이학박사 학위논문

Modified Cholesky Decomposition based Precision Matrix Estimation via Scaled Lasso

2023년 8월

서울대학교 대학원

통계학과

이다혜

Modified Cholesky Decomposition based
Precision Matrix Estimation via Scaled Lasso
수정 촐레스키 분해 기반의
스케일드 라쏘 기법을 활용한 역공분산 행렬 추정

지도교수 임 요 한

이 논문을 이학박사 학위논문으로 제출함
2023년 4월

서울대학교 대학원
통계학과
이 다 혜

이다혜의 이학박사 학위논문을 인준함
2023년 6월

위 원 장	이 재 용	(인)
부위원장	임 요 한	(인)
위 원	원 중 호	(인)
위 원	박 건 웅	(인)
위 원	유 동 현	(인)

Modified Cholesky Decomposition based Precision Matrix Estimation via Scaled Lasso

by

Dahye Lee

A Thesis
submitted in fulfillment of the requirement
for the degree of
Doctor of Philosophy
in
Statistics

Department of Statistics
College of Natural Sciences
Seoul National University
August, 2023

Abstract

Modified Cholesky Decomposition based Precision Matrix Estimation via Scaled Lasso

Dahye Lee

Department of Statistics

The Graduate School

Seoul National University

In the thesis, we propose a method to estimate a high dimensional precision matrix Ω of a random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)^\top$ when the variables $Y_1, Y_2, \ldots, Y_p$ are observed with time-orderings. The precision matrix can be written in a form of modified cholesky decomposition (MCD), $\Omega = L^\top D^{-1} L$, where L is a lower-diagonal matrix with ones on diagonals and D is a diagonal matrix. We propose an estimator SLMCD by applying the scaled lasso regression method (Sun and Zhang, 2012) in MCD settings. Our proposed SLMCD has several advantages. First, by applying the scaled lasso method, it allows us to use a prefixed tuning parameter and it brings the computational efficiency over the cross-validation method. Second, under certain regularity conditions, we show that SLMCD reaches the optimal convergence rate of $O_p\big(\sqrt{s \log p/n}\big)$, where s is the number of non-zero elements in L. Third, we numerically show that SLMCD outperforms existing lasso-based estimators in various situations. We apply SLMCD to perform real data analysis of electric safety data provided by Kyorim Soft coorporation.

Keywords: Convergence rate, modified cholesky decomposition, high dimensional precision matrix, scaled lasso, electric safety data
Student Number: 2016-20271

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In today's modern society, complexity has significantly increased, which is evident through the abundance of data resulting from the advancements in statistical tools and computing machines. This complexity is best represented by the relationships between variables, often captured in the form of covariance matrix or precision matrix. Therefore, accurately estimating these matrices becomes crucial as they serve as representatives of the underlying phenomena in the present-day world.

In certain scenarios, many variables are interrelated and the needs of covariance or precision matrix increases as they can show those interrelated complexity. Also, as the number of variables (p) increases, the instances where the number of variables (p) exceeds the number of available datasets (n) need attention. In such cases, the estimation of covariance matrices or precision matrices becomes more challenging, since it becomes computationally expensive. As a result, sparsity becomes a crucial feature in such cases.

The precision matrix Ω, also known as the inverse of the covariance matrix

Σ^{-1}, becomes a focal point of interest when considering likelihood functions, such as in the Gaussian distribution, as the likelihood function solely relies on the precision matrix rather than the covariance matrix itself.

In this thesis, we focus on estimating the precision matrix Ω under some sparsity conditions. Also, we assume that the data are in time-ordered in order to apply the modified cholesky decomposition (MCD) idea introduced by Gill and Murray (1974) to proposals, which can contribute to sparsity.

The rest of the thesis is as follows. In Chapter 2, we first review some theoretical reviews of methodologies directly applied to our proposal. One is the MCD-based precision matrix estimation method, followed by introducing the work of Huang et al. (2006) which proposed a precision matrix estimation method using MCD. The other is the scaled lasso method with regression method (Sun and Zhang, 2012) and proposal of precision matrix estimator (Sun and Zhang, 2013).

In Chapter 3, we propose a precision matrix estimator named MCD-based precision matrix estimator using the scaled lasso (SLMCD) and this estimator comes from the idea combination of MCD and scaled lasso introduced in previous chapter. We introduce the detailed algorithm of proposal and prove the convergence rate of proposed estimator under certain assumptions. In Chapter 4, we numerically show our proposal has advantages compared to some lasso-based precision matrix estimators under certain situations by using the numerical settings of Huang et al. (2006) and Sun and Zhang (2013) and execute the real data analysis of electric-safety data in Chapter 5.

Chapter 2

Review

2.1 MCD-based covariance matrix estimation

Assume $Y = (Y_1, ..., Y_p)^\top$ is a time-ordered random vector with zero mean and covariance matrix Σ. Then for $t = 1, ..., p$,

$$Y_t = \sum_{j=1}^{t-1} \phi_{tj} Y_j + \epsilon_t = \phi_t^\top Y_{1:(t-1)} + \epsilon_t, \quad Var(\epsilon_t) = \sigma_t^2 \tag{2.1}$$

where $\phi_t = (\phi_{t1}, ..., \phi_{t(t-1)})^\top$ and $Y_{1:(t-1)} = (Y_1, ..., Y_{t-1})^\top$.

By rearranging (2.1), the model (2.1) can be written as

$$(I - \Phi)\boldsymbol{y} = L\boldsymbol{y} = \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \dots \\ \epsilon_p \end{bmatrix} \tag{2.2}$$

where

$$\Phi = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \phi_{2,1} & 0 & 0 & 0 & 0 \\ \phi_{3,1} & \phi_{3,2} & 0 & 0 & 0 \\ & & & \cdots & 0 \\ \phi_{p,1} & \phi_{p,2} & & \phi_{p(p-1)} & 0 \end{bmatrix}, \; L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -\phi_{2,1} & 1 & 0 & 0 & 0 \\ -\phi_{3,1} & -\phi_{3,2} & 1 & 0 & 0 \\ & & & \cdots & 0 \\ -\phi_{p,1} & -\phi_{p,2} & & -\phi_{p(p-1)} & 1 \end{bmatrix} \quad (2.3)$$

and $Cov(\boldsymbol{\epsilon}) = D = diag(\sigma_1^2, ..., , \sigma_p^2)$.

For a positive-definite covariance matrix Σ, MCD of Σ is expressed as (2.4).

$$L\Sigma L^\top = D \quad (2.4)$$

Therefore, Ω, the inverse of Σ, can be written as (2.5).

$$\Omega = \Sigma^{-1} = L^\top D^{-1} L \quad (2.5)$$

MCD has some useful advantages in estimating Ω. First, by applying the MCD to Ω, Ω can be estimated by estimating L and D, where L, D can be estimated via regressing each Y_t on its predecessors $Y_{1:(t-1)}$ for $t = 2, ..., p$ since we assumed time-orderings. MCD has insights that L and D both have its own meanings where L explains the relationships between Y_t and its predecessors $Y_{1:(t-1)}$ and D explains the error variance of Y_t for each $t = 1, ..., p$. Second, estimating Ω using MCD ensures that the number of parameters to be estimated in regression equations matches the number of parameters in the precision matrix. Last, the MCD method inherently guarantees the positive-definiteness of the estimator of Σ and Ω.

There are several papers that explore MCD-based precision matrix estimators with assumptions on both with and without time orderings in data structures. For the cases of time-ordered variables, Huang et al. (2006), Levina et al. (2008) and Jiang (2012) employ information from ordination to estimate the precision matrix. For the unordered-variables, Rothman et al. (2010) and Kang et al. (2020) predict the precision matrix without utilizing ordination information.

In this section, we review the paper of Huang et al. (2006) in detail which proposed precision estimation method combining MCD of time-ordered datasets with penalized likelihood regression methods. Then, we introduce various methods estimating Σ and Ω briefly.

2.1.1 Huang et al. (2006)

Huang et al. (2006) proposed a precision matrix estimator by combining MCD with penalized likelihood regression methods under ℓ_1 and ℓ_2- regularizatons in time-ordered datasets.

Under the multivariate normal assumptions on n observations $y_i = (y_{i1}, ..., y_{ip})^\top$ for $i = 1, ..., n$ of Y of (2.1) with only relevant variables, the loglikehood function $l(\Sigma; y_i)$ is as (2.6).

$$\begin{aligned} -2l(\Sigma; y_i) &= \log|\Sigma| + y_i^\top \Omega y_i \\ &= \log|D| + y_i^\top L^\top D^{-1} L y_i = \log|D| + \epsilon_i^\top D^{-1} \epsilon_i \\ &= \sum_{j=1}^{p} \log \sigma_j^2 + \sum_{j=1}^{p} \frac{\epsilon_{ij}^2}{\sigma_j^2} \end{aligned} \tag{2.6}$$

where $\epsilon_1 = y_1$ and $\epsilon_k = y_k - \sum_{j=1}^{p-1} y_j \phi_{kj}$ for $k = 2, ..., p$.

For n observations $y_i = (y_{i1}, ..., y_{ip})_{i=1,...,n}^{\top}$, the joint loglikehood function with only relevant variables can be rewritten as (2.7).

$$-2l(\Sigma; y_1, ..., y_n) = \sum_{j=1}^{p} \left(n \log \sigma_j^2 + \sum_{i=1}^{n} \frac{\epsilon_{ij}^2}{\sigma_j^2} \right) \tag{2.7}$$

with $\epsilon_{i1} = y_{i1}$ annd $\epsilon_{ik} = y_{ik} - \sum_{j=1}^{p-1} y_{ij} \phi_{kj}$ for $k = 2, ..., p$.

Let $\boldsymbol{y} = (y_1, ..., y_n)^{\top}$, then using the penalized regression method with a given penalty $\lambda > 0$, the penalized loglikelihood is as (2.8).

$$\begin{aligned} &-2l(\Sigma; \boldsymbol{y}) + \lambda J(\boldsymbol{\Phi}) \\ &\quad = \sum_{j=1}^{p} \left(n \log \sigma_j^2 + \sum_{i=1}^{n} \frac{\epsilon_{ij}^2}{\sigma_j^2} \right) + \lambda \sum_{j=2}^{p} \sum_{i=1}^{j-1} |\phi_{ji}|^m \\ &\quad = \left(n \log \sigma_1^2 + \sum_{i=1}^{n} \frac{\epsilon_{i1}^2}{\sigma_1^2} \right) + \sum_{j=2}^{p} \left(n \log \sigma_j^2 + \sum_{i=1}^{n} \frac{\epsilon_{ij}^2}{\sigma_j^2} + \lambda \sum_{i=1}^{j-1} |\phi_{ji}|^m \right) \end{aligned} \tag{2.8}$$

When λ is fixed, (2.8) can be solved via alternating minimization problem over $\{\phi_{kj}\}_{j=1,...,p-1}$ and $\{\sigma_k^2\}$ for each $k = 1, ..., p$.

For penalized likelihood function, (i) the penalty term $J(\Phi)$ and (ii) tuning parameter λ needs to be determined and there are several suggestions. For (i) $J(\Phi)$, Huang et al. (2006) described penalty term as (2.9) for $m = 1, 2$ which is either ℓ_1- regularization (lasso) when $m = 1$ or ℓ_2- regularization (ridge) when $m = 2$.

$$J(\Phi) = \lambda \sum_{j=2}^{p} \sum_{i=1}^{j-1} |\phi_{ji}|^m \tag{2.9}$$

Levina et al. (2008) suggested new nested lasso-based penalty as in (2.10) which reduces the rest term $\phi_{j,j-k}$ to all 0 for $k \geq l$ if $\phi_{j,j-l}$ is 0.

$$\begin{aligned} J(\Phi) &= \lambda \sum_{j=1}^{p} \Big(|\phi_{j,j-1}| + \frac{|\phi_{j,j-2}|}{|\phi_{j,j-1}|} + ... + \frac{|\phi_{j,1}|}{|\phi_{j,2}|} \Big) \\ \text{or } J(\Phi) &= \lambda \sum_{j=1}^{p} \Big(\frac{|\phi_{j,j-1}|}{\hat{\phi}^{\star}_{j,j-1}} + \frac{|\phi_{j,j-2}|}{|\phi_{j,j-1}|} + ... + \frac{|\phi_{j,1}|}{|\phi_{j,2}|} \Big) \end{aligned} \tag{2.10}$$

For the choice of (ii) λ, K-fold cross-validation (KCV) and generalized cross-validation (GCV) method is the representatives.

