JUSTC

Article

http://justc.ustc.edu.cn

Received: January 06, 2022; Accepted: February 28, 2022

# Simultaneous inference for a high-dimensional precision matrix

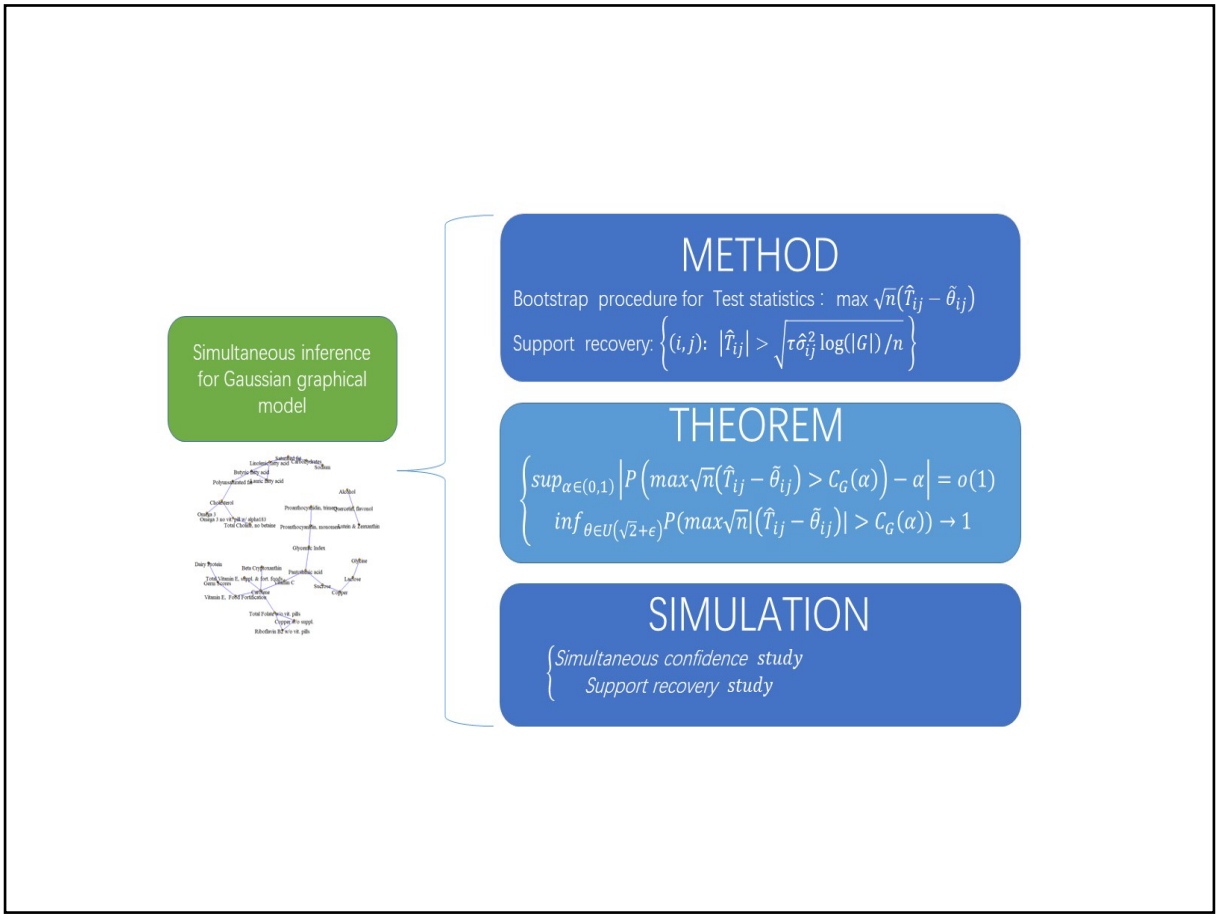
Wenjie Gao, Ruipeng Dong, and Jie Wu ✉

*International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230026, China*

✉Correspondence: Jie Wu, E-mail: wu12jie@mail.ustc.edu.cn



## Graphical abstract




*The method, theorem and simulation study of the simultaneous inference for a high-dimensional precision matrix.*

## Public summary

- We propose a bootstrap procedure to conduct simultaneous inference for Gaussian graphical models.
- The procedure is applied to large-scale graphical models and allows the dimension of the parameter vector of interest to exceed the sample size.
- Both theoretical and simulation results verify the feasibility of our method.

**Citation:** Gao W J, Dong R P, Wu J. Simultaneous inference for a high-dimensional precision matrix. *JUSTC*, **2022**, 52(7): 2. DOI: 10.52396/JUSTC-2022-0004

JUSTC

Article

http://justc.ustc.edu.cn

Received: January 06, 2022; Accepted: February 28, 2022

# Simultaneous inference for a high-dimensional precision matrix

Wenjie Gao, Ruipeng Dong, and Jie Wu ✉

*International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230026, China*

✉Correspondence: Jie Wu, E-mail: wu12jie@mail.ustc.edu.cn



Cite This: *JUSTC*, **2022**, 52(7): 2 (10pp)

Read Online

**Abstract:** Gaussian graphical models have been widely used for network data analysis. Although various methods exist for estimating the parameters, simultaneous inference is essential for graphical models. In this study, we propose a bootstrap procedure to conduct simultaneous inference for Gaussian graphical models. The simultaneous inference procedure is applied to large-scale graphical models and allows the dimension of the parameter vector of interest to exceed the sample size. We prove that the simultaneous test achieves a pre-set significance level asymptotically. Further simulation studies demonstrate the effectiveness of the proposed methods.

**Keywords:** high dimensionality; simultaneous inference; Gaussian graphical models; sparse recovery

**CLC number:** O212 **Document code:** A

**2020 Mathematics Subject Classification:** 65D40

## 1 Introduction

In the era of the rapid development of information technology, vast amounts of information are being introduced for a range of applications, including modern healthcare, online marketing, and quantitative finance. Graphical models have been widely used in the analysis of networks comprising many individuals. In Gaussian graphical models, the conditional independence of two variables is equivalent to that of the corresponding element with a value of zero in the precision matrix (inverse covariance matrix)[1]. Therefore, Gaussian graphical models can accurately describe and estimate functional brain networks. Comparing the functional connectivity of subjects in the two populations calls for the comparison of these estimated Gaussian graphical models. Belilovsky et al.[2] proposed an approach to identify differences within such a model, which is known to have a similar structure.

As Gaussian graphical models find many applications, numerous studies have estimated the precision matrices of such models. A widely used method for estimating the precision matrix is to maximize the penalized Gaussian likelihood or minimize the penalized empirical risk[3–5]. The penalized regression or Dantzig selector-type optimization methods have also attracted interest[6–8]. The above methods have the advantages of rationality and theory, but they are unsatisfactory in terms of computation when the sample dimension $p$ is large. The innovated scalable efficient estimation (ISEE)[9] combined the advantages of high-dimensional sparse modeling and large covariance matrix estimation. This computational problem was solved to a certain extent for large-scale graphical models. The aforementioned methods mainly focused on the point estimation of the precision matrix and did not characterize the uncertainty of the estimated entries.

In recent years, several studies have focused on de-biasing the penalized estimators by inverting the Karush-Kuhn-Tucker (KKT) condition[10] such that the asymptotic normal distributions for linear and generalized linear models can be obtained. The de-bias method is also used to deal with the Gaussian graphical model estimator. For example, Jankov and van de Geer[11] proposed a de-sparsified graphical lasso estimator by inverting the KKT condition based on a penalized empirical risk estimator. A similar study was conducted by them in Ref. [12]. Recently, Zhou et al.[13] adopted a similar approach to achieve asymptotic normality in the elements of the de-sparsified ISEE estimator. The problem with the inference methods of graphical models is that only one element of the precision matrix can be inferred at a time, whereas researchers tend to prioritize the state of a block's connection. Jankov and van de Geer[12] indicated the problem of support recovery for multiple variables, but it was not discussed in depth. Thus, it is necessary to develop new methods for simultaneous inference of the precision matrix. Several studies have been conducted on the simultaneous inference of high-dimensional linear models based on de-sparsified estimators[14]. We consider using a similar method to process simultaneous inferences for Gaussian graphical models.

In this study, we propose a bootstrap procedure to conduct simultaneous inference for Gaussian graphical models based on the ISEE[9] and its de-sparsifying estimator[13]. We overcome the problem wherein the de-bias estimator recovers the precision matrix based only on a single point. This procedure can naturally construct simultaneous confidence intervals; support recovery is carried out on the entire graph. In theory, a simultaneous testing procedure asymptotically achieves a



DOI: 10.52396/JUSTC-2022-0004

*JUSTC*, **2022**, 52(7): 2

pre-specified significance level, and the test set can even overwrite all elements in the precision matrix. Moreover, it inherits the computational advantages of ISEE, which can deal with ultra-large precision matrices with simple tuning. The simulation results also support the reliability of the proposed method.

The remainder of this paper is organized as follows. In Section 2, we introduce the simultaneous inference method for a large precision matrix. The theoretical results are provided in Section 3. Simulation studies are also presented in Section 4. The proofs of the theorems in Section 3 and the technical details are included in the Appendix.

**Notation.** For a vector $\boldsymbol{x} \in R^d$ and $p \in (0,\infty)$, we use the notation $\|\boldsymbol{x}\|_p$ to denote the $p$-norm of $x$ in the classical sense. By $\boldsymbol{e}_i$, we denote a $p$-dimensional vector of zeros, with one at position $i$. For matrix $\boldsymbol{A} \in R^{d\times d}$, we use the notation $\boldsymbol{A}_i = \boldsymbol{A}\boldsymbol{e}_i$, $|||\boldsymbol{A}|||_\infty = \max_i \|\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{A}\|_1$, $\|\boldsymbol{A}\|_\infty = \max_{i,j}|A_{i,j}|$, and $\mathrm{supp}(\boldsymbol{A}) = \{(i,j) : A_{i,j} \neq 0\}$. We define $\mathcal{G}(M,K) = \{\boldsymbol{A}$, where each row has at most $K$ nonzero off-diagonal entries, and $M^{-1} \leqslant \lambda_{\min}(\boldsymbol{A}) \leqslant \lambda_{\max}(\boldsymbol{A}) \leqslant M$, where $K \in \boldsymbol{N}$, $K \leqslant d$ and $M \in \boldsymbol{R}, M > 0\}$. For two numbers, $a$ and $b$, let $a \vee b = \max\{a,b\}$ and $a \wedge b = \min\{a,b\}$. For sequences $f_n, g_n$, we use the notation $f_n = O(g_n)$ if $f_n \leqslant Cg_n$ and $f_n = \Omega(g_n)$ if $f_n \geqslant Cg_n$ for some constant $C > 0$ independent of $n$. We write $f_n \asymp g_n$ if both $f_n = O(g_n)$ and $f_n = \Omega(g_n)$, and $f_n = o(g_n)$ if $\lim_{n\to\infty} f_n/g_n = 0$.

