LARGE MULTIPLE GRAPHICAL MODEL INFERENCE VIA BOOTSTRAP

Yongli Zhang, Xiaotong Shen and Shaoli Wang

*University of Oregon, University of Minnesota, Minneapolis
and Shanghai University of Finance and Economics*

Abstract: Large economic and financial networks may experience stage-wise changes as a result of external shocks. To detect and infer a structural change, we consider an inference problem within a framework of multiple Gaussian Graphical Models when the number of graphs and the dimension of graphs increase with the sample size. In this setting, two major challenges emerge as a result of the bias and uncertainty inherent in the regularization required to treat such overparameterized models. To deal with these challenges, the bootstrap method is utilized to approximate the sampling distribution of a likelihood ratio test statistic. We show theoretically that the proposed method leads to a correct asymptotic inference in a high-dimensional setting, regardless of the distribution of the test statistic. Simulations show that the proposed method compares favorably to its competitors such as the Likelihood Ratio Test. Finally, our statistical analysis of a network of 200 stocks reveals that the interacting units in the financial network become more connected as a result of the financial crisis between 2007 and 2009. More importantly, certain units respond more strongly than others. Furthermore, after the crisis, some changes weaken, while others strengthen.

Key words and phrases: Bootstrap, graphical models, high-dimensional inference, model selection, regularization.

1. Introduction

In economics, network analyses play a fundamental role in studying consumer behavior and international trade. In finance, network analyses help to identify financial contagion and minimize systemic risk, thus preventing future crises (Diebold and Yılmaz (2015)). However, estimations and inferences involving large networks often face problems related to high dimensionality (Fan, Fan and Lv (2008); Fan, Liao and Liu (2016)). For example, exploring a network of 200 stocks involves $(200^2 + 200)/2 = 20{,}100$ pairwise edges when the sample size is in the hundreds. In such cases, classical inference approaches become invalid or break down. Refer to Fan, Lv and Qi (2011) for a discussion on

the problems high dimensions cause in inferences. Even worse, as the number of graphs increases with the sample size or the number of nodes in a multiple graphical model (MGM), which is often used to model networks that experience stage-wise change as a result of external forces, estimations and inferences become more difficult. The additional challenges when estimating an MGM, over and above those of a single graphical model, are studied in Zhu, Shen and Pan (2014); Cai et al. (2016); Qiu et al. (2016). To respond to these challenges, we develop inference tools for large stage-wise networks defined by MGMs.

We employ a multiple Gaussian graphical model (MGGM) to model stage-wise networks. Here, the number of model parameters may greatly exceed the sample size, and the number of stages may increase with the numbers of nodes and/or observations. Two major challenges emerge. First, regularization is often used to treat overparameterized models in a high-dimensional setting, introducing bias into the estimation and, thus a biased inference. Moreover, the usual asymptotic approximations for the sampling distribution of a test statistic become inadequate (Jankova and De Geer (2015)). Second, the selection uncertainty inherent in regularization is mathematically intractable, even in a low-dimensional setting (Zhang and Shen (2015)). For graphical model inferences, Zhu, Shen and Pan (2016) proposed a maximum likelihood inference approach for a simple Gaussian graphical model (GGM). Furthermore, they derived the asymptotic distribution of the constrained likelihood ratio as a chi-square or a normal distribution, depending on the size of the co-dimension of the inference space.

In this study, we develop a likelihood ratio inference approach for MGGMs, where the bootstrap method is utilized to approximate the sampling distribution of a test statistic in order to account for the bias and selection uncertainty caused by the regularization. The benefits of this approach are demonstrated numerically and theoretically. As shown 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 $n^{1/2}$ (Bai et al. (2009); Zhu, Shen and Pan (2016)). However, such approximation becomes inadequate when p is larger, as shown by our simulations in Figures 1–4. In this sense, the bootstrap method offers an attractive alternative to asymptotic high-dimensional inferences.

Relatively few studies have applied GGMs within the fields of economics and

finance. (Fan, Fan and Lv (2008)) developed methods for estimating large co-variance and precision matrices for economic and financial data. To the best of 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 increase. Importantly, our method enables us to identify the type, origin, and evolution of the interactions between nodes. With the help of the proposed inference method, we examine the impact of the Lehman Brothers bankruptcy on financial networks by analyzing the historical prices of 200 stocks traded publicly in the United States between January 1, 2005, and December 31, 2010. In particular, we investigate the changes in the structure and strength of the network over time as a result of the Lehman Brothers collapse. We compare our results with those of Liu, Han and Zhang (2012), who examined a static network of 452 stocks during the boom cycle between January 1, 2003, and January 1, 2008. As suggested by our analysis, the financial network has experienced profound changes since the Lehman Brothers bankruptcy. Overall, connectivity becomes more widespread and strong, although individual sectors exhibit disparate patterns.

The remainder of this paper is organized as follows. Section 2 formulates the problem and proposes our inference method. The theoretical validity of the method is proved in Section 3. Section 4 discusses our simulation results, and Section 5 presents a data analysis. Section 6 concludes the paper.

2. Inference

This section develops a likelihood inference method for large networks based on MGGMs.

2.1. MGGMs

To model a stock network that experiences stages $t = 1, \ldots, T$, we consider an MGGM of T graphs $(GR_1, \ldots, GR_T)$, each of which represents a network at different stages. For an inference, T independent random samples $\boldsymbol{Y} = (\boldsymbol{Y}^t)_{t=1}^T$ are obtained, where $\boldsymbol{Y}^t = (\boldsymbol{Y}_1^t, \ldots, \boldsymbol{Y}_{n_t}^t)$ are n_t independent and identically distributed (i.i.d.) p-dimensional random vectors. Here, $\boldsymbol{Y}_k^t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t)$, for $1 \leq k \leq n_t$, where $\boldsymbol{0}$ is a p-dimensional zero vector, $\boldsymbol{\Sigma}_t = (\sigma_{i,j,t})_{1 \leq i,j \leq p}$ is a $p \times p$ covariance matrix, and the sample size is $N = \sum_{t=1}^T n_t$. Suppose $\max_t \operatorname{tr}(\boldsymbol{\Sigma}_t)/p^2 \to 0$. In this case, the precision matrix $\boldsymbol{\Omega}_t = (\omega_{i,j,t})_{1 \leq i,j \leq p} = (\boldsymbol{\Sigma}_t)^{-1}$ is the inverse of the covariance matrix, which has an off-diagonal entry $\omega_{i,j,t}$ of zero if and only

if the nodes i and j are conditionally independent, given all other $p - 2$ nodes at time t, for $t = 1, \ldots, T$ (Whittaker (1990)). Thus, $\boldsymbol{\Omega}_t$ uniquely encodes an undirected graph at t, for $t = 1, \ldots, T$. Both the number of stages T and the number of nodes p are allowed to increase.