For the K-fold cross-validation (KCV), we split the full dataset into K subsets randomly and use one subset among K subsets as validation set and the rest as the training set. We compute following criterion as (2.11) using the information of each K choice of validation sets. Then choose λ that results in the smallest value of $CV(\lambda)$.

$$CV(\lambda) = \frac{1}{K} \sum_{\nu=1}^{K} \Big(s_{\nu} \log |\hat{\Sigma}_{-\nu}| + \sum_{i \in I_{\nu}} y_i^{\top} \hat{\Sigma}_{-\nu}^{-1} y_i \Big) \tag{2.11}$$

where I_{ν} the index set of the data in S^{ν}, s_{ν} the size of I_{ν} and $\hat{\Sigma}_{-\nu}$ the variance-covariance matrix under the training dataset $S - S^{\nu}$.

For the generalized cross-validation (GCV), the idea comes from the leave-one-out crossvalidation method where $K = n$ in KCV. The calculation of GCV criterion is as (2.12).

$$GCV(\lambda) = \frac{1}{np} \sum_{t=1}^{p} \sum_{i=1}^{n} \Big(\frac{y_{it} - \hat{y}_{it}}{1 - tr(S_t)/n} \Big) \tag{2.12}$$

where $S_t = X_t(H_t + \lambda I_t)^{-1}X_t$ with $H_t = (\sum_{i=1}^{n} y_{i(t)}y_{i(t)}^{\top})/\sigma_t^2$ for $t = 1, ..., p$.

Approximately, we can compute the components of (2.12) as follows.

$$\begin{bmatrix} \hat{y}_{1t} \\ \cdots \\ \hat{y}_{nt} \end{bmatrix} = X_t(H_t + \lambda L_t^{(k)})^{-1}X_t^{\top} \begin{bmatrix} y_{1t} \\ \cdots \\ y_{nt} \end{bmatrix}$$

with

$$L_t^{(k)} = diag\Big(\frac{1}{2|\phi_{t,1}^{(k)}|}, \frac{1}{2|\phi_{t,2}^{(k)}|}, \cdots, \frac{1}{2|\phi_{t(t-1)}^{(k)}|}\Big)$$

and

$$X_t = \frac{1}{\sigma_t}\begin{bmatrix} y_{1(t)}^{\top} \\ \cdots \\ y_{n(t)}^{\top} \end{bmatrix} = \frac{1}{\sigma_t}\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1(t-1)} \\ \cdots & & & \\ y_{n1} & y_{n2} & \cdots & y_{n(t-1)} \end{bmatrix}.$$

Huang et al. (2006) uses both KCV and GCV for the choice of tuning parameter λ, but this idea results in computational burden as n, p increases. The detailed algorithm of Huang et al. (2006) is described in Algorithm 1.

2.1.2 Other methods

In this subsection, we will briefly introduce some papers that have been dedicated to estimate Σ or Ω using various methodologies.

Huang et al. (2006) used the penalized regression methods to estimate MCD-based Σ and Ω. Also, Levina et al. (2008) extended Huang et al. (2006) by utilizing various nested penalties to successfully estimate large covariance

Algorithm 1 Huang et al. (2006)

1: For $t = 1$, $y_1 = \epsilon_1$,

$$\hat{\sigma}_1^2 = \sum_{i=1}^{n} \frac{\epsilon_{i1}^2}{n} = \sum_{i=1}^{n} \frac{y_{i1}^2}{n}$$

2: For $t = 2, ..., p$, $y_t = \sum_{i=1}^{t-1} \phi_{ti} y_i + \epsilon_t$ with some initial σ_t^2 and for m=1 or 2,

(2-1) Estimate ϕ.

$$(\hat{\phi}_{t1}, \hat{\phi}_{t2}, ..., \hat{\phi}_{t(t-1)}) = \underset{\phi_{t1},...,\phi_{t(t-1)}}{argmin} \Big\{ \sum_{i=1}^{n} \Big(\frac{y_{it} - \sum_{j=1}^{t-1} \phi_{ti} y_{ij}}{\sigma_t^2} \Big) \Big\} + \lambda \Big(\sum_{i=1}^{t-1} |\phi_{ti}|^m \Big)$$

(2-2) Estimate σ.

$$(\hat{\sigma}_t^2) = \sum_{i=1}^{n} \frac{\epsilon_{it}^2}{n} = \sum_{i=1}^{n} \frac{(y_{it} - \sum_{j=1}^{t-1} \hat{\phi}_{tj} y_{ij})^2}{n}$$

3: Iterate (2-1) and (2-2) until convergence. □

matrices in sparse settings. Rothman et al. (2008) handled sparse situations using a permutation-invariant estimation method for covariance matrices.

Some studies focus on estimating K-banded Σ and Ω by leveraging smoothing techniques to regularize L. Wu and Pourahmadi (2003) employed a nonparametric estimation method for large covariance matrices in longitudinal settings. Rothman et al. (2010) introduced a novel Cholesky-based covariance regularization approach tailored for high-dimensional data. Meanwhile, Jiang (2012) adopted a group effect in the Cholesky factor to estimate the precision matrix using cholesky decomposition.

To address order-related issues, certain papers explore permutating variables in MCD for estimating Σ and Ω. Kang et al. (2020) investigated permutations of variable orders to optimize Ω prediction using the MCD-based approach. Moreover, Kang and Deng (2020) developed variable ordination

of MCD for estimating time-varying covariance matrices and also proposed Cholesky-based estimation for large-dimensional covariance matrices.

For nonparametric Ω estimation, Lee and Lee (2021) employed a nonparametric K-banded Cholesky prior to estimate bandable MCD-based Ω.

2.2 Scaled Lasso

Scaled lasso method is proposed by Sun and Zhang (2012) that adds a ℓ_1- regularization penalty term to the ordinary least squares (OLS) loss function by combining a scaling factor. This method is proposed to deal with high-dimensional data and to solve the multicollinearity. The objective of the scaled lasso method is to perform variable selection and shrink less important coefficients as in lasso in order to validate sparsity and to deal with situations especially when variables have various scalings by using a scaling factor which is pre-chosen by some suggestions which we will describe in details.

In this section, we will introduce regression method using scaled lasso (Sun and Zhang, 2012) which we applied the idea to our proposals and then introduce the precision estimation method using the idea of scaled lasso (Sun and Zhang, 2013).

2.2.1 Regression method

Sun and Zhang (2012) proposed ℓ_1- penalized linear regression method using a scaling factor. For a design matrix $X \in R^{n\times p}$ and a response vector $y \in R^n$ with a p-dimensional regression coefficients vector β, Sun and Zhang (2012) assume that the regression has a form of $y = X\beta + \epsilon$ with $Var(\epsilon) = \sigma^2$. Sun

and Zhang (2012) combined the l_1- regularization and scaling factor to solve the regression, where the penalty is scaled proportional to the error variance $\hat{\sigma}$ to jointly estimate σ^2 and β. The objective loss function of this method is in (2.13).

$$Ł_{\lambda_0}(\beta, \sigma) = \frac{|y - X\beta|_2^2}{2n\sigma} + \frac{\sigma}{2} + \lambda_0|\beta|_1 \tag{2.13}$$

Then, (2.13) is jointly convex in (β, σ) and therefore it converges to $(\hat{\beta}, \hat{\sigma}) = \underset{\beta,\sigma}{argmin} L_{\lambda_0}(\beta, \sigma)$ by solving the alternatinng minimization algorithm as (2.14).

$$\begin{aligned} &\hat{\sigma} \longleftarrow |y - X\hat{\beta}^{\text{old}}|_2/\sqrt{n} \\ &\lambda \longleftarrow \hat{\sigma}\lambda_0, \qquad\qquad (\lambda = \sigma\lambda_0) \\ &\hat{\beta} \longleftarrow \hat{\beta}^{\text{new}}, \;\; L_\lambda(\hat{\beta}^{\text{new}}) \leq L_\lambda(\hat{\beta}^{\text{old}}) \end{aligned} \tag{2.14}$$

Here, λ_0 is a prefixed tuning parameter. There are several commonly used choices of tuning parameter, such as universal tuning parameter λ_0^U, union bound tuning parameter λ_0^{ub} and probabilistic error bound tuning parameter λ_0^{pb} as in (2.15). Among these parameters, a universal tuning parameter λ_0^U is most commonly used.

$$\begin{aligned} \lambda_0^U &= \sqrt{2n^{-1}\log(p-1)} \\ \lambda_0^{ub} &= \sqrt{4n^{-1}\log p} \\ \lambda_0^{pb} &= \sqrt{2}L_n(k/p) \end{aligned} \tag{2.15}$$

For λ_0^{pb}, k is a real solution of $k = L_1^4(k/p) + 2L_1^2(k/p), L_n(t) = n^{-1/2}\Phi^{-1}(1-t)$ and it can be solved via bisection method.

Scaled lasso regression method has advantages that variables with various

scalings can be handled easily by using scaling factor and it has computational efficiency by using a prefixed tuning parameter λ_0.

2.2.2 Precision matrix estimation

Sun and Zhang (2013) introduced the precision matrix estimation method using the scaled lasso method reviewed in 2.2.1. The algorithm uses a neighborhood selection method to estimate each column of the precision matrix. By using neighborhood selection method, the resulting precision matrix is not symmetric, which contradicts the inherent symmetry of precision matrices. To solve this issue, Yuan (2010) proposed a method to restore the symmetry of the calculated precision matrix.

The algorithm of proposed precision matrix estimator is as follows.

Algorithm 2 Scaled Lasso Matrix Inversion

Let X be $n \times p$ matrix, then for $k = 1, ..., p$,

1: Let X_{-j} be X without jth column. Sum $X_{-j}^T X_{-j}$ columnwise and divide by n, and let S_j be square root of resulted value.

2: Let Y_j be X_{-j} divided by S_j columnwise. Apply scaled lasso method by Sun and Zhang (2012) to Y_j and X_j, where X_j is j-th column of X. Use tuning universal parameter $\lambda_0^{(univ)}$.

3: Let σ^{-2} be diagonals of D and β be coefficient matrix calculated in 2. Then, the resulted precision matrix is $\hat{\Omega}^{(\text{nhd})} = -\beta \times D$.

4: Symmetrize the calculated $\hat{\Omega}^{(\text{nhd})}$ by Yuan (2010).

$$\Omega_F = \underset{\tilde{\Omega}=\tilde{\Omega}^T}{argmin} \|\hat{\Omega}^{(\text{nhd})} - \tilde{\Omega}\|_1$$

The scaled lasso estimator $\hat{\beta}, \hat{\sigma}$ are consistent under certain conditions, (i) the penalty level condition and (ii) the compatibility condition (Van de

Geer and Bühlmann, 2009). The penalty level condition bounds the tuning parameter level as in (2.16).

$$\lambda_0 > A\sqrt{2n^{-1}\log p}, \;\; A > 1 \tag{2.16}$$

The compatibility condition for $\hat{\Sigma}$ requires for all β satisfy $\|\beta_{s_0^c}\|_1 \leq 3\|\beta_{s_0}\|_1$, where β is a $p \times 1$ vector with a subset $S \subseteq \{1, \cdots, p\}$ and define β_s by $\beta_{s,j} := \beta_j \mathbb{1}\{j \in S\}$. Then, the compatibility condition is (2.17).

$$^{\exists}\phi_0 > 0, \quad s.t. \;\; \|\beta_{s_0}\|_1^2 \leq s_0 \beta^T \hat{\Sigma} \beta / \phi_0^2 \tag{2.17}$$

where ϕ_0^2 is the compatibility constant.

Under above conditions (2.16), (2.17), the scaled lasso precision matrix estimator $\hat{\Omega}$ achieves a convergence rate of $O_p(d\sqrt{\log p/n})$ where d implies the maximum sparsity factor among p rows of Ω.

$$d = \max_{1 \leq j \leq p} \#\{k : \Omega_{jk} \neq 0\} \tag{2.18}$$

The scaled lasso precision matrix estimator effectively performs variable selection by using scaled lasso property and is computationally efficient when dealing with data where p is much larger than n.

