Forecast Combination Using High-Dimensional Precision Matrix*

Tae-Hwy Lee† Millie Yi Mao‡ Aman Ullah§

Abstract

The estimation of a large covariance matrix is challenging when the dimension p is large relative to the sample size n. Common approaches to deal with the challenge have been based on thresholding or shrinkage methods in estimating large covariance matrices. We examine the combined forecasts based on the ISEE estimator of Fan and Lv (2016) and compare it with those based on the thresholding and shrinkage methods.

Key Words: High-dimensional precision matrix, ISEE, Forecast combination puzzle.

1. Introduction

In many applications of multivariate statistical analysis, such as forecast combination, the estimation of high-dimensional covariance matrices is a challenging issue especially when the number of random variables is larger than the number of observations. There are two popular regularization techniques used in the literature to overcome the challenge – shrinkage and thresholding methods. See Ledoit and Wolf (2004), Bickel and Levina (2008), Cai and Liu (2011) and Bailey, Pesaran and Smith (2018).

In most applications however, what we need is not a covariance matrix but its inverse, which is known as a precision matrix. Therefore another challenge is to invert a covariance matrix to obtain the precision matrix. Even more challenging is when a covariance matrix is high-dimensional as it may be computationally heavy to invert or infeasible to invert.

To overcome the difficulty one may directly estimate precision matrices rather than indirectly from inverting the covariance matrices. Fan and Lv (2016) propose a method called the 'innovated scalable efficient estimation' (ISEE) for the direct estimation of a large precision matrix through linear transformation, which bypasses inverting a large covariance matrix. We examine the efficiency of the ISEE estimator in comparison with that of several shrinkage methods and thresholding methods by Monte Carlo simulations, which show the advantage of the ISEE estimator in resolving the "forecast combination puzzle".

This paper is organized as follows. In Section 2, we introduce the estimation algorithm of high-dimensional precision matrices. In Section 3, we study forecast combination using the ISEE estimator. Section 4 concludes.

2. The ISEE Algorithm

Consider a p-variate random vector

$$\mathbf{x} = (X_1, \ldots, X_p)' \sim N(\boldsymbol{\mu}, \Sigma) \tag{1}$$

where $\boldsymbol{\mu}$ is a p-dimensional mean vector, $\Sigma = (\sigma_{jk})$ is a $p \times p$ covariance matrix. Define the precision matrix as $\Omega = (\omega_{jk})$, the inverse Σ^{-1} of the covariance matrix Σ. Assume

*We thank the seminar participants at JSM2019 in Denver.

†Department of Economics, University of California, Riverside, CA 92521. E-mail: taelee@ucr.edu

‡Department of Mathematics, Physics and Statistics, Azusa Pacific University, CA 91702. E-mail: ymao004@ucr.edu

§Department of Economics, University of California, Riverside, CA 92521. E-mail: aman.ullah@ucr.edu

the mean vector $\boldsymbol{\mu} = \mathbf{0}$ without loss of generality. Throughout this paper, X represents a random variable, $\mathbf{x}$ represents a vector of the random variables, and $\mathbf{X}$ represents a data matrix.

In most applications (in statistical estimation, forecast combination, optimal portfolio estimation, etc), what we need is not the covariance matrix Σ but its inverse $\Omega = \Sigma^{-1}$ (the precision matrix). For example, consider X_j $(j = 1, \ldots, p)$ as a forecast from model j. The optimal combination $w'\mathbf{x}$ of the p forecasts in $\mathbf{x}$ can be obtained with the optimal weight given by

$$w = \frac{\Sigma^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\Sigma^{-1}\boldsymbol{\iota}} \tag{2}$$

where $\boldsymbol{\iota} = (1, ..., 1)'$ is a $p \times 1$ vector of ones. See Bates and Granger (1969), Stock and Watson (2004), and Timmermann (2006). Note that the precision matrix $\Omega = \Sigma^{-1}$ is needed instead of the covariance matrix Σ. Nevertheless, the literature is largely about estimation of Σ rather than $\Omega = \Sigma^{-1}$.

