Estimation of High-Dimensional Dynamic Conditional Precision Matrices with an Application to Forecast Combination*

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

First version: November 2018
This version: March 2020

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 covariance matrices. However, in many applications (e.g., regression, forecast combination, portfolio selection), what we need is not the covariance matrix but its inverse (the precision matrix). In this paper we introduce a method of estimating the high-dimensional "dynamic conditional precision" (DCP) matrices. The proposed DCP algorithm is based on the estimator of a large unconditional precision matrix by Fan and Lv (2016) to deal with the high-dimension and the dynamic conditional correlation (DCC) model by Engle (2002) to embed a dynamic structure to the conditional precision matrix. The simulation results show that the DCP method performs substantially better than the methods of estimating covariance matrices based on thresholding or shrinkage methods. Finally, inspired by Hsiao and Wan (2014), we examine the "forecast combination puzzle" using the DCP, thresholding, and shrinkage methods.

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

JEL Classifications: C3, C4, C5

*The authors are thankful to the editor, associate editor, two anonymous referees, Ekaterina Seregina, and seminar participants at the conference in honor of Professor Cheng Hsiao in Beihang University in Beijing, the Joint Statistical Meetings (JSM2019) in Denver, and the seminar at Amazon Inc in Seattle, for many valuable comments.

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

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

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

1 Introduction

In many applications of multivariate statistical analysis, such as forecast combination, optimal portfolio selection and social networks, 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 (2019).[1]

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.

The ISEE method is however for the unconditional high-dimensional precision matrix. Extending the ISEE of Fan and Lv (2016), we develop a method to estimate the conditional high-dimensional precision matrices which we will call the "dynamic conditional precision" (DCP) matrix. We first use the ISEE method to obtain a transformed data, to which we apply a dynamic conditional covariance model as in the DCC model of Engle (2002). In this step, we build a large

[1] There are other papers that use the random matrix theory (RMT) to estimate the covariance matrix, such as Karoui (2008) who develops an estimator of eigenvectors and eigenvalues of covariance matrices by discretizing and inverting the Stieltjes transform of a limiting sample spectral distribution. However, as Ledoit and Wolf (LW, 2015) point out, this method does not exploit the natural discreteness of the population spectral distribution for finite number of variables (p). LW (2012) use the same discretization strategy as in Karoui (2008), but they match population eigenvalues to sample eigenvalues on the real line. The drawback of this approach is that they only consider the case when $p < n$. LW (2015) extend LW (2012) and develop an estimator of the population eigenvalues that works also when $p > n$. They use a different discretization strategy and show that their estimator works better than LW (2012) even for $p < n$. Another related paper is by Mestre (2008) who proposes an estimator of eigenvalues and eigenvectors of covariance matrices using contour integration of analytic functions in the complex plane. The review paper by Bun et al (2017) provides a comprehensive overview of the modern techniques in RMT and their usefulness for estimating large correlation matrices. Recently, Engle et al (2019) use non-linear shrinkage method which is based on RMT by LW (2012, 2015) to develop an improved estimation of large dynamic covariance matrices (to be more precise, RMT is used for estimating the unconditional correlation matrix and then use it for correlation targeting).

dynamic conditional precision matrix estimator from the product of its diagonal (scale) terms and its off-diagonal (correlation) terms. The diagonal terms are estimated by univariate GARCH-type conditional variance model (so it is very low-dimensional) and the off-diagonal terms are estimated component-wise (one off-diagonal element at a time) by bivariate (thus also low-dimensional) conditional correlation matrices. Lastly, we combine the diagonal elements and the off-diagonal elements to obtain the high-dimensional DCP estimator. We examine the efficiency of the DCP estimator in comparison with that of several shrinkage methods and thresholding methods by Monte Carlo simulations, which show that the DCP estimator is more efficient in estimating high-dimensional conditional precision matrices. Finally, we demonstrate the advantage of the DCP estimator in resolving the "forecast combination puzzle".

This paper is organized as follows. We introduce the estimation algorithm of large dynamic conditional precision matrices in Section 2. In Section 3, Monte Carlo experiments are presented to examine the performance of the DCP estimator. We study several forecast combination examples using the DCP estimator in Section 4. Section 5 concludes.

2 Estimating Conditional Precision Matrices

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 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 $\mathbf{x}$ (subject to the constraint $w'\boldsymbol{\iota} = 1$) 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), Timmermann (2006), and Hsiao and Wan (2014). Another example where the precision matrix $\Omega = \Sigma^{-1}$ is needed instead of the covariance matrix Σ is to form the optimal portfolio (Markowitz 1952). Consider X_j $(j = 1, \ldots, p)$ as financial return from a financial asset j. The optimal portfolio return $w'\mathbf{x}$ can be obtained with the optimal weight given by

$$w = \frac{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} - \gamma\left(\boldsymbol{\iota}'\Sigma^{-1}\boldsymbol{\mu}\right)}{\left(\boldsymbol{\iota}'\Sigma^{-1}\boldsymbol{\iota}\right)\left(\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}\right) - \left(\boldsymbol{\iota}'\Sigma^{-1}\boldsymbol{\mu}\right)^2}\Sigma^{-1}\boldsymbol{\iota} + \frac{\gamma\left(\boldsymbol{\iota}'\Sigma^{-1}\boldsymbol{\iota}\right) - \left(\boldsymbol{\iota}'\Sigma^{-1}\boldsymbol{\mu}\right)}{\left(\boldsymbol{\iota}'\Sigma^{-1}\boldsymbol{\iota}\right)\left(\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}\right) - \left(\boldsymbol{\iota}'\Sigma^{-1}\boldsymbol{\mu}\right)^2}\Sigma^{-1}\boldsymbol{\mu} \tag{3}$$

where γ is the targeted return on the portfolio. Nevertheless, the literature is largely about estimation of Σ rather than $\Omega = \Sigma^{-1}$.