Chapter 3

Modified Cholesky Decomposition based Precision Matrix Estimation via Scaled Lasso

Let $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ be the n copies of $p-$dimensional vector with mean zero and covariance matrix Σ. We let the data matrix as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1^\top \\ \mathbf{y}_2^\top \\ \vdots \\ \mathbf{y}_n^\top \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1p} \\ Y_{11} & Y_{12} & \cdots & Y_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{np} \end{pmatrix}. \tag{3.1}$$

In this thesis, we assume the $p-$variables are ordered in time or observed in a sequence and, once the order of variables are given, the precision matrix

Ω has the unique decomposition in the form of $L^\top D^{-1} L$, where L is a lower triangular matrix and D is a diagonal matrix. We are interested in estimating Ω from $\mathbf{Y}$ with the use of the order information among variables.

As we review in Section 2.1, the elements of L in (2.3) are related with the regression coefficients of the sequential regression model (2.1) and L can be estimated by solving a set of least square problems. Our proposal in the thesis is the use of scaled lasso by Sun and Zhang (2012). To estimate L and D, we suggest to minimize

$$\mathcal{L}(\Phi, \boldsymbol{\sigma}) = \sum_{k=2}^{p} \left\{ \frac{1}{2n\sigma_k} \|Y_k - \sum_{j=1}^{k-1} \phi_{kj} Y_j\|_2^2 + \frac{1}{2}\sigma_k + \lambda_{0k} \sum_{j=1}^{k-1} |\phi_{kj}| \right\} \tag{3.2}$$

where $Y_k = \big(Y_{1k}, Y_{2k}, \ldots, Y_{nk}\big)^\top$ and $\boldsymbol{\sigma} = \big\{\sigma_k, k = 1, 2, \ldots, p\big\}$, and λ_{0k} is set as $\lambda_{0k} = \sqrt{2n^{-1}\log(k-1)}$ following Sun and Zhang (2012). We let $\phi_k = \big(\phi_{k1}, \phi_{k2}, \ldots, \phi_{k(k-1)}\big)^\top$ below.

To solve (3.2), we propose the following iterative algorithm, the iteration between the estimation of $\boldsymbol{\sigma}$ and the estimation of Φ.

Algorithm 3 SLMCD

For each $k = 1, 2, \cdots, p$

1: Let $\{\hat{\phi}_{ki}^{(0)}\}_{i=1,\ldots,k-1}$ be initial values of $\{\phi_{ki}\}_{i=1,\ldots,k-1}$.

For each iteration $t = 0, 1, \cdots$

2: Calculate $\hat{\sigma}_k^{(t)} = \frac{1}{\sqrt{n}} \|Y_k - \sum_{j=1}^{k-1} \hat{\phi}_{kj}^{(t)} Y_j\|_2$.

3: $\hat{\phi}_k^{\text{new}} = \underset{\phi_k}{\text{argmin}} \left\{ \frac{1}{2n} \|Y_k - \sum_{j=1}^{k-1} \phi_{kj} Y_j\|_2^2 + \lambda_{0k} \sigma_k \sum_{j=1}^{k-1} |\phi_{kj}| \right\}$

$\hat{\phi}_k^{\text{old}} \longleftarrow \hat{\phi}_k^{\text{new}}$, $L\big(\hat{\phi}_k^{\text{new}}\big) \leq L\big(\hat{\phi}_k^{\text{old}}\big)$

4: Iterate 2-3 until convergence. □

3.1 Convergence of SLMCD

We next investigate the convergence rate of our SLMCD estimator. Our results on the rate mainly depends on the convergence rate of the scaled lasso estimator given in Corollary 1 of Sun and Zhang (2012) and we start with the recall of the corollary.

Three assumptions are made for the convergence rate of the scaled lasso estimator. First, (A1) they assume the data $\mathbf{y}_i$ are independently from the normal distribution with mean zero and covariance matrix Σ. Second, (A2) the tuning parameter is set as

$$\lambda_{0p} \sim \sqrt{\log p/n},$$

which approaches 0 as n increases. Third, (A3) they make restricted eigenvalue (RE) condition for the design matrix $\mathbf{X}$ and coefficient vector $\boldsymbol{\beta}$ - $\mathbf{X}$ and $\boldsymbol{\beta}$ are the terms in the linear regression model and we do not define them specifically - that is, if there exists $c > 0$ such that

$$\kappa = \inf \left\{ \frac{\|\hat{\Sigma} \cdot \boldsymbol{\beta}_S\|_2^2}{n\|\boldsymbol{\beta}_S\|_2^2} : \boldsymbol{\beta}_S \in C(S) \setminus \{0\} \right\} \geq c, \tag{3.3}$$

where $I(A)$ is the indicator function of the event A, $\boldsymbol{\beta}_S$ is a vector of $\big\{\beta_j I(\{j \in S\}), j = 1, 2, \ldots, p\big\}$, and

$$C(S) := \Big\{\boldsymbol{\beta} \in \mathcal{R}^p \mid \big\|\boldsymbol{\beta}_{S^c}\big\|_1 \leq 3\big\|\boldsymbol{\beta}_S\big\|_1\Big\},$$

we call $\mathbf{X}$ and $\boldsymbol{\beta}$ satisfy the RE condition.

Theorem A [Corollary 1 of Sun and Zhang (2012)]

Under assumptions (A1)-(A3), suppose we set

$$\lambda_{0p} = A\sqrt{\frac{2\log p}{n}} \quad \textit{with } A > (\xi+1)/(\xi-1), \textit{ for some } \xi > 1$$

then

$$\big\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big\|_2^2 \le C\, s\, \lambda_{0p}^2$$

for appropriately defined $C > 0$ *and* $s = \sum_{j=1}^{p} I\big(\beta_j \neq 0\big)$.

The following theorem shows the convergence rate of the proposed SLMCD.

Theorem 1. *Under the assumption that each regression model in* (2.1) *satisfies the assumptions (A1)-(A3) with certain probabilties, we have*

$$\big\|\hat{\Omega} - \Omega\big\|_F^2 = O_p\Big(\frac{s\log p}{n}\Big), \quad s = \sum_{k=2}^{p}\sum_{\ell=1}^{k-1} I\big(\phi_{k\ell} \neq 0\big). \tag{3.4}$$

Proof of Theorem 1. We will show the following property by using Corollary 1 of Sun and Zhang (2012).

$$\|\hat{\Omega} - \Omega\|_F^2 = \|\hat{L}^T\hat{D}^{-1}\hat{L} - L^T D^{-1} L\|_F^2 = O_p\Big(\frac{s\log p}{n}\Big)$$

Decompose the following equation $\|\hat{\Omega} - \Omega\|_F = \|\hat{L}^T\hat{D}\hat{L}^T - L^T D L\|_F$, here

we used D instead of D^{-1} for simplicity.

$$\|\hat{L}^T\hat{D}\hat{L} - L^TDL\|_F \leq \underbrace{\|(\hat{L}-L)(\hat{D}-D)(\hat{L}^T-L^T)\|_F}_{\textbf{(1)}} + \underbrace{\|\hat{L}\hat{D}L^T - \hat{L}D\hat{L}^T + \hat{L}DL^T - L\hat{D}L^T + L\hat{D}\hat{L}^T - LD\hat{L}^T\|_F}_{\textbf{(2)}}$$

Above inequality can be proved by using the decomposition of **(1)** and the properties of the sum of absolute values.

$$\textbf{(1)} : (\hat{L}-L)(\hat{D}-D)(\hat{L}^T-L^T) = (\hat{L}\hat{D} - L\hat{D} - \hat{L}D + LD)(\hat{L}-L^T)$$
$$= \hat{L}\hat{D}\hat{L}^T + \underbrace{\hat{L}\hat{D}L^T - \hat{L}D\hat{L}^T + \hat{L}DL^T - L\hat{D}L^T + L\hat{D}\hat{L}^T - LD\hat{L}^T}_{(2)} - LDL^T$$

We will decompose the **(2)** equation in followings.

$$\textbf{(2)} : \|\hat{L}\hat{D}L^T - \hat{L}D\hat{L}^T + \hat{L}DL^T - L\hat{D}L^T + L\hat{D}\hat{L}^T - LD\hat{L}^T\|_F = \|\textbf{(2-1)} - \textbf{(2-2)}\|_F$$
$$\leq \|\textbf{(2-1)} + \|\textbf{(2-2)}\|_F$$

$$\begin{aligned}
\textbf{(2-1)} : (\hat{L}-L)(\hat{D}-D)L^T &= \hat{L}\hat{D}L^T - \cancel{\hat{L}DL^T} - L\hat{D}L^T + \cancel{LDL^T} \\
-(\hat{L}-L)D(\hat{L}^T-L) &= \hat{L}D\hat{L}^T - \cancel{\hat{L}DL^T} - \cancel{LD\hat{L}^T} + \cancel{LDL^T} \\
+L(\hat{D}-D)(\hat{L}^T-L^T) &= \underline{L\hat{D}\hat{L}^T - L\hat{D}L^T - \cancel{LD\hat{L}^T} + LDL^T} \\
&= \hat{L}\hat{D}L^T - \hat{L}D\hat{L}^T + L\hat{D}\hat{L}^T - 2L\hat{D}L^T + LDL^T
\end{aligned}$$

$$
\begin{aligned}
\textbf{(2-2)} : (\hat{L}-L)DL^T &= \hat{L}DL^T - LDL^T \\
+ L(\hat{D}-D)L^T &= L\hat{D}L^T - \cancel{LDL^T} \\
- LD(\hat{L}^T - L^T) &= \underline{LD\hat{L}^T - \cancel{LDL^T}} \\
&= \hat{L}DL^T - LDL^T + L\hat{D}L^T - LD\hat{L}^T
\end{aligned}
$$

$$
\begin{aligned}
\textbf{(2-1)} + \textbf{(2-2)} &= \Big(\hat{L}\hat{D}L^T - \hat{L}D\hat{L}^T + L\hat{D}\hat{L}^T - 2L\hat{D}L^T + \cancel{LDL^T}\Big) \\
&\quad + \Big(\hat{L}DL^T - \cancel{LDL^T} + \cancel{L\hat{D}L^T} - LD\hat{L}^T\Big) \\
&= \hat{L}\hat{D}L^T - \hat{L}D\hat{L}^T + L\hat{D}\hat{L}^T + \hat{L}DL^T - L\hat{D}L^T - LD\hat{L}^T = \textbf{(2)}
\end{aligned}
$$

Also, since the property of the multiplication of matrix $\|AB\|_F \le \|A\|_F\|B\|_F$ also can be applied to Frobenius norm, we can do the decompositions as followings.

$$
\begin{aligned}
\|\hat{L}^T\hat{D}\hat{L} - L^TDL\|_F \le \underbrace{2\underbrace{\|\hat{L}-L\|_F}_{(a)}\|D\|_F + \underbrace{\|\hat{D}-D\|_F}_{(b)}\|L\|_F^2}_{(\star)} \\
+ 2\|\hat{L}-L\|_F\|\hat{D}-D\|_F\|L\|_F + \|\hat{L}-L\|_F^2\|D\|_F \\
+ \|\hat{L}-L\|_F^2\|\hat{D}-D\|_F
\end{aligned}
$$

For (a), we will prove the following.

$$
\|\hat{L}-L\|_F^2 = \sum_k \|\hat{\phi}_k - \phi_k\|_2^2 = O_p\Big(s\frac{\log p}{n}\Big) \tag{3.5}
$$

$$
\begin{aligned}
\|\hat{L}-L\|_F = \Big(\sum_k \|\hat{\phi}_k - \phi_k\|_2\Big)^{1/2} = O_p\Big(\sum_{k=1}^{p}\frac{s_k\log k}{n}\Big)^{1/2} &= O_p\Big(\log p\sum_{k=1}^{p}\frac{s_k}{n}\Big)^{1/2} \\
&= O_p\Big(s\frac{\log p}{n}\Big)^{1/2}
\end{aligned}
$$

Here,

$$s_k = \sum_{i=1}^{k-1} I(\phi_{ki} \neq 0), \quad \bar{s} = \sum_k s_k/p, \quad s = \sum_{j=2}^{p}\sum_{i=1}^{j-1} I(\phi_{ji} \neq 0)$$

which are sparsity parameters.

