Statistica Sinica Preprint No: SS-2017-0141	
Title	Large Multiple Graphical Model Inference via Bootstrap
Manuscript ID	SS-2017-0141
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0141
Complete List of Authors	Yongli Zhang Xiaotong Shen and Shaoli Wang
Corresponding Author	Yongli Zhang
E-mail	yongli.zhang@gmail.com
Notice: Accepted version subject to English editing.	

Large Multiple Graphical Model Inference via Bootstrap *

Yongli Zhang[1] Xiaotong Shen[2] Shaoli Wang[3]

Summary

Large economic and financial networks may experience stage-wise change as a result of external shocks. To detect and infer a structural change, we consider an inference problem in the framework of multiple Gaussian graphical models, particularly when the number of graphs and the dimension of a graph expand with the sample size. In such a situation, two major challenges emerge as a result of the bias and uncertainty inherent in regularization, which is required treating such overparameterized models. To deal with these challenges, bootstrap is utilized to approximate the sampling distribution of a likelihood ratio test statistic. Theoretically, we show that the proposed method leads to correct asymptotic inference in a high-dimensional situation regardless of the distribution of the test statistic. Numerically, it compares favorably to its competitors through simulations. Finally, our statistical analysis of the network of 200 stocks reveals that the financial network exhibits a dramatic change in that the interacting units become more connected due to the financial crisis between 2007 and 2009. More importantly, certain units respond more strongly than others, and after the crisis, some alterations fade while others strengthen.

Key words: Bootstrap, Graphical models, High-dimensional inference, Model selection, Regularization

*[1] Yongli Zhang is Assistant Professor, Lundquist College of Business, University of Oregon, 1208 University Ave, Eugene, OR 97403 (E-mail: yongli@uoregon.edu), [2]Xiaotong Shen is Professor, School of Statistics, University of Minnesota, 224 Church Street S.E., Minneapolis, MN 55455 (E-mail: xshen@umn.edu), and Shaoli Wang, Associate Professor, School of Management Science and Statistics, Shanghai Finance and Economics University, China (E-mail swang@shufe.edu.cn). This research supported in part by NSF grant DMS-1415500, DMS-1712564, DMS-1721216, NIH grants 1R01GM081535-01, HL65462 and R01HL105397, and Wang's research is partially supported by National Natural Science Foundation of China NSFC, 11371235. The authors would like to thank the editor, the associate editor and anonymous referees for helpful comments and suggestions.

1 Introduction

In economics, network analysis plays a fundamental role in studying consumer behavior and international trade. In finance, network analysis leads to uncover financial contagion and minimize systemic risk, thus leading to early prevention to future crises [8]. However, estimation and inference of large networks often face challenges of high dimensionality [12, 14]. For example, exploration of a network of 200 stocks involves $(200^2 + 200)/2 = 20100$ pairwise edges while the sample size is in hundreds. In a situation as such, classical inference approaches become invalid or break down. The reader may consult a survey article by [15] for challenges in inference due to high dimensions. Even worse, as the number of graphs involved in a multiple graphical model (MGM), which is often used for modeling networks experiencing stage-wise change due to external forces, grows with the sample size or the number of nodes, estimation and inference become more challenging. The additional challenges in estimating a multiple compared with a single graphical model have been studied in [29, 4, 22]. To respond to these challenges, we develop inference tools for large stage-wise networks defined by multiple graphical models.

This article considers multiple Gaussian graphical model (MGGM) to model stage-wise networks, in which the number of model parameters may greatly exceed the sample size and the number of stages may increase with the size of nodes and/or observations. Tow major challenges emerge. First, a certain form of regularization is often used to treat overparameterized models in a high-dimensional situation, which undoubtedly introduces a bias of estimation thus leading to biased inference. Moreover, usual asymptotic approximations for the sampling distribution of a test statistic become inadequate [18]. Second, the selection uncertainty inherent in regularization is mathematically intractable even in a low-dimensional situation [28]. Concerning graphical model inference, [30] proposed a maximum likelihood inference approach for a simple Gaussian graphical model, and derived an asymptotic distribution of the constrained likelihood ratio, which is the chi-square or normal distributions depending on the size of the co-dimension for the inference space.

In this article, we develop a likelihood ratio inference approach for multiple Gaussian graphical models, where the bootstrap method is utilized to approximate the sampling distribution of a test statistic to account for the bias and selection uncertainty due to regularization. Benefits of this approach are demonstrated numerically and theoretically. As showed in Theorem 1, the bootstrap likelihood ratio test is asymptotically valid when the size of graph p grows in an order slightly smaller than $\exp(cn)$ for some small $c > 0$ in the sample size n. In contrast, an asymptotic chi-square or normal approximation of the sampling distribution of the likelihood ratio may work when p is roughly of smaller order of $n^{1/2}$ [30], and such an approximation becomes inadequate when p is larger, as shown by our simulations in Figures 1-4. In this sense, bootstrap offers an attractive alternative to asymptotic high-dimensional inference.

Concerning the application of Gaussian graphical models (GGMs) in economics and finance, they have not received much attention. [12] developed methods of estimating large covariance and precision matrices for economic and financial data. To our knowledge, the proposed method is the first attempt to infer a large multiple graphical model where both the number of graphs and the number of linkages may expand. Importantly, our method enables us to identify the type, origin and evolution of interactions among nodes. With the help of the proposed inference method, we examine the impact of Lehman Brothers' breakdown on financial networks by analyzing historical prices of 200 stocks publicly traded in North America between January 1, 2005 and December 31, 2010. Particularly, we investigated structural and strength changes of the network of these stocks over time as a consequence of Lehman Brothers' collapse. The end result will be contrasted to that in another study [20] involving a static network of 452 stocks during the booming cycle between January 1, 2003 and January 1, 2008. As suggested by our analysis, the financial network has experienced a profound alteration since Lehman Brothers' breakdown. Overall, connectivity becomes more widespread and strong, while different sectors exhibit disparity patterns.