Furthermore, all the dynamic models in the literature is entirely about the conditional covariance matrix Σ_t, and there is no single research paper (to our knowledge) on the conditional precision matrix Ω_t. The dynamic conditional covariance matrix models have been studied by Bollerslev et al (1988), Engle and Kroner (1995), Engle (2002), Engle et al (2019), Pakel et al (2020), among many others. However, there is no paper on the dynamic conditional precision matrix and we believe this paper is the first on that. Therefore the goal of this paper is to introduce the model of the high-dimensional dynamic conditional precision matrix Ω_t and its estimation method. The new method is based on the ISEE algorithm of Fan and Lv (2016) to deal with the high-dimensionality and it is based on the DCC model of Engle (2002) to embed a dynamic structure in Ω_t (not in Σ_t). Adapting these two approaches produces the proposed DCP matrix estimator.

We will first review the ISEE method for the estimation of the unconditional precision matrix Ω in subsection 2.1 and then we will introduce our new method for estimation of the conditional precision matrix Ω_t in subsection 2.2. Let us establish some notation to begin. 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.

2.1 Unconditional Precision Matrix: the ISEE Algorithm

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

$$\mathbf{z} = \Omega \mathbf{x}, \tag{4}$$

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

$$\begin{aligned} E(\mathbf{z}) &= E(\Omega \mathbf{x}) = 0, \\ \mathrm{COV}(\mathbf{z}) &= \mathrm{COV}(\Omega \mathbf{x}) = \Omega \mathrm{COV}(\mathbf{x}) \Omega = \Omega \Sigma \Omega = \Omega. \end{aligned} \tag{5}$$

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

$$\begin{pmatrix} \mathbf{z}_A \\ \mathbf{z}_{A^c} \end{pmatrix} = \begin{pmatrix} \Omega_{A,A} & \Omega_{A,A^c} \\ \Omega_{A^c,A} & \Omega_{A^C,A^C} \end{pmatrix} \begin{pmatrix} \mathbf{x}_A \\ \mathbf{x}_{A^c} \end{pmatrix} = \begin{pmatrix} \Omega_{A,A}\mathbf{x}_A + \Omega_{A,A^c}\mathbf{x}_{A^c} \\ \Omega_{A^c,A}\mathbf{x}_A + \Omega_{A^C,A^C}\mathbf{x}_{A^c} \end{pmatrix}, \tag{6}$$

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

$$\mathbf{z}_A = \Omega_{A,A}\mathbf{x}_A + \Omega_{A,A^c}\mathbf{x}_{A^c} = \Omega_{A,A}\left(\mathbf{x}_A + \Omega_{A,A}^{-1}\Omega_{A,A^c}\mathbf{x}_{A^c}\right) \equiv \Omega_{A,A}\mathbf{e}_A, \tag{7}$$

where

$$\mathbf{e}_A \equiv \mathbf{x}_A + \Omega_{A,A}^{-1}\Omega_{A,A^c}\mathbf{x}_{A^c}. \tag{8}$$

From equations (1) and (4), we have $\mathbf{z} \sim N(0, \Omega)$. 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.[2] Hence, the conditional distribution of $\mathbf{x}_A$ given $\mathbf{x}_{A^c}$ is

$$\mathbf{x}_A|\mathbf{x}_{A^c} \sim N\left(-\Omega_{A,A}^{-1}\Omega_{A,A^c}\mathbf{x}_{A^c}, \Omega_{A,A}^{-1}\right). \tag{9}$$

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

$$\left(\hat{\beta}_j, \hat{\theta}_j^{1/2}\right) = \underset{\beta_j\in\mathbb{R}^{p-|A|},\sigma>0}{\arg\min}\left\{\frac{\left\|\mathbf{X}_j - \mathbf{X}_{A^c}\beta_j\right\|_2^2}{2n\sigma} + \frac{\sigma}{2} + \lambda||\beta_*||_1\right\}, \tag{10}$$

where β_* is the Hadamard (component-wise) product of two $(p-|A|)$-dimensional vectors β_j and

[2] $\mathrm{COV}(\mathbf{e}_A, \mathbf{x}_{A^c}) = \mathrm{COV}\left(\mathbf{x}_A + \Omega_{A,A}^{-1}\Omega_{A,A^c}\mathbf{x}_{A^c}, \mathbf{x}_{A^c}\right) = \Sigma_{A,A^c} + \Omega_{A,A}^{-1}\Omega_{A,A^c}\Sigma_{A^c,A^c}$. 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, $\mathrm{COV}(\mathbf{e}_A, \mathbf{x}_{A^c}) = 0$. Since $\mathbf{e}_A$ and $\mathbf{x}_{A^c}$ have joint normality, they are independent.

$\left(n^{-1/2}||\mathbf{X}_k||_2\right)_{k\in A^c}$ 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 $q \geq 1$. Here, the minimizer $\hat{\theta}_j^{1/2}$, which is over σ, provides an estimator of the error standard deviation $\hat{\theta}_j^{1/2} = \text{var}^{1/2}(e_j)$, where e_j is a component of $\mathbf{e}_A$ corresponding to node j.

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

$$\hat{\Omega}_{A,A} = (n^{-1}\hat{\mathbf{E}}_A'\hat{\mathbf{E}}_A)^{-1}, \tag{11}$$

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

$$\underset{n\times p}{\hat{\mathbf{Z}}} = \left(\hat{\mathbf{Z}}_A\right)_{\forall A} \tag{12}$$

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

$$\hat{\Omega}_{ISEE,ini} = n^{-1}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}. \tag{13}$$

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 $B = (b_{jk})$ thresholded at 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 $\hat{\Omega}_{ISEE}$. We have implemented both $\hat{\Omega}_{ISEE,ini}$ and $\hat{\Omega}_{ISEE}$ in subsection 2.2 and found similar results. Therefore, the results reported in Sections 4 and 5 are based on $\hat{\Omega}_{ISEE,ini}$.