We can compute the above convergence rate in this following way. $\|\hat{\phi}_k - \phi_k\|_2^2 \leq C_k k \lambda_{0k}^2 \sigma_k^{\star 2}/(1-\tau_{k\star}^2)^2$ has the convergence rate of $\lambda_{0k}^2 = O_p\Big(\frac{\log k}{n}\Big)$. C_k is a constant, $\sigma_k^\star = \|Y_k - Y_{[1:(k-1)]}^T \phi_k\|/\sqrt{n}$ and $(1-\tau_{k\star}^2)$ are constants converge. Also, $k = |s_k|$ implies the number of nonzero $\{\phi_{kj}\}_{j=1,\ldots,(k-1)}$. This factor has to be appeared since $p \geq n \geq \|\phi\|_0 \to \infty$.

For (b),

$$(b) : \|\hat{D} - D\|_F = \Big(\sum_{i=1}^{p} |\hat{\sigma}_i - \sigma_i|^2\Big)^{1/2} = \Big(\sum_{i=1}^{p} \sigma_i^2 O_p\Big(\frac{\log p}{n}\Big)\Big)^{1/2} \cong O_p\Big(\sqrt{\frac{p \log p}{n}}\Big)$$

and

$$\|D\|_F^2 = \sum_{j=1}^{p} \sigma_j^2 = (constant)$$
$$\|L\|_F^2 = \sum_{i,j} |\phi_{ij}^2| = (constant)$$

□

Chapter 4

Numerical study

We numerically compare our method, SLMCD, with other methods in the literature. The chapter is composed of two folds. In the first, we compare SLMCD with the four methods by Huang et al. (2006), which is reviewed in Section 2.1. In the second, we consider more settings of the sample size, dimension, and true precision matrix, and two more estimators of the sparse precision matrix.

4.1 The cases of Huang et al. (2006)

In this section, we numerically compare our proposal SLMCD to the four methods by Huang et al. (2006), where the four methods are the combinations of (i) ℓ_1 and ℓ_2- regularizations and (ii) 5CV and GCV for the selection of the tuning parameter.

For the simulation we consider exactly the same settings as those in Huang et al. (2006). We consider following four precision matrices as the true matrices,

(i) The identity matrix: $\Omega_1 = \mathrm{I}_p$.

(ii) The heterogenous diagonal matrix:

$$\Omega_2 = \mathrm{diag}\left(\frac{1}{p}, \frac{1}{p-1}, \frac{1}{p-2}, \cdots, 1\right).$$

(iii) The 1st order autoregressive AR(1):

$$\Omega_3 = L^\top D^{-1} L,$$

where

$$L^\top = \begin{pmatrix} 1 & -0.8 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -0.8 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \vdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & 1 & -0.8 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and $D = 0.01 \cdot \mathrm{I}_p$.

(iv) The compound symmetry for Σ:

$$\Sigma = \mathrm{I}_p + \rho \mathbf{1}_p \mathbf{1}^\top$$

and

$$\Omega_4 = L^\top D^{-1} L,$$

where $\mathbf{1}_p$ is the $p \times 1$ vector of ones, and

$$L = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ -\frac{\rho}{1+\rho} & 1 & 0 & \cdots & \cdots & 0 \\ -\frac{\rho}{1+2\rho} & -\frac{\rho}{1+2\rho} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{\rho}{1+(p-2)\rho} & -\frac{\rho}{1+(p-2)\rho} & -\frac{\rho}{1+(p-2)\rho} & -\frac{\rho}{1+(p-2)\rho} & 1 & 0 \\ -\frac{\rho}{1+(p-1)\rho} & -\frac{\rho}{1+(p-1)\rho} & -\frac{\rho}{1+(p-1)\rho} & -\frac{\rho}{1+(p-1)\rho} & -\frac{\rho}{1+(p-1)\rho} & 1 \end{pmatrix}.$$

We set $\sigma = 1$ and $\rho = 0.5$.

We remark that the cases (i)-(iii) satisfies our sparsity assumption but (iv) does not.

The data are generated from the normal distribution, and the sample size and dimension are set as $n = 100$ and $p = 30$. In each data set, we compute the four estimators by Huang et al. (2006) and our SLMCD. We replicate the above 100 times and evaluate the empirical error of the estimation.

Again, following Huang et al. (2006), we consider two errors (loss functions) of the estimation, the entropy loss (ℓ_E) and the quadratic loss (ℓ_Q) in (4.1).

$$\begin{aligned} \ell_E(\Sigma, G) &= \mathrm{tr}(\Sigma^{-1}G) - \log|\Sigma^{-1}G| - p \\ \ell_Q(\Sigma, G) &= \mathrm{tr}(\Sigma^{-1}G - I)^2. \end{aligned} \tag{4.1}$$

Both errors above are defined with the covariance matrix estimator G, not the estimator of the precision matrix, and thus, to evaluate it, we need evaluate the inverse of the estimator of the precision matrix. The empirical errors of the estimation over 100 replications are reported in Table 4.1.

		Sample	Huang L_1		Huang L_2		SLMCD
			GCV	5CV	GCV	5CV	
EL	Ω_1 Identity	5.231	0.761 (4)	0.248 (2)	0.850 (5)	0.728 (3)	0.269 (1)
(ℓ_E)	Ω_2 Heter. Diag.	5.258	1.490 (3)	0.462 (1)	1.780 (4)	1.887 (5)	0.466 (2)
	Ω_3 AR(1)	5.376	1.969 (2)	1.976 (3)	3.911 (4)	4.020 (5)	1.389 (1)
	Ω_4 CS	5.863	2.405 (4)	2.452 (5)	2.201 (3)	2.198 (2)	2.036 (1)
QL	Ω_1 Identity	8.824	1.296 (4)	0.489 (1)	1.483 (5)	1.284 (3)	0.521 (2)
(ℓ_Q)	Ω_2 Heter. Diag.	11.513	2.860 (3)	1.017 (1)	3.654 (4)	3.875 (5)	1.029 (2)
	Ω_3 AR(1)	9.672	3.629 (2)	3.644 (3)	6.864 (4)	7.058 (5)	2.392 (1)
	Ω_4 CS	9.789	6.681 (1)	11.072 (5)	7.513 (3)	7.036 (2)	8.036 (4)

Table 4.1: Losses and orders of losses are calculated for cases of Huang et al. (2006) with covariance estimation method of Huang et al. (2006) with the combinations of (i) ℓ_1 and ℓ_2- regularizations and (ii) 5CV and GCV for the selection of the tuning parameter and our method SLMCD with $n = 100, p = 30$ and $N = 100$ replications.

Table 4.1 shows that our SLMCD outperforms over the methods by Huang et al. (2006) when the true precision matrix satisfies the sparsity assumption (the cases of Ω_1, Ω_2, and Ω_3), while it does not for the dense precision matrix (Ω_3), However, even for the dense case, SLMCD performs well in view of ℓ_E error. In addition, although we do not separately report them, SLMCD uses a pre-determined value for the tuning parameter instead of searing it, and is computationally much more efficient than Huang's methods.

4.2 Extended study

In this section, we numerically compare our proposal SLMCD to several lasso-based precision matrix estimation methods. Those compared methods include Huang et al. (2006), with ℓ_1 and ℓ_2- regularizations, scaled lasso precision matrix estimation method by Sun and Zhang (2013), Graphical Lasso by Friedman et al. (2008), CLIME by Cai et al. (2011).

To demonstrate the effectiveness of our method, we conducted an evaluation using a set of example precision matrices labeled (i)-(viii).

(i) The identity matrix: $\Omega_1 = \mathrm{I}_p$.

(ii) The heterogenous diagonal matrix:

$$\Omega_2 = \mathrm{diag}\left(\frac{1}{p}, \frac{1}{p-1}, \frac{1}{p-2}, \cdots, 1\right).$$

(iii), (iv) The 1st order autoregressive AR(1):

$$\Omega_3 = L^\top D_3^{-1} L$$
$$\Omega_4 = L^\top D_4^{-1} L$$

where

$$L^\top = \begin{pmatrix} 1 & -0.8 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -0.8 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \vdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & 1 & -0.8 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and $D_3 = \mathrm{I}_p$, $D_4 = \mathrm{diag}\,(1, ..., 1, 0.01, ..., 0.01)$ with $p/2$ number of 1, 0.01, where p is even number.

(v) The compound symmetry for Σ:

$$\Sigma = \mathrm{I}_p + \rho \mathbf{1}_p \mathbf{1}^\top$$

and

$$\Omega_5 = L^\top D^{-1} L,$$

where $\mathbf{1}_p$ is the $p \times 1$ vector of ones, and

$$L = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ -\frac{\rho}{1+\rho} & 1 & 0 & \cdots & \cdots & 0 \\ -\frac{\rho}{1+2\rho} & -\frac{\rho}{1+2\rho} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{\rho}{1+(p-2)\rho} & -\frac{\rho}{1+(p-2)\rho} & -\frac{\rho}{1+(p-2)\rho} & -\frac{\rho}{1+(p-2)\rho} & 1 & 0 \\ -\frac{\rho}{1+(p-1)\rho} & -\frac{\rho}{1+(p-1)\rho} & -\frac{\rho}{1+(p-1)\rho} & -\frac{\rho}{1+(p-1)\rho} & -\frac{\rho}{1+(p-1)\rho} & 1 \end{pmatrix}.$$

We set $\sigma = 1$ and $\rho = 0.5$.

Following (vi)-(viii) are dense models.

(vi) $\Omega_6 = \{\Omega_{ij}\}$ with $\Omega_{ij} = 0.6^{|i-j|}$

(vii) $\Omega_7 = B + \delta \mathrm{I}_p$

$B = \{B_{ij}\}$ with $2B_{ij} \sim Bin(1, 0.1)$ for $i \neq j$, $B_{ij} = 0$ for $i = j$, with δ a constant chosen s.t. the condition number (κ) of Ω_7 is p.

Then, we rescaled resulted Ω to the unit in diagonal.

(viii) $\Omega_8 = D^{1/2}\Theta D^{1/2}$

$\Theta = \{\Theta_{ij}\}$ with $\Theta_{ij} = 0.6^{|i-j|}$ and $D = \{D_{ii}\}$ is a diagonal with $D_{ii} = (4i + p - 5)/\{5(p-1)\}$.

(i)-(v) are the settings sourced from Huang et al. (2006), where (iii), (iv) include variations of diagonal elements in D of $\Omega = L^\top D^{-1} L$. This adjustment was made to achieve similar levels of numbers compared to other cases in Huang et al. (2006). (v)-(vii) are the exact same settings derived from Sun and Zhang (2013). Among these settings, (i)-(iv) and (vi) were selected as they represent sparse cases that align well with the assumptions of our proposed method. However, (iv)-(vi) present dense cases that deviate from our assumptions. Additionally, cases (v)-(vii) do not consider the time-ordination of data structures, further violating our underlying assumptions.

The data are generated from normal distributions and our numerical study considered various n, p situations, including where the number of samples (n) is less than, equal to, and greater than the number of variables (p). The study involved $n = 100$ samples and three different scenarios for the number of variables: $p = 30$, $p = 100$, and $p = 150$. We performed a total of $N = 100$ replications for each case.

Throughout the study, we sought to evaluate the performance of our method and compare it to other relevant approaches under these diverse scenarios, encompassing both sparse and dense models with varying levels of variables. The repetition of the experiments allowed us to assess the stability and consistency of the results.

Our primary focus is on the estimation of precision matrices. To evaluate the performance of various methods, we employed three errors (loss functions) tailored to precision matrices. These loss functions include the matrix 1-norm (ℓ_1), the matrix 2-norm (ℓ_2), and the Frobenius norm (ℓ_F). In these errors, we used Ω to represent the true precision matrix and F to represent the estimated

precision matrix.

$$\begin{aligned}
\ell_1(\Omega, F) &= \max \sum_{i=1}^{p} |(\Omega_{ij} - F_{ij})| \\
\ell_2(\Omega, F) &= \max\{\lambda_{\text{eigen}}(\Omega - F)^T(\Omega - F)\}^{\frac{1}{2}} \\
\ell_F(\Omega, F) &= \sqrt{\sum_{i,j}(\Omega_{ij} - F_{ij})^2}
\end{aligned} \tag{4.2}$$

For ease of comparison, we have divided the results into two parts: one for Huang et al. (2006) and the other for Sun and Zhang (2013). Interestingly, the results remain consistent whether using ℓ_1, ℓ_2 or ℓ_F norms. For the cases of Huang et al. (2006), in the case of Ω_1, which represents the simplest scenario using an identity matrix for the precision matrix, all methods yield adequately small levels of losses, indicating their proficiency in handling this straightforward case. For Ω_2, Ω_{31} and Ω_{32}, which best align with our underlying assumptions, SLMCD consistently outperforms other methods by yielding the smallest losses across all dimensions of p. This outcome strongly supports the theoretical results we derived. Regarding Ω_4, we observed that every method performed similarly, with comparable levels of losses.