We suppose the sample mean $\bar{\mathbf{Y}}_t = n_t^{-1} \sum_{k=1}^{n_t} \mathbf{Y}_k^t = \mathbf{0}$ $(t = 1, \ldots, T)$. 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 $\boldsymbol{\Omega}_t$ is a positive-definite and symmetric $p \times p$ matrix, for $t = 1, \ldots, T$. Let $\mathbf{Z}_k^t = \boldsymbol{\Omega}_t^{1/2} \mathbf{Y}_k^t$. 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 whether a network stays the same across all T stages. To achieve this, we approximate the test statistic, which is based on the logarithm of the likelihood ratio, by bootstrapping. Then, we establish its theoretical validity in the high-dimensional case, in which regularization is imposed to achieve sparsity, thus introducing bias to the inference.

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

$$\mathcal{L}(\mathbf{S}; \mathbf{G}) = \mathrm{tr}(\mathbf{S}\mathbf{G}) - \log \det(\mathbf{G}) \qquad (2.1)$$

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

$$\mathcal{C}(\mathbf{G}_1, \ldots, \mathbf{G}_T) = \sum_{t=1}^{T} n_t \big(\mathcal{L}(\mathbf{S}_t; \mathbf{G}_t) + F_\lambda^v(\mathbf{G}_t) \big), \qquad (2.2)$$

where $\mathcal{L}(\mathbf{S}_t; \mathbf{G}_t)$, the negative log-likelihood, is the empirical loss function, and $F_\lambda^v(\mathbf{G}_t)$ is a penalty regularizing $(\mathbf{G}_1, \ldots, \mathbf{G}_T)$. Here, we consider the following L_1 $(v = 1)$ and L_2 $(v = 2)$ 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, the optimal choice of the regularization parameters for an inference may depend on the sample size, which must be estimated from the data. Here, we adopt $\lambda = \sqrt{\log p / n}$ (Jankova and De Geer (2015); Zhang and Zhang (2014)) for the L_1-regularization (i.e., graphical Lasso (GLasso)) to guard against introducing an overly large bias in the pursuit of sparsity. The penalized maximum likelihood estimate $(\widehat{\boldsymbol{\Omega}}_1, \ldots, \widehat{\boldsymbol{\Omega}}_T)$ is then defined as $\mathrm{argmin}_{\mathbf{G}_1, \ldots, \mathbf{G}_T} \mathcal{C}(\mathbf{G}_1, \ldots, \mathbf{G}_T)$.

2.2. Inference

High-dimensional inferences may involve hypothesis testing, the construction of confidence intervals or regions, and the derivation of sampling distributions. In our case study, the number of unknown parameters $(p^2 + p)T/2 = 60{,}300$ ($p = 200, T = 3$) 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 a biased inference. Second, a form of regularization is often necessary for estimation, which undoubtedly introduces bias and selection uncertainty to an inference, particularly when L_1-regularization is employed. This is analogous to the problem often encountered in inferences after model selection (Efron (2014); Zhang and Shen (2015)), which, as expected, yields highly biased inferences, either optimistically or pessimistically, depending highly on whether the parameter of interest is included in the final model after regularization. To deal with these challenges, we use the bootstrap method to provide an alternative. As such, we do not need to derive asymptotic approximations as part of the inference, but can still account for the effect of the bias and selection uncertainty caused by regularization.

Our bootstrap procedure proceeds as follows:

Step 1: *Draw B bootstrap samples $(\mathbf{Y}_1^*, \ldots, \mathbf{Y}_B^*)$ from the original sample $\mathbf{Y}$ 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 (2.2) based on $\mathbf{Y}_b^*$ ($b = 1, \ldots, B$; $t = 1, \ldots, T$) for a preselected regularization coefficient λ.*

Step 3: *The original test statistic is*

$$D = \frac{1}{Np^2} \sum_t \left(n_t \mathcal{L}(\mathbf{S}_t; \widehat{\mathbf{\Omega}}_{t,b}^0) - n_t \mathcal{L}(\mathbf{S}_t; \widehat{\mathbf{\Omega}}_{t,b}^1) \right), \quad (2.3)$$

and the bootstrapping test statistic is

$$D_b^* = \frac{1}{Np^2} \sum_t \left(n_t \mathcal{L}(\mathbf{S}_t^*; \widehat{\mathbf{\Omega}}_{t,b}^{0*}) - n_t \mathcal{L}(\mathbf{S}_t^*; \widehat{\mathbf{\Omega}}_{t,b}^{1*}) \right). \quad (2.4)$$

An inference is made based on the empirical distribution of the B bootstrapped test statistics $\{D_1^*, \ldots, D_B^*\}$. The null hypothesis H_0 is rejected when $D > q_{1-\alpha}^*$, with a Type-I error at α, where $q_{1-\alpha}^*$ is the $(1 - \alpha)$ percentile of $\{D_1^*, \ldots, D_B^*\}$. For this test, the P-value is $\#\{D_b^* > D\}/B$ (Shao and Tu (1995, Sec. 4.5)).

Next, we discuss some of the steps in greater detail.

Consider the 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; thus, all parameters are regularized. Our bootstrap method generates B bootstrap samples $(\mathbf{Y}_1^*, \ldots, \mathbf{Y}_B^*)$ of size N from the combined sample $\mathbf{Y} = (\mathbf{Y}^1, \ldots, \mathbf{Y}^T)$ under $H_0 : \boldsymbol{\Omega}_1 = \cdots = \boldsymbol{\Omega}_T$. Under H_0, $\widehat{\boldsymbol{\Omega}}_b^{0*}$ is obtained from the combined bootstrap sample $\mathbf{Y}_b^*$, which is an $N \times p$ matrix, and the negative log-likelihood; that is, the loss is $\mathcal{L}(\mathbf{S}_b^*; \widehat{\boldsymbol{\Omega}}_b^{0*})$, where $\mathbf{S}_b^* = (1/N)(\mathbf{Y}_b^*)'\mathbf{Y}_b^*$. To calculate the likelihood under $H_0 \bigcup H_a$, partition $\mathbf{Y}_b^*$ into T disjoint subsamples $\mathbf{Y}_b^{*1}, \ldots, \mathbf{Y}_b^{*T}$ of size $n_1, \ldots, n_T$, and derive the negative log-likelihood $\mathcal{L}(\mathbf{S}_{t,b}^*; \widehat{\boldsymbol{\Omega}}_{t,b}^{1*})$, where $\widehat{\boldsymbol{\Omega}}_{t,b}^{1*}$ is the estimated precision matrix based on $\mathbf{S}_{t,b}^* = (1/n_t)(\mathbf{Y}_b^{*t})'(\mathbf{Y}_b^{*t})$, for $t = 1 \ldots, T; b = 1, \ldots, B$.