# 2 Simultaneous inference procedures

## 2.1 Model setting

We consider a $p$-variate zero-mean Gaussian random vector,

$$\boldsymbol{x} = (X_1,...,X_p)^{\mathrm{T}} \sim N(\boldsymbol{0},\boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = (\Sigma_{ij})$ denotes the population covariance matrix of $\boldsymbol{x}$. Suppose that $\boldsymbol{X} = (\boldsymbol{x}_1,...,\boldsymbol{x}_n)^{\mathrm{T}}$ is an $n \times p$ sample matrix with $n$ i.i.d. observations $\boldsymbol{x}_i$, $1 \leqslant i \leqslant n$, where $\boldsymbol{x}_i \sim N(\boldsymbol{0},\boldsymbol{\Sigma})$. Let $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}$ be the empirical covariance matrix of $\boldsymbol{X}$, and let $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ be the inverse of the covariance matrix, which is also called the precision matrix. In Gaussian graphical models, an edge exists between nodes $X_i$ and $X_j$ is defined as the conditional dependence between $X_i$ and $X_j$. $X_i$ and $X_j$ are conditionally dependent if and only if $\Theta_{ij} \neq 0$[1].

In Section 1, we introduce several estimation methods for the precision matrix $\boldsymbol{\Theta}$ and an inference method for a single estimation point. However, in high-dimensional graphical models, it is important to analyze the connection of a set of variables. Thus, we focus on the simultaneous inference of the elements of the precision matrix $\boldsymbol{\Theta}$. Consider the hypothesis testing problem:

$$H_{0,G} : \Theta_{ij} = \tilde{\Theta}_{ij}, \text{ for all } (i,j) \in G \subseteq \{(k,l) : 1 \leqslant k \leqslant p, 1 \leqslant l \leqslant p\} \quad (1)$$

versus alternative $H_{a,G} : \Theta_{ij} \neq \tilde{\Theta}_{ij}$ for some $(i,j) \in G$, where $(\tilde{\Theta}_{ij})$ with $(i,j) \in G$ is a vector of pre-specified values. Specifically, the test can be applied to support the recovery of $\boldsymbol{\Theta}$ when $\tilde{\Theta}_{ij} = 0$ for all $(i,j) \in G$. The problem of recovering the Gaussian graph is equivalent to supporting the recovery of $\boldsymbol{\Theta}$ because of the characterization of the edge set based on $\boldsymbol{\Theta}$. Moreover, we expect that a testing set with index set $G$ in $\boldsymbol{\Theta} = (\Theta_{ij})$ can be designed arbitrarily. To solve this problem, we propose a bootstrap method for the simultaneous inference of $\boldsymbol{\Theta}$ based on de-bias estimator.

## 2.2 The de-sparsified estimator

It is essential to obtain a valid initial estimate of the precision matrix to address the hypothesis testing problem. We select the ISEE estimator[9], which converts the precision matrix estimation problem into that of a large covariance matrix estimation as the initial estimator for $\boldsymbol{\Theta}$ because it can deal with ultra-large graphical models and provide high computational efficiency. In the remainder of this study, we denote the ISEE estimator $\widehat{\boldsymbol{\Theta}}_{\mathrm{cini}}$[9] by $\widehat{\boldsymbol{\Theta}}$.

A scalable inference method exists for a single element in $\boldsymbol{\Theta}$. Let

$$\widehat{\boldsymbol{T}} = 2\widehat{\boldsymbol{\Theta}} - \widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Theta}} = (\widehat{T}_{ij})$$

be the de-sparsified ISEE estimator proposed in Ref. [13]. Then, we obtain

$$\sqrt{n}(\widehat{T}_{ij} - \Theta_{ij}) = Z_{ij}^n + o_p(1), \quad (2)$$

where $Z_{ij}^n$ converges to the Gaussian distribution $N(0,\sigma_{ij}^2)$ with $\sigma_{ij} > 0$, $\sigma_{ij}^2 = \Theta_{ii}\Theta_{jj} + \Theta_{ij}^2$, and the negligible term does not depend on $(i,j)$. Although the above conclusion is a good description of the asymptotic properties of individual elements, it is not sufficient for simultaneous inferences.

## 2.3 Bootstrap method for simultaneous inference

Based on the estimator $\widehat{\boldsymbol{T}}$, we propose a test statistic

$$\max_{(i,j)\in G} \sqrt{n}(\widehat{T}_{ij} - \tilde{\Theta}_{ij})$$

for hypothesis testing Eq. (1). Specifically, under the null hypothesis $\tilde{\Theta}_{ij} = \Theta_{ij}$, for every $(i,j) \in G$, we obtain

$$\max_{(i,j)\in G} \sqrt{n}(\widehat{T}_{ij} - \tilde{\Theta}_{ij}) = \max_{(i,j)\in G} \sqrt{n}(\widehat{T}_{ij} - \Theta_{ij}) = \max_{(i,j)\in G}(Z_{ij}^n + o_p(1)). \quad (3)$$

We hope that this method can eliminate the influence of tail item in Eq. (2), such we only need to consider the distribution of $\max_{(i,j)\in G} Z_{ij}^n$, which has been proved in the next section. Note that $\boldsymbol{\Theta}$ is a symmetric matrix. For an efficient calculation, we only consider hypothesis testing for the upper triangular block of the inverse covariance matrix $\boldsymbol{\Theta}$ without loss of generality. Thus, $G \subset \{(i,j) : 1 \leqslant i \leqslant j \leqslant p\}$.

As the distribution of $\max_{(i,j)\in G} \sqrt{n}(\widehat{T}_{ij} - \Theta_{ij})$ cannot be obtained, one alternative is to obtain a reliable approximation of the distribution for Eq. (3). We construct a simple bootstrap method. Note that $Z_{ij}^n$ is asymptotically normal, and the remainder in Eq. (2) is negligible; thus, we use a multivariate normal random vector to simulate the distribution of our test statistic. It can be proven that the sequence $\{Z_{ij}^n\}_{1\leqslant i\leqslant j\leqslant p}$ follows an asymptotic normal distribution with the covariance matrix

$$\boldsymbol{\Xi} = (\Theta_{ik}\Theta_{jl} + \Theta_{il}\Theta_{jk})_{ij,kl}, \quad 1 \leqslant i \leqslant j \leqslant p, 1 \leqslant k \leqslant l \leqslant p,$$

where $\boldsymbol{\Xi}$ is a $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ matrix with the diagonal elements $(\sigma_{ij}^2)_{i\leqslant j}$. Based on this form of the covariance matrix for


DOI: 10.52396/JUSTC-2022-0004
JUSTC, 2022, 52(7): 2

$\{Z_{ij}^n\}_{1\leqslant i\leqslant j\leqslant p}$, we define

$$\widehat{\boldsymbol{\Xi}} = (\widehat{\Theta}_{ik}\widehat{\Theta}_{jl} + \widehat{\Theta}_{il}\widehat{\Theta}_{jk})_{ij,kl},\ \ 1\leqslant i\leqslant j\leqslant p, 1\leqslant k\leqslant l\leqslant p$$

as the estimator of $\boldsymbol{\Xi}$, with diagonal elements $\widehat{\sigma}_{ij}^2 = \widehat{\Theta}_{ii}\widehat{\Theta}_{jj} + \widehat{\Theta}_{ij}^2$. We then generate a random vector $\{\hat{Z}_{ij}\}_{1\leqslant i\leqslant j\leqslant p} \sim N(\mathbf{0}, \widehat{\boldsymbol{\Xi}})$ and define

$$W_G = \max_{(i,j)\in G} \hat{Z}_{ij}.$$

By definition, $W_G$ has an approximate distribution as follows: $\max_{(i,j)\in G} \sqrt{n}(\widehat{T}_{ij} - \Theta_{ij})$. The bootstrap critical value is given by $c_G(\alpha) = \inf\{t\in R : P(W_G\leqslant t|X)\geqslant 1-\alpha\}$. Our method is used to generate a bootstrap sequence of i.i.d. random vectors $\{\hat{Z}_{ij,m}\}_{1\leqslant m\leqslant M}$ and select the $\alpha$-quantile of the sequence $\{\max_{(i,j)\in G}\{\hat{Z}_{ij,m}\}_{1\leqslant m\leqslant M}\}$ as a substitute for $c_G(\alpha)$.

Moreover, we can adjust the test and bootstrap statistics to obtain a similar regularization test. We consider the studentized statistic

$$\max_{(i,j)\in G} \sqrt{n}(\widehat{T}_{ij} - \tilde{\Theta}_{ij})/\widehat{\sigma}_{ij},$$

where $\widehat{\sigma}_{ij} > 0$ and $\hat{\sigma}_{ij}^2 = \widehat{\Theta}_{ii}\widehat{\Theta}_{jj} + \widehat{\Theta}_{ij}^2$. Correspondingly, the bootstrap critical value can be obtained via $\bar{c}_G(\alpha) = \inf\{t\in R : P(\bar{W}_G\leqslant t|X)\geqslant 1-\alpha\}$, defined as

$$\bar{W}_G = \max_{(i,j)\in G} \hat{Z}_{ij}/\widehat{\sigma}_{ij}.$$

We now consider a simultaneous confidence interval. Suppose there is a given set $G\subseteq\{(k,l) : 1\leqslant k\leqslant p, 1\leqslant l\leqslant p\}$, which is the index set of the subset to be tested for $\boldsymbol{\Theta}$. To address the hypothesis testing problem, we generate a $n\times|G|$ bootstrap sequence $\{\widehat{Z}_{ij}\}_{(i,j)\in G,n}$ with $n$ i.i.d. samples, where $(\widehat{Z}_{ij})_{(i,j)\in G} \sim N(\mathbf{0}, \widehat{\boldsymbol{\Xi}}_G)$, $\widehat{\boldsymbol{\Xi}}_G$ is a $|G|\times|G|$ subblock of $\widehat{\boldsymbol{\Xi}}$. The sequence $\{\max_{(i,j)\in G}\hat{Z}_{ij}\}_n$ has an approximate distribution of $\max_{(i,j)\in G}\sqrt{n}(\widehat{T}_{ij} - \Theta_{ij})/\hat{\sigma}_{ij}$. We then obtain the hypercuboid $I_G = I_1\times I_2\times\ldots\times I_{|G|}$ as the confidence region of $\{\tilde{\Theta}_{ij}\}_{(i,j)\in G}$ with a confidence level $1-\alpha$, where $I_k = (\widehat{T}_{i_kj_k} - \hat{\sigma}_{i_kj_k}\bar{c}_G(\alpha/2)/\sqrt{n}, \widehat{T}_{i_kj_k} + \hat{\sigma}_{i_kj_k}\bar{c}_G(\alpha/2)/\sqrt{n})$. The confidence region $I_G$ can also be the acceptance domain of the null hypothesis $\tilde{\Theta} = \Theta$ for the hypothesis problem (1), with confidence level $1-\alpha$.