Overall, these findings provide valuable insights into the effectiveness of our proposed SLMCD method in comparison to existing approaches, especially in scenarios that best adhere to our assumptions.

Table 4.2 - 4.4 are cases of Huang et al. (2006) and Table 4.5 - 4.7 are cases of Sun and Zhang (2013). Table 4.2, 4.3 shows the calculated losses of matrix 1-norm (ℓ_1) , Table 4.3, 4.6 shows the calculated losses of matrix 2-norm (ℓ_2) and Table 4.4, 4.7 shows the calculated losses of frobenius norm (ℓ_F).

In Table 4.2 - 4.4, SLMCD performs small losses for all settings under various n, p cases. Especially, for Ω_{31}, Ω_{32}, SLMCD results in much smaller losses than other precision estimation methods and thoose gaps increase as p increases.

In Table 4.5 - 4.7, specifically Ω_5, Ω_6, and Ω_7, the scaled lasso method yields the smallest losses for Ω_6 and Ω_7. On the other hand, our propoosed SLMCD does not achieve the best results for these cases since they do not consider the time-ordination of data structures. Furthermore, Ω_5, Ω_7 are dense cases, which also contradict assumptions of SLMCD, SLMCD does not perform the smaller losses compared to other methods.

These tables provide a comprehensive overview of the performance of different methods under various norms and cases from Sun and Zhang (2013). The comparison of these results contributes to a thorough understanding of how different methods handle scenarios without time-ordination constraints and helps in identifying their respective strengths and limitations in such contexts.

(i) Matrix 1-norm (ℓ_1)

		SLMCD	Huang L_1 5CV	Huang L_2 5CV	Scaled Lasso	GL	CL
$p = 30$	Ω_1 Identity	**0.430**	**0.394**	1.570	0.451	0.492	0.484
	Ω_2 Heter. Diag	**0.139**	**0.162**	0.928	0.587	0.492	0.479
	Ω_{31} AR(1)	**1.251**	1.293	10.929	**1.168**	2.732	2.910
	Ω_{32} AR(1)	**109.241**	291.722	**283.609**	292.753	323.492	322.643
	Ω_4 CS	6.598	6.989	**6.583**	**6.480**	7.537	6.803
$p = 100$	Ω_1 Identity	0.514	**0.474**	**0.486**	0.521	0.492	0.598
	Ω_2 Heter. Diag	**0.136**	**0.139**	1.485	6.202	0.498	0.720
	Ω_{31} AR(1)	**1.451**	1.456	3.318	**1.448**	2.732	2.533
	Ω_{32} AR(1)	**134.408**	**135.063**	297.914	296.186	323.492	320.916
	Ω_4 CS	**11.009**	11.444	11.430	**10.877**	11.965	11.359
$p = 150$	Ω_1 Identity	0.535	**0.510**	0.527	0.540	**0.492**	0.587
	Ω_2 Heter. Diag	**0.120**	0.121	0.364	**0.096**	0.502	0.762
	Ω_{31} AR(1)	**1.566**	2.992	3.001	**1.717**	2.732	2.240
	Ω_{32} AR(1)	**143.238**	**139.531**	297.065	297.015	323.492	319.829
	Ω_4 CS	**12.603**	13.022	13.005	**12.477**	13.528	12.947

Table 4.2: Matrix 1-norm (ℓ_1) are calculated for cases of Huang et al. (2006) with some lasso-based precision matrix estimation methods and proposed SLMCD for comparison with $n = 100, p = 30, 100, 150$ and $N = 100$ replications.

(ii) Matrix 2-norm (ℓ_2)

		SLMCD	Huang L_1 5CV	Huang L_2 5CV	Scaled Lasso	GL	CL
$p = 30$	Ω_1 Identity	0.398	**0.386**	0.670	0.408	0.492	**0.388**
	Ω_2 Heter. Diag	**0.138**	**0.150**	0.426	0.343	0.492	0.478
	Ω_{31} AR(1)	**0.991**	0.875	5.464	**0.833**	2.723	2.200
	Ω_{32} AR(1)	**85.397**	162.203	**155.045**	282.549	320.241	318.956
	Ω_4 CS	4.656	4.815	**4.382**	**4.621**	5.806	4.783
$p = 100$	Ω_1 Identity	0.486	**0.474**	0.475	0.486	0.492	**0.436**
	Ω_2 Heter. Diag	**0.136**	**0.138**	0.846	2.993	0.498	0.720
	Ω_{31} AR(1)	1.175	**1.013**	2.634	**0.936**	2.731	1.785
	Ω_{32} AR(1)	**107.063**	**99.569**	286.409	289.255	323.183	319.986
	Ω_4 CS	**6.803**	6.928	6.911	**6.791**	8.054	7.034
$p = 150$	Ω_1 Identity	0.519	0.510	0.513	0.519	**0.492**	**0.433**
	Ω_2 Heter. Diag	**0.120**	0.121	0.257	**0.090**	0.502	0.762
	Ω_{31} AR(1)	**1.291**	2.918	2.912	**1.020**	2.731	1.728
	Ω_{32} AR(1)	**113.851**	**99.605**	289.777	289.821	323.353	319.471
	Ω_4 CS	**7.445**	7.552	7.532	**7.434**	8.717	7.679

Table 4.3: Matrix 2-norm (ℓ_2) are calculated for cases of Huang et al. (2006) with some lasso-based precision matrix estimation methods and proposed SLMCD for comparison with $n = 100, p = 30, 100, 150$ and $N = 100$ replications.

(iii) Frobenius norm (ℓ_F)

		SLMCD	Huang L_1 5CV	Huang L_2 5CV	Scaled Lasso	GL	CL
$p = 30$	Ω_1 Identity	**0.859**	**0.834**	1.517	0.877	2.693	1.032
	Ω_2 Heter. Diag	**0.173**	**0.183**	0.530	0.462	2.302	0.519
	Ω_{31} AR(1)	2.323	**1.933**	9.036	**1.825**	8.632	6.581
	Ω_{32} AR(1)	**147.154**	**243.610**	254.398	602.561	750.651	746.583
	Ω_4 CS	**4.971**	5.139	5.015	**4.925**	8.594	5.156
$p = 100$	Ω_1 Identity	1.530	**1.478**	**1.481**	1.539	4.918	2.000
	Ω_2 Heter. Diag	**0.177**	0.178	**0.983**	4.099	4.711	0.789
	Ω_{31} AR(1)	4.769	**4.151**	14.697	**3.385**	15.930	9.116
	Ω_{32} AR(1)	**315.813**	**282.821**	1169.440	1174.672	1395.386	1376.981
	Ω_4 CS	**7.668**	7.783	7.787	**7.618**	15.827	8.212
$p = 150$	Ω_1 Identity	1.874	**1.816**	**1.823**	1.888	6.023	2.410
	Ω_2 Heter. Diag	**0.158**	**0.159**	0.308	13.523	5.891	0.840
	Ω_{31} AR(1)	**6.205**	20.721	20.668	**4.191**	19.540	10.672
	Ω_{32} AR(1)	**405.027**	**332.325**	1458.174	1449.986	1713.302	1685.990
	Ω_4 CS	**8.622**	8.720	8.744	**8.571**	19.245	9.318

Table 4.4: Frobenius norm (ℓ_F) are calculated for cases of Huang et al. (2006) with some lasso-based precision matrix estimation methods and proposed SLMCD for comparison with $n = 100, p = 30, 100, 150$ and $N = 100$ replications.

(i) Matrix 1-norm (ℓ_1)

		SLMCD	Huang L_1 5CV	Huang L_2 5CV	Scaled Lasso	GL	CL
$p = 30$	Ω_5	3.318	**3.157**	4.639	**3.157**	3.490	**2.985**
	Ω_6	1.488	1.510	2.131	**1.355**	2.140	**1.393**
	Ω_7	2.695	**2.350**	3.047	2.513	2.858	**2.494**
$p = 100$	Ω_5	3.469	3.568	3.908	**3.414**	3.492	**3.453**
	Ω_6	3.095	3.316	3.363	**2.938**	3.231	**2.984**
	Ω_7	3.139	3.228	3.401	**3.086**	3.229	**3.056**
$p = 150$	Ω_5	3.500	3.644	3.720	**3.499**	**3.492**	3.566
	Ω_6	3.719	3.801	3.868	**3.447**	3.773	**3.585**
	Ω_7	**3.264**	3.386	3.457	3.316	3.300	**3.229**

Table 4.5: Matrix 1-norm (ℓ_1) are calculated for cases of Sun and Zhang (2013) with some lasso-based precision matrix estimation methods and proposed SLMCD for comparison with $n = 100, p = 30, 100, 150$ and $N = 100$ replications.

2) Matrix 2-norm (ℓ_2)

		SLMCD	Huang L_1 5CV	Huang L_2 5CV	Scaled Lasso	GL	CL
$p = 30$	Ω_5	2.997	**2.644**	2.790	2.768	3.367	**2.428**
	Ω_6	0.945	**0.831**	0.981	**0.796**	1.550	0.836
	Ω_7	2.315	**1.815**	**1.978**	2.030	2.537	2.056
$p = 100$	Ω_5	3.221	3.324	3.405	**2.994**	3.478	**3.039**
	Ω_6	1.857	1.898	1.821	**1.667**	2.162	**1.722**
	Ω_7	2.818	2.905	2.998	**2.441**	3.009	**2.675**
$p = 150$	Ω_5	3.265	3.440	3.502	**3.037**	3.486	**3.152**
	Ω_6	2.270	2.354	2.349	**2.094**	2.507	**2.206**
	Ω_7	2.940	3.096	3.126	**2.495**	3.113	**2.826**

Table 4.6: Matrix 2-norm (ℓ_2) are calculated for cases of Sun and Zhang (2013) with some lasso-based precision matrix estimation methods and proposed SLMCD for comparison with $n = 100, p = 30, 100, 150$ and $N = 100$ replications.

3) Frobenius norm (ℓ_F)

		SLMCD	Huang L_1 5CV	Huang L_2 5CV	Scaled Lasso	GL	CL
$p = 30$	Ω_5	5.230	**4.503**	4.635	4.685	6.265	**4.052**
	Ω_6	2.059	1.849	2.142	**1.834**	3.625	**1.835**
	Ω_7	3.413	**2.749**	2.993	**2.867**	3.886	2.906
$p = 100$	Ω_5	10.218	10.806	11.235	**9.240**	11.616	**9.528**
	Ω_6	5.385	5.533	**4.996**	**4.933**	6.900	5.044
	Ω_7	6.644	7.160	7.345	**5.522**	7.210	**6.154**
$p = 150$	Ω_5	12.713	13.912	14.339	**11.469**	14.257	**12.206**
	Ω_6	6.972	7.260	7.154	**6.469**	8.508	**6.898**
	Ω_7	8.272	9.105	9.131	**6.804**	8.850	**7.825**

Table 4.7: Frobenius norm (ℓ_F) are calculated for cases of Sun and Zhang (2013) with some lasso-based precision matrix estimation methods and proposed SLMCD for comparison with $n = 100, p = 30, 100, 150$ and $N = 100$ replications.

Chapter 5

Data example

In this chapter, we applied our proposed SLMCD to real data analysis. We forecast the intensity of average electric current with 8mA and above collected by Kyolim soft coorporation. Kyolim soft coorpoartion collected electric-safety data from general households, traffic or street lights and local markets etc. which are electrically-vulnerable facilities in order to prevent accidential events that can occur by electric accidents.