For a global inference of a single precision matrix, the log-likelihood ratio converges to a chi-square distribution (Zhu, Shen and Pan (2016)) 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 (Zhu, Shen and Pan (2016)), provided that the rate of growth of the co-dimension is not too fast. As shown in our simulations, the likelihood ratio cannot be well approximated by either the chi-square or the normal distribution in a high-dimensional setting. In this sense, the bootstrap method becomes attractive. Note that our method allows T and p in the MGGMs to increase with the sample size. In contrast, the usual asymptotic chi-square approximation breaks down for a global inference, whereas the bootstrap method continues to work, as illustrated 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 a minimum at $\boldsymbol{\Omega}$, which is the true precision matrix.

Next, we establish the closeness between the likelihood ratio test statistic and its bootstrapped version, where the precision matrices are both estimated using the GLasso method with the regularization parameter $\lambda = \sqrt{\log p/n}$, in terms of the Mallows' distance (Bickel and Freedman (1981); Shao and Tu (1995)).

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

$$\widehat{\mathbf{\Omega}} = \operatorname{argmin}_{\mathbf{\Omega}} \left(\mathcal{L}(\mathbf{S}; \mathbf{\Omega}) + F_\lambda^1(\mathbf{\Omega}) \right); \tag{3.1}$$

$$\widehat{\mathbf{\Omega}}^* = \operatorname{argmin}_{\mathbf{\Omega}} \left(\mathcal{L}(\mathbf{S}^*; \mathbf{\Omega}) + F_\lambda^1(\mathbf{\Omega}) \right). \tag{3.2}$$

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

$$\frac{1}{p^2} d_1 \left(\mathcal{L}(\mathbf{S}; \widehat{\mathbf{\Omega}}), \mathcal{L}(\mathbf{S}^*; \widehat{\mathbf{\Omega}}^*) \right) \to 0, \;\; almost \; surely. \tag{3.3}$$

Theorem 1 implies that the (conditional) distribution of the bootstrap test
statistic (2.4) converges to the distribution of the original test statistic (2.3),
almost surely. In establishing convergence, no assumption is imposed about the
asymptotic distribution of the test statistics. Furthermore, we do not require
sparsity of the true precision matrix $\mathbf{\Omega}$, which can be dense. In addition, the
trace of $\mathbf{\Omega}$ is allowed to go to infinity at a higher order of p, say $\operatorname{tr}(\mathbf{\Omega}) = O(p^2)$.

Theorem 1 states that the bootstrapped test statistic is distributed in the
same way as the sampling distribution of the original test statistic in a high-
dimensional setting (both p and $T \to \infty$). In other words, the bootstrap method
continues to work, even when regularization is imposed, as long as a certain
condition is met for the penalty functions. The proof of Theorem 1 is presented
in the Supplementary Material.

4. Simulations

This section presents the simulation studies to examine the operating char-
acteristics of the proposed method with respect to detecting changes in struc-
ture/strength.

Three methods are compared:

- **Original Likelihood Ratio Test**: This method is only examined for 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

using the GLasso method (Friedman, Hastie and Tibshirani (2008)), where the penalty coefficient is set as $\lambda = \sqrt{\log p/n}$ (Jankova and De Geer (2015)). 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 using the GLasso method in each bootstrapped sample, where the penalty coefficient is set as $\lambda = \sqrt{\log p/n}$ (Jankova and De Geer (2015)). The test is performed following the proposed procedure, described in Section 2. The bootstrap size B is set to 1,000.

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

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 at the beginning of Section 2.1. In all examples, we test $H_0 : \boldsymbol{\Omega}_1 = \boldsymbol{\Omega}_2 = \boldsymbol{\Omega}_3 = \boldsymbol{\Omega}_4$ versus $H_a : H_0$ 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 $\boldsymbol{\Omega}_1 = \boldsymbol{\Omega}_2 = \boldsymbol{\Omega}_3 = \boldsymbol{\Omega}_4$, with $\omega_{i,j,1} = \omega_{i,j,2} = \omega_{i,j,3} = \omega_{i,j,4} = 0.5^{|i-j|}$, for $1 \le i, j \le p$.

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

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 performance, whereas the original LR test does not perform well as $p \ge 10$. Prior studies have shown that the chi-square approximation fails even when p^2 is of the same order as that of n Bai et al. (2009); Zhu, Shen and Pan (2016).

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

Table 1. Example 1: Proportions of rejections based on 100 simulation replications with $T = 4$. $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 = \mathbf{\Omega}_3 = \mathbf{\Omega}_4$ and $\omega_{i,j,t} = 0.5^{|i-j|}$; $t = 1, 2, 3, 4$; $1 \leq i, j \leq p$. Three methods are compared: the original likelihood ratio test, penalized likelihood ratio test, and 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	0.522	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 2. Example 2: Proportions of rejections based on 100 simulation replications with $T = 4$. $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 = \mathbf{\Omega}_3 = \mathbf{\Omega}_4$ and $\omega_{i,j,t} = 0.4$ as $|i - j| = p - 1$ and zero otherwise; $t = 1, 2, 3, 4$; $1 \leq i, j \leq p$. Three methods are compared: the original likelihood ratio test, penalized likelihood ratio test, and 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

4.2. Performance comparison under H_a

Example 3. Let $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 \neq \mathbf{\Omega}_3 = \mathbf{\Omega}_4$, where all diagonal elements are ones. $\omega_{i,j,1} = \omega_{i,j,2} = 0.4$ as $|i - j| = 1$, and zero otherwise; $\omega_{i,j,3} = \omega_{i,j,4} = 0.4$ as $|i - j| = 2$, and zero otherwise, for $1 \leq i, j \leq p$.

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

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

Table 3. Example 3: Proportions 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 ones. $\omega_{i,j,1} = \omega_{i,j,2} = 0.4$ as $|i - j| = 1$, and zero otherwise; $\omega_{i,j,3} = \omega_{i,j,4} = 0.4$ as $|i - j| = 2$, and zero otherwise; $1 \leq i, j \leq p$. Three methods are compared: the original likelihood ratio test, penalized likelihood ratio test, and 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

With denser precision matrices, all three approaches lost power at different measures, as shown in Table 4 (Example 4). In particular, the penalized LR test rejects H_0 with proportion 33% when $p \geq 20$, and with proportion 0% when $p \geq 30$. 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 keeping the Type-I error under control at the nominal level.

Finally, we examine the distribution of the test statistic (2.3) 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 not be chi-square or normal. Evidently, the test statistic is quite different to the chi-square distribution with degrees of freedom $(p^2 + p)(L - 1)/2$, which suggests that the Wilks test is no longer valid.

We also studied the case where the precision matrix is estimated using the GLasso method and an approximation using a normal or chi-square distribution still falls apart. The corresponding Q–Q plots are not included. Note that the test statistic (2.3) estimated using the GLasso method may give negative values; this case is not presented here.

