DIRECTIONAL FALSE DISCOVERY RATE CONTROL VIA DEBIASED AND DISTRIBUTED PROCEDURES IN GAUSSIAN GRAPHICAL MODELS

Ensiyeh Nezakati, Eugen Pircalabelu

LIDAM Discussion Paper ISBA
2023 / 24



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

Directional false discovery rate control via debiased and distributed procedures in Gaussian graphical models

Ensiyeh Nezakati[†] NEZAKATI.ENSIYEH@UCLOUVAIN.BE

Eugen Pircalabelu[†] EUGEN.PIRCALABELU@UCLOUVAIN.BE

[†]*Institute of Statistics, Biostatistics and Actuarial Sciences*
UCLouvain
Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium

Abstract

In this paper, a multiple testing procedure is established for the conditional dependence in Gaussian graphical models. In practice, it is important to determine whether the conditional dependence between variables is positive or negative. However, there are several challenges to building statistics for testing using sample data. For instance, due to privacy concerns, one is not able to aggregate different datasets from several locations in one single location. In this study, different test statistics are constructed using debiased and distributed estimators to address this problem in a multiple testing framework. It is shown that, under mild conditions, the proposed procedure can control asymptotically the directional false discovery rate, which focuses on the sign of the estimation, at a prespecified level. An asymptotic power equal to one is also attainable under mild conditions on the non-zero entries of the precision matrix. Different simulation scenarios and real data examples are used to investigate the performance of the proposed procedure and to confirm the theoretical results.

Keywords: Multiple hypotheses testing, directional false discovery rate, Gaussian graphical models, debiased estimator, distributed estimator.

1. Introduction

Gaussian graphical models (GGMs), which represent the conditional dependence relationships among a set of random variables through a graph structure, have many applications in the real world, including brain connectivity based on FMRI measurements (Ng et al., 2013), gene regulatory network discovery (Schäfer and Strimmer, 2005), social network analysis, and many more. A large body of literature studied the estimation of GGMs, such as Yuan and Lin (2007), Friedman et al. (2008), Cai et al. (2011) and Ravikumar et al. (2011) among many others.

The estimation of GGMs often leads to a high false discovery rate (FDR) when testing the hypothesis $H_{0,ab} : \mathbf{\Theta}_{ab} = 0$ vs $H_{1,ab} : \mathbf{\Theta}_{ab} \neq 0$, where $\mathbf{\Theta}_{ab}$ is the (a, b)-th entry of the inverse covariance matrix $\mathbf{\Theta}$ in GGMs. This is typically due to the presence of high-dimensional data and small sample sizes in many practical applications. Moreover, due to the unknown distribution of the graphical Lasso estimator, inference for GGMs is always challenging. Liu (2013) considered inference for GGMs based on a nodewise regression approach and introduced a multiple testing procedure on partial correlations with false discovery proportion (FDP) and FDR control. Xia et al. (2015) studied simultaneous testing for differential networks. Recently, Li et al. (2022) constructed a multiple testing procedure

based on a debiased CLIME estimator for edge detection with FDR control. They have shown theoretically and numerically that this procedure controls the FDR in the multiple testing under a pre-specified level. In this work, we propose a procedure based on the debiased graphical Lasso estimator which controls the directional FDR as well as the FDR at a pre-specified level.

The directional FDR, which was first introduced by Gelman and Tuerlinckx (2000), is a useful metric for identifying type s errors (where "s" stands for sign). A type s error (also known as type III error) occurs when one confidently claims that a comparison goes one way, but it actually goes the other way. This measure is relevant for GGMs since by identifying the connections between the nodes, one is able to also state with confidence that their conditional dependence is positive or negative. In fact, by controlling the type s error in graphical models, one can control type I error as well as the error that occurs when estimating the sign. In the framework of regression, Zhao and Yu (2006) studied the sign consistency of Lasso based on the Irrepresentable condition. Jia et al. (2013) studied sign consistency of a sparse Poisson-like heteroscedastic model. Barber and Candès (2019) introduced a procedure using knockoff filters to control the directional FDR for high-dimensional linear models. Recently, Javanmard and Javadi (2019) and Cui et al. (2021) proposed different procedures to control the directional FDR under a pre-specified level using a debiased estimator for high-dimensional linear regression models with a random design and high-dimensional generalized linear models, respectively. In this paper, our goal is to maintain under control the directional FDR at a pre-specified level for Gaussian graphical models using a debiased estimator when data are distributed on multiple machines.

The process of collecting all sample data at a single location for estimation and inference is becoming challenging these days due to the growing size of the datasets, limited capacity of machines and security and privacy concerns. For example, suppose that one wants to estimate a genetic network for a specific disease using datasets that come from different hospitals. To maintain the privacy of patients' information, one cannot aggregate all the datasets in one location. To address these issues, distributed statistical approaches, which involve aggregating multiple estimators from different locations, have gained popularity in recent years for solving various statistical problems. The reader can refer to Chen and Xie (2014), Battey et al. (2018), Jordan et al. (2018) and Tang et al. (2020) for some relevant references on this topic. Nezakati and Pircalabelu (2023b) developed a new aggregated estimator for GGMs based on debiased local estimators which use non-overlapping parts of the dataset with unbalanced sizes. A similar approach has been explored in Nezakati and Pircalabelu (2023a) for covariate adjusted GGMs. However, to the best of our knowledge, the control of the (directional) FDR in the distributed setting has not yet been investigated.

The rest of the paper is organized as follows. Notation and preliminaries are presented in Section 2. The debiased estimator using the full dataset and distributed estimators are reviewed in Section 3 and their properties are presented. A procedure for controlling the directional FDR is proposed in Section 4 based on the debiased estimator using the full dataset and distributed estimators and their statistical guarantees are investigated. The performance of this procedure is evaluated via a simulation study and real data examples in Sections 5 and 6, respectively. We close with a discussion on the method and possible extensions in Section 7. Proofs of the supporting lemmas and theorems can be found in the Appendix.

2. Preliminaries and notation

Consider the Gaussian graphical models framework, where a p-dimensional random vector $\dot{\mathbf{X}} = (X^1, \ldots, X^p)^\top$ follows a multivariate Gaussian distribution with mean vector zero and covariance matrix $\mathbf{\Sigma}$. The components of $\dot{\mathbf{X}}$ are mapped to the node set $\mathcal{V} = \{1, \ldots, p\}$ of an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the edge set $\mathcal{E}$ describes the conditional dependence between every pair of components $X^1, \ldots, X^p$. A pair (a, b) is not included in the edge set $\mathcal{E}$ if and only if the variables X^a and X^b are independent given all remaining variables. Under the multivariate Gaussian distribution, a pair of variables is conditionally independent given all remaining variables if and only if the corresponding entry in the precision matrix $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$ is equal to zero. Elements of the precision matrix may thus be interpreted as the edge weights. Denote the maximum number of non-zero (active) off-diagonal entries of $\mathbf{\Theta}$ per row, i.e., the maximal node degree, with d and the index set of the non-zero upper diagonal entries of $\mathbf{\Theta}$ with S. Formally,

$$S = \{(a, b) \in \mathcal{V} \times \mathcal{V}, a < b : \quad \mathbf{\Theta}_{ab} \neq 0\},$$

having cardinality $s = \#S$, where $\mathbf{\Theta}_{ab}$ is the (a, b)-th element of $\mathbf{\Theta}$. The set of non-active components is denoted by S^c.