However, (13) is an estimator of the unconditional static precision matrix. In the case where information is updating, the large precision matrices can be updated and time-varying. In subsection 2.2, we introduce the dynamic conditional precision (DCP) matrix estimator, for which we will need to use $\hat{\mathbf{Z}}$ in (12) from the ISEE algorithm (but $\hat{\Omega}$ in (13) will not be needed).

2.2 Conditional Precision Matrix: the DCP Algorithm

Consider a p-variate random vector

$$\mathbf{x}_t = (X_{1t}, ..., X_{pt})' \, | \mathcal{F}_{t-1} \sim N(\mathbf{0}, \Sigma_t), \qquad t = 1, ..., n$$

$\Sigma_t = E(\mathbf{x}_t \mathbf{x}_t' | \mathcal{F}_{t-1})$ is the $p \times p$ conditional covariance matrix. $\Omega_t := \Sigma_t^{-1}$ is the $p \times p$ conditional precision matrix. If we knew $\mathbf{z}_t$ such that

$$\underset{p \times 1}{\mathbf{z}_t} = \Omega_t \mathbf{x}_t, \tag{14}$$

then we could estimate Ω_t directly from the conditional covariance matrix of $\mathbf{z}_t$ since

$$E(\mathbf{z}_t \mathbf{z}_t' | \mathcal{F}_{t-1}) = E(\Omega_t \mathbf{x}_t \mathbf{x}_t' \Omega_t | \mathcal{F}_{t-1}) = \Omega_t \Sigma_t \Omega_t = \Omega_t.$$

The problem of estimating Σ_t and then inverting it is now transformed into obtaining $\mathbf{z}_t$ and getting its conditional covariance matrix.

As we do not know $\mathbf{z}_t$ in reality, we will estimate it from $\hat{\mathbf{z}}_t$ which is the tth row of $\underset{n \times p}{\hat{\mathbf{Z}}} = \left(\hat{\mathbf{Z}}_A\right)_{\forall A}$ in (12) from the ISEE algorithm. The conditional precision matrix Ω_t is then estimated from using $\hat{\mathbf{z}}_t$. Inspired by the dynamic conditional correlation (DCC) procedure in Engle (2002), we decompose Ω_t into

$$\Omega_t = V_t W_t V_t, \tag{15}$$

where $V_t^2 \equiv \text{diag}(\Omega_t)$ is the conditional variances of $\hat{\mathbf{z}}_t$ and W_t is the conditional correlation matrix

of $\hat{\mathbf{z}}_t$. We propose that the diagonal elements of V_t are estimated from the univariate GARCH models for $\hat{\mathbf{z}}_t$ and the off-diagonal elements of W_t are estimated from the pair-wise bivariate conditional correlation models as in Engle (2002). Once V_t and W_t are estimated, our estimator of Ω_t is obtained.

Summarizing, the estimation of dynamic conditional precision matrix Ω_t is conducted by the following algorithm. Note that the DCP algorithm is based on the transformed data matrix obtained in (12) from the ISEE algorithm.

The DCP Algorithm:

1. The DCP uses the transformed data matrix $\underset{n\times p}{\hat{\mathbf{Z}}} = \left(\hat{\mathbf{Z}}_A\right)_{\forall A}$ in (12) from the ISEE algorithm.

2. $V_t^2 \equiv \text{diag}(\Omega_t)$. Ω_t has diagonals $\hat{\omega}_{j,t}^2$, $j = 1, \ldots, p$ estimated from a univariate GARCH model for the jth variable in $\hat{\mathbf{Z}}$. Let W_t be the conditional correlation matrix of $\mathbf{z}_t$.

3. Let $\boldsymbol{\varepsilon}_t \equiv \hat{V}_t^{-1}\hat{\mathbf{z}}_t$, where $\hat{\mathbf{z}}_t$ is the tth row of $\hat{\mathbf{Z}}$.

4. For each pair $B = \{j, k\} \subset \{1, ..., p\}$,

$$\begin{aligned} \hat{Q}_{B,t} &= \Gamma_B \Gamma_B' + \Psi_B \left(\boldsymbol{\varepsilon}_{B,t-1}\boldsymbol{\varepsilon}_{B,t-1}'\right) \Psi_B' + \Phi_B \hat{Q}_{B,t-1} \Phi_B', \\ \hat{W}_{B,t} &= \text{diag}\{\hat{Q}_{B,t}\}^{-1/2}\ \hat{Q}_{B,t}\ \text{diag}\{\hat{Q}_{B,t}\}^{-1/2}. \end{aligned} \tag{16}$$

   Γ_B, Ψ_B and Φ_B are 2×2. $\hat{W}_{B,t}$ has diagonals of ones. $\hat{W}_t$ is constructed from plugging the off-diagonal term in $\hat{W}_{B,t}$ into the (j, k) position in $\hat{W}_t$.[3]

5. $\hat{\Omega}_t$ is obtained by combining the diagonal terms $\hat{V}_t$ and the off-diagonal term $\hat{W}_t$

$$\hat{\Omega}_t = \hat{V}_t \hat{W}_t \hat{V}_t. \tag{17}$$

   The resulting $\hat{\Omega}_t$ is the dynamic conditional precision (DCP) matrix estimator.[4]

[3] The DCP algorithm is *component-wise* for the estimation of the conditional precision matrix Ω_t, i.e., element by element for each pair $B = \{j, k\} \subset \{1, ..., p\}$. We recently found that a similar method was used in Pakel et al (2020) who did the component-wise estimation of the conditional covariance matrix Σ_t.