In summary, the proposed procedure achieves high power under the nominal level and shows advantages over other methods.
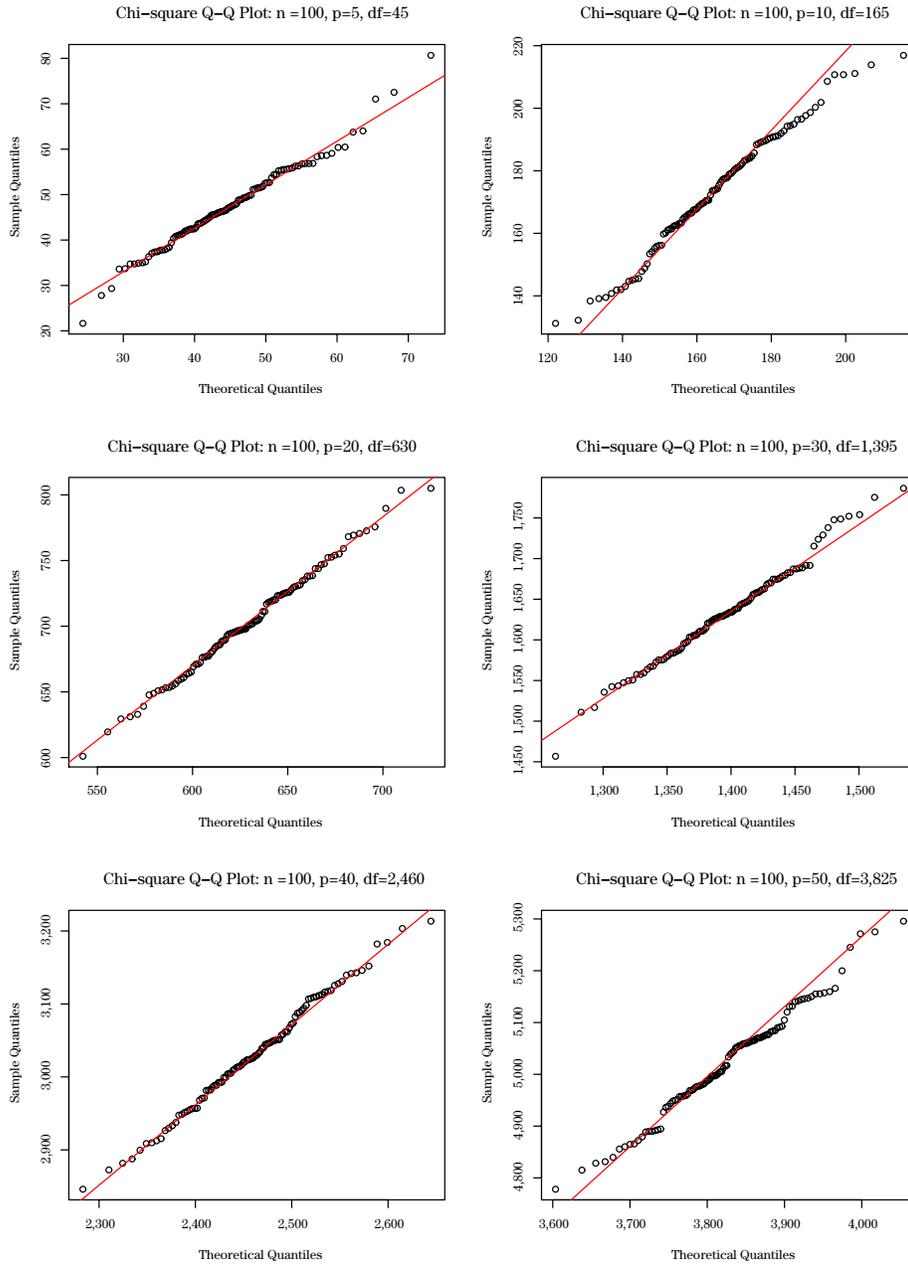
Figure 1. The Q-Q plots of test statistics of 100 independent repetitions as of Example 1 under $H_0 : \mathbf{\Omega}_1 = \cdots = \mathbf{\Omega}_4$ holds. The precision matrix is estimated by by inverting the sample covariance matrix. In the chi-square Q-Q plots, the degrees of freedom are determined by the number of constraints, $(p^2 + p)(T - 1)/2$.
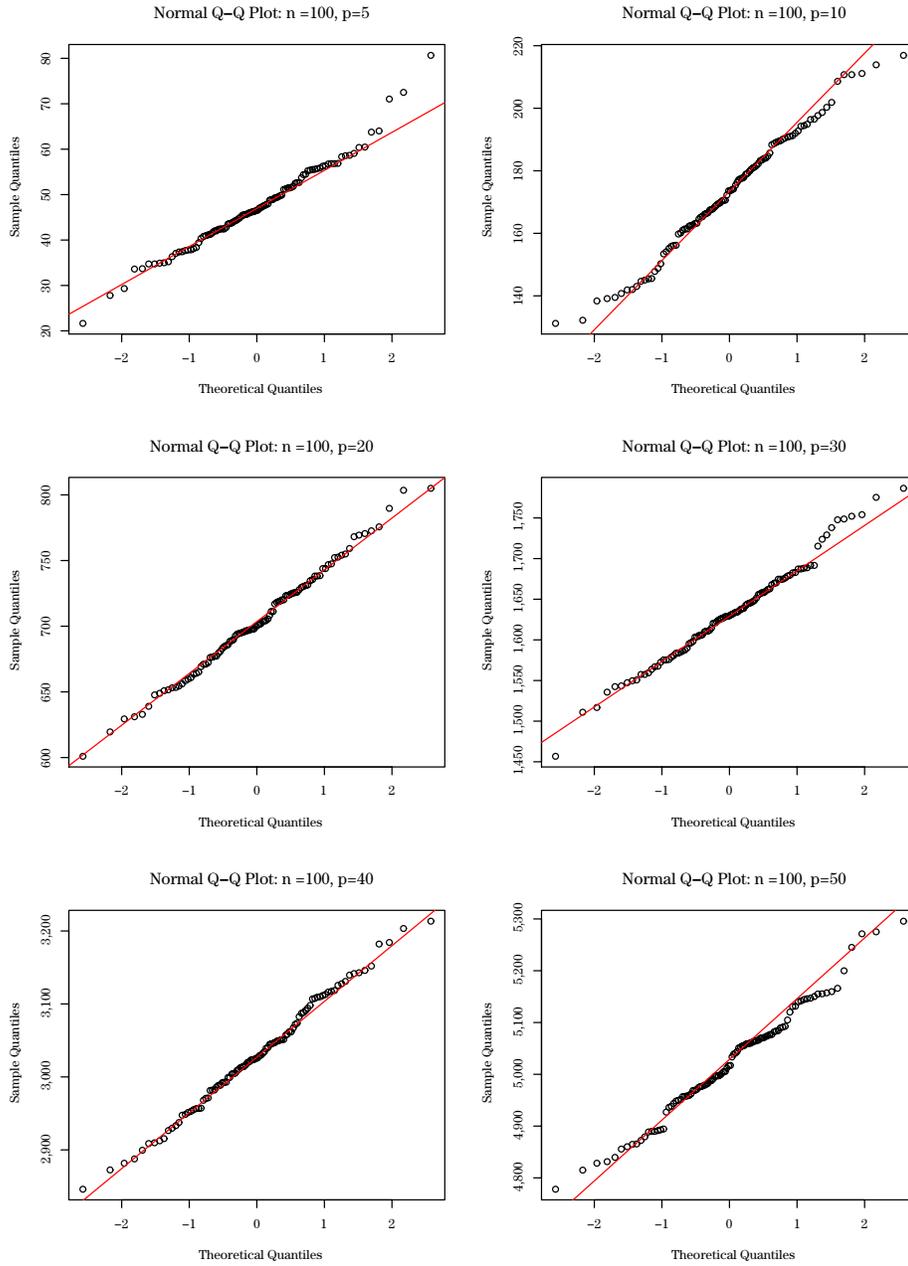
Figure 2. The Q-Q plots of test statistics of 100 independent repetitions as of Example 1 under $H_0 : \boldsymbol{\Omega}_1 = \cdots = \boldsymbol{\Omega}_4$ holds. The precision matrix is estimated by by inverting the sample covariance matrix.