This article is organized as follows. Section 2 formulates the problem and proposes our inference method and its theoretical validity is proved in Section 3. Section 4 is devoted to simulation results, followed by data analysis in Section 5. Section 6 concludes this paper.

2 Inference

This section develops a likelihood inference method for large networks on the ground of multiple Gaussian graphical models.

2.1 Multiple Gaussian graphical models (MGGMs)

To model the stock network experiencing stages $t = 1, \cdots, T$, we consider a multiple Gaussian graphical model (MGGM) of T graphs $(GR_1, \cdots, GR_T)$, representing a network at different stages. For inference, T independent random samples $\boldsymbol{Y} = (\mathbf{Y}^t)_{t=1}^T$ are obtained, where $\mathbf{Y}^t = (\mathbf{Y}_1^t, \cdots, \mathbf{Y}_{n_t}^t)$ are n_t independent and identically distributed p-dimensional random vectors, with $\mathbf{Y}_k^t \sim N(\mathbf{0}, \mathbf{\Sigma}_t)$; $1 \leqslant k \leqslant n_t$, $\mathbf{0}$ is a p-dimensional zero vector, $\mathbf{\Sigma}_t = (\sigma_{i,j,t})_{1 \leqslant i,j \leqslant p}$ a $p \times p$ covariance matrix, and the sample size is $N = \sum_{t=1}^T n_t$. Suppose $\max_t \operatorname{tr}(\mathbf{\Sigma}_t)/p^2 \to 0$. In this situation, the precision matrix $\mathbf{\Omega}_t = (\omega_{i,j,t})_{1 \leqslant i,j \leqslant p} = (\mathbf{\Sigma}_t)^{-1}$ is the inverse of the covariance matrix, whose off-diagonal entry $\omega_{i,j,t}$ is zero if and only if the nodes i and j are conditionally independent given all other p-2 nodes at time t; $t = 1, \cdots, T$ [26]. Thus, $\mathbf{\Omega}_t$ uniquely encodes undirected graph at t; $t = 1, \cdots, T$. In this article, the number of stages T and the number of nodes p are both allowed to expand.

In this article, we suppose the sample mean $\bar{\mathbf{Y}}_t = n_t^{-1} \sum_{k=1}^{n_t} \mathbf{Y}_k^t = \mathbf{0}$ $(t = 1, \cdots, T)$. The sample covariance matrix of stage t is defined as $\mathbf{S}_t = n_t^{-1} \sum_{k=1}^{n_t} (\mathbf{Y}_k^t)(\mathbf{Y}_k^t)'$, and $\mathbf{\Omega}_t$ is a positive definite and symmetric $p \times p$ matrix; $t = 1, \cdots, T$. Let $\mathbf{Z}_k^t = \mathbf{\Omega}_t^{1/2} \mathbf{Y}_k^t$. Then $\mathbf{Z}_k^t \sim MVN(\mathbf{0}, \mathbf{I}_p)$, where $\mathbf{I}_p$ is a $p \times p$ identity matrix and $\|\mathbf{Z}_k^t\|_2^2 \sim \chi_p^2$.

Our goal is to infer if a network stays the same across all T stages. Our major approach is to approximate the test statistic, which is based on the logarithm of likelihood ratio, by bootstrapping and establish its theoretical validity in high dimensional case, in which regularization is imposed to achieve sparsity, introducing the bias to inference.

In low-dimensional case, the precision matrix is estimated by minimizing the negative log-likelihood or the empirical loss function $\mathcal{L}(\mathbf{S}; \mathbf{G})$

$$\mathcal{L}(\mathbf{S}; \mathbf{G}) = \operatorname{tr}(\mathbf{S}\mathbf{G}) - \log\det(\mathbf{G}) \tag{1}$$

over $\mathbf{G}$ that is positive definite for a given sample covariance matrix $\mathbf{S}$, which may be singular. In the context of multiple graphical models and a high-dimensional situation (i.e., large p), the negative log-likelihood is often regularized collectively or individually to yield a unique solution for such an overparameterized model:

$$\mathcal{C}(\mathbf{G}_1, \cdots, \mathbf{G}_T) = \sum_{t=1}^{T} n_t \big(\mathcal{L}(\mathbf{S}_t; \mathbf{G}_t) + F_\lambda^{\upsilon}(\mathbf{G}_t)\big), \tag{2}$$

where $\mathcal{L}(\mathbf{S}_t; \mathbf{G}_t)$ is the negative log-likelihood or the empirical loss function $F_\lambda^{\upsilon}(\mathbf{G}_t)$ is a penalty regularizing $(\mathbf{G}_1, \cdots, \mathbf{G}_T)$. In this article, we consider the L_1 ($\upsilon = 1$) and L_2 ($\upsilon = 2$) penalties: (1) $F_\lambda^1(\mathbf{G}_t) = \lambda \|\mathbf{G}_t\|_1 = \lambda \sum_{i,j} |g_{i,j,t}|$; (2) $F_\lambda^2(\mathbf{G}_t) = \lambda \|\mathbf{G}_t\|_F^2 = \lambda \sum_{i,j} g_{i,j,t}^2$. Third, an optimal choice of regularization parameters for inference may depend on the sample size, which must be estimated based on data. In this article, we adopt $\lambda = \sqrt{\log p/n}$ [18, 27] for L_1-regularization (i.e., Graphical Lasso) to guard against overly large bias caused by the pursue of sparsity. The penalized maximum likelihood estimate $(\widehat{\mathbf{\Omega}}_1, \cdots, \widehat{\mathbf{\Omega}}_T)$ is then defined as $\operatorname{argmin}_{\mathbf{G}_1, \cdots, \mathbf{G}_T} \mathcal{C}(\mathbf{G}_1, \cdots, \mathbf{G}_T)$.