## 3 Theoretical results

This section presents theoretical support for the procedures in Section 2. The theorems demonstrate the consistency of the asymptotic distribution between $W_G$ and $\sqrt{n}(\widehat{T}_{ij} - \Theta_{ij})$. Thus, the bootstrap method is established as reasonable. We also prove the validity of the support recovery process.

The following four assumptions are necessary to support this result:

**Assumption 3.1.** Assume that $\boldsymbol{\Theta}\in\mathcal{G}(M,K)$, with $M = O(1)$, $M > 1$ and $p > n, \log p/n = o(1)$. For every $i\neq j$, there exists a positive constant $c$, in which $|\Theta_{ij}|/\sqrt{\Theta_{ii}\Theta_{jj}}\leqslant c < 1$.

**Assumption 3.2.** There exists a certain sparsity level $s\geqslant K$, $s = O(d) = o(\sqrt{n}/\log p)$ and some constants $0\leqslant\alpha\leqslant 1/2$ and $\xi > 1$, such that the $l1$-norm cone invertibility factor

$$F_\infty = \inf\{\frac{\|\boldsymbol{\Sigma}\boldsymbol{u}\|_\infty}{\|\boldsymbol{u}\|_\infty} : \|\boldsymbol{u}_{S^c}\|_1\leqslant\xi\|\boldsymbol{u}_S\|_1\neq 0 \text{ for some } S\in\{1,\ldots,p\} \text{ with } |S|\leqslant O(K)\}$$

of covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Theta}^{-1}$ satisfies $F_\infty^{-1} = O(K^\alpha)$.

**Assumption 3.3.** We assume that $(\log(pn))^7/n\leqslant C_1 n^{-c_1}$ for constants $c_1, C_1 > 0$.

**Assumption 3.4.** There exists a sequence of positive numbers $\alpha_n\to\infty$, such that $\alpha_n/p = o(1)$ and $\alpha_n(\log p)^2 K\sqrt{\frac{\log p}{n}} = o(1)$, $(\log p)^2\max_{1\leqslant i,j\leqslant p}\Delta_{ij} = o_p(1)$.

Based on Assumptions 3.1 and 3.2, We obtain the asymptotic normality of a single element of the de-sparsified estimator $\widehat{\boldsymbol{T}}$ (Lemma A.1). We then propose a theorem to ensure the theoretical feasibility of simultaneous inference. Assumptions 3.3 and 3.4 are necessary to prove this theorem, as they control for the bias of the covariance of the bootstrap statistics.

The following theorem justifies the validity of the bootstrap procedure for the studentized-type statistic $\max_{(i,j)\in G}\sqrt{n}(\widehat{T}_{ij} - \Theta_{ij})/\widehat{\sigma}_{ij}$. This is an important theoretical basis for the simultaneous confidence interval inference.

**Theorem 3.1.** Suppose that Assumptions 3.1–3.4 hold. Then, for any $G\subseteq\{(k,l) : 1\leqslant k\leqslant l\leqslant p\}$,

$$\sup_{\alpha\in(0,1)}|P(\max_{(i,j)\in G}\sqrt{n}(\widehat{T}_{ij} - \Theta_{ij})/\widehat{\sigma}_{ij} > \bar{c}_G(\alpha)) - \alpha| = o(1). \tag{4}$$

The theorem demonstrates the reasonability of using the bootstrap quantile $\bar{c}_G(\alpha)$ as the quantile of the distribution for $\max_{(i,j)\in G}\sqrt{n}(\widehat{T}_{ij} - \Theta_{ij})/\widehat{\sigma}_{ij}$. A crucial feature of this bootstrap-assisted testing procedure is that it explicitly accounts for the effect of $|G|$ because the bootstrap critical value $c_G(\alpha)$ depends on $G$. Thus, the result is readily applicable to the construction of simultaneous confidence intervals for $(\tilde{\Theta}_{ij})$ with $(i,j)\in G$.

Next, we consider the inference of support recovery. The major goal of support recovery is to identify nonzero signal locations in a pre-specified set $G$. The procedure involves setting an appropriate threshold $\tau$ in the set

$$\widehat{\boldsymbol{S}}_0(\tau) = \{(i,j)\in G : |\widehat{T}_{ij}| > \sqrt{\tau\hat{\sigma}_{ij}^2\log(|G|)/n}\},$$

where $\tau$ is the appropriate threshold. We then define the real support

$$\boldsymbol{S}_0 = \{(i,j)\in G : \boldsymbol{\Theta}_{ij}\neq 0\}.$$

Theorem 3.2 in the following section shows that the above support recovery procedure is consistent and effective if $\tau\geqslant 2$ and provides the theoretical optimum separation parameter $\tau_0 = 2$. Next, we conduct a power analysis for the above procedure. When $|G|$ is fixed, the test is consistent (based on the asymptotic normality of a single element in $\widehat{T}$). Thus, we focus on the case where $|G|\to\infty$. Note that no signal strength assumption is required in simultaneous confidence interval inference. However, this assumption is necessary to support recovery.

We define the separation set as follows:

$$\mathcal{U}_G(c) = \{\Theta^* : \max_{(i,j)\in G}|\Theta_{ij}^* - \Theta_{ij}|/\widehat{\sigma}_{ij} > c\sqrt{\log(|G|)/n}\}.$$


DOI: 10.52396/JUSTC-2022-0004
*JUSTC*, **2022**, 52(7): 2

The separation set $\mathcal{U}_G(c)$ shows that there is at least one element in $G$, such that $|\Theta^*_{ij}-\Theta_{ij}|/\widehat{\sigma}_{ij} > c\sqrt{\log(|G|)/n}$. The following theorem shows that the test rejects the null hypothesis as long as one element in $G$ satisfies the above condition.

**Theorem 3.2.** We assume that $G \subset \{(i,j): 1 \leqslant i \leqslant j \leqslant p\}$. Under the assumptions of Theorem 3.1, for any $\epsilon_0 > 0$,

$$\inf_{\Theta \in \mathcal{U}_G(\sqrt{2}+\epsilon_0)} P(\max_{(i,j)\in G} \sqrt{n}|\widehat{T}_{ij}-\Theta_{ij}|/\widehat{\sigma}_{ij} > \bar{c}^*_G(\alpha)) \to 1. \tag{5}$$

Theorem 3.2 supports the conclusion that the proposed procedure is highly sensitive in terms of detecting sparse alternatives. This is the theoretical basis for the support recovery procedure.

From the proof of Theorem 3.2, we also observe that the separation rate $(\sqrt{2}+\epsilon_0)\sqrt{\log(|G|)/n}$ for any $\epsilon_0 > 0$ derived in Theorem 3.2 is minimax optimal under suitable assumptions. The following theorem shows that the support recovery procedure is consistent and effective if the threshold value is set as $\tau_0 = 2$ and further justifies the optimality of $\tau_0$. A similar support recovery test procedure was presented by Ref. [12]. We show the relationship between the optimal threshold parameter and the number of test variables in the following corollary.

**Corollary 3.1**.Under the assumptions in Theorem 3.2, we obtain the following:

$$\inf_{\boldsymbol{\Theta}\in\Psi(2\sqrt{2})} P(\hat{\boldsymbol{S}}_0(2) = \boldsymbol{S}_0) \to 1, \tag{6}$$

where $\Psi(c_0) = \{\boldsymbol{\Theta} = (\Theta_{ij}) : \min_{(i,j)\in S_0} |\Theta_{ij}|/\sigma_{ij} > c_0\sqrt{\log(|G|)/n}\}$. Moreover, for any $0 < \tau < 2$,

$$\sup_{\boldsymbol{\Theta}\in\boldsymbol{S}^*(s_0)} P(\hat{\boldsymbol{S}}_0(\tau) = \boldsymbol{S}_0) \to 0, \tag{7}$$

where $\boldsymbol{S}^*(s_0) = \{\boldsymbol{\Theta} = (\Theta_{ij}) : \sum_{1\leqslant i<j\leqslant p} \boldsymbol{I}\{\Theta_{ij} \neq 0\} = s_0\}$.

This observation, in part, has motivated the consideration of the numerical study procedure in Section 4.

# 4 Numerical studies

In Section 4, we present the simulation results for the applications of the simultaneous inference process, including simultaneous confidence interval and support recovery.

## 4.1 Simultaneous confidence interval

We consider the following model setting to demonstrate the proposed method. $\boldsymbol{\Theta} = \text{tridiag}(\rho; 1; \rho)$ where a given $\rho = 0.45$ is a tri-diagonal matrix. For every data set, the rows of sample $X$ are sampled as i.i.d. copies of $N(\boldsymbol{0}, \boldsymbol{\Theta}^{-1})$. We consider various sample sizes, $n$, and dimensions, $p$, where $\boldsymbol{\Theta} \in R^{p\times p}$. The index set $G$ of the inverse covariance matrix for the test is set as follows: (I) $G$ contains the top 50 lines of a column of $\boldsymbol{\Theta}$ with $|G| = 50$. (II) $G$ is a column of $\boldsymbol{\Theta}$, with $|G| = p$. For both cases, we test the top 10 columns for each set of data. (III) $G$ is an upper tri-diagonal subblock of $\boldsymbol{\Theta}$ without the diagonal elements. The subblock is an $11\times 11$ submatrix such that $|G| = 55$. (IV) $G$ is defined as the above case with a $d\times d$ submatrix, where $d = \min\{p/10, \lceil\sqrt{2p}+1/2\rceil\}$, such that $|G|$ is close to $p$. For both cases, we test 10 disjoint subblocks for each set of data. The tuning parameter is set as the value suggested by Ref. [9], and the confidence interval is shown in Section 2. All the performance measures are averaged over $N = 50$ replications.