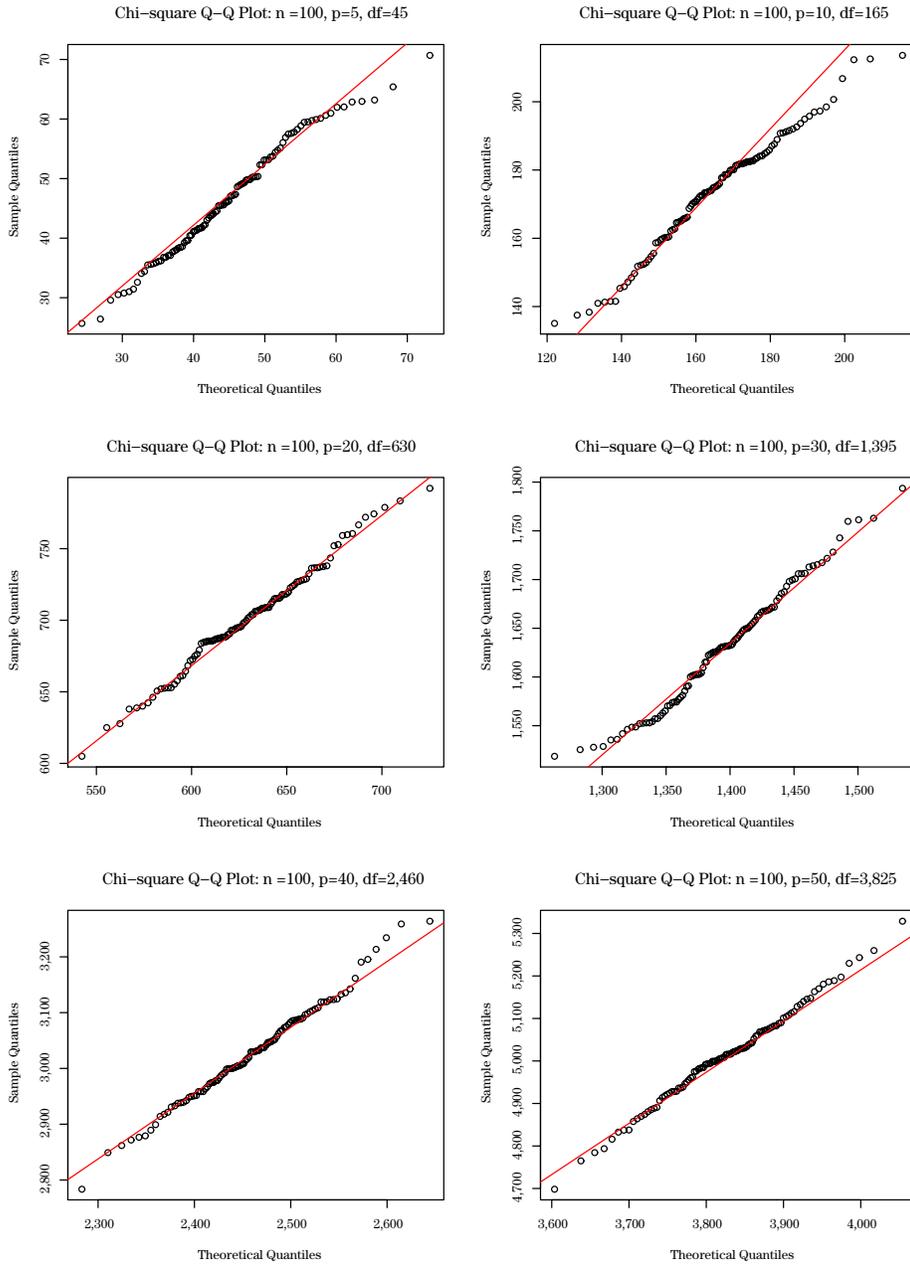
Figure 3. The Q-Q plots of test statistics of 100 independent repetitions as of Example 2 under $H_0 : \mathbf{\Omega}_1 = \cdots = \mathbf{\Omega}_4$ holds. The precision matrix is estimated by by inverting the sample covariance matrix. In the chi-square Q-Q plots, the degrees of freedom are determined by the number of constraints, $(p^2 + p)(T - 1)/2$.
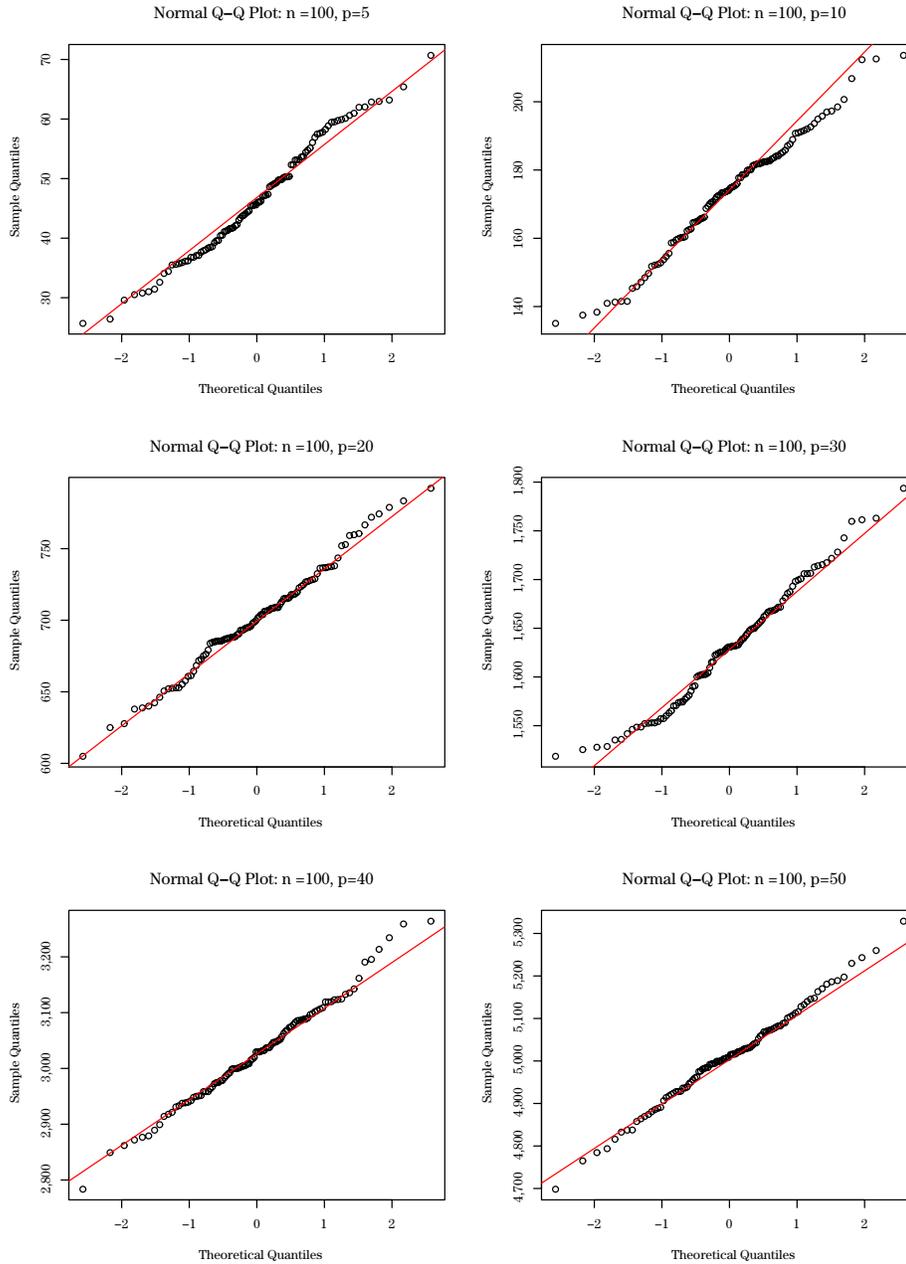
Figure 4. The Q-Q plots of test statistics of 100 independent repetitions as of Example 2 under $H_0 : \mathbf{\Omega}_1 = \cdots = \mathbf{\Omega}_4$ holds. The precision matrix is estimated by inverting the sample covariance matrix.