2.2 Inference

High-dimensional inference may involve hypothesis testing and construction of confidence intervals or regions and its sampling distribution. In our case study, the number of unknown parameters $(p^2 + p)T/2 = 60300$ ($p = 200, T = 3$) is much greater than the grand sample size $N = 750$ ($n_1 = n_2 = n_3 = 250$), which imposes several challenges. First, an asymptotic approximation is usually inadequate in a context of "large p but small n", which could lead to biased inference. Second, a form of regularization is often necessary for estimation, which undoubtedly introduces the bias and selection uncertainty to inference, particularly when the L_1-regularization is employed. This is an analogy of the commonly encountered issue in inference after model selection [10, 28], which, as expected, yields highly biased inference, either optimistically or pessimistically, depending highly on if the parameter of interest is included in the final model after regularization. To deal with these challenges, we use bootstrap to offer an alternative without requiring deriving asymptotic approximations in inference, while accounting for the effect of bias and selection uncertainty brought by regularization.

Our bootstrap procedure proceeds as follows:

Step 1: *Draw B bootstrap samples $(\boldsymbol{Y}_1^*, \cdots, \boldsymbol{Y}_B^*)$ from the original sample $\boldsymbol{Y}$ under H_0.*

Step 2: *Derive the estimates of the precision matrices, $\widehat{\mathbf{\Omega}}_{t,b}^{0*}$ and $\widehat{\mathbf{\Omega}}_{t,b}^{1*}$ under H_0 and $H_0 \bigcup H_a$ by minimizing (2) based on $\mathbf{Y}_b^*$ ($b = 1, \cdots, B; t = 1, \cdots, T$) for a preselected regularization coefficient*

λ.

Step 3: *The original test statistic is*

$$D = \frac{1}{Np^2} \sum_t \Big(n_t \mathcal{L}(\mathbf{S}_t; \widehat{\mathbf{\Omega}}^0_{t,b}) - n_t \mathcal{L}(\mathbf{S}_t; \widehat{\mathbf{\Omega}}^1_{t,b}) \Big), \tag{3}$$

and the bootstrapping test statistic is

$$D^*_b = \frac{1}{Np^2} \sum_t \Big(n_t \mathcal{L}(\mathbf{S}^*_t; \widehat{\mathbf{\Omega}}^{0*}_{t,b}) - n_t \mathcal{L}(\mathbf{S}^*_t; \widehat{\mathbf{\Omega}}^{1*}_{t,b}) \Big). \tag{4}$$

Inference is made based on the empirical distribution of B bootstrapped test statistics $\{D^*_1, \cdots, D^*_B\}$. The null hypothesis H_0 is rejected when $D > q^*_{1-\alpha}$ with Type-I error at α, where $q^*_{1-\alpha}$ is the $(1-\alpha)$ percentile of $\{D^*_1, \cdots, D^*_B\}$. For this test, the P-value is $\#\{D^*_b > D\}/B$ (Section 4.5 of [23]).

Next we provide some specific steps.

Consider a null hypothesis $H_0 : \mathbf{\Omega}_1 = \cdots = \mathbf{\Omega}_T$ versus its alternative H_a : H_0 is not true. There are $T(p^2 + p)/2$ parameters and $(T - 1)(p^2 + p)/2$ constraints involved in H_0. This is a composite test, and thus all the parameters are regularized. Our bootstrap method generates B bootstrap samples $(\boldsymbol{Y}^*_1, \cdots, \boldsymbol{Y}^*_B)$ of size N from the combined sample $\boldsymbol{Y} = (\mathbf{Y}^1, \cdots, \mathbf{Y}^T)$ under H_0 : $\mathbf{\Omega}_1 = \cdots = \mathbf{\Omega}_T$. Under H_0, $\widehat{\mathbf{\Omega}}^{0*}_b$ is obtained from the combined bootstrap sample $\boldsymbol{Y}^*_b$, which is an $N \times p$ matrix, and the negative log-likelihood, i.e., the loss is $\mathcal{L}(\mathbf{S}^*_b; \widehat{\mathbf{\Omega}}^{0*}_b)$ where $\mathbf{S}^*_b = (1/N)(\mathbf{Y}^*_b)'\mathbf{Y}^*_b$. To calculate the likelihood under $H_0 \bigcup H_a$, partition $\boldsymbol{Y}^*_b$ into T disjoint subsamples $\mathbf{Y}^{*1}_b, \cdots, \mathbf{Y}^{*T}_b$ of size $n_1, \cdots, n_T$ and derive the negative log-likelihood $\mathcal{L}(\mathbf{S}^*_{t,b}; \widehat{\mathbf{\Omega}}^{1*}_{t,b})$ where $\widehat{\mathbf{\Omega}}^{1*}_{t,b}$ is the estimated precision matrix based on $\mathbf{S}^*_{t,b} = (1/n_t)(\mathbf{Y}^{*t}_b)'(\mathbf{Y}^{*t}_b)$; $t = 1 \cdots, T$; $b = 1, \cdots, B$.

For global inference of a single precision matrix, the log-likelihood ratio converges to a chi-square distribution [30] when the co-dimension of the test is fixed. However, when the co-dimension varies with the sample size n, the log-likelihood ratio can be approximated by a normal distribution [30] provided that the rate of growth of the co-dimension is not too fast. As showed in our simulations next, the likelihood ratio can not be well approximated by either the chi-square or normal distributions in a higher-dimensional situation. In this sense, the bootstrap method becomes attractive. Note that our method allows T and p in MGGMs to grow with the sample size. In contrast, the usual asymptotic chi-square approximation breaks down for global inference whereas the bootstrap method continues to work, as demonstrated in Figure 8. Finally, the proposed method is justified theoretically in Section 3 and demonstrated numerically in Section 4.