[4] We conjecture that the consistency of the DCP estimator for Ω_t may be established under some assumptions that the maximum number of nonzeros in a row in Ω_t (the degree of non-sparsity) and the dimensionality p grow at certain rates slow enough relative to the sample size n. We leave this for future research.

3 Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to compare performance of the high-dimensional precision matrix estimator by the DCP algorithm and other shrinkage and thresholding estimators which compute the covariance matrix and invert it to obtain the precision matrix. This section has two parts – unconditional and conditional. Subsection 3.1 compares unconditional precision matrix estimators. Subsection 3.2 compares conditional precision matrix estimators. 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 (2019)

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.

Computational loss of all the estimators is analyzed in terms of the Frobenius norm of $\left(\hat{\Omega} - \Omega\right)$. For any $p \times p$ matrix A, the Frobenius norm is defined as

$$\|A\|_F = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{p} |a_{ij}|^2}, \tag{18}$$

where a_{ij} is the element on the ith row and jth column in matrix A, for $i = 1, \ldots, p$ and $j = 1, \ldots, p$.

3.1 Unconditional Precision Matrix

We fix the sample size n to be 100 and increase the number of covariates p from 30, 100 to 500. This subsection contains two data-generating processes (DGP):

1. DGP 1: $\Sigma \sim$ Toeplitz (0.9).
2. DGP 2: $\Sigma \sim$ Tridiagonal (0.5).

Note that we generate data from a covariance matrix Σ (not from its inverse) so that the two DGPs coincide with the models of the four methods (LW, UT, AT, MT). Nevertheless, we are able to show that ISEE performs more efficiently than all the other four methods in computing the precision matrices.

Under DGP 1, we use the Toeplitz (0.9) matrix as the true data generating process of covariance matrix Σ. The diagonal elements of Toeplitz (0.9) matrix are 1. The off-diagonal element with distance d from the diagonal has the value 0.9^d. DGP 1 is a case where it is harder to invert Σ than DGP 2. In Table 1, we report the loss in terms of matrix norms of $\hat{\Omega} - \Omega$. When the dimensionality p is increasing, the Frobenius norms under each estimation procedure are increasing. Among all the five unconditional precision matrix estimators, ISEE has the smallest Frobenius norms, indicating that it outperforms all the other four methods in terms of the matrix norm loss.

Under DGP 2, we use the Tridiagonal (0.5) matrix as the true covariance matrix Σ. Σ is a band matrix with diagonal elements being one and the off-diagonal elements which have distance one from the diagonals are set to 0.5. The rest elements are 0. This DGP 2 is a case where it is easier to invert Σ. In Table 1, we also report the matrix norms of $\hat{\Omega} - \Omega$ for the five different estimation methods. As the dimensionality p increases, ISEE performs as good as the other four estimators and the Frobenius norms under each estimation procedure are increasing.

3.2 Conditional Precision Matrix

We now examine the performance of the DCP estimator for conditional precision matrices. We generate data from the following data generating processes

$$
\begin{aligned}
\mathbf{x}_t|\mathcal{F}_{t-1} &\sim N\left(\boldsymbol{\mu}_t, \Sigma_t\right),\ t=1,\ldots,n \\
\mu_{j,t} &= \phi_j x_{j,t-1}, \quad \phi_j = 0, \quad j=1,\ldots,p \\
\sigma^2_{j,t} &= \gamma_j + \alpha_j e^2_{j,t-1}, \quad \gamma_j = 1-\alpha_j \\
e_{j,t} &\sim N\left(0, \sigma^2_{e,j}\right), \quad \sigma^2_{e,j} = 1 \\
\Sigma_t &= D_t R_t D_t \\
D_t &= \operatorname{diag}\left(\sigma_{1,t}, \cdots, \sigma_{p,t}\right) \\
\boldsymbol{\varepsilon}_t &= D_t^{-1}\mathbf{x}_t|\mathcal{F}_{t-1} \sim N\left(0, R_t\right) \\
R_t &= S\left(1-a-b\right) + a\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}\right) + bR_{t-1} \\
S &= \text{Toeplitz}(0.9).
\end{aligned}
$$

We set $\omega_j = \sigma^2\left(1-\alpha_j\right)$, $\sigma^2_{e,j} = 1$, $\alpha_j = 0.5$, $\phi_j = 0$. For the correlation component, we consider two different DGPs.

3. DGP 3: $\alpha_j = 0.5$, $a = 0.0$, $b = 0.0$. Constant conditional correlation (CCC, Bollerslev 1990).

4. DGP 4: $\alpha_j = 0.5$, $a = 0.5$, $b = 0.0$. Dynamic conditional correlation (DCC, Engle 2002).

We regularize the sample correlation matrix $\hat{S}$ using LW, UT, AT and MT methods in subsection 4.1 and then embed a CCC or DCC process to estimate the dynamic precision matrix $\hat{\Omega}_t$. The new approaches with an embedded dynamic process are call DLW, DUT, DAT and DMT respectively. As a comparison we compute the matrix norm loss of $\hat{\Omega}_t - \Omega_t$ using DLW, DUT, DAT and DMT estimators besides the DCP estimator.