Table 4. Example 4: Proportions 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 ones. $\omega_{i,j,1} = \omega_{i,j,2} = 0.5^{|i-j|}$; $\omega_{i,j,3} = \omega_{i,j,4} = 0.2^{|i-j|}$; $1 \leq i, j \leq p$. Three methods are compared: the original likelihood ratio test, penalized likelihood ratio test, and 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

5. Financial Network Inference

This section investigates the effect of the Lehman Brothers bankruptcy on a 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 are the changes in the financial networks over 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 for the period January 1, 2005 to December 31, 2010; see Table 5.

The top 20 of the $p = 200$ stocks by market capitalization, as of December 31, 2010, are selected from each of 10 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 three major industries: major banks, investment banks, and insurance. The *energy* sector is composed of oil and gas production, consumer electronics, and other industries.

Three periods ($T = 3$) are considered: *Pre-crisis (1/1/2005–12/31/2005):* before the Lehman Brothers collapse; *Crisis (7/1/2008–6/30/2009)*, the period

Table 5. Stock list from 10 sectors.

Basic	PG	DOW	DD	MON	PCP	ECL	PX	APD	PPG	GLW
Industries	SCCO	IP	VMC	NUE	CHD	SRCL	EMN	IFF	LEN	NEM
Consumer	BA	UTX	HON	LMT	DHR	TMO	F	GD	RTN	CAT
Durables	NOC	ILMN	DE	ROP	PCAR	APH	SWK	A	PH	ROK
Consumer	KO	PEP	MO	NKE	RAI	MDLZ	CL	GIS	EL	STZ
Non-Durables	MNST	K	VFC	SYY	HRL	ADM	TSN	HSY	CAG	CPB
Consumer	AMZN	WMT	DIS	HD	CMCSA	MCD	SBUX	COST	LOW	SPG
Services	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	T	VZ	DUK	NEE	SO	D	AEP	EXC	PCG	WM
Utilities	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

during which Lehman Brothers filed for bankruptcy; and *Post-crisis (1/1/2010–12/31/2010)*, the recovery after the Lehman Brothers collapse. Accordingly, a network is assumed to be constant within each 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, owing to the effect of dominant 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 MGGMs with three periods. This seems appropriate in that the normality assumption is approximately satisfied, as indicated by the Q–Q plots in Figure 5.