In this paper, we propose a framework to select a set $\hat{S}$ of pairs of nodes, while controlling the directional false discovery rate (FDR_{dir}) for the selected pairs. We define FDR_{dir} for a selected set $\hat{S}$ of pairs of nodes with estimated $\widehat{\text{sign}}_{ab} \in \{-1, 1\}$ for the sign of $\mathbf{\Theta}_{ab}$ as

$$\text{FDR}_{\text{dir}} = \mathbb{E}[\text{FDP}_{\text{dir}}], \quad \text{FDP}_{\text{dir}} = \frac{\#\{(a, b) \in \hat{S} : \quad \widehat{\text{sign}}_{ab} \neq \text{sign}_{ab}\}}{\max\{\hat{s}, 1\}}, \tag{1}$$

where $\hat{s}$ denotes the cardinality of $\hat{S}$, i.e., $\hat{s} = \#\hat{S}$, $\text{sign}_{ab} = \text{sign}(\mathbf{\Theta}_{ab})$ and by convention $\text{sign}(0) = 0$. Based on this definition, FDR_{dir} represents the expected fraction of false discoveries among the selected edges, where a false discovery is measured with respect to type s and type I errors. For example, if $\widehat{\text{sign}}_{ab} = +1$ while $\mathbf{\Theta}_{ab} = 0$, we consider that the procedure which selects $\hat{S}$ made a type I error. On the other hand, if $\widehat{\text{sign}}_{ab} = +1$ while $\mathbf{\Theta}_{ab} < 0$, we consider that the procedure made a type s error. In both cases, it is considered as a false discovery. From (1) and based on the definition of the classical FDR, i.e., $\text{FDR} = \mathbb{E}\big[\frac{\#\{(a,b)\in\hat{S}: \quad \mathbf{\Theta}_{ab}=0\}}{\max\{\hat{s},1\}}\big]$, it holds that $\text{FDR}_{\text{dir}} \geq \text{FDR}$. As such, by controlling the FDR_{dir}, one can provide a control on the FDR, too. It is noteworthy that controlling the FDR_{dir} is more challenging. The reason is that when one works with type I errors, the false discovery rate is built under the null hypothesis $H_{0,ab} : \mathbf{\Theta}_{ab} = 0$. In contrast, the directional FDR is not built under the null hypothesis and the element $\mathbf{\Theta}_{ab}$ might be nonzero with a positive or negative sign. Similarly, we define the directional power of a selected set $\hat{S}$ as

$$\text{Power}_{\text{dir}} = \mathbb{E}\left[\frac{\#\{(a, b) \in \hat{S} : \quad \widehat{\text{sign}}_{ab} = \text{sign}_{ab}\}}{\max\{s, 1\}}\right]. \tag{2}$$

In this paper, we provide a theoretical procedure, where we control the directional false discovery rate and we achieve asymptotic power equal to one using a debiased distributed estimator.

Before starting the discussion, we introduce some notation which is needed later in the paper. For two sequences $\{a_n; n \geq 1\}$ and $\{b_n; n \geq 1\}$, $b_n = O(a_n)$ if there exist positive numbers M_0 and N_0 such that $|\frac{b_n}{a_n}| \leq M_0$ for all $n \geq N_0$. We write $b_n \asymp a_n$ if both $b_n = O(a_n)$ and $a_n = O(b_n)$ hold. Similarly, for a random sequence $\{X_n; n \geq 1\}$, we write $X_n = O_p(a_n)$ if for every $\epsilon > 0$, there exist finite numbers $M_0 > 0$ and $N_0 > 0$ such that $\mathbb{P}(|\frac{X_n}{a_n}| > M_0) < \epsilon$ for all $n \geq N_0$. Furthermore, $b_n = o(a_n)$ if $\lim_{n\to\infty} \frac{b_n}{a_n} = 0$. In the case of a random sequence $\{X_n; n \geq 1\}$, we write $X_n = o_p(a_n)$ if $\frac{X_n}{a_n} \xrightarrow{p} 0$, as $n \to \infty$, where the notation $\xrightarrow{p}$ denotes convergence in probability. For a matrix $\mathbf{A}$, we use the notation $|||\mathbf{A}|||_\infty = \max_a \sum_b |\mathbf{A}_{ab}|$ and $||\mathbf{A}||_\infty = \max_{a,b} |\mathbf{A}_{ab}|$ for the matrix and elementwise ℓ_∞ norms, respectively. Moreover, $||\mathbf{A}||_1 = \sum_{a,b} |\mathbf{A}_{ab}|$ and $||\mathbf{A}||_F = \sqrt{\sum_{a,b} \mathbf{A}_{ab}^2} = \sqrt{\text{trace}(\mathbf{A}\mathbf{A}^\top)}$ are used for the elementwise ℓ_1 and Frobenius norms of a matrix $\mathbf{A}$, respectively. Finally, $a \vee b = \max\{a, b\}$.

3. Debiased Lasso estimator

Consider i.i.d. samples $\dot{\mathbf{X}}_1, \ldots, \dot{\mathbf{X}}_n$ from $\dot{\mathbf{X}}$, each of length p, and arrange them in matrix form denoted as $\mathbf{X}$ of size $n \times p$. The goal is to control the false discovery rate for the procedure estimating the structure of the graph $\mathcal{G}$, or equivalently, finding the zero pattern of $\boldsymbol{\Theta}$. The usual assumption in this context is that the underlying graph is sparse, which is translated in imposing a certain bound on the maximal node degree (d) of the precision matrix. To impose this sparsity condition on the estimation, one common approach is to add an ℓ_1 penalty to the negative log-likelihood function. There exists a wide variety of methods making use of ℓ_1 regularization, see for example Friedman et al. (2008), Hsieh et al. (2014), Cai et al. (2016), among others. The graphical Lasso estimator $\hat{\boldsymbol{\Theta}}$, defined by Friedman et al. (2008), is the solution to the optimization problem

$$\hat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta} \in S_{++}^p} \Big\{ \text{trace}\big(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}\big) - \log\det(\boldsymbol{\Theta}) + \lambda ||\boldsymbol{\Theta}||_{1,\text{off}} \Big\}, \tag{3}$$

where $\hat{\boldsymbol{\Sigma}} = \mathbf{X}^\top\mathbf{X}/n$ is the sample covariance, $\lambda > 0$ is the penalty term, $||\cdot||_{1,\text{off}}$ is the ℓ_1 off-diagonal norm of the matrix, defined as $||\boldsymbol{\Theta}||_{1,\text{off}} = \sum_{a\neq b} |\boldsymbol{\Theta}_{ab}|$ and S_{++}^p is the space of positive definite matrices of dimension $p \times p$. The graphical Lasso estimator (3) is biased due to the ℓ_1 penalty which is added to the negative log-likelihood function. Using the Karush-Kuhn-Tucker (KKT) condition (see for example, Ravikumar et al., 2011), and certain regularity assumptions (assumptions A1-A2 which are provided in Appendix A for the reader's convenience), Jankova and van de Geer (2015) proposed the debiased graphical Lasso estimator as

$$\hat{\boldsymbol{\Theta}}^d = 2\hat{\boldsymbol{\Theta}} - \hat{\boldsymbol{\Theta}}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\Theta}}. \tag{4}$$

We call this estimator as "debiased full", since it uses the full dataset of size $n \times p$. Under regularity assumptions, presented in Appendix A, with $\lambda \asymp \sqrt{\log(p)/n}$ and the sparsity condition $d^{3/2} = o(\sqrt{n}/\log(p))$, it has been shown in Jankova and van de Geer (2015) that for every $(a, b) \in \mathcal{V} \times \mathcal{V}$,

$$\frac{\sqrt{n}}{\sigma_{ab}}(\hat{\boldsymbol{\Theta}}_{ab}^d - \boldsymbol{\Theta}_{ab}) = \frac{1}{\sqrt{n}\sigma_{ab}} \sum_{l=1}^n \mathbf{Z}_{ab,l} + \frac{\sqrt{n}}{\sigma_{ab}} \mathbf{R}_{ab}, \tag{5}$$

where $\frac{1}{\sqrt{n}\sigma_{ab}}\sum_{l=1}^{n}\mathbf{Z}_{ab,l}$ converges weakly to $\mathcal{N}(0,1)$, and $|\mathbf{R}_{ab}| = o_p(1/\sqrt{n})$. Note that $\hat{\mathbf{\Theta}}^d_{ab}$ and $\mathbf{R}_{ab}$ are the (a,b)-th element of $\hat{\mathbf{\Theta}}^d$ and a remainder term matrix, called $\mathbf{R}$, respectively. Moreover, $\sigma^2_{ab} = \mathbf{\Theta}_{aa}\mathbf{\Theta}_{bb} + \mathbf{\Theta}^2_{ab}$ and $\mathbf{Z}_{ab,l} = \mathbf{\Theta}_a^\top \dot{\mathbf{X}}_l \dot{\mathbf{X}}_l^\top \mathbf{\Theta}_b - \mathbf{\Theta}_{ab}$, where $\mathbf{\Theta}_a$ is the a-th column of $\mathbf{\Theta}$. Using Jankova and van de Geer (2015), it can be shown that

$$||\hat{\mathbf{\Theta}}^d - \mathbf{\Theta}||_\infty = O_p\bigg(\max\big\{d\sqrt{\log(p)/n}, d^{3/2}\log(p)/n, d^2(\log(p)/n)^{3/2}\big\}\bigg),$$

which is an $o_p(1)$ quantity under the mentioned sparsity assumption on d.