Data was collected from 20th Sep to 28th Nov in 2021 and total number of date is 70 days, 10 weeks. For adequate data dimension, we transformed data to 15 min basis from 1 min basis by averaging. Therefore, total number of data per day is $24 \times 4 = 96$. Figure 5.1 describes the structure of data used for analysis.

Let the average electric current be Y_t with a implying week, b the day of the week (DOW) and k an order among 15-minute interval term within a day. Here, k indicates time interval from $15(k-1)$ min to $15k$ min. For example, $\{k = 1\}$ implies time from 00:00 am to 00:15 am and $\{k = 96\}$ implies time

	Date	DOW(b)	k	Y
1	2021-09-20	2	1	1.980078
2	2021-09-20	2	2	2.043701
3	2021-09-20	2	3	2.063529
4	2021-09-20	2	4	2.005906
5	2021-09-20	2	5	1.946032
6	2021-09-20	2	6	1.920866
7	2021-09-20	2	7	1.936471
8	2021-09-20	2	8	1.876471

Figure 5.1: Data preview including the intensity of average electric current with 8mA and above collected by Kyolim soft coorporation used in real data analysis.

from 11:45 pm to 12:00 am.

$$\begin{aligned}(Y_t)_{1\le t\le 96*70} &= (Y_{\{(a)-\text{week }(b)-\text{th day }(k)-\text{th interval}\}}) \\ &= (Y_{\{((a-1)*7+(b-1))*96+k\}}) \\ \text{(or simply)} \ &= (Y_{(a,b,k)}),\ 1\le a\le 10,\ 1\le b\le 7,\ 1\le k\le 96\end{aligned} \tag{5.1}$$

To measure the performance of the forecast, we split total 70 days data into training and testing sets and Table 5.1 describes notations used in data analysis for training and testing sets. First 63 days were used as training set and the rest 7 days (22rd to 28th of November 2021) were used as test set.

Also, $\mu^{(b,k)}$ and $\Sigma^{(b,k)}, \Omega^{(b,k)}$ of $(Z_{(b,k)}^{(test)}) = (Z_{[1:(k-1)]}^{(b)(test)}, Z_{[k:k]}^{(b)(test)})$ are partitioned along with $b = 1, ..., 7,\ k = 2, ..., 96$ as described in equation (5.1).

Notations	**Meaning**
$Y_{\text{training}},\ Y_{\text{test}}$	Training and test set
$N_{\text{training}}, N_{\text{test}}$	Number of observation in training, test set
$Y_t^{(\text{training})} = Y_{(a,b,k)}^{(\text{training})}$ $Y_t^{(\text{test})} = Y_{(b,k)}^{(\text{test})}$	$t-$th 15-min interval data for $t \in \{1, ..., N_{\text{training}} \text{ or } N_{\text{test}}\}$
$\{\mu_{(b,k)}\} = \{\sum_{a=1}^{9} Y_{(a,b,k)}^{(\text{training})}/7\}$	Mean values of $Y_t^{(\text{training})}$ for b , k
$Z_t^{(\text{training})} = \{Y_t^{(\text{training})} - \mu_{(b,k)}\}$	Centralized $Y_t^{(\text{training})}$ for DOW basis
$Z_t^{(\text{test})} = \{Y_t^{(test)} - \mu_{(b,k)}\}$	Centralized $Y_t^{(\text{test})}$ by training set for DOW basis

Table 5.1: Notations used in data analysis for training and testing sets.

$$\mu^{(b,k)} = \begin{pmatrix} \mu_{[1:k-1]}^{(b)} \\ \\ \mu_{[k:k]}^{(b)} \end{pmatrix}, \quad \Sigma^{(b,k)} = \begin{pmatrix} \Sigma_{[1:(k-1),1:(k-1)]}^{(b)} & \Sigma_{[1:(k-1)],k}^{(b)} \\ \\ \Sigma_{k,[1:(k-1)]}^{(b)} & \Sigma_{k,k}^{(b)} \end{pmatrix},$$
$$\Omega^{(b,k)} = \begin{pmatrix} \Omega_{[1:(k-1),1:(k-1)]}^{(b)} & \Omega_{[1:(k-1)],k}^{(b)} \\ \\ \Omega_{k,[1:(k-1)]}^{(b)} & \Omega_{k,k}^{(b)} \end{pmatrix} \tag{5.2}$$

Here, the best linear predictor of $Z_{(b,k)}^{(test)}$ from $Z_{[1:(k-1)]}^{(b)(test)}$ is described in equation (5.3).

$$\text{BLP}(Z_{(b,k)}^{(test)}) = \Sigma_{k,[1:(k-1)]}^{(b)} \Omega_{[1:(k-1),1:(k-1)]}^{(b)} Z_{[1:(k-1)]}^{(b)(test)} \tag{5.3}$$

To remove the weekly periodicity, we measured the weekly basis mean values $\{\mu_{(b,k)}\}_{b=1,...,7,\ k=1,...,96}$ of $Y_{training}$ for each b, k and subtracted calculated

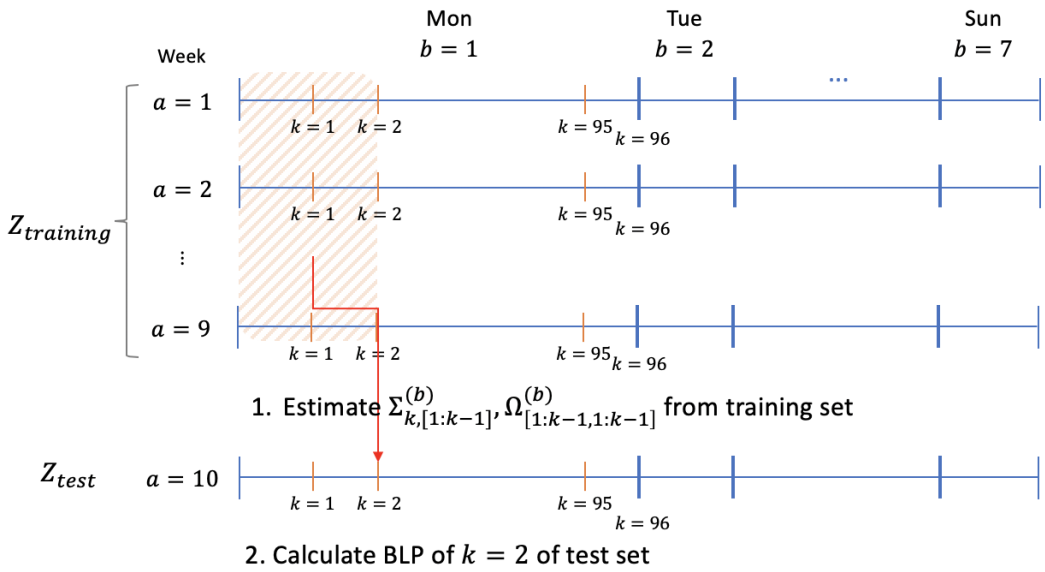


Figure 5.2: Explaining how data analysis was done by using training and test sets by best linear prediction method by using example of $k = 2$.

mean values from the original $Y_t^{(training)} = Y_{(a,b,k)}^{(training)}$, which is $Z_t^{(training)} = Z_{(a,b,k)}^{(training)}$ and also for test sets $Y_t^{(test)} = Y_{(b,k)}^{(test)}$, also subtracted mean values $\{\mu_{(b,k)}\}$ calculated by training sets to remove the estimated weekly periodicity.

For the estimation of $\Omega_{[1:(k-1),1:(k-1)]}$, $\Sigma_{k,[1:(k-1)]}$, following several lasso-based methods were used for each estimator. For the estimation of $\Omega_{[1:(k-1),1:(k-1)]}$, we compared our proposed method SLMCD, SLMCD-all, Huang et al. (2006), scaled lasso, GL, CL methods. Here, SLMCD-all estimates 96×96 precision matrix $\hat{\Omega}_{96} = \hat{L}_{96}^T \hat{D}_{96}^{-1} \hat{L}_{96}$ once and calculates $\{\hat{\Omega}_k\}_{k \leq 95}$ using the information of $\hat{\Omega}_{96}$. For the estimation of $\hat{\Sigma}_{k,[1:(k-1)]}$, we compared SLMCD, SLMCD-all, scaled lasso, Huang et al. (2006) and sample covariance. Here, sample covariance estimator was used for the estimating $\hat{\Sigma}_{k,[1:(k-1)]}$ in cases of GL and CL in order to reduce the computational cost.

We compared the performance by estimating average absolute forecast error of the test set. For the performance measure, we assumed that each DOW is independent to each other. So we calculated 7 different errors for each DOW, E_1 to E_7 as in equation (5.4).

$$\begin{aligned}
E_1 &= \frac{1}{96}\sum_{k=1}^{96}\left|\hat{Y}_{(1,k)}^{(\text{test})} - Y_{(1,k)}^{(\text{test})}\right| = \frac{1}{96}\sum_{k=1}^{96}\left|\hat{Z}_{(1,k)}^{(\text{test})} - Z_{(1,k)}^{(\text{test})}\right| \\
E_2 &= \frac{1}{96}\sum_{k=1}^{96}\left|\hat{Y}_{(2,k)}^{(\text{test})} - Y_{(2,k)}^{(\text{test})}\right| = \frac{1}{96}\sum_{k=1}^{96}\left|\hat{Z}_{(2,k)}^{(\text{test})} - Z_{(2,k)}^{(\text{test})}\right| \\
&\vdots \\
E_7 &= \frac{1}{96}\sum_{k=1}^{96}\left|\hat{Y}_{(7,k)}^{(\text{test})} - Y_{(7,k)}^{(\text{test})}\right| = \frac{1}{96}\sum_{k=1}^{96}\left|\hat{Z}_{(7,k)}^{(\text{test})} - Z_{(7,k)}^{(\text{test})}\right|
\end{aligned} \tag{5.4}$$

For total average error E_{total}, we averaged seven day error terms.

$$\begin{aligned}
E_{\text{total}} &= \frac{1}{7}\Big\{E_1 + \cdots + E_7\Big\} \\
&= \frac{1}{96*7}\sum_{b=1}^{7}\sum_{k=1}^{96}\left|\hat{Y}_{(b,k)}^{(test)} - Y_{(b,k)}^{(test)}\right| \\
&= \frac{1}{96*7}\sum_{b=1}^{7}\sum_{k=1}^{96}\left|\hat{Z}_{(b,k)}^{(test)} - Z_{(b,k)}^{(test)}\right|
\end{aligned} \tag{5.5}$$

Table 5.2 displays the results of comparing the performance of SLMCD with other lasso-based estimators across seven different datasets (DOWs) in terms of total average losses. Consistently, SLMCD, SLMCD-all, and Huang et al. (2006) estimator with ℓ_1 regularization demonstrate the smallest losses for the total average, showing their superiority over the other methods across all DOWs.

In Table 5.3, we present the standard deviations of losses calculated among various methods. These standard deviations provide insight into the variability of the estimation results for each method across the different datasets.

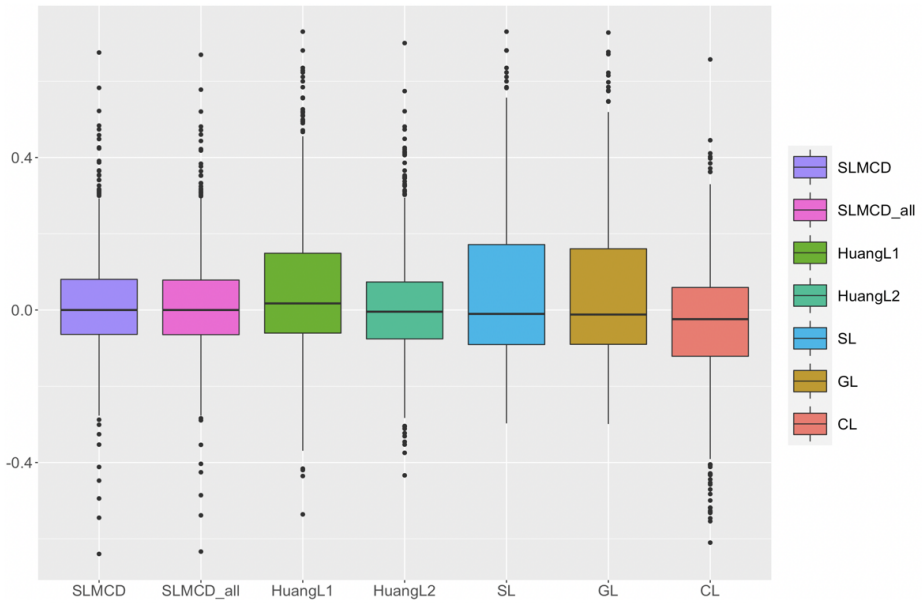
Additionally, Figure 5.3 illustrates the boxplots of losses calculated through the seven DOWs, comparing various lasso-based precision matrix estimation methods with SLMCD. The width of each boxplot represents the deviation, and a boxplot closer to zero indicates better performance of the estimator. As observed, SLMCD, SLMCD-all, and Huang et al. (2006) estimator with ℓ_1 regularization demonstrate the best performance based on these boxplots.