3 Theory

The validity of our procedure is summarized in the following theorem. Theorem 1 establishes a consistency result for the sampling distribution of bootstrapped likelihood ratios.

The empirical loss function $\mathcal{L}(\mathbf{S};\mathbf{G})$ as a function of a positive definite and symmetric matrix $\mathbf{G}$ is convex. Its expectation $h(\mathbf{G}) = E\mathcal{L}(\mathbf{S};\mathbf{G}) = \mathrm{tr}(\boldsymbol{\Sigma};\mathbf{G}) - \log\det(\mathbf{G})$ is also convex for a positive definite $p \times p$ matrix $\mathbf{G}$, and $h(\mathbf{G})$ achieves the minimum at $\boldsymbol{\Omega}$, which is the true precision matrix.

Next we establish the closeness between the likelihood ratio test statistic and its bootstrapping version, in which the precision matrices are both estimated by Graphical Lasso (GLasso) with the regularization parameter $\lambda = \sqrt{\log p/n}$, in terms of Mallow's distance [2, 23].

Theorem 1 *(Distribution of bootstrapped penalized likelihood ratios)*
Let $(\mathbf{Y}_1, \cdots, \mathbf{Y}_n)$ be a random sample from the multivariate normal distribution $MVN(\mathbf{0}, \boldsymbol{\Sigma})$, where $\mathbf{0}$ is a p-dimensional vector of zeros and the $p \times p$ covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}$ is positive definite. Suppose $\overline{\mathbf{Y}} = \mathbf{0}$. Let $(\mathbf{Y}_1^*, \cdots, \mathbf{Y}_n^*)$ be a boostrapping random sample from $(\mathbf{Y}_1, \cdots, \mathbf{Y}_n)$. Let $\mathbf{S} = (1/n)\sum_{k=1}^n \mathbf{Y}_k\mathbf{Y}_k'$ and $\mathbf{S}^* = (1/n)\sum_{k=1}^n \mathbf{Y}_k^*(\mathbf{Y}_k^*)'$. Suppose $p^{-2}\,\mathrm{tr}\,\boldsymbol{\Sigma}$, λ and $p^{-1}\log(1/\lambda)$ all converge to 0 as $n \to \infty$. Let

$$\hat{\boldsymbol{\Omega}} = \mathrm{argmin}_{\boldsymbol{\Omega}} \left(\mathcal{L}(\mathbf{S};\boldsymbol{\Omega}) + F_\lambda^1(\boldsymbol{\Omega})\right); \tag{5}$$

$$\hat{\boldsymbol{\Omega}}^* = \mathrm{argmin}_{\boldsymbol{\Omega}} \left(\mathcal{L}(\mathbf{S}^*;\boldsymbol{\Omega}) + F_\lambda^1(\boldsymbol{\Omega})\right). \tag{6}$$

Then the scaled L_1 Mallows' distance between $\mathcal{L}(\mathbf{S};\hat{\boldsymbol{\Omega}})$ and $\mathcal{L}(\mathbf{S}^*;\hat{\boldsymbol{\Omega}}^*)$

$$\frac{1}{p^2} d_1\left(\mathcal{L}(\mathbf{S};\hat{\boldsymbol{\Omega}}), \mathcal{L}(\mathbf{S}^*;\hat{\boldsymbol{\Omega}}^*)\right) \to 0 \text{ almost surely.} \tag{7}$$

Automatically, Theorem 1 implies the (conditional) distribution of the bootstrap test statistic (4) converges to the distribution of the original test statistic (3) almost surely. In establishments of convergence, no assumption about the asymptotic distribution of the test statistics is imposed.

Furthermore, we do not require sparsity of the true precision matrix $\mathbf{\Omega}$, which can be dense and the trace of $\mathbf{\Omega}$ is allowed to go to infinite at the higher order of p, say $\mathrm{tr}(\mathbf{\Omega}) = O(p^2)$.

Theorem 1 justifies that the bootstrapped test statistic has the same distribution as that of the sampling distribution of the original test statistic in a high-dimensional situation (both p and $T \to \infty$). In other words, bootstrap continues to work even if regularization is imposed as long as a certain condition is met for the penalty functions. The proof of Theorem 1 is presented in the supplemental file.

4 Simulations

This section is devoted to simulation studies of operating characteristics of the proposed method with respect to detection of the structure/strength change.

Three methods are compared:

- **Original Likelihood Ratio Test**: This method is only examined in the case $p < n$. The precision matrix is estimated by inverting the sample covariance matrix. A chi-square test is employed and the degrees of freedom is set as the number of constraints in the hypotheses.

- **Penalized Likelihood Ratio Test**: The precision matrix is estimated by GLasso [16] where the penalty coefficient is set as $\lambda = \sqrt{\log p/n}$ [18]. A chi-square test is employed and the degrees of freedom is set as the number of constraints in the hypotheses.

- **Bootstrapped Penalized Likelihood Ratio Test**: The precision matrix is estimated by GLasso in each bootstrapped sample where the penalty coefficient is set as $\lambda = \sqrt{\log p/n}$ [18]. The test is performed following the proposed procedure in Section 2. The bootstrap size B is set to be 1000.

Several simulation examples are examined. In each example, 100 simulation replications are performed. The averaged empirical nominal levels and rejection rates at $\alpha = 0.05$ are compared.

Now consider a network of p nodes with $T = 4$ stages. A random sample of size $n_1 = \cdots = n_T = n$ is generated according to the Gaussian graphical model introduced in the beginning of Section 2.1. In all examples, we test $H_0 : \mathbf{\Omega}_1 = \mathbf{\Omega}_2 = \mathbf{\Omega}_3 = \mathbf{\Omega}_4$ v.s. $H_a : H_0$ is not true.