For any subsets $A, B \subset \{1, ..., p\}$, denote $\mathbf{x}_A$ a subvector of $\mathbf{x}$ formed by its components with indices in A, and $\Omega_{A,B} = (\omega_{jk})_{j\in A,k\in B}$ a submatrix of Ω with rows in A and columns in B. Denote the cardinality of the set A by $|A|$. In this paper we make $|A| = 2$ when the number of nodes p is even and $|A| = 2$ or 3 when p is an odd number.

Inverting the sample covariance matrix is difficult or infeasible. To avoid this problem, Fan and Lv (2016) suggest a new approach – the innovated scalable efficient estimation (ISEE) of large precision matrices based on the following linear transformation

$$\mathbf{z} = \Omega\mathbf{x}, \tag{3}$$

where the mean and variance of $\mathbf{z}$ are

$$\begin{aligned} E(\mathbf{z}) &= E(\Omega\mathbf{x}) = 0, \\ \text{COV}(\mathbf{z}) &= \text{COV}(\Omega\mathbf{x}) = \Omega\text{COV}(\mathbf{x})\Omega = \Omega\Sigma\Omega = \Omega. \end{aligned} \tag{4}$$

If the transformed vector $\mathbf{z}$ can be obtained, then estimating the precision matrix Ω is equivalent to estimating the covariance matrix of $\mathbf{z}$. Obtaining $\mathbf{z}$ by the two parts Ω and $\mathbf{x}$ is not feasible since it depends on the unknown precision matrix Ω. Instead, Fan and Lv (2016) break the long vector $\mathbf{z}$ into small subvectors with each subvector corresponding to a partition of the index set $\{1, ..., p\}$.

For any subset $A \subset \{1, ..., p\}$, write $\mathbf{z} = \Omega\mathbf{x}$ in partition

$$\begin{pmatrix} \mathbf{z}_A \\ \mathbf{z}_{A^c} \end{pmatrix} = \begin{pmatrix} \Omega_{A,A} & \Omega_{A,A^c} \\ \Omega_{A^c,A} & \Omega_{A^C,A^C} \end{pmatrix} \begin{pmatrix} \mathbf{x}_A \\ \mathbf{x}_{A^c} \end{pmatrix} = \begin{pmatrix} \Omega_{A,A}\mathbf{x}_A + \Omega_{A,A^c}\mathbf{x}_{A^c} \\ \Omega_{A^c,A}\mathbf{x}_A + \Omega_{A^C,A^C}\mathbf{x}_{A^c} \end{pmatrix}, \tag{5}$$

with A^c denoting the complement of the subset A, to obtain

$$\mathbf{z}_A = \Omega_{A,A}\mathbf{x}_A + \Omega_{A,A^c}\mathbf{x}_{A^c} = \Omega_{A,A}\left(\mathbf{x}_A + \Omega_{A,A}^{-1}\Omega_{A,A^c}\mathbf{x}_{A^c}\right) \equiv \Omega_{A,A}\mathbf{e}_A, \tag{6}$$

where

$$\mathbf{e}_A \equiv \mathbf{x}_A + \Omega_{A,A}^{-1}\Omega_{A,A^c}\mathbf{x}_{A^c}. \tag{7}$$

Under the assumption of the normality of $\mathbf{x}$, $\mathbf{z} \sim N(0, \Omega)$. The subset $\mathbf{z}_A = \Omega_{A,A}\mathbf{e}_A \sim N(0, \Omega_{A,A})$, which means $\mathbf{e}_A \sim N\left(0, \Omega_{A,A}^{-1}\right)$. Note that

$$E(\mathbf{e}_A|\mathbf{x}_{A^c}) = E\left(\mathbf{x}_A + \Omega_{A,A}^{-1}\Omega_{A,A^c}\mathbf{x}_{A^c}|\mathbf{x}_{A^c}\right) = E(\mathbf{x}_A|\mathbf{x}_{A^c}) + \Omega_{A,A}^{-1}\Omega_{A,A^c}\mathbf{x}_{A^c} = 0$$

which implies that the conditional mean of $\mathbf{x}_A$ is

$$E\left(\mathbf{x}_A|\mathbf{x}_{A^c}\right) = -\Omega_{A,A}^{-1}\Omega_{A,A^c}\mathbf{x}_{A^c}.$$