Now, consider K i.i.d. samples from $\mathbf{X}$ and denote the k-th sample by $\mathbf{X}_k$ with size $n_k \times p$, $k = 1, \ldots, K$, such that $n_k/n \to c_k \in (0,1)$ as $n_k \to \infty$, and $\lim_{K\to\infty}\sum_{k=1}^{K} c_k = 1$. For any fixed K, one can utilize each sample in parallel, and using (4), we have access to K debiased estimators $\hat{\mathbf{\Theta}}^d_{ab,k}$, $k = 1, \ldots, K$, for every element $(a,b) \in \mathcal{V}\times\mathcal{V}$. Due to the asymptotic normal distribution of the local estimators, a final aggregated estimator has been derived in Nezakati and Pircalabelu (2023a,b) by maximizing a pseudo log-likelihood function. This function is constructed using the asymptotic normal density of the local estimators $\hat{\mathbf{\Theta}}^d_{ab,k}$, $k = 1, \ldots, K$, and the aggregated estimator is of the form

$$\tilde{\mathbf{\Theta}}_{ab} = \frac{1}{\sum_{k=1}^{K}\frac{n_k}{\hat{\sigma}^2_{ab,k}}} \times \sum_{k=1}^{K}\frac{n_k}{\hat{\sigma}^2_{ab,k}}\hat{\mathbf{\Theta}}^d_{ab,k}. \tag{6}$$

The reader can refer to the mentioned references above for more details on the derivations and theoretical properties of the estimator in (6). Note that $\hat{\sigma}^2_{ab,k}$ in (6) is a consistent estimator of σ^2_{ab} obtained based on the k-th sample and is of the form $\hat{\sigma}^2_{ab,k} = \hat{\mathbf{\Theta}}_{aa,k}\hat{\mathbf{\Theta}}_{bb,k} + \hat{\mathbf{\Theta}}^2_{ab,k}$, where $\hat{\mathbf{\Theta}}_{ab,k}$ is the (a,b)-th element of the graphical Lasso estimator in (3). It has been shown in Nezakati and Pircalabelu (2023b) that for every $(a,b) \in \mathcal{V}\times\mathcal{V}$,

$$\sqrt{\sum_{k=1}^{K}\frac{n_k}{\hat{\sigma}^2_{ab,k}}}\bigg(\tilde{\mathbf{\Theta}}_{ab} - \mathbf{\Theta}_{ab}\bigg) = \mathbf{W}_{ab,\dagger} + \mathbf{R}_{ab,\dagger},$$

where $\mathbf{W}_{ab,\dagger} \xrightarrow{d} \mathcal{N}(0,1)$ and $|\mathbf{R}_{ab,\dagger}| = o_p(1)$, as K grows at the rate $K = O\big(n^{1/3}/(d\log(p))\big)$, under the sparsity condition $d^{3/2} = o(\sqrt{n_\dagger}/\log(p))$ and $\lambda_k \asymp \sqrt{\log(p)/n_k}$, where $n_\dagger = \min_{1\le k\le K} n_k$. Note that, $\mathbf{R}_{ab,\dagger}$ is the (a,b)-th element of a remainder term matrix, called $\mathbf{R}_\dagger$. It can be shown that

$$||\tilde{\mathbf{\Theta}} - \mathbf{\Theta}||_\infty = O_p\bigg(K\max\big\{d\sqrt{\log(p)/n}, d^{3/2}\log(p)/n, d^2(\log(p))^{3/2}/(n\sqrt{n_\dagger})\big\}\bigg),$$

where $\tilde{\mathbf{\Theta}}$ is the matrix form of the aggregated estimator with elements $\tilde{\mathbf{\Theta}}_{ab}$, $(a,b) \in \mathcal{V}\times\mathcal{V}$. Using the triangle inequality, it is deduced that also $||\tilde{\mathbf{\Theta}} - \hat{\mathbf{\Theta}}^d||_\infty$ is equal to the bound above. This implies that the convergence rate at which the distributed estimator approaches the full estimator $\hat{\mathbf{\Theta}}^d$ is the same convergence rate at which it approaches the true matrix $\mathbf{\Theta}$. As such, it is noteworthy that $\tilde{\mathbf{\Theta}}$ approximates the debiased full estimator $\hat{\mathbf{\Theta}}^d$ well, and it exhibits a similar statistical error as $\hat{\mathbf{\Theta}}^d$ as long as the number of machines K is not too large.

In the next section, we provide a procedure using the debiased full and distributed estimator to control the directional false discovery rate.

4. Controlling the directional false discovery rate

In order to perform model selection for high-dimensional GGMs, we consider the multiple testing problem

$$H_{0,ab} : \mathbf{\Theta}_{ab} = 0 \quad \text{vs} \quad H_{1,ab} : \mathbf{\Theta}_{ab} \neq 0, \quad 1 \leq a < b \leq p.$$

Here the $p(p-1)/2$ hypotheses are tested at the same time and it is equivalent to selecting the set $\hat{S}$ of estimated non-zero entries of $\mathbf{\Theta}$. By controlling the directional FDR, one not only cares about type I error, but also about the sign of the entries of the selected set $\hat{S}$ and whether they match with the sign of the true entries.

4.1 Directional FDR control via the debiased full estimator

To control the directional FDR using the debiased full estimator from (4), for every $1 \leq a < b \leq p$, we define the test statistic

$$\hat{T}_{ab} := \frac{\sqrt{n}\hat{\mathbf{\Theta}}^d_{ab}}{\hat{\sigma}_{ab}}, \tag{7}$$

where $\hat{\sigma}_{ab}$ is a consistent estimator for σ_{ab} defined as $\hat{\sigma}^2_{ab} = \hat{\mathbf{\Theta}}_{aa}\hat{\mathbf{\Theta}}_{bb} + \hat{\mathbf{\Theta}}^2_{ab}$ (for more details, see Lemma 2 of Jankova and van de Geer, 2015). For a given threshold $t > 0$, we reject $H_{0,ab} : \mathbf{\Theta}_{ab} = 0$ in favor of $H_{1,ab} : \mathbf{\Theta}_{ab} \neq 0$, if $|\hat{T}_{ab}| \geq t$ and we return the sign of $\hat{T}_{ab}$ as the estimated sign of $\mathbf{\Theta}_{ab}$. With this test statistic, the directional FDP at a given threshold t is defined as

$$\text{FDP}_{\text{dir}}(t) = \frac{\sum_{(a,b)\in S_{\geq 0}} \mathbb{I}(\hat{T}_{ab} \leq -t) + \sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(\hat{T}_{ab} \geq t)}{R(t) \vee 1}, \tag{8}$$

where

$$S_{\leq 0} = \{(a,b) \in \mathcal{V} \times \mathcal{V}, a < b : \quad \mathbf{\Theta}_{ab} \leq 0\},$$
$$S_{\geq 0} = \{(a,b) \in \mathcal{V} \times \mathcal{V}, a < b : \quad \mathbf{\Theta}_{ab} \geq 0\}.$$

Moreover, $R(t) := \sum_{1\leq a<b\leq p} \mathbb{I}(|\hat{T}_{ab}| \geq t)$ is the total number of rejections at threshold t, with $\mathbb{I}(\cdot)$ being the indicator function. In words, the term $\sum_{(a,b)\in S_{\geq 0}} \mathbb{I}(\hat{T}_{ab} \leq -t)$ in (8) counts the number of times that $\mathbf{\Theta}_{ab}$ is truly non-negative, but one mistakenly estimates its sign as negative. A similar argument holds for $\sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(\hat{T}_{ab} \geq t)$.