Although our procedure is applied to achieve the simultaneous confidence interval of the entire precision matrix $\boldsymbol{\Theta}$, it is not computationally feasible because of the complexity of

**Table 1.** The simultaneous confidence interval result in Cases I and II. "C" denotes the coverage rate of the confidence interval and "L" denotes the length of the confidence interval. They are similarly defined in the following table as well.

| $n$ | $p$ | Case I | | | | Case II | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 90%C | 90%L | 95%C | 95%L | 90%C | 90%L | 95%C | 95%L |
| 200 | 150 | 0.87 | 0.1126 | 0.916 | 0.1352 | 0.814 | 0.0974 | 0.892 | 0.1175 |
| 200 | 300 | 0.878 | 0.1129 | 0.93 | 0.1354 | 0.862 | 0.0905 | 0.918 | 0.1092 |
| 200 | 450 | 0.892 | 0.1124 | 0.934 | 0.1350 | 0.866 | 0.0873 | 0.918 | 0.1056 |
| 350 | 450 | 0.9 | 0.0851 | 0.946 | 0.1023 | 0.862 | 0.0659 | 0.92 | 0.0794 |
| 500 | 450 | 0.9 | 0.0710 | 0.936 | 0.0853 | 0.858 | 0.0551 | 0.922 | 0.0663 |
| 500 | 600 | 0.898 | 0.0712 | 0.94 | 0.0856 | 0.856 | 0.0537 | 0.918 | 0.0648 |
| 500 | 750 | 0.894 | 0.0712 | 0.94 | 0.0855 | 0.872 | 0.0526 | 0.932 | 0.0636 |

**Table 2.** The simultaneous confidence interval result in Cases III and IV.

| $n$ | $p$ | Case III | | | | Case IV | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 90%C | 90%L | 95%C | 95%L | 90%C | 90%L | 95%C | 95%L |
| 200 | 150 | 0.848 | 0.1155 | 0.916 | 0.1383 | 0.856 | 0.1053 | 0.902 | 0.1267 |
| 200 | 300 | 0.86 | 0.1153 | 0.914 | 0.1384 | 0.868 | 0.0930 | 0.926 | 0.1119 |
| 200 | 450 | 0.868 | 0.1156 | 0.922 | 0.1388 | 0.876 | 0.0887 | 0.94 | 0.1068 |
| 350 | 450 | 0.86 | 0.0875 | 0.922 | 0.1050 | 0.896 | 0.0671 | 0.938 | 0.0808 |
| 500 | 450 | 0.882 | 0.0731 | 0.944 | 0.0877 | 0.872 | 0.0562 | 0.934 | 0.0677 |
| 500 | 600 | 0.878 | 0.0732 | 0.948 | 0.0877 | 0.884 | 0.0546 | 0.946 | 0.0658 |
| 500 | 750 | 0.886 | 0.0731 | 0.942 | 0.0877 | 0.9 | 0.0535 | 0.944 | 0.0645 |


DOI: 10.52396/JUSTC-2022-0004
*JUSTC*, **2022**, 52(7): 2

computing the square-rooting matrix of size $p^2 \times p^2$.

We report the results at two confidence levels: $\alpha = 0.1, 0.05$. The results are summarized in Tables 1 and 2. The length of the interval is the confidence interval of $\max_{(i,j)\in G}(\widehat{T}_{ij} - \Theta_{ij})/\widehat{\sigma}_{ij}$. From Tables 1 and 2, we observe that simultaneous confidence interval inference is effective in a variety of situations. In particular, the procedure performs satisfactorily and stably under high-dimensional conditions, where $n < p$. Notably, the length of the confidence intervals decreases with an increase in the size of sample $n$ and test set $|G|$. This result is consistent with the theoretical results.

### 4.2 Support recovery

We consider the most challenging and valuable scenario, where $G = \{(i,j)\}_{1\leqslant i<j\leqslant p}$, because the test for non-diagonal elements is of practical significance. We consider two model settings to demonstrate the proposed method: (Type I) $\boldsymbol{\Theta} = \text{tridiag}(\rho; 1; \rho)$ with a given $\rho > 0$ being a tri-diagonal matrix; we set $\rho = 0.45$. (Type II) $\boldsymbol{\Theta} = \text{five-diag}(\rho_0;\rho_1;\rho_2)$, which is defined by

$$\boldsymbol{\Theta}_{ij} = \begin{cases} \rho_0, & \text{if } i = j; \\ \rho_1, & \text{if } |i-j| = 1; \\ \rho_2, & \text{if } |i-j| = 2; \\ 0, & \text{else;} \end{cases}$$

and set $\boldsymbol{\Theta}$= five-diag$(1, 0.5, 0.4)$. The setting of sample $X$ is the same as that in simultaneous confidence interval inference. To assess the performance of support recovery, we consider the power and number of false positives (FP) of the procedure. We simultaneously compare the result with the estimated support set by ISEE, which is defined as $\{(i,j) : |\widehat{\Theta}_{ij}^{\text{isee}}| > 10^{-3}\}$. Moreover, in line with Ref. [14], we consider the following similarity measure:

$$d(\hat{\boldsymbol{S}}_0(\tau), \boldsymbol{S}_0) = \frac{|\hat{\boldsymbol{S}}_0(\tau) \cap \boldsymbol{S}_0|}{\sqrt{|\hat{\boldsymbol{S}}_0(\tau)| \times |\boldsymbol{S}_0|}}.$$

We also compare the effect and computation speed with De-GLasso and De-NLasso, as recommended by Ref. [12]. All performance measures were averaged over $N = 50$ replications.

The results are summarized in Tables 3, 4, and 5. Although ISEE has a remarkably high recognition rate for nonzero elements, some zero elements are picked out incorrectly. Thus, our approach provides a better way to support recovery. Tables 3 and 4 show that the sample size $n$ has a significant influence on the result compared with the dimension $p$. To ensure that the sample size $n$ is sufficiently large, our procedure performs well even when $p$ is larger than $n$. Based on Table 5, we see that although the effects are similar among the different methods for estimating $\Theta$, De-ISEE en-

**Table 3.** The support recovery result, where $\boldsymbol{\Theta}$ follows Type I, $d = d(\hat{\boldsymbol{S}}_0(2), \boldsymbol{S}_0)$.

| $n$ | $p$ | De-ISEE | | | ISEE | | |
|---|---|---|---|---|---|---|---|
| | | FP | Power | $d$ | FP | Power | $d$ |
| 200 | 150 | 1.1 | 0.9729 | 0.9826 | 8.94 | 1 | 0.9714 |
| 200 | 300 | 0.88 | 0.9490 | 0.9726 | 11.26 | 1 | 0.9817 |
| 200 | 450 | 0.8 | 0.9264 | 0.9616 | 16.3 | 1 | 0.9823 |
| 350 | 450 | 0.94 | 1 | 0.9990 | 23 | 1 | 0.9754 |
| 500 | 450 | 0.76 | 1 | 0.9992 | 36.48 | 1 | 0.9618 |
| 500 | 600 | 0.74 | 1 | 0.9994 | 41.46 | 1 | 0.9671 |
| 500 | 750 | 0.74 | 1 | 0.9995 | 44.04 | 1 | 0.9719 |

**Table 4.** The support recovery result, where $\boldsymbol{\Theta}$ follows Type II.

| $n$ | $p$ | De-ISEE | | | ISEE | | |
|---|---|---|---|---|---|---|---|
| | | FP | Power | $d$ | FP | Power | $d$ |
| 350 | 450 | 2.88 | 0.9992 | 0.9980 | 16.42 | 1 | 0.9909 |
| 500 | 450 | 2.42 | 1 | 0.9987 | 24.86 | 1 | 0.9864 |
| 500 | 600 | 2.38 | 1 | 0.9990 | 27.1 | 1 | 0.9889 |
| 500 | 750 | 2.3 | 1 | 0.9992 | 29.26 | 1 | 0.9904 |

**Table 5.** The support recovery result and computation time for the different estimators, where $\boldsymbol{\Theta}$ follows Type I.

| $n$ | $p$ | De-ISEE | | De-GLasso | | De-NLasso | |
|---|---|---|---|---|---|---|---|
| | | $d$ | CPU-time | $d$ | CPU-time | $d$ | CPU-time |
| 200 | 150 | 0.9826 | 2.783 | 0.9997 | 19.382 | 0.9959 | 60.057 |
| 200 | 300 | 0.9726 | 14.548 | 0.9993 | 167.97 | 0.9945 | 148.46 |
| 200 | 450 | 0.9616 | 31.54 | 0.9988 | 635.22 | 0.9920 | 267.19 |
| 350 | 450 | 0.9990 | 40.49 | 1 | 730 | 0.9997 | 539 |
| 500 | 450 | 0.9992 | 59.91 | 1 | 792.67 | 0.9998 | 1102.3 |


DOI: 10.52396/JUSTC-2022-0004
*JUSTC*, **2022**, 52(7): 2

ables much faster calculation than De-GLasso and De-NLasso.

## 5 Conclusions

In this study, we propose a bootstrap method to deal with simultaneous inferences for Gaussian graphical models. The method is based on de-sparsified ISEE so it has asymptotic normality and computational advantages. We then demonstrate the effectiveness of the proposed method both theoretically and through simulation. Notwithstanding our findings, the problem of simultaneous confidence interval inference for the entire graph continues to persist because of the high computational cost involved. This is a direction for future research.

## Appendix

In the appendix, we provide proofs of the main results of this study. First, we list the notation used in the proofs. Let $\tilde{\boldsymbol{x}} = \boldsymbol{\Theta}\boldsymbol{x} = (\tilde{X}_1, ..., \tilde{X}_p)$. It is evident that $\tilde{\boldsymbol{x}} \sim N(\boldsymbol{0}, \boldsymbol{\Theta})$. Let $\{z_{ij}\}_{1\leqslant i\leqslant j\leqslant p}$ be a $p(p+1)/2$ sequence of the mean zero independent Gaussian vector with $Ez_{ij}z_{ij}^{\mathrm{T}} = \boldsymbol{\Xi}$. We use the notation $\Xi_{ij,kl}$ to show the $ij$th row and $kl$th column element of $\boldsymbol{\Xi}$, which can be regarded as the covariance of the variance $\tilde{x}^i\tilde{x}^j - \Theta_{ij}$ and $\tilde{x}^k\tilde{x}^l - \Theta_{kl}$. We define

$$\bar{\boldsymbol{\Theta}} = (\Theta_{ij}/\sqrt{\Theta_{ii}\Theta_{jj}})_{1\leqslant i,j\leqslant p}$$

as the centralized result of matrix $\boldsymbol{\Theta}$. Let $\xi_{ijk} = \boldsymbol{\Theta}_i^{\mathrm{T}}\boldsymbol{x}_k\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{\Theta}_j - \Theta_{ij}$. As each element in an i.i.d. sequence $\{\boldsymbol{\Theta}\boldsymbol{x}_k\}_{k=1}^n$ has the same distribution as $\tilde{\boldsymbol{x}}$, we observe that $\{\xi_{ijk}\}_{1\leqslant i\leqslant j\leqslant p}$ and $\{z_{ij}\}_{1\leqslant i\leqslant j\leqslant p}$ have the same mean and covariance matrices.