The conditional covariance of $\mathbf{x}_A$ is

$$V\left(\mathbf{x}_A|\mathbf{x}_{A^c}\right) = E\left[\left(\mathbf{x}_A - E\left(\mathbf{x}_A|\mathbf{x}_{A^c}\right)\right)'\left(\mathbf{x}_A - E\left(\mathbf{x}_A|\mathbf{x}_{A^c}\right)\right)|\mathbf{x}_{A^c}\right] = E\left(\mathbf{e}_A'\mathbf{e}_A|\mathbf{x}_{A^c}\right) = \Omega_{A,A}^{-1}.$$

The last equality holds because $\mathbf{e}_A$ and $\mathbf{x}_{A^c}$ are independent.[1] Hence, the conditional distribution of $\mathbf{x}_A$ given $\mathbf{x}_{A^c}$ is

$$\mathbf{x}_A|\mathbf{x}_{A^c} \sim N\left(-\Omega_{A,A}^{-1}\Omega_{A,A^c}\mathbf{x}_{A^c}, \Omega_{A,A}^{-1}\right). \tag{8}$$

Thus, we can obtain $\mathbf{e}_A$ as the error term from the linear regression of $\mathbf{x}_A$ on $\mathbf{x}_{A^c}$. Accordingly, the multivariate linear regression of $\mathbf{x}_A$ on $\mathbf{x}_{A^c}$ has the form

$$\mathbf{x}_A = C_A^T\mathbf{x}_{A^c} + \mathbf{e}_A,$$

where $C_A = -\Omega_{A^c,A}\Omega_{A,A}^{-1}$ represents the coefficient matrix and $\mathbf{e}_A$ is the vector of regression errors.

In matrix form, regress a submatrix $\mathbf{X}_A$ on the rest of the data $\mathbf{X}_{A^c}$

$$\mathbf{X}_A = \mathbf{X}_{A^c}C_A + \mathbf{E}_A,$$

where $\mathbf{X}_A$ and $\mathbf{X}_{A^c}$ are submatrices of $\mathbf{X}$ with columns in A and its complements A^c, C_A is the regression coefficient matrix and $\mathbf{E}_A$ is an $n \times |A|$ matrix of errors. For each node $j \in A$, Fan and Lv (2016) consider the univariate linear regression model for response $\mathbf{X}_j$, which is the jth column of data matrix $\mathbf{X}$

$$\underset{n\times 1}{\mathbf{X}_j} = \underset{n\times(p-|A|)}{\mathbf{X}_{A^c}}\ \underset{(p-|A|)\times 1}{\beta_j} + \underset{n\times 1}{\mathbf{E}_j}$$

which is estimated by the penalized least squares with the scaled Lasso

$$\left(\hat{\beta}_j, \hat{\theta}_j^{1/2}\right) = \underset{\beta_j\in\mathbb{R}^{p-|A|},\sigma>0}{\arg\min}\left\{\frac{\|\mathbf{X}_j - \mathbf{X}_{A^c}\beta_j\|_2^2}{2n\sigma} + \frac{\sigma}{2} + \lambda||\beta_*||_1\right\}, \tag{9}$$

where β_* is the Hadamard (component-wise) product of two $(p-|A|)$-dimensional vectors β_j and $\left(n^{-1/2}||\mathbf{X}_k||_2\right)_{k\in A^c}$ with $\mathbf{X}_k$ the kth column of $\mathbf{X}$, $\lambda \geq 0$ is a regularization parameter associated with the weighted L_1-penalty, and $\|v\|_q$ denotes the L_q-norm of a given vector v for $q \geq 1$.