Define next the random variable

$$\tilde{T}_{ab} := \frac{1}{\sqrt{n}\hat{\sigma}_{ab}} \sum_{l=1}^{n} \mathbf{Z}_{ab,l} + \frac{\sqrt{n}}{\hat{\sigma}_{ab}} \mathbf{R}_{ab}. \tag{9}$$

Using the definition of $\hat{T}_{ab}$ in (7), it can be shown that $\hat{T}_{ab} = \frac{\sqrt{n}\mathbf{\Theta}_{ab}}{\hat{\sigma}_{ab}} + \frac{\sqrt{n}}{\hat{\sigma}_{ab}}\mathbf{R}_{ab} + \frac{1}{\sqrt{n}\hat{\sigma}_{ab}} \sum_{l=1}^{n} \mathbf{Z}_{ab,l}$, with $\mathbf{R}_{ab}$ and $\mathbf{Z}_{ab,l}$ the same quantities as in (5). Then on the set $S_{\leq 0}$ we get $\hat{T}_{ab} \leq \tilde{T}_{ab}$ and as such, $\mathbb{I}(\hat{T}_{ab} \geq t) \leq \mathbb{I}(\tilde{T}_{ab} \geq t)$. On the other hand, on $S_{\geq 0}$, we have $\hat{T}_{ab} \geq \tilde{T}_{ab}$ and as such $\mathbb{I}(\hat{T}_{ab} \leq -t) \leq \mathbb{I}(\tilde{T}_{ab} \leq -t)$. As a result, in (8), we can write

$$\text{FDP}_{\text{dir}}(t) \leq \frac{\sum_{(a,b)\in S_{\geq 0}} \mathbb{I}(\tilde{T}_{ab} \leq -t) + \sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(\tilde{T}_{ab} \geq t)}{R(t) \vee 1}. \tag{10}$$

In the sequel, we propose a procedure for which one can control the directional FDR by bounding the right hand side of (10) at a pre-specified level $0 \leq q \leq 1$.

As it is shown in Lemma 1, the tail probability $\mathbb{P}(|\tilde{T}_{ab}| > t)$ can be approximated by $G(t)$ uniformly in $1 \leq a < b \leq p$, where $G(t) = 2(1 - \Phi(t))$ and $\Phi(t)$ is the cumulative distribution function of the standard normal distribution. As such, the expected value of the numerator in (10) can be approximated by $p_1 G(t)$, where $p_1 := p(p-1)/2$. To find the threshold t for a pre-specified level $0 \leq q \leq 1$, let $t_p = \big(2\log(p_1) - 2\log(\log(p_1))\big)^{1/2}$ and define

$$t_0 = \inf\big\{0 \leq t \leq t_p : \quad \frac{p_1 G(t)}{R(t) \vee 1} \leq q\big\}. \tag{11}$$

If the value in (11) does not exist, let $t_0 = \sqrt{2\log(p_1)}$. For every $1 \leq a < b \leq p$, we reject $H_{0,ab}$ if $|\hat{T}_{ab}| \geq t_0$ and we return $\widehat{\text{sign}}_{ab} = \text{sign}(\hat{T}_{ab})$ as the estimate for $\text{sign}(\boldsymbol{\Theta}_{ab})$.

To provide statistical guarantees on controlling the directional FDR using the defined procedure, we need to introduce two technical sets. For every $\gamma > 0$ and some constant $c_0 > 0$, define first the set

$$\Gamma(\gamma, c_0) = \big\{1 \leq a < b \leq p : \quad |\boldsymbol{\Theta}^0_{ab}| \geq c_0(\log(p))^{-1-\gamma/2}\big\}, \tag{12}$$

with cardinality $\Upsilon = \#\Gamma(\gamma, c_0)$, and define next the set

$$\Psi(\rho) = \big\{1 \leq a < b \leq p : \quad |\boldsymbol{\Theta}^0_{ab}| \geq \frac{1}{\sqrt{2}}\big(\frac{1-\rho}{1+\rho}\big)\big\},$$

for every $\rho \in (0,1)$, with cardinality $\psi = \#\Psi(\rho)$, where $\boldsymbol{\Theta}^0_{ab}$ is the normalized form of $\boldsymbol{\Theta}_{ab}$, defined as $\boldsymbol{\Theta}^0_{ab} = \boldsymbol{\Theta}_{ab}/\sqrt{\boldsymbol{\Theta}_{aa}\boldsymbol{\Theta}_{bb}}$. We call the set $\Gamma(\gamma, c_0)$ as the set of 'highly correlated pair nodes', and its complement set, which is denoted by $\Gamma^c(\gamma, c_0)$, as the set of 'weakly correlated pair nodes'. These two sets $\Gamma(\gamma, c_0)$ and $\Psi(\rho)$, involving the elements of the precision matrix $\boldsymbol{\Theta}$, are similar to the sets defined in Javanmard and Javadi (2019), are needed to guarantee the control of the directional FDR in GGMs, and are used explicitly in the proofs of Lemma 1 and Theorem 2.

Since the threshold t_0 tends typically to infinity with a growing number of nodes p, to control the directional FDP in (10), the elementwise asymptotic normality result in (5) is not enough to provide statistical guarantees for the procedure. As such we use a stronger, Cramér-type moderate deviation result similar to Liu (2013). This is provided in Lemma 1.

Lemma 1 Let $p \leq n^r$ for some $r > 0$ and suppose that assumptions (A1)-(A2) in Appendix A hold. Consider the random variable $\tilde{T}_{ab}$ defined in (9), and denote by

$$T_{ab} := \frac{1}{\sqrt{n}\sigma_{ab}} \sum_{l=1}^{n} \mathbf{Z}_{ab,l}, \tag{13}$$

for every $1 \leq a < b \leq p$, where $\mathbf{Z}_{ab,l}$ is defined in (5). Under the sparsity condition $d^{3/2} = o\big(\sqrt{n}/(\log(p))^{3/2}\big)$, and considering the set $\Gamma(\gamma, c_0)$ with cardinality $\Upsilon = o(p^{\delta\rho})$, $\delta \in (0, 1/2]$, $\rho \in (0,1)$, we have

$$\max_{a,b} \big|\tilde{T}_{ab} - T_{ab}\big| = o_p\big(1/\sqrt{\log(p)}\big) \tag{14}$$

and

$$\max_{a,b} \sup_{0\leq t\leq t_p} \left| \frac{\mathbb{P}(|T_{ab}| > t)}{G(t)} - 1 \right| \to 0.$$

The proof is given in Appendix B.1.

Theorem 2 next provides statistical guarantees for the control of the directional FDR using the threshold introduced in (11).

Theorem 2 *Let $p \leq n^r$, for some $r > 0$, and suppose that assumptions* (A1)-(A2) *in Appendix A hold. Consider the test statistic $\hat{T}_{ab}$ defined in* (7) *based on the debiased estimator from* (4)*. Under the sparsity condition $d^{3/2} = o(\sqrt{n}/(\log(p))^{3/2})$, and considering the sets $\Gamma(\gamma, c_0)$ and $\Psi(\rho)$ with cardinalities $\Upsilon = o(p^{\delta\rho})$, $\rho \in (0,1)$, $\delta \in (0, 1/2]$ and $\psi = O(p^{\delta\rho}/\sqrt{\log(p)})$, respectively, for the threshold t_0 defined in* (11) *and every $\epsilon > 0$, we have*

$$\limsup_{(n,p)\to\infty} \text{FDR}_{\text{dir}} \leq q, \quad \textit{and} \quad \lim_{(n,p)\to\infty} \mathbb{P}(\text{FDP}_{\text{dir}} \leq q + \epsilon) = 1. \tag{15}$$

The proof is given in Appendix B.2.

Comparing the sparsity condition of Theorem 2 with that of Jankova and van de Geer (2015) for the pointwise asymptotic normality of the test statistic, it is observed that the number of non-zero entries per row of the precision matrix $\boldsymbol{\Theta}$ should grow at a slower rate in this case, such that the matrix be sparser for the proposed procedure to provide control on the directional FDR. Moreover, the conditions on the cardinality of two sets $\Gamma(\gamma, c_0)$ and $\Psi(\rho)$ confirm that the test statistics should have mild correlations for the directional FDR to be under control. In the sequel, Theorem 3 provides guarantees for the statistical power of the proposed procedure.