4.1 Performance comparison under H_0

Example 1: The following cases are considered: $n = 100$ and $p = 5, 10, 20, 30, 40, 50, 100, 200$. We set that $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 = \mathbf{\Omega}_3 = \mathbf{\Omega}_4$ with $\omega_{i,j,1} = \omega_{i,j,2} = \omega_{i,j,3} = \omega_{i,j,4} = 0.5^{|i-j|}$; $1 \leqslant i, j \leqslant p$.

Example 2: The parameters (n, p, T) are the same as in Example 1. However, $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 = \mathbf{\Omega}_3 = \mathbf{\Omega}_4$ and $\omega_{i,j,t} = 0.4$ as $|i - j| = p - 1$ and 0, otherwise; $t = 1, 2, 3, 4$; $1 \leqslant i, j \leqslant p$.

Tables 1-2 about here

As shown in Table 1 (Example 1), The Type-I error of the bootstrapped test statistic is close to or below the nominal level under H_0. The penalized LR test and penalized bootstrap LR test exhibit similar performances, while the original LR test does not perform well as $p \geqslant 10$.. In literature it has been testified that the chi-square approximation fails even as p^2 is of the same order of n.

As shown in Table 2 (Example 2), all four precision matrices are the same and sparse. The results are similar to Example 1, though the penalized LR and Bootstrapped penalized LR tests perform slightly better than Example 1 where the precision matrix is dense.

4.2 Performance comparison under H_a

Example 3: Let $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 \neq \mathbf{\Omega}_3 = \mathbf{\Omega}_4$: All diagonal elements are 1's. $\omega_{i,j,1} = \omega_{i,j,2} = 0.4$ as $|i - j| = 1$ and 0 otherwise; $\omega_{i,j,3} = \omega_{i,j,4} = 0.4$ as $|i - j| = 2$ and 0 otherwise; $1 \leqslant i, j \leqslant p$.

Example 4: Let $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 \neq \mathbf{\Omega}_3 = \mathbf{\Omega}_4$ where all diagonal elements are 1's and $\omega_{i,j,1} = \omega_{i,j,2} = 0.5^{|i-j|}$; $\omega_{i,j,3} = \omega_{i,j,4} = 0.2^{|i-j|}$; $1 \leqslant i, j \leqslant p$. The precision matrix is denser than Example 3, and thus brings more challenges.

Tables 3-4 about here

As demonstrated in Table 3 (Example 3), the proposed method achieves high power in that H_0 is rejected 100% in the all cases when H_0 is not true.

With the denser precision matrix, all three approaches lost their power at different measure as shown in Table 4 (Example 4). In particular, the performance of penalized LR test rejects H_0 with proportion 33% when $p \geqslant 20$ and with proportion 0 when $p \geqslant 30$. The bootstrap LR test rejects H_0 with proportion close to 100% (above 95% in all cases).

Overall, the bootstrapped log-likelihood ratio test performs well in that it yields high power while the Type-I error is under control at the nominal level.

Finally, we examine the distribution of the test statistic (3) in Examples 1 and 2, where the precision matrix is estimated by inverting the sample covariance matrix. As shown in Figures 1-4, the distribution of D may be neither the chi-square nor normal. Evidently, the test statistic is far from the chi-square distribution of degrees of freedom $(p^2 + p)(L - 1)/2$, which suggests that the Wilks test is no longer valid.

Figures 1-4 about here

We also studied the case where the precision matrix is estimated by GLasso and the approximation by normal or chi-square distributions still falls apart. The corresponding qqplots are not included. Note that the corresponding test statistic (3) estimated by GLasso may give negative values, which is not presented here.

In summary, the proposed procedure achieves high power under the nominal level and shows advantages over other methods.

5 Financial network inference

This section investigates the effect of Lehman Brothers' collapse on the network of 200 publicly-traded stocks. As in the foregoing discussion, a network is described by the corresponding precision matrix. Let $\mathbf{\Omega}_1$, $\mathbf{\Omega}_2$ and $\mathbf{\Omega}_3$ be the precision matrices corresponding to the three phases. Of particular interest is the change of financial networks over the three periods, pre-crisis, crisis and post-crisis.

5.1 Background

As described in the introduction, we extract log-returns of daily adjusted closing prices of 200 US stocks from 1/1/2005 to 12/31/2010, as listed in Table 5.

Table 5 about here

Among these $p = 200$ stocks, top 20 stocks by market capitalization as of 12/31/2010, are selected from each of the ten sectors, *basic industries, consumer durables, consumer nondurables, consumer services, energy, finance, health care, public utilities, technology* and *transportation* (http://www.nasdaq.com). Each sector is divided further into several industries. For example, the *finance* sector comprises of three major industries: *Major Banks*, *Investment Banks* and *Insurance*. The *energy* sector is composed of *Oil & Gas Production*, *Consumer Electronics* and other industries.

Three periods ($T = 3$) are considered: *Pre-crisis (1/1/2005-12/31/2005):* before Lehman Brothers' collapse, *Crisis (7/1/2008-6/30/2009):* the period during the process of Lehman Brothers' filing bankruptcy, and *Post-crisis (1/1/2010 -12/31/2010):* the recovery after Lehman Brothers' collapse. Accordingly, a network is assumed to be constant within each time period with $n_t = 250$ ($t = 1, 2, 3$) observations.

Our preliminary analysis of the sample covariance/correlation matrices of each of the three periods suggests that the pairwise correlations of these stocks became strong during the crisis period, but weak in the other two periods, because of the effect of dominating systemic factors such as market panic. However, it remains unclear how the pairwise associations may behave after the effects of systemic or common factors are removed. To make a formal inference, we employ multiple Gaussian graphical models (MGGMs) with three time periods, which seems suitable in that the normality assumption is approximately satisfied, as indicated by the Q-Q plots in Figure 5.