Based on the regression step, for each node j in the index set A, define

$$\begin{aligned}\hat{\mathbf{E}}_j &= \mathbf{X}_j - \mathbf{X}_{A^c}\hat{\beta}_j\\ \hat{\mathbf{E}}_A &= (\hat{\mathbf{E}}_j)_{j\in A}.\end{aligned}$$

Then $\Omega_{A,A}$ and $\hat{\mathbf{Z}}_A$ are estimated by

$$\hat{\Omega}_{A,A} = (n^{-1}\hat{\mathbf{E}}_A'\hat{\mathbf{E}}_A)^{-1}, \tag{10}$$

[1] $\text{COV}(\mathbf{e}_A, \mathbf{x}_{A^c}) = \text{COV}\left(\mathbf{x}_A + \Omega_{A,A}^{-1}\Omega_{A,A^c}\mathbf{x}_{A^c}, \mathbf{x}_{A^c}\right) = \Sigma_{A,A^c} + \Omega_{A,A}^{-1}\Omega_{A,A^c}\Sigma_{A^c,A^c}$. Based on the property of the inverse of a partitioned matrix, $\Sigma_{A,A^c} = -\Omega_{A,A}^{-1}\Omega_{A,A^c}\Sigma_{A^c,A^c}$. Thus, $\text{COV}(\mathbf{e}_A, \mathbf{x}_{A^c}) = 0$. Since $\mathbf{e}_A$ and $\mathbf{x}_{A^c}$ have joint normality, they are independent.

The unobservable submatrix $\mathbf{Z}_A$ is estimated by

$$\hat{\mathbf{Z}}_A = \hat{\mathbf{E}}_A \hat{\Omega}_A.$$

Stacking $\hat{\mathbf{Z}}_A$ for all the partitions A's, the ISEE estimates the empirical matrix $\hat{\mathbf{Z}}$ as the $n \times p$ matrix

$$\underset{n\times p}{\hat{\mathbf{Z}}} = \left(\hat{\mathbf{Z}}_A\right)_{\forall A} \tag{11}$$

Therefore, the initial ISEE estimator of the precision matrix of $\mathbf{X}$ is the sample covariance matrix of $\hat{\mathbf{Z}}$, which is computed as

$$\hat{\Omega}_{ISEE,ini} = n^{-1}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}. \tag{12}$$

Remark: In Fan and Lv (2016), they refine the initial ISEE estimator by thresholding. For a given threshold $\tau \geq 0$, define the new estimator with thresholding as

$$\hat{\Omega}_{ISEE,g} = T_\tau\left(\hat{\Omega}_{ISEE,ini}\right),$$

where $T_\tau(B) = \left(b_{jk}1_{\{|b_{jk}|\geq\tau\}}\right)$ denotes the matrix $B = (b_{jk})$ with threshold τ. The choice of the threshold τ is made through a cross-validation method in Fan and Lv (2016). Based on $\hat{\Omega}_{ISEE,g}$, one can update the (j,k) entry of $\hat{\Omega}_{ISEE,g}$ when the nodes j and k are from different index sets A's by replacing the off-diagonal entry of the 2×2 matrix $\hat{\Omega}_{A,A}$ with A being $\{j,k\}$. The resulting updated precision matrix estimator is $\hat{\Omega}_{ISEE}$.

3. Application to Forecast Combination

In this section, we conduct Monte Carlo simulations to compare performance of the high-dimensional precision matrix estimator by the ISEE algorithm and other shrinkage and thresholding estimators in computing the precision matrices. Four other high-dimensional matrix regularization approaches are compared:

LW: Ledoit and Wolf (2004)

UT: Bickel and Levina (2008)

AT: Cai and Liu (2011)

MT: Bailey, Pesaran and Smith (2018)

LW is a shrinkage estimator of a large covariance matrix, which is a weighted average between the large sample covariance matrix and the identity matrix multiplying the mean of diagonal elements. UT is a "universal thresholding" estimator of a large covariance matrix, where the threshold is chosen by cross-validation. The elements whose absolute values are smaller than the threshold are shrunk to 0. AT is an "adaptive thresholding" estimator of a large covariance matrix, where each element has a different threshold value. MT is a "multiple testing" estimator of a large covariance matrix, where the sample covariance matrix is decomposed into the diagonal and correlation matrix. The correlation matrix is regularized by the universal thresholding method.