	SLMCD	SLMCD all	Huang L1 GCV	Huang L1 5CV	Huang L2 GCV	Huang L2 5CV	GL	CL
E1	0.103	0.104	0.146	0.149	0.111	0.104	0.159	0.121
E2	0.088	0.090	0.096	0.097	0.083	0.079	0.107	0.085
E3	0.099	0.099	0.129	0.130	0.109	0.106	0.142	0.121
E4	0.109	0.109	0.105	0.106	0.126	0.127	0.130	0.138
E5	0.127	0.150	0.151	0.152	0.126	0.127	0.144	0.191
E6	0.108	0.106	0.162	0.167	0.115	0.108	0.188	0.129
E7	0.100	0.101	0.147	0.146	0.113	0.107	0.165	0.134
Etotal	**0.105**	**0.109**	**0.134**	**0.135**	**0.112**	**0.108**	**0.148**	**0.132**

Table 5.2: Mean values of absolute errors calculated through 7 DOWs compared with various lasso-based precision matrix estimation method and SLMCD.

	SLMCD	SLMCD all	Huang L1 GCV	Huang L1 5CV	Huang L2 GCV	Huang L2 5CV	GL	CL
E1	0.107	0.105	0.145	0.146	0.145	0.146	0.148	0.106
E2	0.083	0.081	0.113	0.113	0.092	0.089	0.110	0.091
E3	0.096	0.094	0.137	0.138	0.109	0.104	0.134	0.107
E4	0.246	0.245	0.247	0.247	0.246	0.247	0.246	0.258
E5	0.137	0.264	0.140	0.140	0.105	0.113	0.118	0.146
E6	0.098	0.096	0.149	0.150	0.117	0.108	0.146	0.108
E7	0.091	0.090	0.137	0.138	0.108	0.103	0.131	0.129
Etotal	**0.132**	**0.157**	**0.156**	**0.157**	**0.133**	**0.132**	**0.152**	**0.146**

Table 5.3: Standard deviations of absolute errors calculated through 7 DOWs compared with various lasso-based precision matrix estimation method and SLMCD.



	SLMCD	SLMCD_all	HuangL1	HuangL2	SL	GL	CL
E Total	**0.1015**	**0.1005**	0.1319	0.1049	0.1469	0.1445	0.1285
s.d.	**0.0979**	**0.0965**	0.1291	0.0984	0.1263	0.1236	0.1179

Figure 5.3: Boxplot of losses calculated through 7 DOWs compared with various lasso-based precision matrix estimation method and SLMCD.

Chapter 6

Conclusion

In this thesis, we developed a precision matrix estimation method in situations where the data is time-ordered and precision matrix achieved a certain sparsity. In our proposed estimator, the methods and advantages of MCD and scaled lasso by Sun and Zhang (2012) were combined and applied.

We demonstrated that our proposed SLMCD achieves a certain convergence rate of $O_p(\sqrt{s \log p/n})$, which is optimal convergence rate. Also, our proposed method has a computational efficiency by using a pre-fixed tuning parameter by applying the algorithm of Sun and Zhang (2012).

In numerical studies with various n, p situations and precision matrix settings, we successfully demonstrated that SLMCD provides a reasonable loss in estimating the true precision matrix, within the lowest computational costs compared to other lasso-based methods. Also, we proved the utility of SLMCD by the real data analsis analyzing the electric-safety data.

6.1 Discussion

In the thesis, we focused on estimating the precision matrix using MCD method, which decomposes Ω as $\Omega = L^\top D^{-1} L$. Among the components L, D of Ω, our algorithm estimates L by utilinzing scaled lasso method and computes D, which is composed of error variances by sample error variance calculations. In this discussion, we will mention the importance of estimating the error variances that compose diagonals of D.

Methods for estimating the error variance such as RCV (Friedman et al., 2008), MM (Dicker, 2014), MLE (Dicker and Erdogdu, 2016), EigenPrism (Janson et al., 2017), and RidgeVar (Liu et al., 2020) will be introduced briefly in Appendix B. Each of these methods offers a unique approach to estimate the error variance and plays a crucial role in robust covariance or precision matrix estimation. By briefly introducing these methods, we aim to provide insights into the various approaches available for handling this important aspect of MCD-based estimation.

Appendix A

Appendix

A.1 Running time of R codes

Table A.1: Running time of R codes are calculated for cases of Huang et al. (2006) and Sun and Zhang (2013) with some lasso-based precision matrix estimation methods and our method SLMCD in comparison with $n = 100, p = 30, 100, 150$ and $N = 100$ replications.

(i) p=30

	Ω_1	Ω_2	Ω_{31}	Ω_{32}	Ω_4	Ω_5	Ω_6	Ω_7
SLMCD	**0.11**	**0.09**	**0.09**	0.12	**0.10**	**0.11**	**0.08**	**0.10**
Huang L_1	142.11	86.72	264.95	269.71	176.58	268.92	131.73	282.53
Huang L_2	199.07	152.09	191.40	187.31	165.05	261.11	134.50	279.28
Scaled Lasso	**0.09**	**0.10**	**0.11**	**0.09**	**0.09**	**0.11**	**0.10**	**0.12**
GL	0.14	0.14	0.12	**0.11**	0.12	0.12	0.12	**0.12**
CL	34.71	31.77	28.22	26.70	30.26	31.78	29.02	31.92

(ii) p=100

	Ω_1	Ω_2	Ω_{31}	Ω_{32}	Ω_4	Ω_5	Ω_6	Ω_7
SLMCD	**0.41**	**0.40**	**0.49**	**0.50**	**0.42**	**0.52**	**0.54**	**0.52**
Huang L_1	383.57	400.60	4623.90	204.59	392.82	386.48	387.33	381.88
Huang L_2	598.50	803.99	597.38	556.73	417.65	486.55	483.54	558.02
Scaled Lasso	**1.28**	**1.50**	**1.51**	**1.58**	**1.27**	**1.54**	**1.53**	**1.58**
GL	15.32	15.23	15.24	15.35	15.23	15.10	14.88	14.71
CL	1620.00	1385.65	1068.91	711.79	1549.46	1510.90	1316.73	1379.98

(iii) p=150

	Ω_1	Ω_2	Ω_{31}	Ω_{32}	Ω_4	Ω_5	Ω_6	Ω_7
SLMCD	**1.43**	**1.32**	**1.47**	**1.42**	**1.34**	**1.48**	**1.58**	**1.47**
Huang L_1	297.10	327.54	2983.27	3260.04	639.94	614.65	598.17	613.18
Huang L_2	1035.76	1976.27	1826.34	1688.03	14870.45	1166.95	1152.45	1321.95
Scaled Lasso	**3.91**	**5.26**	**4.50**	**4.62**	**4.04**	**4.41**	**4.29**	**4.59**
GL	18.65	18.58	18.66	18.65	17.76	18.52	18.49	18.40
CL	7929.57	7119.63	4882.25	2835.36	7794.85	8154.80	6302.52	6998.25

Throughout $p = 30, 100, 150$, SLMCD performs the shortest running time of R codes. As p increases, SLMCD performs with the shorter running time compared to other methods, which is the strong point of SLMCD.

A.2 Ratio of zeros/non-zeros identified among zeros/non-zeros of Ω

A.2.1 p=30

Sparse cases

(i) Identifying zeros / zeros of Ω (0-0)

	Ω_1	Ω_2	Ω_{31}	Ω_{32}	Ω_6
SLMCD	0.97	0.97	0.93	0.94	0.98
Huang L_1	1.00	1.00	0.78	0.72	1.00
Huang L_2	1.00	1.00	1.00	1.00	1.00
Scaled Lasso	0.98	0.14	0.83	0.97	0.98
GL	1.00	1.00	1.00	1.00	1.00
CL	0.00	0.00	0.00	0.00	0.00

Table A.2: Ratios of identifying zeros of true zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 30$ and $N = 100$ replications.

(ii) Identifying non-zeros / zeros of Ω (1-0)

	Ω_1	Ω_2	Ω_{31}	Ω_{32}	Ω_6
SLMCD	0.03	0.03	0.07	0.06	0.02
Huang L_1	0.00	0.00	0.22	0.28	0.00
Huang L_2	0.00	0.00	0.00	0.00	0.00
Scaled Lasso	0.02	0.86	0.17	0.03	0.02
GL	0.00	0.00	0.00	0.00	0.00
CL	1.00	1.00	1.00	1.00	1.00

Table A.3: Ratios of identifying non-zeros of true zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 30$ and $N = 100$ replications.

Dense cases

(i) Identifying non-zeros / non-zeros of Ω (1-1)

	Ω_4	Ω_5	Ω_7
SLMCD	0.06	0.15	0.14
Huang L_1	0.04	0.44	0.71
Huang L_2	0.03	0.03	0.03
Scaled Lasso	0.04	0.31	0.66
GL	0.03	0.03	0.03
CL	1.00	1.00	1.00

Table A.4: Ratios of identifying non-zeros of true non-zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 30$ and $N = 100$ replications.

(ii) Identifying zeros / non-zeros of Ω (0-1)

	Ω_4	Ω_5	Ω_7
SLMCD	0.94	0.85	0.86
Huang L_1	0.96	0.56	0.29
Huang L_2	0.97	0.97	0.97
Scaled Lasso	0.96	0.69	0.34
GL	0.97	0.97	0.97
CL	0.00	0.00	0.00

Table A.5: Ratios of identifying zeros of true non-zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 30$ and $N = 100$ replications.

A.2.2 p=100

Sparse cases

(i) Identifying zeros / zeros of Ω (0-0)

	Ω_1	Ω_2	Ω_{31}	Ω_{32}	Ω_6
SLMCD	0.99	0.99	0.99	0.99	0.95
Huang L_1	1.00	1.00	0.94	0.97	0.95
Huang L_2	1.00	1.00	1.00	1.00	1.00
Scaled Lasso	0.99	0.05	0.86	0.97	0.82
GL	1.00	1.00	1.00	1.00	1.00
CL	0.00	0.00	0.00	0.00	0.00

Table A.6: Ratios of identifying zeros of true zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 100$ and $N = 100$ replications.

(ii) Identifying non-zeros / zeros of Ω (1-0)

	Ω_1	Ω_2	Ω_{31}	Ω_{32}	Ω_6
SLMCD	0.01	0.01	0.01	0.01	0.05
Huang L_1	0.00	0.00	0.06	0.03	0.05
Huang L_2	0.00	0.00	0.00	0.00	0.00
Scaled Lasso	0.01	0.95	0.14	0.03	0.18
GL	0.00	0.00	0.00	0.00	0.00
CL	1.00	1.00	1.00	1.00	1.00

Table A.7: Ratios of identifying non-zeros of true zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 100$ and $N = 100$ replications.

Dense cases

(i) Identifying non-zeros / non-zeros of Ω (1-1)

	Ω_4	Ω_5	Ω_7
SLMCD	0.02	0.04	0.04
Huang L_1	0.01	0.03	0.02
Huang L_2	0.01	0.01	0.01
Scaled Lasso	0.01	0.17	0.54
GL	0.01	0.01	0.01
CL	1.00	1.00	1.00

Table A.8: Ratios of identifying non-zeros of true non-zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 100$ and $N = 100$ replications.