Under DGP 3 (CCC), $R_t = R =$ Toeplitz (0.9). We set the sample size $n \in \{20, 50, 100, 200\}$ and the number of variables $p \in \{30, 50, 100\}$. We compare the performance of the DCP estimator with four other methods – DLW, DUT, DAT and DMT, using the Frobenius norm as an evaluation criterion. In Table 2, the Frobenius norms are reported under different methods. As the sample size n increases, Frobenius norms of the DCP estimator become smaller. On the other hand, as the number of variables p increases, Frobenius norms of the DCP method become larger. It shows

that the estimation errors of the DCP estimator are decreasing in n and increasing in p. Among the five different estimators, the DCP estimator has the smallest error matrix norms and thus is the most advantageous in estimating conditional precision matrices, regardless of the sample size n and the number of variables p. On the other hand, according to the simulation results, there is no clear convergence pattern for the shrinkage and thresholding estimators. The regularized sample covariance matrices are sometimes near singular, which makes the average norm loss quite erratic for these methods. It indicates that the shrinkage and thresholding methods for the covariance matrices may not be stable to estimate precision matrices.

Under DGP 4 (DCC), the conditional correlation matrix is modeled as $R_t = S(1-a-b) + a\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'\right) + bR_{t-1}$, where we set $S = \text{Toeplitz}\ (0.9)$, $a = 0.5$, $b = 0$. We compare the performance of the DCP method with other four methods in terms of matrix norms. In Table 3, the Frobenius norms are reported. It is shown that matrix norms under the DCP estimator are increasing in p and decreasing in n. For all the sample size n and the number of variables p, DCP has the smallest estimation error loss when estimating conditional precision matrices. It is shown to be the most efficient among all the estimation methods.

To continue exploring the properties of the DCP estimator, we consider an application to the "forecast combination puzzle" in the next section.

4 Forecast Combination Puzzle

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 $w = \frac{1}{p}\iota = \left(\frac{1}{p}, ..., \frac{1}{p}\right)'$ (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. Hsiao and Wan (2014) consider several geometric approaches for combining forecasts in large samples – a simple eigenvector approach, a mean corrected eigenvector and trimmed eigenvector approach. They also consider a mean and scale corrected simple average of all predictive models for finite sample and give conditions where

the simple average is an optimal combination. Claeskens, Magnus, Vasnev, and Wang (2016) show this puzzle can happen if w is assumed to be fixed. Sun et al (2018) consider time-varying model averaging weights in forecast combination problems.

When information is updating from time to time, the optimal forecast combination weight is also time-varying and w_t can be computed by Σ_t. Given $\Omega_t = \Sigma_t^{-1}$, the optimal forecast combination weight (subject to the constraint $w_t'\boldsymbol{\iota} = 1$ for each t) becomes

$$w_t = \frac{\Omega_t \boldsymbol{\iota}}{\boldsymbol{\iota}' \Omega_t \boldsymbol{\iota}}, \tag{19}$$

which is the time-varying version of (2) based on the conditional precision matrix. We apply the DCP estimator as well as the four other shrinkage and thresholding methods to estimate Ω_t and obtain the optimal forecast combination weight w_t. Our results show that the DCP estimator is much more advantageous than the alternative methods including the equal weights $1/p$. Under the optimal forecast combination weight w_t obtained using the DCP estimator as in (19), 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 DCP estimator resolves the forecast combination puzzle.

DGP: In order to demonstrate that the optimal forecast combination weight using the DCP estimator of Ω_t outperforms the equally-weighted combined forecast, Monte Carlo simulation is conducted using the following moving average (MA) process of infinite order

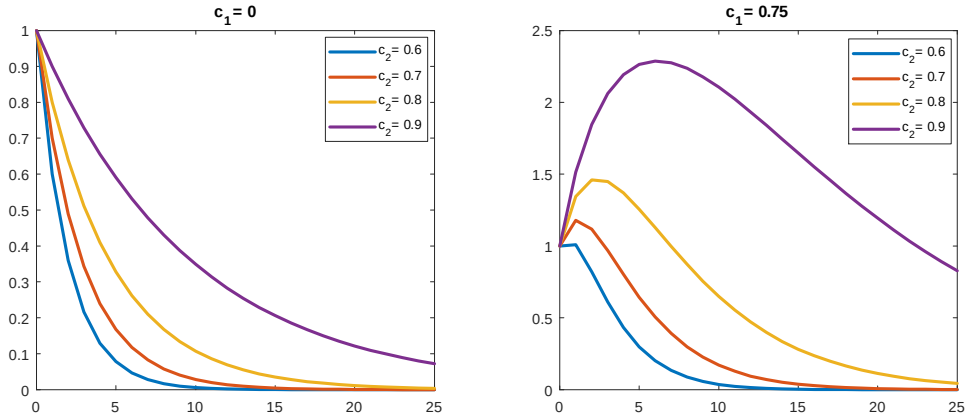
$$y_t = \sum_{k=0}^{\infty} \theta_k e_{t-k}, \tag{20}$$

to generate the data y_t with the MA coefficients θ_k using the rule

$$\theta_k = (1+k)^{c_1} c_2^k, \tag{21}$$

according to Hansen (2008). We consider various combinations of $c_1 \in \{0,\ 0.75\}$ and $c_2 \in \{0.6,\ 0.7,\ 0.8,\ 0.9\}$. We show the decay of the coefficients θ_k over k with different c_1 and c_2 parameters in Figure 1.

Figure 1: Plot of $\theta_k = (1+k)^{c_1} c_2^k$



When $c_1 = 0$, θ_k is gradually decaying when k increases. When $c_1 = 0.75$, θ_k first increases and then decreases.