MGGMs are used to model pairwise conditional dependencies, where each stock corresponds one node in a graph at one time point; that is, a lack of a connecting edge between two nodes implies pairwise independence, conditioning on all other $p - 2$ nodes (Lauritzen (1996)). Furthermore, the strength and sign of a pairwise conditional dependency between two nodes, given all other $p - 2$ nodes, is measured by the partial correlation coefficient in a GGM (Whittaker (1990)). 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 the macroeconomic environment." Of particular interest is es-
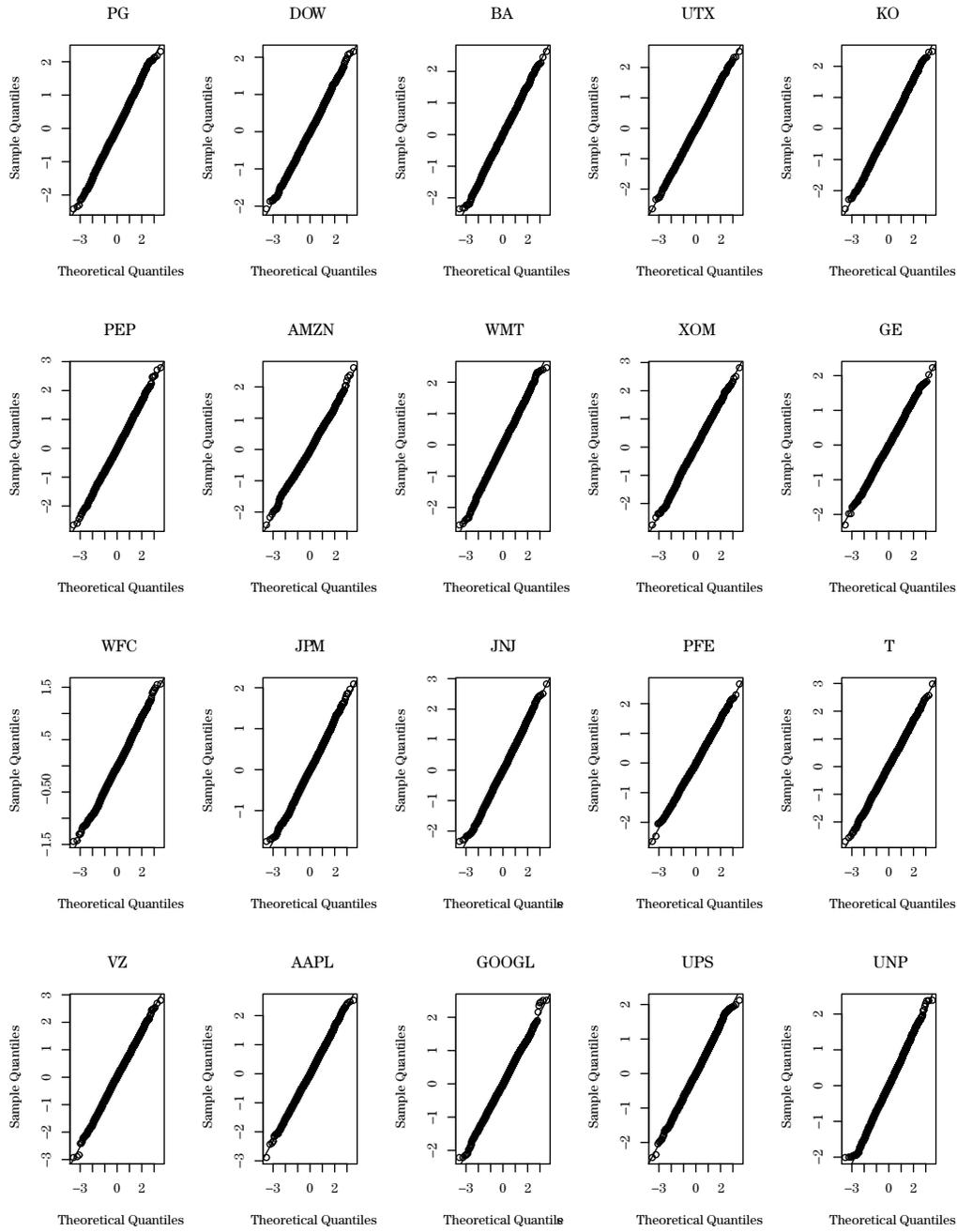
Figure 5. Q-Q plots of log return of daily closing prices of twenty stocks.

Pre−Crisis (1/1/2005−12/31/2005)   Crisis (7/1/2008−6/30/2009)   Post−Crisis (1/1/2010−12/31/2010)
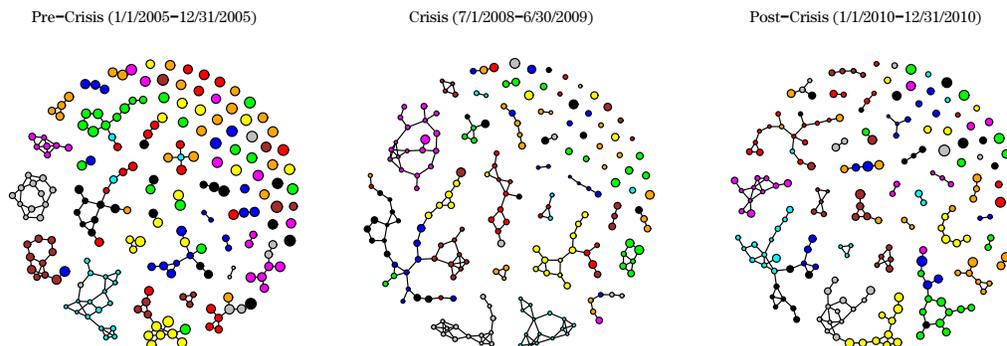
Figure 6. Estimated network of all sectors by graphical Lasso from 200 stock's log return of daily closing prices during three periods. The nodes are colored according to their sectors.

tablishing how the network structure evolves after 2005 by inferring an MGGM. First, we test whether the financial network remains the same across the three periods, which is an example of a global inference in a high-dimensional setting. 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 using the GLasso method (Friedman, Hastie and Tibshirani (2008)) for each period. In the following graphs, the size of a vertex denotes the estimated conditional (or partial) variance (Whittaker (1990)) of each stock, and an edge denotes an estimated partial correlation coefficient (Whittaker (1990)) with an absolute value above 0.1.

5.2. Major findings

We perform a global hypothesis test to determine whether our financial network became more interconnected after the Lehman Brothers collapse.

As shown in Figure 6, the pre-crisis node size is larger than those of the following two periods, indicating that the conditional variance of most stocks decreased during the crisis and post-crisis periods owing to increasing correlations between 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 period differs from the other two periods. There are many small and pure sub-communities in the pre-crisis period, which implies that the connections occur mainly between stocks of the same sector, whereas inter-sector linkages are rare and weak. Strong intra-sector linkages exist in just four sectors: public utilities (gray), energy (cyan), finance (magenta), and capital goods (red), which
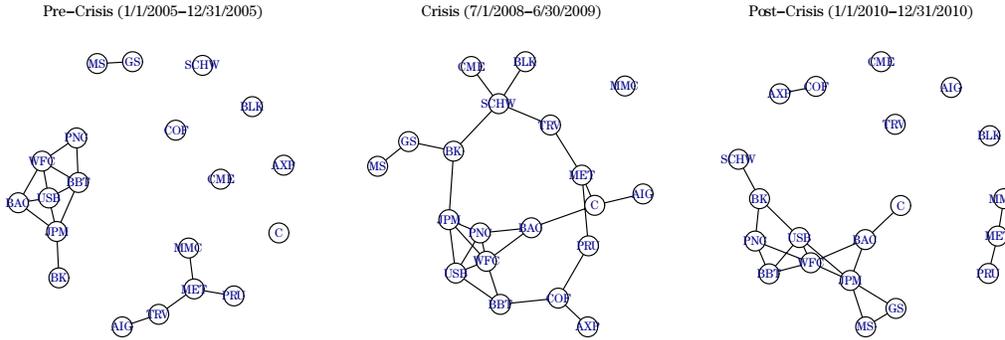
Figure 7. Estimated network of the *Finance* sector by graphical Lasso from 200 stock's log return of daily closing prices during three periods.

are more or less related to raw materials and infrastructure. During the crisis and post-crisis periods, one large diversified community is formed, in addition to several small sub-communities. Thus, strong inter-sector connections occur much more frequently during a period of economic contraction.