We then present several lemmas that are used in the proof of the theorem in Section 3.

**Lemma A.1.** [13, Theorem 3.1] Let $\widehat{\boldsymbol{T}}$ be defined in Section 2. Under Assumptions 3.1 and 3.2 and $\|\boldsymbol{\Sigma}\|_\infty = O(1)$, we obtain

$$\sqrt{n}(\hat{T}_{ij} - \Theta_{ij}) = \sum_{k=1}^{n}(\boldsymbol{\Theta}_i^{\mathrm{T}}\boldsymbol{x}_k\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{\Theta}_j - \Theta_{ij})/\sqrt{n} + \Delta_{ij},$$

where $\Delta_{ij} = o(1)$ for all $1 \leqslant i, j \leqslant p$, Moreover, the de-sparsified estimator has the following asymptotic distribution.

$$\sqrt{n}(\hat{T}_{ij} - \Theta_{ij})/\sigma_{ij} = Z_{ij}^n + o_p(1),$$

where $\sigma_{ij}^2 = \Theta_{ii}\Theta_{jj} + \Theta_{ij}^2$ and $Z_{ij}^n \rightsquigarrow N(0,1)$.

**Lemma A.2.** Under the assumptions in Theorem 3.1, for any $G \in \{(i,j) : 1 \leqslant i \leqslant p, 1 \leqslant j \leqslant p\}$ and $c > 0$,

$$\sup_{x\in R}|P(\max_{(i,j)\in G}\sum_{k=1}^{n}\xi_{ijk}/\sqrt{n} \leqslant x) - P(\max_{(i,j)\in G} z_{ij} \leqslant x)| \leqslant n^{-c}. \tag{A.1}$$

**Lemma A.3.** Under Assumption 3.1, for $i \neq j$, we obtain

$$\max_{1\leqslant i\leqslant j\leqslant p}\sum_{1\leqslant k\leqslant l\leqslant p}\Xi_{ij,kl}^2 \leqslant C_0$$

and

$$\max_{1\leqslant i\leqslant j\leqslant p, 1\leqslant k\leqslant l\leqslant p}|\Xi_{ij,kl}|/\sqrt{\Xi_{ij,ij}\Xi_{kl,kl}} \leqslant c_0 < 1, \tag{A.2}$$

where $(i,j) \neq (k,l)$ for constants $C_0$ and $c_0 > 0$.

**Lemma A.4.** Suppose Assumptions 3.1–3.4 hold. Subsequently, for any $G \subseteq \{(k,l) : 1 \leqslant k \leqslant l \leqslant p\}$,

$$\sup_{\alpha\in(0,1)}|P(\max_{(i,j)\in G}\sqrt{n}(\widehat{T}_{ij} - \Theta_{ij}) > c_G(\alpha)) - \alpha| = o(1). \tag{A.3}$$

### Appendix A.1 The proof of Lemma A.4

Let $\pi(v) = C_1 v^{1/3}(1 \vee (\log(p/v))^{2/3}$ with $C_1 > 0$, and

$$\Gamma = \max_{(i,j),(k,l)}|\widehat{\Xi}_{ij,kl} - \Xi_{ij,kl}|. \tag{A.4}$$

From Ref. [9], we obtain

$$|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}|_\infty = O\left(K\sqrt{\frac{\log p}{n}}\right),$$

which implies that

$$\begin{aligned}|\widehat{\Xi}_{ij,kl} - \Xi_{ij,kl}| =& |\widehat{\Theta}_{ik}\widehat{\Theta}_{jl} + \widehat{\Theta}_{il}\widehat{\Theta}_{jk} - \Theta_{ik}\Theta_{jl} - \Theta_{il}\Theta_{jk}| \leqslant \\ & |\widehat{\Theta}_{ik}\widehat{\Theta}_{jl} - \Theta_{ik}\Theta_{jl}| + |\widehat{\Theta}_{il}\widehat{\Theta}_{jk} - \Theta_{il}\Theta_{jk}| \leqslant \\ & |\widehat{\Theta}_{ik}(\widehat{\Theta}_{jl} - \Theta_{jl})| + |\Theta_{jl}(\Theta_{ik} - \widehat{\Theta}_{ik})| + \\ & |\widehat{\Theta}_{il}(\widehat{\Theta}_{jk} - \Theta_{jk})| + |\Theta_{jk}(\widehat{\Theta}_{il} - \Theta_{il})| = \\ & O\left(K\sqrt{\frac{\log p}{n}}\right)\end{aligned}$$

for all $1 \leqslant i, j, k, l \leqslant p$, where the triangle inequality is used, and $|\widehat{\boldsymbol{\Theta}}|_\infty = O(1), |\boldsymbol{\Theta}|_\infty = O(1)$. Thus, we obtain

$$\Gamma = O\left(K\sqrt{\frac{\log p}{n}}\right).$$

Let $c_{z,G}(\alpha) = \inf\{t \in R : P(\max_{(i,j)\in G} z_{ij} \leqslant t) \geqslant 1-\alpha\}$. Following the arguments in the proof of Ref. [15, Lemma 3.2],

$$\begin{aligned}P(c_G(\alpha) \leqslant c_{z,G}(\alpha + \pi(v))) \geqslant 1 - P(\Gamma > v),\\ P(c_{z,G}(\alpha) \leqslant c_G(\alpha + \pi(v))) \geqslant 1 - P(\Gamma > v).\end{aligned}$$

As $\log p \max_{1\leqslant i,j\leqslant p}\Delta_{ij} = o(1)$, let $\xi_1 = 1/\log p$; then, we obtain

$$P(\max_{(i,j)\in G}|\sqrt{n}(\widehat{T}_{ij} - \Theta_{ij}) - \sum_{k=1}^{n}\xi_{ijk}/\sqrt{n}| \geqslant \xi_1) = P(\max_{(i,j)\in G}|\Delta| \geqslant \xi_1) < \xi_2, \tag{A.5}$$

where $\xi_2 = o(1)$. Based on the arguments in the proof of Ref. [15, Theorem 3.2], for every $v > 0$,

$$\begin{aligned}&\sup_{\alpha\in(0,1)}|P(\max_{(i,j)\in G}\sqrt{n}(\widehat{T}_{ij} - \Theta_{ij}) > c_G(\alpha)) - \alpha| \leqslant \\ &\sup_{\alpha\in(0,1)}|P(\max_{(i,j)\in G}\sum_{k=1}^{n}\xi_{ijk}/\sqrt{n} > c_G(\alpha)) - \alpha| + \xi_1\sqrt{1 \vee \log(p/\xi_1)} + \xi_2 \leqslant \\ &\sup_{\alpha\in(0,1)}|P(\max_{(i,j)\in G} z_{ij} > c_G(\alpha)) - \alpha| + n^{-c} + \xi_1\sqrt{1 \vee \log(p/\xi_1)} + \xi_2 \leqslant \\ &\pi(v) + P(\Gamma > v) + n^{-c} + \xi_1\sqrt{2\log p} + \xi_2.\end{aligned}$$

Under Assumption 3.4, by choosing $v = 1/(\alpha_n(\log p)^2)$, we deduce that

$$\sup_{\alpha\in(0,1)}|P(\max_{(i,j)\in G}\sqrt{n}(\widehat{T}_{ij} - \Theta_{ij}) > c_G(\alpha)) - \alpha| = o(1).$$

This completes the proof of Lemma A.4.

### Appendix A.2 The proof of Theorem 3.1

Given a probability tending to one, we obtain


DOI: 10.52396/JUSTC-2022-0004
JUSTC, 2022, 52(7): 2

$$|\widehat{\Theta}_{ij}-\Theta_{ij}|=O\left(K\sqrt{\frac{\log p}{n}}\right)=o(1),\ |\Theta_{ij}|=O(1),$$

for all $1\leqslant i\leqslant j\leqslant p$ values. Thus, we can conclude the following:

$$\begin{aligned}
&\widehat{\Theta}_{ij}\asymp\Theta_{ij}=O(1),\\
&|\widehat{\sigma}_{ij}^2|=|\widehat{\Theta}_{ii}\widehat{\Theta}_{jj}+\widehat{\Theta}_{ij}^2|=O(1),\\
&|\sigma_{ij}^2|=|\Theta_{ii}\Theta_{jj}+\Theta_{ij}^2|=O(1),\\
&|\widehat{\sigma}_{ij}^2-\sigma_{ij}^2|=|\widehat{\Theta}_{ii}\widehat{\Theta}_{jj}+\widehat{\Theta}_{ij}^2-(\Theta_{ii}\Theta_{jj}+\Theta_{ij}^2)|\leqslant\\
&|\widehat{\Theta}_{ii}\widehat{\Theta}_{jj}-\Theta_{ii}\Theta_{jj}|+|\widehat{\Theta}_{ij}^2-\Theta_{ij}^2|\leqslant\\
&|\widehat{\Theta}_{ii}||\widehat{\Theta}_{jj}-\Theta_{jj}|+|\Theta_{jj}||\widehat{\Theta}_{ii}-\Theta_{ii}|+|\widehat{\Theta}_{ij}+\Theta_{ij}||\widehat{\Theta}_{ij}-\Theta_{ij}|=\\
&O\left(K\sqrt{\frac{\log p}{n}}\right),\\
&|\widehat{\sigma}_{ij}-\sigma_{ij}|=|\widehat{\sigma}_{ij}^2-\sigma_{ij}^2|/|\widehat{\sigma}_{ij}+\sigma_{ij}|=O\left(K\sqrt{\frac{\log p}{n}}\right),\\
&|\widehat{\sigma}_{ij}\widehat{\sigma}_{kl}-\sigma_{ij}\sigma_{kl}|\leqslant|\widehat{\sigma}_{ij}||\widehat{\sigma}_{kl}-\sigma_{kl}|+|\sigma_{kl}||\widehat{\sigma}_{ij}-\sigma_{ij}|=\\
&O\left(K\sqrt{\frac{\log p}{n}}\right),
\end{aligned}$$

for each $1\leqslant i\leqslant j\leqslant p$. Then we have

$$\begin{aligned}
&\max_{(i,j)\in G}|\sqrt{n}(\widehat{T}_{ij}-\Theta_{ij})/\widehat{\sigma}_{ij}-\sum_{k=1}^{n}\xi_{ijk}/(\sqrt{n}\sigma_{ij})|\leqslant\\
&\max_{(i,j)\in G}|\Delta_{ij}|/\widehat{\sigma}_{ij}+\max_{(i,j)\in G}|\sum_{k=1}^{n}\xi_{ijk}|\max_{(i,j)\in G}|\frac{1}{\hat{\sigma}_{ij}}-\frac{1}{\sigma_{ij}}|=\\
&O(|\Delta|_\infty)+O\left(\sqrt{\log p}K\sqrt{\frac{\log p}{n}}\right).
\end{aligned}$$