Figure 5 about here

MGGMs are used to model pairwise conditional dependencies with each stock corresponding one node in a graph at one time point, that is, no connecting edge between two nodes implies pairwise independence conditioning on all other $p - 2$ nodes [19]. Furthermore, the strength and sign of pairwise conditional dependencies between two nodes given all other $p - 2$ nodes are measured by the partial correlation coefficient in a Gaussian graphical model [26]. Roughly, "conditioning on all other $p - 2$ nodes" can be interpreted as "conditioning on the overall performance of all these stocks", or in some sense, "conditioning on macroeconomic environment". Of particular interest is how the network structure evolves since 2005 by inferring a MGGM. First, we test if the financial network remains the same across the three time periods, which is an example of the global inference in a high-dimensional situation. Second, we study how the linkage between two specific stocks, such as an investment bank and a retail bank, evolves.

The precision matrices are estimated by the graphical Lasso [16] for each period. In the following graphs, the vertex size stands for the estimated conditional (or partial) variance [26] of each stock and the edge stands for the estimated partial correlation coefficient [26] whose absolute values are above 0.1.

5.2 Major Findings

We perform a global hypothesis test with regard to the question about whether our financial network has become more interconnected after Lehman Brothers' collapse.

Figure 6 about here

As shown in Figure 6, the node size of pre-crisis is larger than the following two periods, indicating that the conditional variance of most stocks shrank during crisis and post-crisis periods because of increasing correlations among the 200 stocks. Consequently, the systemic risk plays a more dominant role in financial risk during recessions. Moreover, as displayed in Figure 6, the pre-crisis is prominently different from the other two periods. There are many small and pure sub-communities in the era of pre-crisis, which implies that the connection occurs mainly among stocks of the same sector while inter-sector linkage is rare and weak. Even the strong intra-sector linkage only exists in four sectors: *Public Utilities* (gray), *Energy* (cyan), *Finance* (magenta) and *Capital Goods* (red), which are more and less related to raw materials and infrastructure. During the crisis and post-crisis periods, one large diversified community is formed besides some small sub-communities. Thus, strong inter-sector connection happens much more frequently during the period of economic contraction.

Finally, it is worth noting that in the post-crisis period the economy is expanding and stock market is a bull market, but the topological structure is still similar to the crisis era but greatly different from the pre-crisis era.

Next we perform hypotheses tests at a significance level $\alpha = 0.05$ using the method developed in Section 2 with $B = 1000$ for bootstrap. Consider a null hypothesis of no changes $H_0 : \mathbf{\Omega}_1 = \mathbf{\Omega}_2 = \mathbf{\Omega}_3$

versus its alternative H_a : not H_0. The P-value for this test is 0.000, so H_0 is rejected to favor the hypothesis that a change has occurred. To further identify where a change occurs, consider a simultaneous test for three hypotheses: $H_0 : \mathbf{\Omega}_2 = \mathbf{\Omega}_3$ versus $H_a : \mathbf{\Omega}_2 \neq \mathbf{\Omega}_3$ to see if a change occurs after period two, $H_0 : \mathbf{\Omega}_1 = \mathbf{\Omega}_3$ versus $H_a : \mathbf{\Omega}_1 \neq \mathbf{\Omega}_3$, and $H_0 : \mathbf{\Omega}_1 = \mathbf{\Omega}_2$ versus $H_a : \mathbf{\Omega}_1 \neq \mathbf{\Omega}_2$ similarly. The corresponding Empirical nominal levels are 0.000,0.000 and 0.000. After adjusted for multiplicity, all three tests are simultaneously rejected at the overall level 0.05. This says that the collapse event impacts the post collapse period as well as the recover period in that the financial network's structure varies significantly.

Next we focus two sectors, *Finance* and *Energy*, to see if they are more susceptible than others.

Figure 7 about here

As shown in Figure 7, the *Finance* sector is more fragmented in booming economy than recession. In pre-crisis, there are two major sub-communities "Major Banks" including WFC, BAC,BK $\cdots$ and "Insurance" including MMC, PRL, AIC $\cdots$, but investment banks are well-separated from them. As the crisis culminated and Lehman Brother filed bankruptcy in September, 2008 almost all banks are merged into a single network. When the economy entered the recovery period in 2010, the insurance industry and major banks broke up again, but some investment banks connected with major banks and formed a larger sub-community. Before the financial crisis 2007-2009 investment banks including Lehman Brothers' raised capital and invested much like major banks but escaped regulation which enabled them to over-leverage and exacerbated system-wide contagion. During the financial crisis, however, the pure investment banks had to transform themselves to bank holding companies (BHC) to get government bailout money and the BHC status now subjects them to the additional oversight. As these investment banks restructured their assets and have to act under stricter scrutiny of the government, they are being forced to operate like a full-service bank. Thus, investment banks start to correlate with major banks.

Figure 8 about here

In all three phases, as displayed in Figure 8, there is always a large sub-community comprising of "oil & natural gas production" companies, as well as some small sub-communities and isolated points of the other industries. Undoubtedly, the major driving force in *Energy* is the crude oil price, and therefore, it leads to strong intra-sector correlation regardless of the macroeconomic performance. Different from the *Finance* sector, the shake-up and reorganization did not occur in the *Energy* sector.

5.3 Conclusion

The financial network has experienced a substantial transmission and connectivity has greatly expanded and intensified since Lehman Brothers' collapse. In a booming economy, sectorial factors mainly drive price movement, so inter-sector connection seldom happens. In a financial crisis, two types of schemes are contributing to the expansion and intensification of both intra- and inter-sector connections: systemic factors and partial correlation coefficients. However, the partial correlation coefficient of stocks of some sectors like *Energy* does not change as other sectors. When the economy recovers again in the wake of Lehman Brothers' bankruptcy, the network does not return to the status of pre-crisis. The effect of Lehman Brothers' breakdown has branded the network because the far-range and extreme financial and monetary measures imposed by the government have become a dominating force in the stock market, especially for stocks in the *Finance* sector.