Models: While the DGP of $\{y_t\}$ is MA(∞), we make forecasts of y_{t+1} (one-step ahead) based on the AR(l) models:

$$\hat{y}_t = \hat{\mu} + \hat{\phi}_1 y_{t-1} + \cdots + \hat{\phi}_l y_{t-l}. \tag{22}$$

We set the number of regression periods (train sample) $m = 100$ and consider two scenarios: (i) the number of out-of-sample prediction periods (test sample) $n = 100$, number of lags $l \in \{0, 1, \ldots, 12\}$, where the dimensionality $p = 13$ for 13 forecast models and (ii) the number of prediction periods $n = 30$, number of lags $l \in \{0, 1, \ldots, 49\}$, where the dimensionality $p = 50$. For each n and p combination, we evaluate the MSFE performance of the 11 estimators which are: $1/p$, LW, UT, AT, MT, ISEE, DLW, DUT, DAT, DMT and DCP. We report the mean-squared forecast errors (MSFE) in Tables 4, 5, 6 and 7.

We consider two DGPs depending on whether the conditional covariance matrix of the error term e_t is time-varying or not:

5. DGP 5: $e_t \sim i.i.d.N(0, 1)$.

6. DGP 6: e_t follows ARCH(1) with the ARCH parameter 0.5.

Under DGP 5, the variance-covariance structure of the error terms is static. Comparing the static precision matrix estimator ISEE and the dynamic precision matrix estimator DCP, the mean-

squared forecast errors with the optimal forecast combination weights obtained using the conditional DCP estimator are slightly larger than that the MSFEs using static unconditional ISEE estimator. The MSFEs of both ISEE and DCP estimators are much smaller than the MSFE of the simple $1/p$ averaging combined forecasts, as shown in Tables 4 and 5. It indicates that the ISEE and DCP estimators are the potential solutions to the forecast combination puzzle. When $n = 30$, $p = 50$, that is when $n < p$, the regularized sample covariance matrices are sometimes hard to invert under the LW, UT, AT and MT approaches, which leads to unstable and erratic MSFEs as shown in Table 5.

Under DGP 6, the variance-covariance structure of the error terms is dynamic. The MSFEs under DGP 6 are reported in Tables 6 and 7. Table 7 contains some erratic MSFEs of the LW, UT, AT and MT approaches when $n < p$. After shrinkage or thresholding, the covariance matrices are sometimes nearly singular, which results in ill-behaved precision matrix estimation and the erratic values in MSFEs. Comparing the static ISEE estimator and the dynamic DCP estimator, the MSFEs with the optimal forecast combination weights obtained using dynamic DCP estimator are much smaller than the MSFEs of using the static ISEE estimator.

5 Conclusions

In this paper, based on the ISEE algorithm by Fan and Lv (2016), we propose the DCP algorithm for estimating a high-dimensional conditional precision matrix Ω_t. We show the consistency of the DCP estimator and examine its efficiency compared with several shrinkage and thresholding methods by simulation experiments and an application to the forecast combination. We show that the DCP estimator can address the forecast combination puzzle better than the shrinkage and thresholding methods for estimating a conditional covariance matrix. The DCP estimator will be useful in many other applications, such as financial portfolio and network theory.

We consider a couple of extensions. One is an application. Besides the forecast combination application considered in this paper, the estimation of large dynamic precision matrices is critical in obtaining the optimal portfolio weights. Another is a semiparametric extension, as in Long et al (2011), which nonparametrically adjusts the parametric DCP estimator and will be useful for real world applications.

References

Bailey, N., Pesaran, M.H. and Smith, L.V. (2019). "A Multiple Testing Approach to the Regularization of Large Sample Correlation Matrices." *Journal of Econometrics* 208(2), 507-534.

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

Bollerslev, T. (1990). "Modelling the Coherence in Short Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model." *Review of Economics and Statistics* 72, 498-505.

Bollerslev, T., Engle, R., and Wooldridge, J. (1988). "A Capital Asset Pricing Model with Time-Varying Covariances." *Journal of Political Economy* 96(1), 116-131.

Bun, J., Bouchaud, J-P., and Potter, M. (2017). "Cleaning Large Correlation Matrices: Tools from Random Matrix Theory." *Physics Reports* 666, 1-109.

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

Claeskens, G., Magnus, J.R., Vasnev, A.L. and Wang, W. (2016). "The Forecast Combination Puzzle: A Simple Theoretical Explanation." *International Journal of Forecasting* 32, 754-762.

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

Engle, R. (2002). "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models." *Journal of Business and Economic Statistics* 20**,** 339-350.

Engle, R. and Kroner, K. (1995). "Multivariate Simultaneous Generalized ARCH." *Econometric Theory* 11(1), 122-150.

Engle, R., Ledoit, O. and Wold, M. (2019). "Large Dynamic Covariance Matrices." *Journal of Business and Economic Statistics* 37(2), 363-375.

Engle, R. and Sheppard, K. (2001). "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH." NBER Working Paper 8554.

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

Hansen, B. (2008). "Least-squares Forecast Averaging." *Journal of Econometrics* 146, 342-350.

Hsiao, C. and Wan, S.K. (2014). "Is there an Optimal Forecast Combination?" *Journal of Econometrics* 178, 294-309.

Karoui (2008). "Spectrum Estimation for Large Dimensional Covariance Matrices Using Random Matrix Theory." *The Annals of Statistics* 36(6), 2757-2790.

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

Ledoit, O. and Wolf, M. (2012). "Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices." *The Annals of Statistics* 40(2), 1024-1060.

Ledoit, O. and Wolf, M. (2015). "Spectrum Estimation: A Unified Framework for Covariance Matrix Estimation and PCA in Large Dimension." *Journal of Multivariate Analysis* 139, 360-384.

Long, X., Su, L. and Ullah, A. (2011). "Estimation and Forecasting of Dynamic Conditional Covariance: A Semiparametric Multivariate Model." *Journal of Business and Economic Statistics* 29, 109-125.

Mestre, X. (2008). "Improved Estimation of Eigenvalues and Eigenvectors of Covariance Matrices Using Their Samples Estimates." *IEEE Transactions on Information Theory* 54(11), 5113-5129.