Under Assumptions 3.3 and 3.4, there exist $\xi_1$ and $\xi_2$ that satisfy $\xi_1\sqrt{1\vee\log(p/\xi_1)}=o(1)$ and $\xi_2=o(1)$; then, we obtain

$$P(|\max_{(i,j)\in G}\sqrt{n}(\widehat{T}_{ij}-\Theta_{ij})/\widehat{\sigma}_{ij}-\max_{(i,j)\in G}\sum_{k=1}^{n}\xi_{ijk}/(\sqrt{n}\sigma_{ij})|\geqslant\xi_1)<\xi_2.$$

Let $\bar{\Gamma}=\max_{(i\leqslant j),(k\leqslant l)}|\widehat{\Xi}_{ij,kl}/(\hat{\sigma}_{ij}\hat{\sigma}_{kl})-\Xi_{ij,kl}/(\sigma_{ij}\sigma_{kl})|$, which then yields

$$\begin{aligned}
\bar{\Gamma}\leqslant&\max_{(i\leqslant j),(k\leqslant l)}|\Xi_{ij,kl}|\max_{(i\leqslant j),(k\leqslant l)}|\frac{1}{\widehat{\sigma}_{ij}\widehat{\sigma}_{kl}}-\frac{1}{\sigma_{ij}\sigma_{kl}}|+\\
&\max_{(i\leqslant j),(k\leqslant l)}|\hat{\Xi}_{ij,kl}-\Xi_{ij,kl}|\max_{(i\leqslant j),(k\leqslant l)}|\frac{1}{\widehat{\sigma}_{ij}\widehat{\sigma}_{kl}}|\leqslant\\
&O(1)\max_{(i\leqslant j),(k\leqslant l)}|\widehat{\sigma}_{ij}\widehat{\sigma}_{kl}-\sigma_{ij}\sigma_{kl}|+O\left(K\sqrt{\frac{\log p}{n}}\right)=\\
&O\left(K\sqrt{\frac{\log p}{n}}\right)
\end{aligned}$$

to arrive at the following conclusion:

$$\max_{(i\leqslant j),(k\leqslant l)}|\widehat{\Xi}_{ij,kl}-\Xi_{ij,kl}|=O\left(K\sqrt{\frac{\log p}{n}}\right).$$

Using the above arguments, we can show that $P(\bar{\Gamma}>v)=o(1)$ for $v=1/(\alpha_n(\log p)^2)$. The remaining proofs are similar to the proof of Lemma A.1; as such, we have omitted the details.

## Appendix A.3 The proof of Theorem 3.2

We define $Z_{ij}=z_{ij}/\sigma_{ij}$ for $1\leqslant i\leqslant j\leqslant p$. Thus, $\{Z_{ij}\}_{1\leqslant i\leqslant j\leqslant p}\sim N(\mathbf{0},\bar{\Xi})$ with

$$\bar{\Xi}=(\Xi_{ij,kl}/\sqrt{\Xi_{ij,ij}\Xi_{kl,kl}})_{i\leqslant j,\,k\leqslant l},$$

where is the regularization result of the matrix $\Xi$. As we have

$$P(|\max_{(i,j)\in G}\sqrt{n}(\widehat{T}_{ij}-\Theta_{ij})/\hat{\sigma}_{ij}-\max_{(i,j)\in G}\sum_{k=1}^{n}\xi_{ijk}/\sqrt{n}\sigma_{ij}|\geqslant\xi_1)<\xi_2,$$

where $\xi_1\sqrt{1\vee\log(p/\xi_1)}=o(1)$ and $\xi_2=o(1)$ from the proof of Theorem 3.1, the distribution of $\max_{(i,j)\in G}\sqrt{n}|\hat{T}_{ij}-\Theta_{ij}|/\widehat{\sigma}_{ij}$ can be approximated by $\max_{(i,j)\in G}|Z_{ij}|$. Under Lemma A.2, by Ref. [16, Lemma 6] for any $x\in R$ and as $|G|\to\infty$, we obtain

$$\begin{aligned}
&P(\max_{(i,j)\in G}|Z_{ij}|^2-2\log(|G|)+\log\log(|G|)\leqslant x)\to\\
&F(x):=\exp\left\{-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{x}{2}\right)\right\}.
\end{aligned}\tag{A.6}$$

Thus, we have

$$P(\max_{(i,j)\in G}n|\widehat{T}_{ij}-\Theta_{ij}|^2/\widehat{\sigma}_{ij}^2\leqslant 2\log(|G|)-\log\log(|G|)/2)\to 1.$$

The bootstrap statistic is a Gaussian variable; thus,

$$|(\bar{c}_G^*)^2-2\log(|G|)+\log\log(|G|)-q_\alpha|=o(1).$$

where $q_\alpha$ is the $100(1-\alpha)\%$ quantile of $F(x)$. There is a $(k,l)\in G$ such that $|\tilde{\Theta}_{kl}-\Theta_{kl}|/\widehat{\sigma}_{kl}>(\sqrt{2}+\epsilon_0)\sqrt{\log(|G|)/n}$. For any $\delta>0$, using the AM-GM inequality, we have

$$\begin{aligned}
n|\tilde{\Theta}_{kl}-\Theta_{kl}|^2/\widehat{\sigma}_{kl}^2\leqslant&(1+\delta^{-1})n|\widehat{T}_{kl}-\Theta_{kl}|^2/\widehat{\sigma}_{kl}^2+\\
&(1+\delta)n|\widehat{T}_{kl}-\tilde{\Theta}_{kl}|^2/\widehat{\sigma}_{kl}^2,
\end{aligned}$$

where

$$n|\widehat{T}_{kl}-\Theta_{kl}|^2/\widehat{\sigma}_{kl}^2=O(1)=o(\log(|G|))$$

as $(k,l)$ is fixed and $|G|$ is sufficiently large. From the proof of Theorem 3.1, we know that the difference between $n|\tilde{\Theta}_{kl}-\Theta_{kl}|^2/\widehat{\sigma}_{kl}^2$ and $n|\tilde{\Theta}_{kl}-\Theta_{kl}|^2/\sigma_{kl}^2$ is asymptotically negligible. Thus, given that $\boldsymbol{\Theta}\in\mathcal{U}_G(\sqrt{2}+\epsilon_0)$, we obtain

$$\max_{(i,j)\in G}n|\widehat{T}_{ij}-\tilde{\Theta}_{ij}|^2/\widehat{\sigma}_{ij}^2\geqslant\frac{1}{1+\delta}(\sqrt{2}+\epsilon_0)^2(\log|G|)-o(\log|G|).$$

Thus, for $\epsilon\to 0$, we have

$$\inf_{\boldsymbol{\Theta}\in\mathcal{U}_G(\sqrt{2}+\epsilon_0)}P(\max_{(i,j)\in G}\sqrt{n}|\widehat{T}_{ij}-\tilde{\Theta}_{ij}|/\hat{\sigma}_{ij}>\bar{c}_G^*(\alpha))\to 1.$$

## Appendix A.4 The proof of Corollary 3.1

Similar to the proof of Theorem 3.2, the distribution of $\max_{(i,j)\in G}\sqrt{n}|\hat{T}_{ij}-\Theta_{ij}|/\widehat{\sigma}_{ij}$ can be approximated by $\max_{(i,j)\in G}|Z_{ij}|$. Under Lemma A.2, by Ref. [16, Lemma 6], for any $x\in R$ and as $|G|\to\infty$, we obtain

$$\begin{aligned}
&P(\max_{(i,j)\in G}|Z_{ij}|^2-2\log(|G|)+\log\log(|G|)\leqslant x)\to\\
&F(x):=\exp\left\{-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{x}{2}\right)\right\}.
\end{aligned}$$

It implies that

$$P(\max_{(i,j)\in S_0^c}n|\widehat{T}_{ij}|^2/\widehat{\sigma}_{ij}^2\leqslant 2\log(|G|)-\log\log(|G|)/2)\to 1.\tag{A.7}$$


DOI: 10.52396/JUSTC-2022-0004
JUSTC, 2022, 52(7): 2

However, it is evident that

$$\min_{(i,j)\in S_0} n|T_{ij}|^2/\widehat{\sigma}_{ij}^2 \leqslant 2\max_{(i,j)\in S_0} n|\widehat{T}_{ij}-T_{ij}|^2/\widehat{\sigma}_{ij}^2 + 2\min_{(i,j)\in S_0} n|\widehat{T}_{ij}|^2/\widehat{\sigma}_{ij}^2.$$

As the difference between $\min_{(i,j)\in S_0} n|T_{ij}|^2/\widehat{\sigma}_{ij}^2$ and $\min_{(i,j)\in S_0} n|T_{ij}|^2/\sigma_{ij}^2$ is asymptotically negligible, and $P(2\max_{(i,j)\in S_0} n|\widehat{T}_{ij}-T_{ij}|^2/\widehat{\sigma}_{ij}^2 \leqslant 4\log(|G|)-\log\log(|G|)) \to 1$, we obtain

$$P(\min_{(i,j)\in S_0} n|T_{ij}|^2/\widehat{\sigma}_{ij}^2 > 2\log|G|) \geq$$
$$P(2\min_{(i,j)\in S_0} n|T_{ij}|^2/\widehat{\sigma}_{ij}^2 + 4\log|G| - \log\log(|G|) > 8\log|G|) \to 1. \tag{A.8}$$

Hence, Eq. (7) follows from Eqs. (A.7) and (A.8).