6 Discussion

Globalization and advancement of information sharing technology have weaved financial markets and institutions everywhere into a large network, which presents challenges to all aspects of risk management and policy-making. Inference concerning large networks is in high demand as ever-increasing connectivity and complexity are seen in the financial system, particularly when the network structure experiences sharp change and exhibits stage-wise patterns due to unexpected external shocks. Nowadays, graphical models provide new means for dealing with these challenges.

The proposed inferential tools allow us to study how the financial network evolved across three periods, pre-crisis, crisis and post-crisis, based on data of daily prices of 200 stocks from 10 sectors. Theoretically, we justify that the bootstrap approximation of the sampling distribution is valid in a high-dimensional situation where the number of stages T and the number of nodes p grow with the sample size. In simulations, we demonstrate that the proposed method compares favorably against its competitors in terms of type 1 and 2 errors. To make the proposed method useful in practice, more investigation is necessary.

References

[1] Abadir, K.M. and Magnus, J.R. (2005) *Matrix algebra*. Cambridge University Press.

[2] Bickel, P. J. and Freedman, D. A. (1981) Some asymptotic theory for the bootstrap. *Annals of Statistics*, **9**, 11961217

[3] Bühlmann, P. and van de Geer, S. (2011) *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer-Verlag: Berlin Heidelberg.

[4] Cai T.T., Li, H., Liu, W. and Xie, J. (2016) Joint estimation of multiple high-dimensional precision matrices. *Statistica Sinica*, **26**, 445-464

[5] Camponovo, L. (2015) On the validity of the pairs bootstrap for lasso estimators. *Biometrika*, **102**, 981-987

[6] Chatterjee, A. and Lahiri, S. N. (2012) Bootstrapping Lasso estimators. *Journal of the American Statistical Association*, **106**, 608-625

[7] Chernozhukov V., Chetverikov, D. and Kato, K. (2015) Central limit theorems and bootstrap in high dimensions. *manuscript*

[8] Diebold, F. X., and Yilmaz, K. (2015): *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press

[9] Efron, B (1981) Nonparametric standard errors and confidence intervals (with discussions). *Canadian Journal of Statistics*, **9**, 139-172

[10] Efron, B. (2014) Estimation and accuracy after model selection. *Journal of the American Statistical Association*, **109**, 991-1007

[11] Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*. Chapman & Hall/CRC.

[12] Fan, J., Fan, Y. and Lv, J. (2008) High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, **147**, 186-197.

[13] Fan, J., Feng, Y. and Wu, Y. (2009) Network exploration via the adaptive lasso and scad penalities. *The Annals of Applied Statistics*, **3**, 521-541.

[14] Fan, J., Liao, Y. and Liu, H. (2016) An overview on the estimation of large covariance and precision matrices. *Econometrics Journal*, published online.

[15] Fan, J., Lv, J. and Qi, L. (2011) Sparse high-dimensional models in economics. *Annual Review of Economics*, **3**, 291-317.

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

[17] Hall, P. (1987) On the bootstrap and likelihood-based confidence regions. *Biometrika*, **74**, 481-493

[18] Jankova J. and van de Geer, S. (2015) Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, **9**, 1205-1229.

[19] Lauritzen, S. L. (1996) *Graphical Models*. Clarendon Press: Oxford.

[20] Liu, H., Han, F. and Zhang, C.-H. (2012) Transelliptical graphical models. *Advances in Neural Information Processing Systems*, 809-817

[21] Li, S., Hsu, L., Peng, J. and Wang, P. (2013) Bootstrap inference for network construction with an application to a breast cancer microarray study. *Annals of Applied Statistics*, **7**, 391-417.

[22] Qiu, H., Han, F., Liu, H. and Caffo, B. (2016) Joint estimation of multiple graphical models from high dimensional time series. *Journal of the Royal Statistical Society: Series B*, **78**, 487-504

[23] Shao, J. and Tu, D. (1995) *The Jackknife and Bootstrap*. Springer-Verlag: New York.

[24] Shen, X., Pan, W., and Zhu, Y. (2012) Likelihood-based selection and sharp parameter estimation. *Journal of the American Statistical Association*, **107**, 223-232

[25] Roman Vershynin, *High Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge University Press, 2018.

[26] Whittaker, J. (1990) *Graphical Models in Applied Multivariate Statistics.* John Wiley and Sons: Chichester.

[27] Zhang, C.-H. and Zhang, S. (2014) Confidence intervals for low-dimensional parameters in high-dimensional linear models. *Journal of the Royal Statistical Society: Series B*, **76**, 217-242.

[28] Zhang, Y. and Shen, X. (2015) Adaptive modeling procedure selection by data perturbation. *Journal of Business & Economic Statistics*, **33**, 541-551.

[29] Zhu, Y., Shen, X. and Pan, W. (2014) Structural pursuit over multiple undirected graphs. *Journal of the American Statistical Association*, **109**, 1683-1696.

[30] Zhu, Y., Shen, X. and Pan, W. (2016). Maximum likelihood inference for a large precision matrix. *manuscript*

parameters		penalized Lr		original Lr		bootstrap Lr	
n	p	Empirical level	Rej	Empirical level	Rej	Empirical level	Rej
100	5	0.522	0.03	0.422	0.05	0.442	0.04
100	10	0.821	0	0.377	0.12	0.495	0.03
100	20	0.999	0	0.087	0.63	0.498	0.01
100	30	1	0	0.002	0.99	0.55	0
100	40	1	0	0	1	0.563	0
100	50	1	0	0	1	0.544	0
100	100	1	0	NA	NA	0.368	0
100	200	1	0	NA	NA	0.227	0