Theorem 3 *Suppose that the conditions of Theorem 2 hold. If $|\boldsymbol{\Theta}_{ab}| > \sqrt{\frac{2\sigma_{ab}^2 \log(p_1/s)}{n}}$ for $(a,b) \in S$, where $p_1 = p(p-1)/2$, $s = \#S$ the cardinality of the support of $\boldsymbol{\Theta}$, and $\sigma_{ab}^2 = \boldsymbol{\Theta}_{aa}\boldsymbol{\Theta}_{bb} + \boldsymbol{\Theta}_{ab}^2$, then we have*

$$\liminf_{n\to\infty} \frac{\text{Power}_{\text{dir}}}{\frac{1}{s}\sum_{(a,b)\in S} \left(1 - \Phi\big(\Phi^{-1}(1 - \frac{qs}{2p_1}) - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\big)\right)} \geq 1, \tag{16}$$

where q is the pre-specified level in the hypotheses testing procedure. Moreover, if $\frac{\sqrt{n}\boldsymbol{\Theta}_{\min}}{\sigma_{\max}} - \Phi^{-1}(1 - \frac{qs}{2p_1}) \to \infty$, where $\boldsymbol{\Theta}_{\min} = \min_{(a,b)\in S} |\boldsymbol{\Theta}_{ab}|$ and $\sigma_{\max} = \max_{(a,b)\in S} \sigma_{ab}$, then $\text{Power}_{\text{dir}} \to 1$.

The proof is given in Appendix B.3.

In the next section, we show that the directional FDR can still be under control using the distributed estimator.

4.2 Directional FDR control via the distributed estimator

Similarly to (7), one can construct the test statistic

$$\hat{U}_{ab} = \sqrt{\sum_{k=1}^{K} \frac{n_k}{\hat{\sigma}^2_{ab,k}}}\tilde{\mathbf{\Theta}}_{ab}, \tag{17}$$

for every $1 \leq a < b \leq p$, based on the distributed estimator. This statistic will be used for the hypothesis test $H_{0,ab} : \mathbf{\Theta}_{ab} = 0$ vs $H_{1,ab} : \mathbf{\Theta}_{ab} \neq 0$. Using this test statistic, one can reject $H_{0,ab}$ in favor of $H_{1,ab}$ if $|\hat{U}_{ab}| \geq t$ for a given threshold t. As such, the directional FDP at a given threshold t based on the statistic defined in (17) will be similar to (8), by replacing $\hat{T}_{ab}$ with $\hat{U}_{ab}$, that is

$$\mathrm{FDP}_{\mathrm{dir}}(t) = \frac{\sum_{(a,b)\in S_{\geq 0}} \mathbb{I}(\hat{U}_{ab} \leq -t) + \sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(\hat{U}_{ab} \geq t)}{R'(t) \vee 1},$$

with $R'(t) := \sum_{1\leq a<b\leq p} \mathbb{I}(|\hat{U}_{ab}| \geq t)$. Similarly to (9), by defining the random variable

$$\tilde{U}_{ab} = \mathbf{W}_{ab,\dagger} + \mathbf{R}_{ab,\dagger}, \tag{18}$$

one can show that

$$\mathrm{FDP}_{\mathrm{dir}}(t) \leq \frac{\sum_{(a,b)\in S_{\geq 0}} \mathbb{I}(\tilde{U}_{ab} \leq -t) + \sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(\tilde{U}_{ab} \geq t)}{R'(t) \vee 1}.$$

Lemma 4 shows that the difference between $\tilde{U}_{ab}$ and T_{ab} is negligible, and as such, using Lemma 1, the tail probability $\mathbb{P}(|\tilde{U}_{ab}| > t)$ can be approximated by $G(t)$. As a result, for a pre-specified level $0 \leq q \leq 1$, the threshold t'_0 can be defined as

$$t'_0 = \inf\big\{0 \leq t \leq t_p : \quad \frac{p_1 G(t)}{R'(t) \vee 1} \leq q\big\}, \tag{19}$$

where relative to (11), we substitute the total number of rejections $R(t)$ by $R'(t)$ based on the distributed test statistic $\hat{U}_{ab}$ from (17).

Lemma 4 Let $p \leq n^r$ for some $r > 0$ and suppose that assumptions (A1)-(A2) in Appendix A hold. Consider the random variable $\tilde{U}_{ab}$ defined in (18) for every $1 \leq a < b \leq p$. Under the sparsity condition $d^{3/2} = o\big(\sqrt{n_\dagger}/(\log(p))^{3/2}\big)$, we have

$$\max_{a,b} \big|\tilde{U}_{ab} - T_{ab}\big| = o_p\big(\max\{\frac{n_\dagger^{1/6}}{n^{1/4}\sqrt{\log(p)}}, \frac{1}{n^{1/3}}, \frac{1}{\log(p)}\}\big),$$

as K grows at the rate $K = O\big(n^{1/4}/(d\log(p))\big)$.

The proof is given in Appendix B.4.

Theorem 5 provides guarantees for the control over the directional FDR using the threshold introduced in (19).

Theorem 5 *Let $p \leq n^r$, for some $r > 0$, and suppose that assumptions* (A1)-(A2) *in Appendix A hold. Consider the test statistic $\hat{U}_{ab}$ defined in* (17) *based on the distributed estimator. Suppose that the sparsity condition $d^{3/2} = o(\sqrt{n_\dagger}/(\log(p))^{3/2})$ holds, and consider the sets $\Gamma(\gamma, c_0)$ and $\Psi(\rho)$ with cardinalities $\Upsilon = o(p^{\delta\rho})$, $\rho \in (0,1)$, $\delta \in (0, 1/2]$ and $\psi = O(p^{\delta\rho}/\sqrt{\log(p)})$, respectively. As long as K grows at the rate $K = O\big(n^{1/4}/(d\log(p))\big)$, the threshold defined in* (19) *based on the test statistic $\hat{U}_{ab}$ controls the directional* FDR *at level q, i.e., for every $\epsilon > 0$,*

$$\limsup_{(n,p)\to\infty} \mathrm{FDR}_{\mathrm{dir}} \leq q, \quad \textit{and} \quad \lim_{(n,p)\to\infty} \mathbb{P}(\mathrm{FDP}_{\mathrm{dir}} \leq q + \epsilon) = 1.$$

The proof is given in Appendix B.5.

Comparing the upper bound provided for the growing number of machines K in Theorem 4 with that of Nezakati and Pircalabelu (2023b) for the pointwise asymptotic normality of the distributed test statistic, it is observed that K should grow at a slower rate here, in order for the procedure to provide a control on the directional FDR. Furthermore, the conditions on the cardinality of the two sets $\Gamma(\gamma, c_0)$ and $\Psi(\rho)$ are the same as in Theorem 2. These conditions are different, but similar, to those of Javanmard and Javadi (2019) for controlling the directional FDR for hypothesis testing in high-dimensional regression models.

Remark 6 By the same technique as in Theorem 3 and the provided assumptions on the entries of $\boldsymbol{\Theta}_{ab}$, under the conditions of Theorem 5, it can be shown that the directional power of the proposed distributed procedure tends to 1 with increasing n.

In the next section, we investigate the performance of the proposed procedure via a simulation study.

5. Simulation study

In this section, we analyze the performance of the proposed procedure in controlling the FDR by conducting a simulation study. Two simulation settings are considered for this purpose. In the first setting, we compare the performance of the statistic defined in (7), which is based on the debiased full estimator, with that of the other existing methods. In the second setting, we compare the performance of the distributed test statistic defined in (17) with that of the debiased full estimator.

To generate a dataset, we simulated n samples from a multivariate normal distribution $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, such that $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ is graph-structured and has the following sparse structures

(i) random with probability of connection 0.05;

(ii) block diagonal with 10 blocks and 0.2 as the probability of connection in each block;

(iii) chain or tridiagonal with 1 on the diagonal and 0.2 on the upper and lower diagonals.

All the simulation results are empirical averages over 100 replications from the specified DGPs. Consider $\hat{S}_r = \{(a,b) \in \mathcal{V} \times \mathcal{V}, a < b : |\hat{T}_{ab,r}| \geq t_{0,r}\}$ as the set of selected pairs (a,b) in the r-th replication, using the proposed procedure, where $\hat{T}_{ab,r}$ and $t_{0,r}$ are the

test statistic defined in (7) and the threshold defined in (11), respectively, for the r-th replication. The empirical FDP for the r-th replication, $r = 1, \ldots, 100$, is defined as

$$\text{FDP}_r = \frac{\sum_{(a,b)\in\hat{S}_r} \mathbb{I}(\boldsymbol{\Theta}_{ab} = 0)}{\hat{s}_r \vee 1},$$

where $\hat{s}_r$ is the cardinality of the selected set $\hat{S}_r$, i.e., $\hat{s}_r = \#\hat{S}_r$. Then, the empirical FDR over 100 replications, is of the form $\text{FDR} = (1/100)\sum_{r=1}^{100} \text{FDP}_r$. Moreover, the empirical power is calculated as

$$\text{Power} = \frac{1}{100}\sum_{r=1}^{100}\left\{\frac{\sum_{(a,b)\in\hat{S}_r} \mathbb{I}(\boldsymbol{\Theta}_{ab} \neq 0)}{s \vee 1}\right\},$$

where s is the cardinality of the set S, which is the set of active components of $\boldsymbol{\Theta}$. Similar quantities can be defined based on different test statistics such as the distributed one.