Next, we prove the optimality of $\tau = 2$. For a sufficiently large $M$, we can choose the set $G$: such that $\Theta_{ij} = 0$ for $(i,j) \in G$, and $|G| = M$. Based on the above arguments, we know that the distribution of $\max_{(i,j)\in G} \sqrt{n}|\hat{T}_{ij}-\Theta_{ij}|/\widehat{\sigma}_{ij}$ can be approximated by $\max_{(i,j)\in G}|Z_{ij}|$. From (A.7), we obtain the following:

$$P(\max_{(i,j)\in G} n|\widehat{T}_{ij}|^2/\widehat{\sigma}_{ij}^2 \geqslant c\log|G|) \to 1,$$

where $c < 2$. Thus, the conclusions follow.

## Appendix A.5 The proof of Lemma A.1

The proof of Lemma A.1 is shown in Ref. [13], and we use this conclusion in this study.

## Appendix A.6 The proof of Lemma A.2

As we have $(\log(pn))^7/n \leq C_1 n^{-c_1}$, let $B = O(1)$ be a positive constant. Thus,

$$B^2(\log(p^2n))^7/n \leqslant B^2 2^7(\log(pn))^7/n \leqslant C_1' n^{-c_1},$$

where $C_1' = 2^7B^2C_1 = O(1)$. Hence, we consider the application of Ref. [15, Corollary 2.1] to the i.i.d. sequence $\{\xi_{ijk}\}_{k=1}^n$. Only its Condition (E.1) must be verified. For clarity, we state the following condition:

$$c_0 \leqslant E\xi_{ijk}^2 \leqslant C_0, \ \max_{l=1,2} E|\xi_{ijk}|^{2+l}/B^l + E\mathrm{e}^{|\xi_{ijk}|/B} \leqslant 4 \tag{A.9}$$

uniformly over $j$, where $c_0, C_0 > 0$ and $B = O(1)$. As $\tilde{\boldsymbol{x}}$ is a normal random vector with $\tilde{\boldsymbol{x}} = \boldsymbol{x}\boldsymbol{\Theta}$, we obtain the following:

$$E\xi_{ijk}^2 = \mathrm{var}(\tilde{x}^i\tilde{x}^j) = \sigma_{ij}^2 = \Theta_{ii}\Theta_{jj} + \Theta_{ij}^2.$$

Thus,

$$E\xi_{ijk}^2 \geqslant \Theta_{ii}\Theta_{jj} \geqslant \lambda_{\min}^2(\Theta),$$
$$E\xi_{ijk}^2 \leqslant 2\Theta_{ii}\Theta_{jj} \leqslant 2\lambda_{\max}^2(\Theta).$$

It is straightforward to see that $\max_{l=1,2} E|\xi_{ijk}|^{2+l} = O(1)$; therefore, $B_1 = O(1)$, such that $\max_{l=1,2} E|\xi_{ijk}|^{2+l}/B_1^l \leqslant 2$.

Let $B_2 := |\Theta_{ij}|/\log(2)$. As $\tilde{X}$ is normal, there exists a constant $B_3$, such that $E\mathrm{e}^{\tilde{X}_i^2/B_3} \leqslant 1$ and $E\mathrm{e}^{\tilde{X}_j^2/B_3} \leqslant 1$. Now, let

$$B = \max\{B_1, B_2, B_3\} = O(1),$$

we then obtain

$$\max_{l=1,2} E|\xi_{ijk}|^{2+l}/B^l + E\mathrm{e}^{|\xi_{ijk}|/B} \leqslant 2 + E\mathrm{e}^{(|\tilde{X}_i\tilde{X}_j|+|\Theta_{ij}|)/B} \leqslant$$
$$2 + 2E\mathrm{e}^{|\tilde{X}_i\tilde{X}_j|/B} \leqslant 2 + 2E\mathrm{e}^{(\tilde{X}_i^2+\tilde{X}_j^2)/(2B)} \leqslant 2 + 2E\mathrm{e}^{\tilde{X}_i^2/B}E\mathrm{e}^{\tilde{X}_j^2/B} \leqslant 4,$$

where we use the average and Cauchy inequalities in the third and fourth inequations, respectively. This completes the proof of Lemma A.2.

## Appendix A.7 The proof of Lemma A.3

First, we prove that there is a constant $0 < c < 1$ such that $|\bar{\Theta}_{ij}| < c$ for all $i \neq j$. As $\boldsymbol{\Theta} \in \mathcal{G}(M,K)$, with $M = O(1)$, we have

$$M^{-1} \leqslant \lambda_{\min}(\boldsymbol{\Theta}) \leqslant \lambda_{\max}(\boldsymbol{\Theta}) \leqslant M.$$

Let $\boldsymbol{z} = \boldsymbol{e}_i/\sqrt{\Theta_{ii}} - \boldsymbol{e}_j/\sqrt{\Theta_{jj}}$. Then, we have

$$2 - 2\bar{\Theta}_{ij} = E(\tilde{x}_i/\sqrt{\Theta_{ii}} - \tilde{x}_j/\sqrt{\Theta_{jj}})^2 = E(\boldsymbol{z}^{\mathrm{T}}\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{z}) =$$
$$\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\Theta}\boldsymbol{z} \geqslant \lambda_{\min}(\boldsymbol{\Theta})\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z} = \left(\frac{1}{\Theta_{ii}} + \frac{1}{\Theta_{jj}}\right)\lambda_{\min}(\boldsymbol{\Theta}) \geqslant$$
$$\frac{2}{M}\lambda_{\min}(\boldsymbol{\Theta}).$$

Thus, $\bar{\Theta}_{ij} \leqslant 1 - \frac{1}{M^2}$. Similarly, we prove that $-\bar{\Theta}_{ij} \leqslant 1 - \frac{1}{M^2}$. Let $c = 1 - \frac{1}{M^2} < 1$, we have $|\bar{\Theta}_{ij}| < c$ for all $i \neq j$.

Notably,

$$\|\Theta_i\|_2^2 = \sum_{j=1}^p \Theta_{ij}^2 \leqslant \lambda_{\max}^2(\Theta) \leqslant M^2 = O(1).$$

For all $1 \leqslant i \leqslant j \leqslant p$,

$$\sum_{1\leqslant k\leqslant l\leqslant p} \Xi_{ij,kl}^2 = \sum_{1\leqslant k\leqslant l\leqslant p} (\Theta_{ik}\Theta_{jl} + \Theta_{il}\Theta_{jk})^2 \leqslant$$
$$\sum_{k=1}^p\sum_{l=1}^p (\Theta_{ik}\Theta_{jl} + \Theta_{il}\Theta_{jk})^2 \leqslant$$
$$2\sum_{k=1}^p\sum_{l=1}^p (\Theta_{ik}^2\Theta_{jl}^2 + \Theta_{il}^2\Theta_{jk}^2) =$$
$$4\sum_{k=1}^p\sum_{l=1}^p \Theta_{ik}^2\Theta_{jl}^2 = 4\sum_{k=1}^p \Theta_{ik}^2\sum_{l=1}^p \Theta_{jl}^2 \leqslant 4M^4.$$

Let $C_0 = 4M^4$; then,

$$\sum_{1\leqslant k\leqslant l\leqslant p} \Xi_{ij,kl}^2 \leqslant C_0.$$

Now, we consider two cases and prove that in each case

$$\max_{(i,j)\neq(k,l)} |\Xi_{ij,kl}|/(\sqrt{\Xi_{ij,ij}\Xi_{kl,kl}}) \leqslant c_0 < 1$$

holds. As $\max_{1\leqslant i<j\leqslant p} |\bar{\Theta}_{ij}| \leqslant c < 1$, for each $i < j, k < l$, we obtain the following:

$$\Xi_{ij,kl}^2/(\Xi_{ij,ij}\Xi_{kl,kl}) = \frac{(\Theta_{ik}\Theta_{jl} + \Theta_{il}\Theta_{jk})^2}{(\Theta_{ii}\Theta_{jj} + \Theta_{ij}^2)(\Theta_{kk}\Theta_{ll} + \Theta_{kl}^2)} = \frac{(\bar{\Theta}_{ik}\bar{\Theta}_{jl} + \bar{\Theta}_{il}\bar{\Theta}_{jk})^2}{(1+\bar{\Theta}_{ij}^2)(1+\bar{\Theta}_{kl}^2)}.$$

**Case 1**. The indices in $(i,j)$ and $(k,l)$ are the same.

Owing to the symmetry of $(i,j)$, $(k,l)$ in $\Xi_{ij,kl} = \Theta_{ik}\Theta_{jl} + \Theta_{il}\Theta_{jk}$, for convenience, we assume that $l = j$. Then, we have that $i \neq k$ and $|\bar{\Theta}_{ik}| \leqslant c$ because $(i,j) \neq (k,l)$. Thus,

$$\frac{(\bar{\Theta}_{ik}\bar{\Theta}_{jl} + \bar{\Theta}_{il}\bar{\Theta}_{jk})^2}{(1+\bar{\Theta}_{ij}^2)(1+\bar{\Theta}_{kl}^2)} \leqslant \frac{(\bar{\Theta}_{ik} + \bar{\Theta}_{ij}\bar{\Theta}_{jk})^2}{(1+\bar{\Theta}_{ij}\bar{\Theta}_{jk})^2} \leqslant \left(\frac{c+\bar{\Theta}_{ij}\bar{\Theta}_{jk}}{1+\bar{\Theta}_{ij}\bar{\Theta}_{jk}}\right)^2 \leqslant \left(\frac{c+1}{2}\right)^2,$$

where the Cauchy inequality is used in the first inequation.


DOI: 10.52396/JUSTC-2022-0004
JUSTC, 2022, 52(7): 2

Let $c_1 = 1 - \frac{1}{2M^2} < 1$. Then, $|\Xi_{ij,kl}|/\sqrt{\Xi_{ij,ij}\Xi_{kl,kl}} < c_1$ for all $(i,j) \neq (k,l)$.