Pakel, C., Shephard, N., Sheppard, K. and Engle, R. (2020). "Fitting Vast Dimensional Time-Varying Covariance Models." *Journal of Business and Economic Statistics*. DOI: 10.1080/07350015.2020.1713795

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.

Sun, Y., Zhang, X., Lee, T.-H., Hong, Y. and Wang, S. (2018). "Time-Varying Model Averaging." CAS-Cornell-UCR.

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

Table 1: Estimation of Unconditional Precision Matrix

Under DGP 1: Toeplitz (0.9)					
p	ISEE	LW	UT	AT	MT
30	16.37	25.49	56.82	94.69	149.42
100	47.49	67.71	121.34	154.74	156.22
500	159.68	214.84	384.20	300.87	566.46
Under DGP 2: Tridiagonal (0.5)					
p	ISEE	LW	UT	AT	MT
30	195.94	199.35	201.26	254.01	229.99
100	2148.12	2148.44	2176.79	2155.23	2187.17
500	5291.20	5291.40	5291.70	5292.20	5292.00

Notes: Reported is the Frobenius norm of $\left(\hat{\Omega} - \Omega\right)$. The Frobenius norm of a $p \times p$ matrix A is $\|A\|_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p} |a_{ij}|^2}$ with a_{ij} being the (i,j)th element of A. In this table, $n = 100$.

Table 2: Estimation of Conditional Precision Matrix

Under DGP 3: CCC					
	DCP	DLW	DUT	DAT	DMT
			$p = 30$		
$n = 20$	520.01	4733.20	9058.50	8930.40	7116.80
$n = 50$	86.04	4660.10	3.13×10^4	2.03×10^4	1.16×10^4
$n = 100$	42.12	4475.00	3.51×10^4	5.10×10^4	3.63×10^4
$n = 200$	22.32	4091.40	9247.30	2.22×10^4	3.50×10^4
			$p = 50$		
$n = 20$	703.90	8339.80	1.07×10^4	1.27×10^4	1.12×10^4
$n = 50$	191.41	7413.00	1.57×10^4	4.47×10^4	6.15×10^6
$n = 100$	68.66	7852.70	8.60×10^4	6.14×10^6	4.38×10^5
$n = 200$	48.71	7598.20	9.55×10^4	6.13×10^4	4.54×10^7
			$p = 100$		
$n = 20$	1971.00	1.99×10^4	2.16×10^4	2.42×10^4	2.36×10^4
$n = 50$	439.31	1.82×10^4	3.30×10^4	1.28×10^5	6.05×10^5
$n = 100$	182.40	1.63×10^4	8.19×10^4	5.58×10^6	3.10×10^7
$n = 200$	84.46	1.51×10^4	7.41×10^4	1.07×10^5	3.22×10^5

Note: Reported is the time-average of the Frobenius norms, $n^{-1}\sum_{t=1}^{n}\left\|\hat{\Omega}_t - \Omega_t\right\|_F$. The Frobenius norm of a $p \times p$ matrix A is $\|A\|_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}|a_{ij}|^2}$ with a_{ij} being the (i,j)th element of A.

Table 3: Estimation of Conditional Precision Matrix

Under DGP 4: DCC					
	DCP	DLW	DUT	DAT	DMT
			$p = 30$		
$n = 20$	1179.50	2.31×10^4	3.07×10^4	4.24×10^6	2.46×10^4
$n = 50$	425.69	1.94×10^4	3.77×10^6	2.47×10^6	5.31×10^4
$n = 100$	258.82	1.80×10^4	3.26×10^4	6.24×10^4	9.50×10^4
$n = 200$	89.74	1.88×10^4	9.42×10^4	6.02×10^6	6.05×10^5
			$p = 50$		
$n = 20$	2529.20	3.78×10^4	6.73×10^4	5.65×10^4	4.27×10^4
$n = 50$	753.14	3.53×10^4	5.30×10^4	6.61×10^4	7.84×10^4
$n = 100$	428.80	3.57×10^4	5.15×10^4	2.08×10^7	8.73×10^4
$n = 200$	201.65	3.18×10^4	3.61×10^6	1.07×10^7	3.59×10^5
			$p = 100$		
$n = 20$	7124.00	8.46×10^4	1.26×10^5	1.68×10^5	9.37×10^4
$n = 50$	1637.60	7.88×10^4	2.45×10^5	7.33×10^5	2.74×10^5
$n = 100$	961.44	7.78×10^4	1.53×10^5	2.04×10^5	3.93×10^5
$n = 200$	418.71	7.18×10^4	3.71×10^5	2.31×10^7	7.51×10^7

Note: Reported is the time-average of the Frobenius norms, $n^{-1}\sum_{t=1}^{n}\left\|\hat{\Omega}_t - \Omega_t\right\|_F$. The Frobenius norm of a $p \times p$ matrix A is $\|A\|_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}|a_{ij}|^2}$ with a_{ij} being the (i,j)th element of A.