5.1 Performance of the debiased full test statistic

To compare the performance of the threshold introduced in (11) based on the debiased full estimator defined in (4) with other competitors, we considered the procedures of Li et al. (2022) and Liu (2013) based on the debiased Clime estimator and the nodewise Lasso procedure, respectively. We set the number of variables (nodes) to $p = 200$ and the sample size to $n = 150, 200, 250$ and 2000. The regularization parameter is considered equal to $\lambda = \sqrt{\log(p)/n}$ for all procedures and the results are presented in Table 1. The proposed statistic and the ones from Li et al. (2022) and Liu (2013) have been denoted by "dGL", "dCl" and "GFC", respectively in this table.

It is observed that dGL and dCl can control FDR at the pre-specified levels $q = 0.1$ and 0.2 for all three scenarios. However, the results for dGL are slightly better than those based on the dCl test statistic. The FDR results for the GFC procedure are high, especially in the random graph structure. In this structure, when n is small, GFC is not able to control FDR at the pre-specified levels, but as n increases, it reaches eventually the pre-specified levels 0.1 and 0.2. However, its power is the highest compared to the two other methods. Moreover, it is observed that the powers in model (ii) are the highest relative to those in models (i) and (iii). The reason is that the non-zero entries in the true precision matrix $\boldsymbol{\Theta}$ have the largest magnitude in model (ii). The results for the directional FDR corresponding to the dGL test statistic were also similar to the ones presented in Table 1, just slightly higher than the reported ones. However, since dCl and GFC do not provide any procedure for controlling the directional FDR, we did not report these results here. The FDPs over 100 replications are also plotted in Figure 1 for the random graph structure, with different values of n with $p = 200$ at the pre-specified level $q = 0.1$. It is observed that most FDPs corresponding to the dGL and dCl procedures are concentrated around the FDRs and under the pre-specified level. Moreover, the median of the dGL is smaller than that of the dCl. The FDP values corresponding to the GFC are larger and are not under the pre-specified level 0.1. However, with increasing n, they are getting closer to 0.1.

Table 1: Empirical false discovery rate and power for the debiased full test statistic and two competitors, when $p = 200$.

			$n = 150$			$n = 200$			$n = 250$			$n = 2000$		
			dGL	dCl	GFC	dGL	dCl	GFC	dGL	dCl	GFC	dGL	dCl	GFC
$q = 0.1$	Random	FDR	.00	.05	.16	.01	.04	.14	.01	.03	.14	.02	.05	.10
		Power	.03	.11	.37	.13	.25	.54	.29	.40	.67	1	1	1
	Bdiag.	FDR	.01	.04	.10	.01	.04	.09	.01	.04	.09	.02	.06	.09
		Power	.38	.49	.65	.62	.71	.82	.78	.85	.91	1	1	1
	Tridiag.	FDR	.03	.05	.11	.02	.08	.10	.03	.08	.10	.06	.08	.09
		Power	.01	.05	.10	.12	.23	.25	.28	.40	.42	1	1	1
$q = 0.2$	Random	FDR	.01	.09	.28	.02	.07	.26	.02	.07	.26	.04	.10	.20
		Power	.06	.19	.51	.22	.35	.66	.39	.51	.78	1	1	1
	Bdiag.	FDR	.02	.08	.20	.02	.08	.19	.02	.08	.19	.05	.13	.19
		Power	.46	.58	.74	.69	.78	.87	.83	.89	.94	1	1	1
	Tridiag.	FDR	.04	.16	.23	.06	.16	.20	.06	.16	.19	.12	.17	.19
		Power	.05	.15	.18	.19	.33	.36	.37	.51	.53	1	1	1

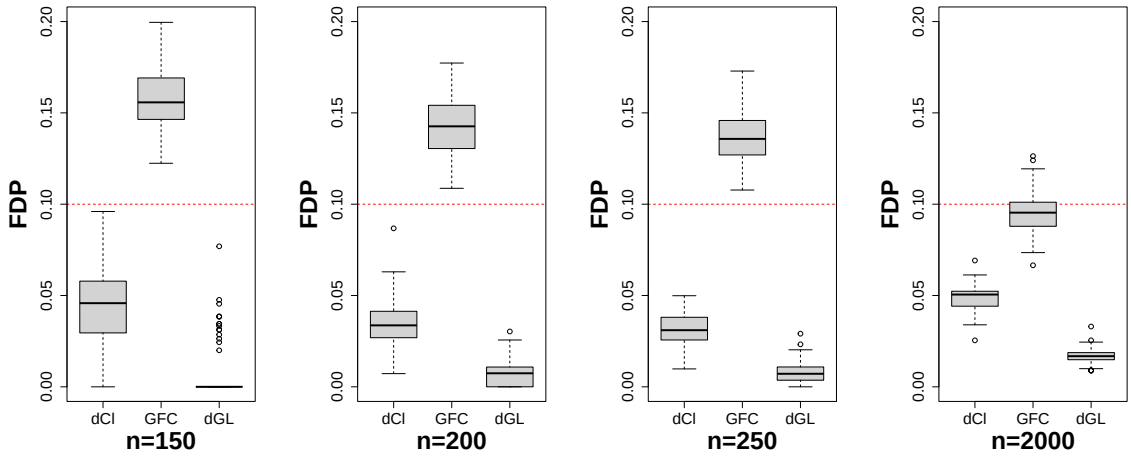


Figure 1: Boxplots of FDPs over 100 replications for the random graph structure with $p = 200$, at the pre-specified level $q = 0.1$.

5.2 Performance of the distributed test statistic

In this section, we investigate the performance of the distributed test statistic and compare it with that of the full ones in controlling the FDR. To this end, we set the total sample size to $n = 50000$, the number of nodes to $p = 1000$ and changed the number of machines from $K = 5, 10$ to 20. To show the performance of the distributed estimator in the unbalanced setting, we considered the following unbalanced data splitting procedure similar to Nezakati and Pircalabelu (2023b).

Suppose that among all available machines, two of them are powerful. The first one is the most powerful one and $(55 - K)\%$ of the dataset is distributed on this machine.

The second one is less powerful than the first one and $(60 - K)\%$ of the remaining dataset is distributed on this one. The remaining dataset is distributed roughly equally on the remaining machines.

In this study, the regularization parameter λ_k in the k-th graphical Lasso problem has been set to $\lambda_k = \lambda = \sqrt{\log(p)/n}$ for all simulation settings except for the tridiagonal case. Due to the high sparsity of the precision matrix in the tridiagonal structure, we increased slightly the regularization parameter to $\lambda_k = \sqrt{\log(p)/n_k}$. The empirical false discovery rate results are shown in Table 2. All the powers were equal to 1 and they have not been presented in the table. Similarly to the previous simulation study, the results for the directional FDR corresponding to the distributed test statistic were slightly higher than the false discovery rates presented in Table 2 and have not been shown in this table.

From Table 2, it is observed that all the false discovery rates in the distributed case are under control at the pre-specified levels $q = 0.1$ and 0.2. Moreover, the results are close to those of the debiased full one and they confirm that one does not lose much information when the dataset is distributed on different locations and estimators are aggregated. The results concerning the FDPs for the three graph structures with $q = 0.1$ over 100 replications are shown in Figure 2 for dCl, GFC, dGL and the distributed procedure when $K = 20$. It is observed that all empirical FDPs for dGL and $K = 20$ are under the pre-specified level $q = 0.1$. However, in the tridiagonal structure, GFC and dCl have values larger than 0.1 for some replications.

Table 2: Empirical false discovery rate for the distributed and debiased full test statistics and two competitors, when $n = 50000$ and $p = 1000$.

