE(.) .

result <- GLSE(x,y,print.info=TRUE,parallel=TRUE,clusters=cl)

# Stop the cluster

stopCluster(cl)

# Results

results$Graph;
results$Beta;
results$SSE;
results$Error;

```

HGLSE

Hybridisation of graphical least square and ridge regression

Description

The function estimates some of the prediction in linear regression models in an unbiased way. It supports both serial and parallel computations. However, the available computational resources should be considered before choosing one of the two options. For any given data, an estimated graph, regression parameters and significant variables are returned.

Usage

```
HGLSE(x, y, centering = TRUE, lambda = 0, parallel = FALSE,  
      clusters = NULL, print.info = TRUE, c.size=NULL  
      x, y, centering=FALSE, lambda=1, parallel=FALSE,  
      clusters=NULL, c.size=NULL, k=5, sig.level=.5, n.boot=500)
```

Arguments

x	A matrix of $n \times v$ dimension consisting of the predictors values.
y	A vector of length n of the response variable.
centering	Logical. If equal to TRUE then all the variables in x and the response y are centered. The default is TRUE.
lambda	The amount of penalty. The default is 1.
parallel	Logical. Whether or not to do parallel computation. If set to TRUE, then <code>clusters</code> must also be provided. The default is FALSE.
clusters	The number of child nodes that must be created in the cluster if the <code>parallel</code> option is set to TRUE.
c.size	Size of the clique to be allowed. The default is n .
k	Number of folds. The default is 5.
sig.level	Significance level. Must be pre-specified.
n.boot	Number of bootstrap samples. The default is 500.

Details

The parallel options also work for Windows. Care should be taken in creating the cluster as all types of clusters do not work for Windows, e.g. FORK.

Value

A list of three features:

Graph	The estimated graph for the given data.
Beta	The vector of estimated regression parameters based on the estimated graph.
Significant variables	Variables Selected by HGLSE metho.

Author(s)

Saeed Aldahmani and Hongsheng Dai

References

- Aldahmani, S. and Dai, H. (2015). Unbiased Estimation for Linear Regression When $n < v$. International Journal of Statistics and Probability, 4(3), p61.
- Aldahmani, S. Dai, H. and Zhang, Q., Hybrid Graphical Least Square Estimation and its application in Portfolio Selection, submitte.
- Dethlefsen, C., and Højsgaard, S. (2005). A common platform for graphical models in R: The gR-base package. Journal of Statistical Software, 14(17), 1-12. URL <http://www.jstatsoft.org/v14/i17/>.
- Jerome F, Trevor H, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22. URL <http://www.jstatsoft.org/v33/i01/>.

See Also

[GLSE](#)

Examples

```
#####

# Example with serial precessing.

#####
library(gRbase)
library(parallel)
library(glmnet)
#load the data
data(qsbralks)
x <- as.matrix(qsbralks[,1:21])
y <- as.matrix(qsbralks[,22])

# Get the graph

# Call HGLSE(.) .

results <- HGLSE (x,y,centering=FALSE,lambda=1,parallel=FALSE,
clusters=NULL,c.size=NULL,k=5,sig.level=.5,n.boot=500)

# Results

results$Graph
results$Beta
results$Significant

#####

# Example with parallel precessing.

#####

#load the data
data(qsbralks)
x <- qsbralks[,1:21]
y <- qsbralks[,22]

# Create cluster

cl <- makeCluster(detectCores())

# Pass the gRbase and glmnet packages to all the nodes

clusterEvalQ(cl, library(gRbase,glmnet))

# Call HGLSE(.) .
```

```

results <- HGLSE (x,y,centering=FALSE,lambda=1,parallel=TRUE,
clusters=cl,c.size=NULL,k=5,sig.level=.5,n.boot=500)

# Stop the cluster

stopCluster(cl)

# Results

results$Graph
results$Beta
results$Significant

```

qsbralks

The qsbralks regression dataset.

Description

This is one of 41 drug design datasets taken from <http://www.openml.org/d/431>. The dataset consists of 21 variables and 13 observations with a continuous response.

Usage

```
data(qsbralks)
```

Format

A data frame with 13 observations on 22 variables including the response.

Examples

```

library(gRbase)
library(parallel)
data(qsbralks)
x <- as.matrix(qsbralks[,1:21])
y <- as.matrix(qsbralks[,22])

# using serial computing.
results <- GLSE(x,y,print.info=TRUE,parallel=FALSE)

#using parallel computing.
# Create cluster

cl <- makeCluster(detectCores())

# Pass the gRbase package to all the nodes.

clusterEvalQ(cl, library(gRbase))

# Call GLSE(.) .

result <- GLSE(x,y,print.info=TRUE,parallel=TRUE,clusters=cl)

```

```
# Stop the cluster
```

```
stopCluster(cl)
```

```
# Results
```

```
results$Graph;
```

```
results$Beta;
```

```
results$SSE;
```

```
results$Error;
```

Index

*Topic **datasets**

qsbralks, [9](#)

*Topic **package**

GLSE, [4](#)

HGLSE, [6](#)

High-dimensional linear
regression via graphical
models-package, [2](#)

Boot.Error, [3](#), [5](#)

GLSE, [3](#), [4](#), [8](#)

HGLSE, [6](#)

High-dimensional linear regression
via graphical models
(High-dimensional linear
regression via graphical
models-package), [2](#)

High-dimensional linear regression via
graphical models-package, [2](#)

qsbralks, [9](#)