**Case 2**. The indices in $(i,j)$ and $(k,l)$ are not the same.

The target formula can be rewritten in the following form:

$$\Xi^2_{ij,kl}/(\Xi_{ij,ij}\Xi_{kl,kl}) = \frac{(E(\tilde{x}_i\tilde{x}_j - \Theta_{ij})(\tilde{x}_k\tilde{x}_l - \Theta_{kl}))^2}{E(\tilde{x}_i\tilde{x}_j - \Theta_{ij})^2 E(\tilde{x}_k\tilde{x}_l - \Theta_{kl})^2}.$$

We divide the Gaussian variable $\tilde{x}_i$ into two independent Gaussian parts $\tilde{x}'_i$ and $\tilde{x}''_i$ such that $\tilde{x}_i = \tilde{x}'_i + \tilde{x}''_i$, with $\tilde{x}''_i$ being the projection of $\tilde{x}_i$ onto the linear space $\mathrm{span}\{\tilde{x}_k : 1 \leqslant k \leqslant p, k \neq i, j\}$. This implies that with an appropriate $p \times 1$ parameter vector $\boldsymbol{u}$ with $u_i = u_j = 0$, we obtain $\tilde{x}''_i = \boldsymbol{u}^{\mathrm{T}}\tilde{\boldsymbol{x}}$. This also implies that $E\tilde{x}'_i\tilde{x}''_i = 0$ and $E\tilde{x}'_i\tilde{x}_k = E\tilde{x}'_i\tilde{x}_l = 0$. Let $\boldsymbol{z}' = \boldsymbol{e}_i - \boldsymbol{u}$; we then have

$$E\tilde{x}'^2_i = E\boldsymbol{z}'^{\mathrm{T}}\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{z}' = \boldsymbol{z}'^{\mathrm{T}}E(\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^{\mathrm{T}})\boldsymbol{z}' = \boldsymbol{z}'^{\mathrm{T}}\boldsymbol{\Theta}\boldsymbol{z}' \geqslant \lambda_{\min}(\boldsymbol{\Theta})\boldsymbol{z}'^{\mathrm{T}}\boldsymbol{z}' \geqslant \lambda_{\min}(\boldsymbol{\Theta}).$$

Therefore, $E\tilde{x}'^2_i \geqslant \frac{1}{M}$. We then use the same partition method to divide $\tilde{x}_j$ into $\tilde{x}'_j$ and $\tilde{x}''_j$. It is evident that $\tilde{x}'_i$, $\tilde{x}'_j$ are independent of $\tilde{x}''_i$, $\tilde{x}''_j$ and are all zero-mean Gaussian variables.

Defining $E\tilde{x}'_i\tilde{x}'_j = \Theta'_{ij}$ and $E\tilde{x}''_i\tilde{x}''_j = \Theta''_{ij}$, we have

$$E\tilde{x}_i\tilde{x}_j = E(\tilde{x}'_i + \tilde{x}''_i)(\tilde{x}'_j + \tilde{x}''_j) = \Theta'_{ij} + \Theta''_{ij} = \Theta_{ij},$$

where we use the independence of $\tilde{x}'_i, \tilde{x}'_j$ and $\tilde{x}'_i, \tilde{x}''_j$. Thus, we have

$$\tilde{x}_i\tilde{x}_j - \Theta_{ij} = (\tilde{x}'_i\tilde{x}'_j - \Theta'_{ij}) + (\tilde{x}''_i\tilde{x}''_j - \Theta_{ij}) + (\tilde{x}'_i\tilde{x}''_j + \tilde{x}''_i\tilde{x}'_j).$$

Notably,

$$\begin{aligned}&E(\tilde{x}_i\tilde{x}_j - \Theta_{ij})(\tilde{x}_k\tilde{x}_l - \Theta_{kl}) = \\ &E(\tilde{x}'_i\tilde{x}'_j - \Theta'_{ij})E(\tilde{x}_k\tilde{x}_l - \Theta_{kl}) + E\tilde{x}'_iE\tilde{x}''_j(\tilde{x}_k\tilde{x}_l - \Theta_{kl}) + \\ &E\tilde{x}'_jE\tilde{x}''_i(\tilde{x}_k\tilde{x}_l - \Theta_{kl}) + E(\tilde{x}''_i\tilde{x}''_j - \Theta''_{ij})(\tilde{x}_k\tilde{x}_l - \Theta_{kl}) = \\ &E(\tilde{x}''_i\tilde{x}''_j - \Theta''_{ij})(\tilde{x}_k\tilde{x}_l - \Theta_{kl}),\end{aligned}$$

and

$$\begin{aligned}&\Theta_{ii}\Theta_{jj} + \Theta^2_{ij} = E(\tilde{x}_i\tilde{x}_j - \Theta_{ij})^2 = \\ &E((\tilde{x}'_i\tilde{x}'_j - \Theta'_{ij}) + (\tilde{x}''_i\tilde{x}''_j - \Theta_{ij}) + (\tilde{x}'_i\tilde{x}''_j + \tilde{x}''_i\tilde{x}'_j))^2 = \\ &E(\tilde{x}'_i\tilde{x}'_j - \Theta'_{ij})^2 + E(\tilde{x}''_i\tilde{x}''_j - \Theta''_{ij})^2 + E(\tilde{x}'_i\tilde{x}''_j + \tilde{x}''_i\tilde{x}'_j)^2 \geqslant \\ &E(\tilde{x}'_i\tilde{x}'_j - \Theta'_{ij})^2 + E(\tilde{x}''_i\tilde{x}''_j - \Theta''_{ij})^2,\end{aligned}$$

where we use the independence of these Gaussian variables. As

$$E(\tilde{x}_i\tilde{x}_j - \Theta_{ij})^2 = \Theta_{ii}\Theta_{jj} + \Theta^2_{ij} \leqslant 2\lambda^2_{\max}(\Theta) \leqslant 2M^2,$$

this implies that

$$\begin{aligned}\frac{\Xi^2_{ij,kl}}{\Xi_{ij,ij}\Xi_{kl,kl}} &= \frac{(E(\tilde{x}_i\tilde{x}_j - \Theta_{ij})(\tilde{x}_k\tilde{x}_l - \Theta_{kl}))^2}{E(\tilde{x}_i\tilde{x}_j - \Theta_{ij})^2 E(\tilde{x}_k\tilde{x}_l - \Theta_{kl})^2} = \\ &\frac{(E(\tilde{x}''_i\tilde{x}''_j - \Theta''_{ij})(\tilde{x}_k\tilde{x}_l - \Theta_{kl}))^2}{E(\tilde{x}_i\tilde{x}_j - \Theta_{ij})^2 E(\tilde{x}_k\tilde{x}_l - \Theta_{kl})^2} \leqslant \\ &\frac{E(\tilde{x}''_i\tilde{x}''_j - \Theta''_{ij})^2 E(\tilde{x}_k\tilde{x}_l - \Theta_{kl})^2}{E(\tilde{x}_i\tilde{x}_j - \Theta_{ij})^2 E(\tilde{x}_k\tilde{x}_l - \Theta_{kl})^2} \leqslant \\ &1 - \frac{E(\tilde{x}'_i\tilde{x}'_j - \Theta'_{ij})^2}{E(\tilde{x}_i\tilde{x}_j - \Theta_{ij})^2} \leqslant \\ &1 - \frac{E\tilde{x}'^2_iE\tilde{x}'^2_j + \Theta'^2_{ij}}{2M^2} \leqslant 1 - \frac{1}{2M^4},\end{aligned}$$

where we use the Cauchy inequality in the first inequation and $E\tilde{x}'^2_i \geqslant \frac{1}{M}$. If $c_2 = \sqrt{1 - \frac{1}{2M^4}} < 1$, we have $|\Xi_{ij,kl}|/\sqrt{\Xi_{ij,ij}\Xi_{kl,kl}} \leqslant c_2$.

Combining Cases 1 and 2, let $c = c_1 \vee c_2$; we then have

$$|\Xi_{ij,kl}|/\sqrt{\Xi_{ij,ij}\Xi_{kl,kl}} \leqslant c < 1$$

for all $(i,j) \neq (k,l)$.

This completes the proof of Lemma A.3.

## Conflict of interest

The authors declare that they have no conflict of interest.

## Biographies

**Wenjie Gao** is currently a master student at the School of Management, University of Science and Technology of China (USTC). He received his B.S. degree from USTC in 2019. His research interests focus on high-dimensional variable selection and inference.

**Jie Wu** is currently a Ph.D. student at the School of Management, University of Science and Technology of China. She received her B.S. degree from Anhui University of Technology in 2017. Her research mainly focuses on high-dimensional variable selection and classification.

## References


[1] Lauritzen S L. Graphical Models. London: Clarendon Press, 1996.

[2] Belilovsky E, Varoquaux G, Blaschko M B. Testing for differences in Gaussian graphical models: Applications to brain connectivity. https://arxiv.org/abs/1512.08643.

[3] Yuan M, Lin Y. Model selection and estimation in the Gaussian graphical model. *Biometrika,* **2007**, *94*: 19–35.

[4] Fan J Q, Yang F, Wu Y. Network exploration via the adaptive lasso and scad penalties. *The Annals of Applied Statistics,* **2009**, *3* (2): 521–541.

[5] Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics,* **2007**, *9*: 432–441.

[6] Meinshausen N, Bühlmann P. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics,* **2006**, *34*: 1436–1462.

[7] Cai T T, Liu W, Zhou H H. Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *The Annals of Statistics,* **2016**, *44*: 455–488.

[8] Peng J, Wang P, Zhou N, et al. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association,* **2009**, *104*: 735–746.

[9] Fan Y, Lv J. Innovated scalable efficient estimation in ultra-large Gaussian graphical models. *The Annals of Statistics,* **2016**, *44*: 2098–2126.

[10] Zhang C H, Zhang S S. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society,* **2014**, *76*: 217–242.

[11] Jankov J, van de Geer S. Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics,* **2015**, *9*: 1205–1229.

[12] Jankov J, van de Geer S. Honest confidence regions and optimality in high-dimensional precision matrix estimation. *Test,* **2017**, *26*: 143–162.

[13] Zhou J, Zheng Z, Zhou H, et al. Innovated scalable efficient inference for ultra-large graphical models. *Statistics and Probability Letters,* **2021**, *173*: 109085.

[14] Zhang X, Cheng G. Simultaneous inference for high-dimensional




DOI: 10.52396/JUSTC-2022-0004
*JUSTC*, **2022**, 52(7): 2

linear models. *Journal of the American Statistical Association,* **2017**, *112*: 757–768.

[15] Chernozhukov V, Chetverikov D, Kato K. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics,* **2013**, *41*: 2786–2819.

[16] Cai T T, Liu W, Xia Y. Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society, Series B(Statistical Methodology),* **2014**, *76*: 349–372.



DOI: 10.52396/JUSTC-2022-0004
*JUSTC*, **2022**, 52(7): 2