		$q = 0.1$				$q = 0.2$			
		K				K			
		1	5	10	20	1	5	10	20
Random	dGL	.03				.08			
	dCl	.07				.14			
	GFC	.09				.19			
	distributed		.04	.04	.05		.08	.09	.11
Bdiag.	dGL	.04				.09			
	dCl	.07				.15			
	GFC	.10				.19			
	distributed		.04	.05	.07		.09	.10	.14
Tridiag.	dGL	.08				.16			
	dCl	.09				.18			
	GFC	.10				.19			
	distributed		.05	.04	.03		.11	.09	.07

As it is mentioned in Theorems 2 and 4, the theoretical guarantee on controlling the directional FDR depends on several assumptions e.g. the sparsity and the magnitude of the entries of $\mathbf{\Theta}$. To investigate the effect of the sparsity on the directional FDR, we considered the random graph structure and changed the probability of connection between every pair nodes from 0.01 to 0.1. The results are presented in the left panel of Figure 3 for the dGL

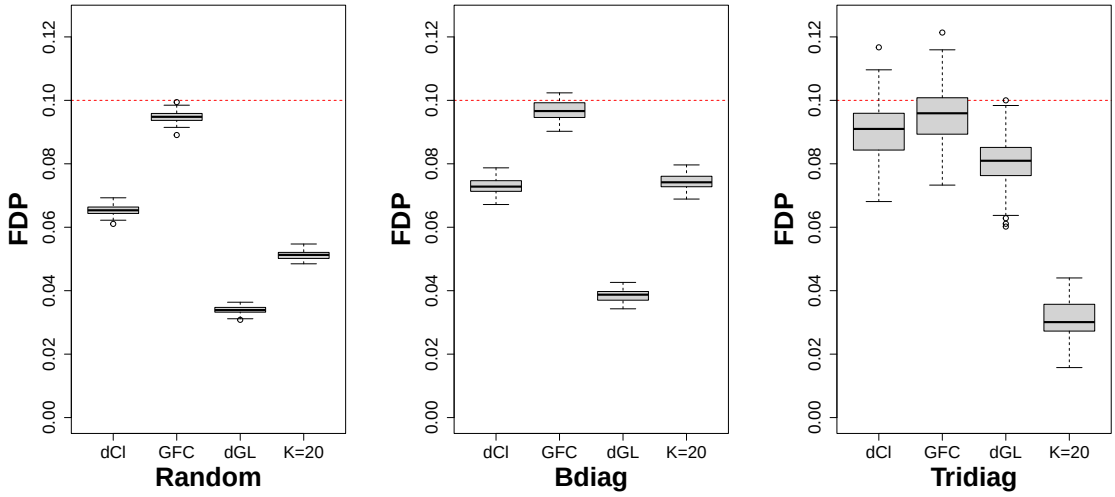


Figure 2: Boxplots of FDPs over 100 replications for three graph structures with $p = 1000$ and $n = 50000$, at the pre-specified level $q = 0.1$.

and the distributed procedure with $K = 5, 10$ and 20. It is observed that the results are not sensitive to the probability of connection, as they are stable in the interval $(0.02, 0.06)$ and under the pre-specified level $q = 0.1$.

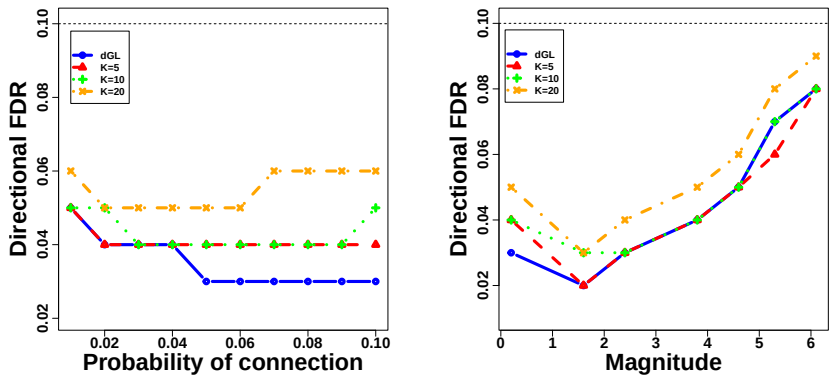


Figure 3: The left panel presents FDR_{dir} for the debiased full procedure labeled dGL and the distributed procedure with $K = 5, 10, 20$ as a function of the probability of connection in the random graph structure. The right panel presents FDR_{dir} for the mentioned procedures as a function of the magnitude of the off-diagonal entries of $\boldsymbol{\Theta}$. For both panels, $n = 50000$, $p = 1000$ and the pre-specified level $q = 0.1$ are considered.

Moreover, to investigate the effect of the magnitude of the off-diagonal entries on the directional FDR, we considered different random graph structures, such that the magnitude of the largest off-diagonal entry changes from 0.2 to 6.5. The results are presented in the right panel of Figure 3. It is observed that as the magnitude of the off-diagonal entries of $\mathbf{\Theta}$ increases, the directional FDR increases for both debiased full and distributed procedures. This confirms that the cardinality of the highly correlated pair nodes should be under control, such that most of the test statistics have mild correlation as only in this case it provides a guarantee on controlling the directional FDR at the pre-specified level q.

6. Real data

In this section, we present two real data examples that illustrate the performance of our methodology.

6.1 Text dataset

The publicly available "4 university web-pages" dataset was collected in January 1997 by the World Wide Knowledge Base project of the CMU text learning group. The original dataset has been pre-processed by standard text mining methods and it is available in Cardoso-Cachopo (2007). This dataset contains web-pages collected from computer science departments of four US universities Cornell, Texas, Washington and Wisconsin for seven categories: student, faculty, staff, department, course, project and other. The "other" category is a collection of web-pages that were not considered in the six main categories. In this study, we considered the four largest categories which are 'student', 'faculty', 'course' and 'project' containing 1641, 1124, 930 and 504 web-pages, respectively. To calculate the term-document matrix $\mathbf{X}$, we used the log-entropy weighting method of Dumais (1991) which was also implemented in Guo et al. (2011). The final dataset contains $n = 4199$ web-pages as the observations and 7686 distinct terms as variables. We considered $p = 500$ terms with the highest log-entropy weights for the analysis.

We applied dGL, dCl, GFC, and the distributed procedure with $K = 3, 4$ and 5 to identify the connections between 500 distinct terms. The splitting setting is considered the same as in the simulation study. The regularization parameter and the pre-specified level are set to $\lambda = \sqrt{\log(p)/n}$ and $q = 0.1$, respectively, for all procedures. The proportion of detected edges, which is obtained as the percentage of the number of detected edges divided by $p(p-1)/2$, is presented in the left panel of Figure 4. It is observed that the proportion of detected edges is low, meaning that the estimated graph is sparse. The distributed procedure identified the sparsest graph, while the graph identified by GFC is quite dense. The proportion of detected edges using dGL and the distributed procedure are close to each other, which confirms the similarity of these two procedures.

To evaluate the performance of the estimated graphs, we computed the out-of-sample prediction error using the negative log-likelihood. To this end, we sampled 10 percent of the dataset as the test set and the remaining data were used as the training set. After estimating the precision matrix using the training dataset by the proposed hypothesis testing procedures, denoted as $\check{\mathbf{\Theta}}$, we symmetrized it and computed the positive definite projection of $\check{\mathbf{\Theta}}$, denoted as $\check{\mathbf{\Theta}}_+$. Similarly to Li et al. (2022), the out-of-sample prediction error of

$\check{\boldsymbol{\Theta}}_+$ is evaluated via

$$\text{PredErr}(\check{\boldsymbol{\Theta}}_+) = \frac{1}{2p}\Big\{\text{trace}(\hat{\boldsymbol{\Sigma}}_{\text{test}}\check{\boldsymbol{\Theta}}_+) - \log\det(\check{\boldsymbol{\Theta}}_+)\Big\},$$

where $\hat{\boldsymbol{\Sigma}}_{\text{test}}$ is the sample covariance using the test dataset. The average over 10 replications of the out-of-sample prediction errors is presented in the upper panel in Table 3. It is observed that all prediction errors are close to each other, indicating that the estimated graphs are relatively similar. As the GFC procedure does not provide any estimates for the diagonal entries of the precision matrix, it is not possible to compute this prediction error for it.

Table 3: Out-of-sample prediction error of the proposed procedure

	Text dataset					
	dCl	GFC	dGL	$K=3$	$K=4$	$K=5$
PredErr	.43	/	.42	.44	.45	.46
	Pan-cancer dataset					
	dCl	GFC	dGL	$K=3$	$K=5$	$K=10$
PredErr	.27	/	-.16	-.15	-.14	-.11

Afterwards, to investigate the performance of the estimators for completely independent variables, we permuted randomly sample data for each variable, thus breaking up the correlation structure in order to construct a dataset with mutually independent variables. As such, we expect that there are no edges in the estimated graph. By applying the mentioned procedures on these dataset, we identified no edges in the estimated graph which confirms this assertion.