Table 1: Example 1: Proportion of rejections based on 100 simulation replications with $T = 4$. $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 = \mathbf{\Omega}_3 = \mathbf{\Omega}_4$ and $\omega_{i,j,t} = 0.5^{|i-j|}$; $t = 1, 2, 3, 4$; $1 \leqslant i, j \leqslant p$. Three methods are compared, including original likelihood ratio test, penalized likelihood ratio test, penalized bootstrapped likelihood ratio test.

parameters		penalized Lr		original Lr		bootstrap Lr	
n	p	Empirical level	Rej	Empirical level	Rej	Empirical level	Rej
100	5	0.661	0.01	0.455	0.07	0.517	0.01
100	10	0.954	0	0.367	0.08	0.633	0.01
100	20	1	0	0.083	0.62	0.783	0
100	30	1	0	0.001	1	0.88	0
100	40	1	0	0	1	0.94	0
100	50	1	0	0	1	0.961	0
100	100	1	0	NA	NA	0.986	0
100	200	1	0	NA	NA	0.958	0

Table 2: Example 2: Proportions of rejection based on 100 simulation replications with $T = 4$. $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 = \mathbf{\Omega}_3 = \mathbf{\Omega}_4$ and $\omega_{i,j,t} = 0.4$ as $|i - j| = p - 1$ and 0, otherwise; $t = 1, 2, 3, 4$; $1 \leqslant i, j \leqslant p$. Three methods are compared, including original likelihood ratio test, penalized likelihood ratio test, penalized bootstrapped likelihood ratio test.

parameters		penalized Lr		original Lr		bootstrap Lr	
n	p	Empirical level	Rej	Empirical level	Rej	Empirical level	Rej
100	5	0	1	0	1	0	1
100	10	0	1	0	1	0	1
100	20	0	1	0	1	0	1
100	30	0	1	0	1	0	1
100	40	0	1	0	1	0	1
100	50	0	1	0	1	0	1
100	100	1	0	NA	NA	0	1
100	200	1	0	NA	NA	0	1

Table 3: Example 3: Proportions of rejection based on 100 simulation replications with $T = 4$. $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 \neq \mathbf{\Omega}_3 = \mathbf{\Omega}_4$: All diagonal elements are 1's. $\omega_{i,j,1} = \omega_{i,j,2} = 0.4$ as $|i-j| = 1$ and 0 otherwise; $\omega_{i,j,3} = \omega_{i,j,4} = 0.4$ as $|i-j| = 2$ and 0 otherwise; $1 \leqslant i, j \leqslant p$. Three methods are compared, including original likelihood ratio test, penalized likelihood ratio test, penalized bootstrapped likelihood ratio test.

parameters		penalized Lr		original Lr		bootstrap Lr	
n	p	Empirical level	Rej	Empirical level	Rej	Empirical level	Rej
100	5	0.003	0.99	0.001	1	0.002	1
100	10	0.002	0.99	0	1	0	1
100	20	0.181	0.33	0	1	0.001	0.99
100	30	0.956	0	0	1	0.001	1
100	40	1	0	0	1	0.005	0.98
100	50	1	0	0	1	0.011	0.96
100	100	1	0	NA	NA	0.019	0.99
100	200	1	0	NA	NA	0.018	1

Table 4: Example 4: Proportion of rejections based on 100 simulation replications with $T = 4$. $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 \neq \mathbf{\Omega}_3 = \mathbf{\Omega}_4$: All diagonal elements are 1's. $\omega_{i,j,1} = \omega_{i,j,2} = 0.5^{|i-j|}$; $\omega_{i,j,3} = \omega_{i,j,4} = 0.2^{|i-j|}$; $1 \leqslant i, j \leqslant p$. Three methods are compared, including original likelihood ratio test, penalized likelihood ratio test, penalized bootstrapped likelihood ratio test.

Basic Industries	PG	DOW	DD	MON	PCP	ECL	PX	APD	PPG	GLW
	SCCO	IP	VMC	NUE	CHD	SRCL	EMN	IFF	LEN	NEM
Consumer Durables	BA	UTX	HON	LMT	DHR	TMO	F	GD	RTN	CAT
	NOC	ILMN	DE	ROP	PCAR	APH	SWK	A	PH	ROK
Consumer Non-Durables	KO	PEP	MO	NKE	RAI	MDLZ	CL	GIS	EL	STZ
	MNST	K	VFC	SYY	HRL	ADM	TSN	HSY	CAG	CPB
Consumer Services	AMZN	WMT	DIS	HD	CMCSA	MCD	SBUX	COST	LOW	SPG
	TWX	FOX	TJX	NFLX	TGT	PSA	CCL	AMT	KR	FOXA
Energy	XOM	GE	CVX	OXY	COP	EOG	VLO	EMR	HAL	BHI
	APC	PXD	CMI	APA	NBL	NOV	CAM	HES	TSO	DVN
Finance	WFC	JPM	BAC	C	AIG	USB	GS	AXP	BLK	MS
	MET	PNC	BK	SCHW	COF	TRV	PRU	CME	MMC	BBT
Healthcare	JNJ	PFE	MRK	MDT	GILD	AMGN	AGN	UNH	BMY	CVS
	LLY	MMM	CELG	BIIB	ABT	REGN	ESRX	MCK	AET	ALXN
Public Utilities	T	VZ	DUK	NEE	SO	D	AEP	EXC	PCG	WM
	SRE	PPL	PEG	ED	EIX	XEL	LVLT	WEC	RSG	WMB
Technology	AAPL	MSFT	ORCL	INTC	IBM	CSCO	QCOM	TXN	EMC	ADBE
	ADP	CTSH	ITW	YHOO	ATVI	INTU	EA	FISV	CERN	AMAT
Transportation	UPS	UNP	FDX	LUV	CSX	NSC	CHRW	ALK	EXPD	JBHT
	KSU	JBLU	ODFL	GWR	LSTR	HA	WERN	AIRM	HTLD	BCO

Table 5: Stock list from ten sectors.