Finally, note that in the post-crisis period, the economy is expanding and the stock market is a bull market; however, the topological structure is similar to that of the crisis era, but significantly different to that of the pre-crisis era.

Next, we perform the hypothesis tests at a significance level $\alpha = 0.05$ using the method developed in Section 2, with $B = 1,000$ for the 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; thus H_0 is rejected in favor of 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 determine whether 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 $H_a : \mathbf{\Omega}_1 \neq \mathbf{\Omega}_2$. The corresponding empirical nominal levels are 0.000, 0.000 and 0.000, respectively. After adjusting for multiplicity, all three tests are simultaneously rejected at the overall level of 0.05. Therefore, the bankruptcy event affects the post-crisis period and the recovery period, because the financial network's structure varies significantly.

Next we focus on two sectors, namely, finance and energy.

As shown in Figure 7, the finance section is more fragmented in a booming economy than it is in a recession. The pre-crisis period includes two major sub-communities: "major banks," including WFC, BAC, and BK, among others, and "insurance," including MMC, PRL, and AIC, among others. However,
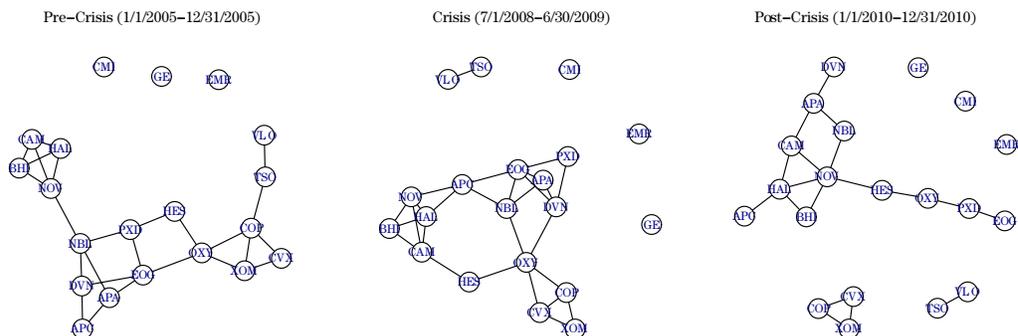
Figure 8. Estimated network of the *Energy* sector by graphical Lasso from 200 stock's log return of daily closing prices during three periods.

the investment banks are well separated from these communities. As the crisis unfolded and Lehman Brothers filed for bankruptcy in September, 2008, almost all banks 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, forming a larger sub-community. Before the financial crisis of 2007-2009, investment banks, including Lehman Brothers, raised capital and invested much as other major banks did. However, by avoiding regulation, they were able to over-leverage, thus exacerbating the system-wide contagion. During the financial crisis, however, the pure investment banks had to transform themselves to bank holding companies (BHC) in order to obtain government bailout money, and their BHC status means they are now subject to the additional oversight. Because these investment banks restructured their assets and have to act under stricter scrutiny by the government, they are being forced to operate as full-service banks. Thus, investment banks began to correlate with major banks.

In all three phases, as displayed in Figure 8, there is always a large sub-community comprising "oil and natural gas production" companies, in addition to some small sub-communities and isolated points for the other industries. Undoubtedly, the major driving force in energy is the crude oil price, which leads to strong intra-sector correlation, regardless of the macroeconomic performance. In contrast to the finance sector, the energy sector was not affected by reorganization.

5.3. Conclusion

The financial network has experienced a substantial transition, and the level of connectivity has increased and intensified significantly since the Lehman Brothers collapse. In a booming economy, sectorial factors mainly drive price movements, which means inter-sector connections seldom happen. In a financial crisis, two types of schemes contribute to the expansion and intensification of both intra- and inter-sector connections: systemic factors, and partial correlation coefficients. However, the partial correlation coefficients of the stocks of some sectors, such as *Energy*, do not change as other sectors do. When the economy recovered in the wake of the Lehman Brothers bankruptcy, the network did not return to its pre-crisis status. The Lehman Brothers collapse has deeply affected the network, because far-ranging and extreme financial and monetary measures have been imposed by the government on the stock market, especially for stocks in the finance sector.

6. Discussion

Globalization and the advancement of information-sharing technology have resulted in financial markets and institutions merging into a large network, which presents challenges to all aspects of risk management and policy-making. Inferences on large networks are becoming increasingly important as the financial system becomes characterized by greater connectivity and complexity, particularly when the network structure experiences sharp changes and exhibits stage-wise patterns owing to unexpected external shocks. Today, graphical models provide new ways of dealing with these challenges.

The proposed inferential tools allow us to study how the financial network evolved over three periods (i.e., pre-crisis, crisis, and post-crisis), based on daily prices of 200 stocks from 10 sectors. We justify theoretically that the bootstrap approximation of the sampling distribution is valid in a high-dimensional setting where the number of stages T and the number of nodes p increase with the sample size. Using simulations, we demonstrate that the proposed method compares favorably to its competitors in terms of Type-I and Type-II errors. To make the proposed method useful in practice, further investigation is necessary.

Supplementary Material

The proof of Theorem 1 is presented in the online Supplementary Material.

ZHANG, SHEN AND WANG

Acknowledgements

References

Bai, Z., Jiang, D., Yao, J. and Zheng, S. (2009). Corrections to lrt on large-dimensional covariance matrix by rmt. *The Annals of Statistics* **37**, 3822–3840.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics* **9**, 1196–1217.

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

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

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

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

Fan, J., Liao, Y. and Liu, H. (2016). An overview of the estimation of large covariance and precision matrices. *Econometrics Journal* **19**, 1–32.

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

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

Jankova, J. and De Geer, S. V. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics* **9**, 1205–1229.

Lauritzen, S. L. (1996). *Graphical Models*. Vol. 17. Clarendon Press.

Liu, H., Han, F. and Zhang, C. (2012). Transelliptical graphical models, 800–808.

Qiu, H., Han, F., Liu, H. and Caffo, B. S. (2016). Joint estimation of multiple graphical models from high dimensional time series. *Journal of The Royal Statistical Society Series B (Statistical Methodology)* **78**, 487–504.

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

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley Publishing.

Zhang, C. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of The Royal Statistical Society Series B (Statistical Methodology)* **76**, 217–242.

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

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

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

Lundquist College of Business, University of Oregon, 1208 Univer-sity Ave, Eugene, OR 97403, USA.

E-mail: yongli@uoregon.edu

School of Statistics, University of Minnesota, 224 Church Street S.E., Minneapolis, MN 55455, USA.

E-mail: xshen@umn.edu

School of Statistics and Management, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, China.

E-mail: swang@shufe.edu.cn