Table 4: MSFE of Forecast Combinations

DGP 5 $e_t \sim i.i.d.N(0,1)$, $n = 100$, $p = 13$											
	$1/p$	LW	UT	AT	MT	ISEE	DLW	DUT	DAT	DMT	DCP
c_2	$c_1 = 0$										
0.6	1.06	1.00	0.93	0.91	0.91	0.87	1.03	1.11	1.24	1.73	0.89
0.7	1.06	0.99	0.90	0.92	0.92	0.89	1.12	1.56	1.28	1.30	0.91
0.8	1.07	0.99	0.92	0.93	0.91	0.86	1.15	1.60	1.65	1.31	0.93
0.9	1.08	1.01	0.93	1.01	0.94	0.89	1.13	1.32	1.43	1.39	0.95
c_2	$c_1 = 0.75$										
0.6	1.07	1.03	0.97	0.94	0.98	0.86	1.13	1.21	1.33	1.71	0.92
0.7	1.08	1.05	0.99	0.99	1.00	0.90	1.11	1.61	1.65	2.06	0.97
0.8	1.16	1.05	1.02	1.02	1.14	0.92	1.23	1.31	1.32	2.45	0.95
0.9	1.55	1.13	1.21	1.07	1.05	0.93	1.87	2.84	1.49	2.84	0.97

Notes: We set the number of regression periods (train sample) $m = 100$ and the number of out-of-sample prediction periods (test sample) $n = 100$. The $p = 13$ forecasts from AR(l) models with $l \in \{0,\ 1,\ \ldots,\ 12\}$ are combined using one of the 11 different combination weights. Each column corresponds to each of the 11 different methods.

Table 5: MSFE of Forecast Combinations

DGP 5 $e_t \sim i.i.d.N(0,1)$, $n = 30$, $p = 50$											
	$1/p$	LW	UT	AT	MT	ISEE	DLW	DUT	DAT	DMT	DCP
c_2	$c_1 = 0$										
0.6	1.25	0.90	3.21	2.99	5.03	0.85	1.21	6.12	1.39	4.83	1.04
0.7	1.25	0.90	3.19	2.39	9.21	0.84	1.24	9.05	3.17	1.95	1.03
0.8	1.26	0.94	7.21	2.56	3.69	0.83	1.22	163.11	126.11	9.64	1.05
0.9	1.27	0.95	7.50	3.11	4.80	0.87	1.32	147.36	3.23	4.11	1.01
c_2	$c_1 = 0.75$										
0.6	1.19	0.84	11.93	16.93	6.87	0.79	1.39	875.18	12.65	576.32	1.07
0.7	1.30	0.94	16.49	15.29	5.38	0.89	1.65	959.81	258.16	61.40	1.01
0.8	1.30	1.04	18.66	15.62	5.56	0.93	1.33	284.19	27.10	7.36	1.13
0.9	1.34	1.32	13.47	12.33	4.06	1.15	1.49	223.79	1.32	537.24	1.10

Notes: We set the number of regression periods (train sample) $m = 100$ and the number of out-of-sample prediction periods (test sample) $n = 30$. The $p = 50$ forecasts from AR(l) models with $l \in \{0,\ 1,\ \ldots, 49\}$ are combined using one of the 11 different combination weights. Each column corresponds to each of the 11 different methods.

Table 6: MSFE of Forecast Combinations

DGP 6 $e_t \sim$ ARCH (1), $n = 100$, $p = 13$											
	$1/p$	LW	UT	AT	MT	ISEE	DLW	DUT	DAT	DMT	DCP
c_2	$c_1 = 0$										
0.6	1.11	1.06	0.97	0.98	0.93	0.90	1.04	1.69	1.65	1.26	0.86
0.7	1.21	1.14	0.94	1.01	1.01	0.89	1.14	1.34	1.84	1.55	0.85
0.8	1.12	1.02	0.94	0.99	1.00	0.90	1.22	1.94	1.28	1.51	0.85
0.9	1.37	1.06	0.98	1.03	0.99	0.91	2.81	1.74	1.55	1.48	0.88
c_2	$c_1 = 0.75$										
0.6	1.09	1.05	1.25	1.04	0.93	0.93	1.30	1.41	1.48	1.38	0.91
0.7	1.19	1.03	3.37	1.09	1.06	0.93	1.29	2.33	1.69	1.22	0.92
0.8	1.22	1.04	2.01	1.18	0.99	0.94	1.44	2.00	1.74	2.20	0.90
0.9	1.73	1.09	3.63	1.10	1.04	0.98	1.99	3.22	1.83	3.94	0.91

Notes: We set the number of regression periods (train sample) $m = 100$ and the number of out-of-sample prediction periods (test sample) $n = 100$. The $p = 13$ forecasts from AR(l) models with $l \in \{0,\ 1,\ \ldots,\ 12\}$ are combined using one of the 11 different combination weights. Each column corresponds to each of the 11 different methods.

Table 7: MSFE of Forecast Combinations

DGP 6 $e_t \sim$ ARCH (1), $n = 30$, $p = 50$											
	$1/p$	LW	UT	AT	MT	ISEE	DLW	DUT	DAT	DMT	DCP
c_2	$c_1 = 0$										
0.6	1.24	1.01	7.62	15.84	2.33	0.98	1.20	83.77	2.23	81.09	0.96
0.7	1.30	0.99	3.47	3.14	13.09	0.98	1.34	271.51	15.12	494.76	0.97
0.8	1.27	1.00	17.75	29.30	2.44	0.99	1.29	958.55	25.63	223.75	0.96
0.9	1.31	1.15	1.82	46.11	15.37	1.00	1.34	535.33	6.84	1138.44	0.98
c_2	$c_1 = 0.75$										
0.6	1.39	1.01	2.44	4.24	21.16	0.99	1.20	104.40	275.27	11.66	0.97
0.7	1.30	1.01	13.48	4.02	7.98	1.00	1.57	175.95	3.39	573.59	0.98
0.8	1.48	1.07	12.51	22.13	2.10	1.01	1.38	67.58	37.17	65.93	0.97
0.9	1.37	1.37	8.56	1.93	7.46	1.04	1.30	709.93	151.23	651.00	0.98

Notes: We set the number of regression periods (train sample) $m = 100$ and the number of out-of-sample prediction periods (test sample) $n = 30$. The $p = 50$ forecasts from AR(l) models with $l \in \{0,\ 1,\ \ldots, 49\}$ are combined using one of the 11 different combination weights. Each column corresponds to each of the 11 different methods.