The optimal forecast combination weight w is given by equation (2), where Σ^{-1} is the precision matrix of the forecast errors of p forecasts. However, in practice the optimal forecast combination is often found to be outperformed by the equally-weighted combined forecast with $w = \frac{1}{p}\iota = (\frac{1}{p}, ..., \frac{1}{p})'$ (which will be referred to as the $1/p$ model below). For

example, the $1/p$ model is optimal if $\Sigma = I_{p\times p}$ in (2). Stock and Watson (2004) call this the "forecast combination puzzle". See also Timmermann (2006) and Elliott (2011). Smith and Wallis (2009) show this puzzle can happen due to estimation error of the combining weights.

We apply the ISEE estimator as well as the four other shrinkage and thresholding methods to obtain the optimal forecast combination weight w. Under the optimal forecast combination weight w obtained using the ISEE estimator, the mean-squared forecast errors are much smaller than those of the other methods and also smaller than the simple average combined forecast. Hence the ISEE estimator resolves the forecast combination puzzle.

For Monte Carlo simulation, we generate the data y_t from MA(∞) :

$$y_t = \sum_{k=0}^{\infty} \beta^k e_{t-k}. \tag{13}$$

where $\beta \in \{0.6,\ 0.7,\ 0.8,\ 0.9\}$ and $e_t \sim i.i.d.N(0,1)$. Forecasts of y_{t+1} (one-step ahead) are based on AR(l) models:

$$\hat{y}_t = \hat{\mu} + \hat{\phi}_1 y_{t-1} + \cdots + \hat{\phi}_l y_{t-l}. \tag{14}$$

We set the number of regression periods $m = 100$, the number of prediction periods $n = 100$, number of lags $l \in \{0,\ 1,\ \ldots,\ 12\}$, where the dimensionality $p = 13$ for 13 forecast models. We evaluate the performance of 6 estimators – $1/p$, LW, UT, AT, MT, and ISEE. The mean-squared forecast errors (MSFE) of ISEE estimator are much smaller than the MSFE of the simple $1/p$ averaging combined forecasts, as shown in Table 1. It indicates that the ISEE estimator resolves the forecast combination puzzle.

Table 1: Forecast Combination

β	$1/p$	LW	UT	AT	MT	ISEE
0.6	1.06	1.00	0.93	0.91	0.91	0.87
0.7	1.06	0.99	0.90	0.92	0.92	0.89
0.8	1.07	0.99	0.92	0.93	0.91	0.86
0.9	1.08	1.01	0.93	1.01	0.94	0.89

4. Conclusions

In this paper, based on the ISEE algorithm by Fan and Lv (2016), we conduct Monte Carlo simulations to estimate a high-dimensional precision matrix Ω in constructing the optimal forecast combination. We show that the ISEE estimator can address the forecast combination puzzle better than the shrinkage and thresholding methods.

REFERENCES

Bailey, N., Pesaran, M.H. and Smith, L.V. (2018). "A Multiple Testing Approach to the Regularization of Large Sample Correlation Matrices." USC.

Bickel, P.J. and Levina, E. (2008). "Covariance Regularization by Thresholding." *The Annuals of Statistics* 36: 2577-2604.

Cai, T. and Liu, W. (2011). "Adaptive Thresholding for Sparse Covariance Matrix Estimation." *Journal of the American Statistical Association* 106: 672-684.

Elliott, G. (2011). "Averaging and the Optimal Combination of Forecasts." UCSD.

Fan, Y. and Lv, J. (2016). "Innovated Scalable Efficient Estimation in Ultra-Large Gaussian Graphical Models." *The Annuals of Statistics* 44: 2098-2126.

Ledoit, O. and Wolf, M. (2004). "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices." *Journal of Multivariate Analysis* 88: 365-411.

Smith, J. and Wallis, K. (2009). "A Simple Explanation of the Forecast Combination Puzzle." *Oxford Bulletin of Economics and Statistics* 71: 331-355.

Stock, J.H. and Watson, M.W. (2004). "Combination Forecasts of Output Growth in a Seven-country Data Set." *Journal of Forecasting* 23: 405-430.

Timmermann, A. (2006). "Forecast Combinations." *Handbook of Economic Forecasting* 1: 135-196.