(ii) Identifying zeros / non-zeros of Ω (0-1)

	Ω_4	Ω_5	Ω_7
SLMCD	0.98	0.96	0.96
Huang L_1	0.99	0.97	0.98
Huang L_2	0.99	0.99	0.99
Scaled Lasso	0.99	0.83	0.46
GL	0.99	0.99	0.99
CL	0.00	0.00	0.00

Table A.9: Ratios of identifying zeros of true non-zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 100$ and $N = 100$ replications.

A.2.3 p=150

Sparse cases

(i) Identifying zeros / zeros of Ω (0-0)

	Ω_1	Ω_2	Ω_{31}	Ω_{32}	Ω_6
SLMCD	0.997	0.996	0.988	0.990	0.976
Huang L_1	1.000	1.000	1.000	0.969	0.997
Huang L_2	1.000	1.000	1.000	1.000	1.000
Scaled Lasso	0.996	0.030	0.861	0.979	0.886
GL	1.000	1.000	1.000	1.000	1.000
CL	0.000	0.000	0.000	0.000	0.000

Table A.10: Ratios of identifying zeros of true zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 150$ and $N = 100$ replications.

(ii) Identifying non-zeros / zeros of Ω (1-0)

	Ω_1	Ω_2	Ω_{31}	Ω_{32}	Ω_6
SLMCD	0.003	0.004	0.012	0.010	0.024
Huang L_1	0.000	0.000	0.000	0.031	0.003
Huang L_2	0.000	0.000	0.000	0.000	0.000
Scaled Lasso	0.004	0.970	0.139	0.021	0.114
GL	0.000	0.000	0.000	0.000	0.000
CL	1.000	1.000	1.000	1.000	1.000

Table A.11: Ratios of identifying non-zeros of true zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 150$ and $N = 100$ replications.

Dense cases

(i) Identifying non-zeros / non-zeros of Ω (1-1)

	Ω_4	Ω_5	Ω_7
SLMCD	0.011	0.024	0.027
Huang L_1	0.007	0.014	0.009
Huang L_2	0.007	0.007	0.007
Scaled Lasso	0.008	0.130	0.559
GL	0.007	0.007	0.007
CL	1.000	1.000	1.000

Table A.12: Ratios of identifying non-zeros of true non-zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 150$ and $N = 100$ replications.

(ii) Identifying zeros / non-zeros of Ω (0-1)

	Ω_4	Ω_5	Ω_7
SLMCD	0.989	0.976	0.973
Huang L_1	0.993	0.986	0.991
Huang L_2	0.993	0.993	0.993
Scaled Lasso	0.992	0.870	0.441
GL	0.993	0.993	0.993
CL	0.000	0.000	0.000

Table A.13: Ratios of identifying zeros of true non-zeros in Ω with SLMCD to some lasso-based precision matrix estimation methods for comparison with $n = 100, p = 150$ and $N = 100$ replications.

Appendix B

Appendix

B.1 Refitted cross-validation method (RCV)

Fan et al. (2012) proposed a two-stage refitted cross-validation (RCV) method for ultrahigh dimensional models. The method initially identifies the appropriate model using a specific fitting method and then calculates the error variance using ordinary least squares (OLS) only on the variables selected in the model. When the number of variables (p) is small, even if some irrelevant variables are selected, the error variance may not be significantly biased. However, in high-dimensional cases where p is large, the likelihood of including many irrelevant variables in the model increases. As a result, the error variance estimate ($\hat{\sigma}^2$) can be substantially biased.

$$\hat{\sigma}^2 = (1 - \gamma_n^2)\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2,$$

where γ_n^2 represents the sample correlation between the spurious variable and

the response variable. As more irrelevant variables are included, the term γ_n^2 tends to decrease, leading to a downward bias in the estimated error variance.

B.2 Method of moments method (MM)

Dicker (2014) proposed the method of moments (MM) estimator for error variance. By leveraging the method of moments properties, Dicker derived non-degenerate linear-combination forms of certain statistics that provide estimates for the quantities of interest, such as σ^2 and $\tau^2 = \beta^T\Sigma\beta = |\Sigma^{1/2}\beta|^2$, which represents the measure of l^2-signal strength. In cases where the number of variables (p) is less than the number of observations (n), estimating σ^2 and τ^2 via the MM method is straightforward. The estimates are as follows:

$$\hat{\sigma}_0^2 = \frac{1}{n-p}\|y - X\hat{\beta}\|^2 = \frac{1}{n-p}\|y\|^2 - \frac{1}{n-p}y^T X(X^T X)^{-1}X^T y$$

$$\hat{\tau}_0^2 = \frac{1}{n}\|y\|^2 - \hat{\sigma}_0^2 = \frac{1}{n-p}y^T X(X^T X)^{-1}X^T y - \frac{p}{n(n-p)}\|y\|^2$$

where $\hat{\sigma}_0^2$, $\hat{\tau}_0^2$ are unbiased estimators.

However, in cases where the number of variables (p) exceeds the number of observations (n), the inverse of $X^T X$ does not exist, presenting challenges for estimation. In such situations, Dicker (2014) suggested a different MM method that is applicable in $p > n$ cases, without relying on any sparsity assumptions on β. This approach utilizes properties of the moments of the normal and Wishart distribution to derive feasible estimators for σ^2 and τ^2 under high-dimensional settings.

B.3 Maximum likelihood method (MLE)

Dicker and Erdogdu (2016) introduced an error variance estimator via maximum likelihood (ML) under the fixed-effects model, expanding on the ML estimator that was already suggested for the random-effects model. The proposed ML estimator for error variance was initially designed for the random-effects model, but Dicker and Erdogdu (2016) extended its application to the fixed-effects model as well. They accomplished this by utilizing coupling arguments to connect the random-effects model to the fixed-effects model when p and n are large.

B.4 EigenPrism

Janson et al. (2017) proposed the EigenPrism method to estimate confidence intervals for the error variance of $|\Sigma^{1/2}\beta|_2^2$ (or θ). This method helps answer the question of how close the estimated coefficient vector $\hat{\beta}$ is to the true coefficient vector β. Additionally, EigenPrism addresses the estimation of the noise level σ^2 in a linear model, which is crucial for prediction accuracy and model selection.

B.5 RidgeVar

Liu et al. (2020) proposed an error variance estimator based on ridge regression and random matrix theory.

B.6 The natural lasso estimator

Yu and Bien (2019) proposed the natural lasso estimator for the error variance, with maximizing the penalized likelihood function by using the natural parameters of the Gaussian multi-parameter exponential family, $\phi = \frac{1}{\sigma^2}$, $\theta = \frac{\beta}{\sigma^2}$. Also, the organic lasso estimator was proposed which uses the modified penalty from the natural lasso estimator. Especially, the proposed method does not require any assumptions on design matrix X or regression coefficients β, which is a strong-point among the estimation methods of error variance.

B.7 The adpative method

Verzelen and Gassiat (2018) proposed the adaptive method for estimating the error variance, with the impact of not knowing the sparsity of the regression parameter or the distribution of the design matrix. This method suggested the adaption to unknown sparsity when Σ is known by suggesting $\hat{\eta}(\hat{\Sigma}^{-1})$ as U-type estimator with $\sqrt{p}/n$-consistent. The idea is that if the real regressionn parameter is sparse, then the suggested method and the previously known k-sparse estimator will be close enough, and if dense then two will be quite different.

Bibliography

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1), 199–227.

Bickel, P. J., Ritov, Y. and Tsybakov, A. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics* 37, 1705–1732.

Cai, T. T., Liu, W. and Luo, X. (2011). A Constrained l_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494), 594-607.

Dicker, L. (2014). Variance estimation in high-dimensional linear models. *Biometrika*, 101(2), 269-284.

Dicker, L. and Erdogdu, M. (2016). Maximum Likelihood for Variance Estimation in High-Dimensional Linear Models. *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, 41, 159-167.

Fan, J., Guo, S. and Hao, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *Journal of the Royal Statistical Society. Series B*, 74(1), 37-65.

Friedman, J., Hastie, J. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), 432-441.

van de Geer, S. A. and Bühlmann, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electronic Journal of Statistics*, 3, 1360-1392.

Gill, P. E. and Murray, W. (1974). Newton-type methods for unconstrained and linearly constrainedoptimization. *Math. Programming*, 28, 311-350.

Huang, Z., Liu, N., Pourahmadi, M. and Liu, L. (2006). Covariance selection and estimation via penalised normal likelihood. *Biometrika*, 93(1), 85-98.

Janson, L., Barber, R. F., Candès, E. (2017). EigenPrism: inference for high dimensional signal-to-noise ratios. *J R Stat Soc Series B Stat Methodol*, 79(4), 1037-1065.

Jiang, X. (2012). Sparse estimation of large covariance matrices. *National University of Singapore*.

Kang, X., Xie, C., Wang, M. (2020). A Cholesky-based estimation of large-dimensional covariance matrices. *Journal of applied statistics*, 47(6), 1017-1030.

Kang, X., Deng, X. (2020). An improved modified cholesky decomposition approach for precision matrix estimation. *Journal of Statistical Computation and Simulation*, 90(3), 443-464.

Kang, X., Deng, X., Tsui, K., Pourahmadi, M. (2020). On variable ordination of modified cholesky decomposition for estimating time-varying covariance matrices. *International Statistical Review*, 90(3), 443-464.

Kuhn, H. W. and Tucker, A. W. (1951). "Nonlinear programming". Proceedings of 2nd Berkeley Symposium. *Berkeley: University of California Press*, 481–492.

Lee, K. and Lee, J. (2021). Estimating large covariance matrices via modified cholesky decomposition. *Statistica Sinica*, 31(1), 173-196.

Levina, E., Rothman, A. and Zhu, J. (2008). Sparse estimation of large covariance matrices. *The Annals of Applied Statistics*, 88(3), 616-641.

Liu, X., Zheng, S. and Feng, X. (2020). Estimation of error variance via ridge regression. *Biometrika*, 107(2), 481-488.

Pourahmadi, M. (2013). High-dimensional covariance estimation. *Wiley, New York*.

Rothman, A., Levina, E. and Zhu, J. (2008). A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika*, 97(3), 539-550.

Rothman, A., Bickel, P., Levina, E. and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2, 494-515.

Sun, T. and Zhang, C.-H. (2012). Scaled sparse linear regression. *Biometrika*, 99, 879–898.

Sun, T. and Zhang, C.-H. (2013). Sparse Matrix Inversion with Scaled Lasso. *Journal of Machine Learning*, 14, 3385-3418.

Verzelen, N. and Gassiat, E. (2018). Adaptive estimation of high-dimensional signal-to-noise ratios. *Bernoulli*, 24(4B), 3683-3710.

Wu, W. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, 90(4), 831-844.

Yu, G. and Bien, J. (2019). Estimating the error variance in a high-dimensional linear model. *Biometrika*, 106(3), 533–546.

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research*, 11, 2261–2286.

국문초록

수정 촐레스키 분해 기반의 스케일드 라쏘 기법을 활용한 역공분산 행렬 추정

본 논문은 시간 순서를 가진 고차원 다변량 자료의 역공분산 행렬 Ω 추정의 새로운 방법론을 제시한다. 역공분산 행렬은 수정된 촐레스키 분해법을 활용하면 $\Omega = L^T D^{-1} L$로 분해가 가능하다. 본 논문에서는 Sun and Zhang (2012)의 스케일드 라쏘방법을 활용한 MCD 기반의 역공분산행렬 추정량 SLMCD를 제안한다. 추정량 SLMCD는 몇 가지 이점을 가지고 있다. 먼저, 스케일드 라쏘 방법론을 적용시킴으로써 사전에 결정된 조절변수를 사용하게 되어 조절변수의 계산비용에 큰 이점이 있다. 두번째로, 적절한 조건 하에서 SLMCD 추정량이 $O_p\big(\sqrt{s \log p/n}\big)$의 최적 수렴 속도를 달성한다. 여기서 s는 L의 0이 아닌 원소의 개수를 의미한다. 마지막으로, 다양한 상황에서 SLMCD 추정량이 다른 라쏘 기반의 추정량들과 비교하여 훌륭한 추정 결과를 가져오는 것을 수치적으로 보여준다. 또한, 교림소프트 사에서 제공한 전기 안전 데이터를 활용하여 실제 데이터 분석을 수행한다.

주요어: 역공분산 행렬, 고차원, 수정 촐레스키 기법, 스케일드 라쏘, 전기 안전 데이터

학번: 2016-20271