The first ten terms with the highest node degrees, identified by the proposed procedures, are presented in Figure 5. Some terms such as "project" and "memori" are identified by almost all the procedures. Based on the node degrees provided in this figure, it is confirmed that the estimated graph by GFC is quite dense compared to the estimated graph by the other procedures.

6.2 Pan-Cancer dataset

The process of collecting the Pan-Cancer dataset from The Cancer Genome Atlas (TCGA) project (available at https://xenabrowser.net/datapages/) was started in 2006 and in 10 years time, TCGA network investigators had characterized the molecular landscape of tumors from more than 11000 patients across 33 cancer types. This particular TCGA molecular dataset has been studied in multiple works to understand the cancer biology, including those of glioblastoma multiforme (GBM), ovarian, breast, lung, prostate, bladder and others (see for instance, Akbani et al., 2015; Cancer Genome Atlas Research Network, 2017; Liu et al., 2018).

We applied the proposed methodology on 743 microRNA mature strand expressions to estimate the underlying graph among the microRNAs. The final dataset consists of $n = 9986$ subjects having $p = 743$ nodes. The splitting setting is considered the same as in the

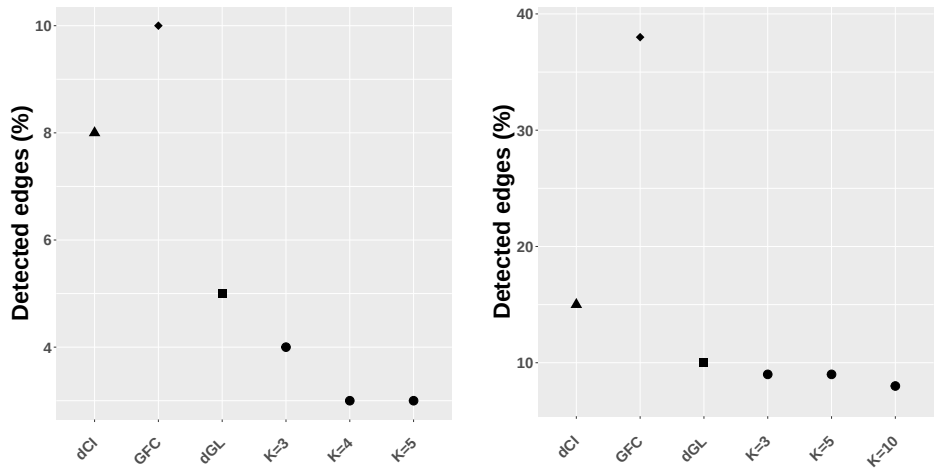


Figure 4: The left panel presents the proportion of detected edges for the text dataset using dCl, GFC, dGL and the distributed procedure with $K = 3, 4$ and 5, at the pre-specified level $q = 0.1$. The right panel presents this proportion for the Pan-Cancer dataset using the mentioned procedures and the distributed one with $K = 3, 5$ and 10, at the pre-specified level $q = 0.05$.

simulation study and the number of machines is set to $K = 3, 5$ and 10. The regularization parameter and the pre-specified level are set to $\lambda = \sqrt{\log(p)/n}$ and $q = 0.05$, respectively, for all procedures. The proportion of detected edges is presented in the left panel of Figure 4. It is observed that the estimated graph by the distributed, dGL and dCl methods are sparse, however the distributed procedure identified the sparsest graph and the estimated graph by the GFC method is much denser than the other presented methods. Moreover, it is observed in the lower panel in Table 3 that the dGL has the smallest prediction error, while dCl has the largest one. The prediction errors of the distributed method are also close to that of the debiased full one which confirms the similarity of these two procedures. Note that, after applying the mentioned procedures on the randomly permuted sample data, we identified no edges between the genes.

As it is mentioned in Wang (2015), the genes hsa-miR-136, hsa-miR-376a and hsa-miR-377 are the most important genes in identifying GBM cancer. Figure 6 presents the estimated sub-graph between these genes using the proposed procedures at the pre-specified level $q = 0.1$. As it is observed, dGL and the distributed procedure with $K = 3, 5$ and 10 identified the same edges between these genes, while dCl and GFC identified a few more edges.

7. Conclusion

In this paper, we studied the directional false discovery rate in a multiple testing framework for GGMs. First, a test statistic has been built using the debiased graphical Lasso estimator

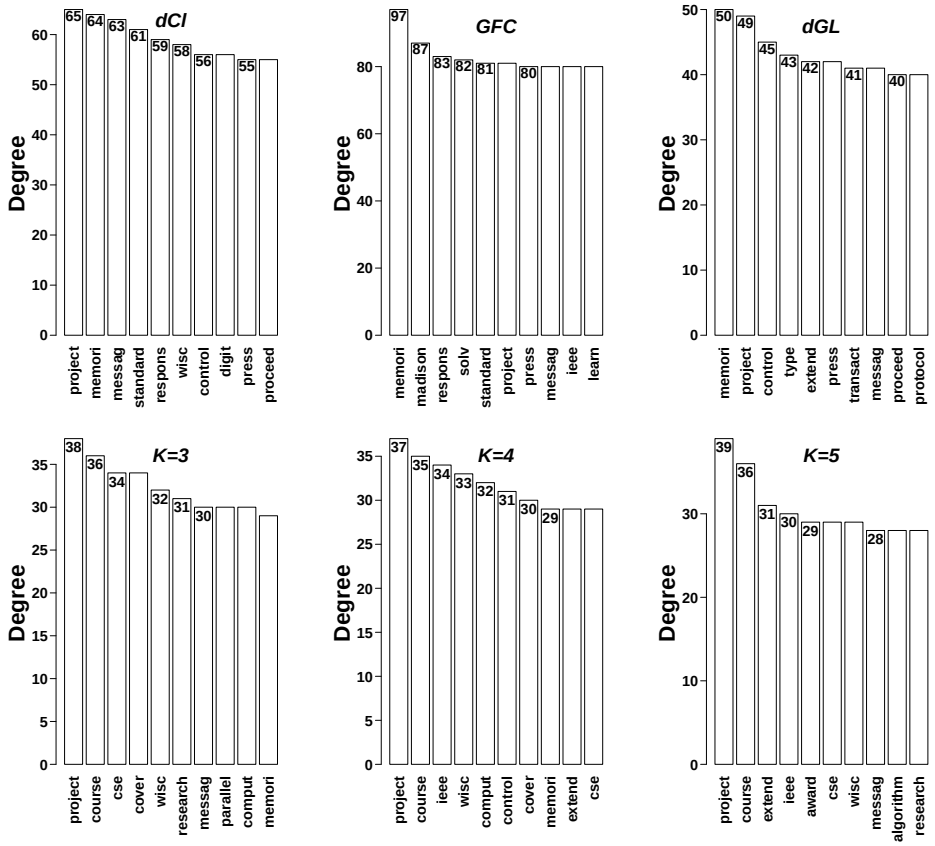


Figure 5: The first 10 terms with the highest node degrees, identified by different procedures, at the pre-specified level $q = 0.1$.

and it has been shown that as long as the maximum node degree of the precision matrix grows at the rate $d^{3/2} = o\big(\sqrt{n}/(\log(p))^{3/2}\big)$, and the cardinality of the highly correlated pair nodes is under control, the directional FDR will be asymptotically below a pre-specified level.

Furthermore, related to modern problems such as data privacy, heterogeneous different datasets and parallelization, a distributed test statistic has been proposed to control the directional FDR in multiple testing. It has been shown that as long as the number of multiple datasets grows at the rate $K = O\big(n^{1/4}/(d\log(p))\big)$ with the sparsity condition $d^{3/2} = o\big(\sqrt{n_\dagger}/(\log(p))^{3/2}\big)$ and the cardinality of highly correlated pair nodes is under control, the directional FDR in the multiple testing of GGMs is asymptotically under a pre-specified level.

The finite sample performance of the proposed procedures has also been evaluated with a simulation study and real data examples. It is observed that both debiased and distributed procedures have competitive results relative to the state-of-the-art methods. The FDR, as well as the directional FDR for both procedures are under the pre-specified level. This con-

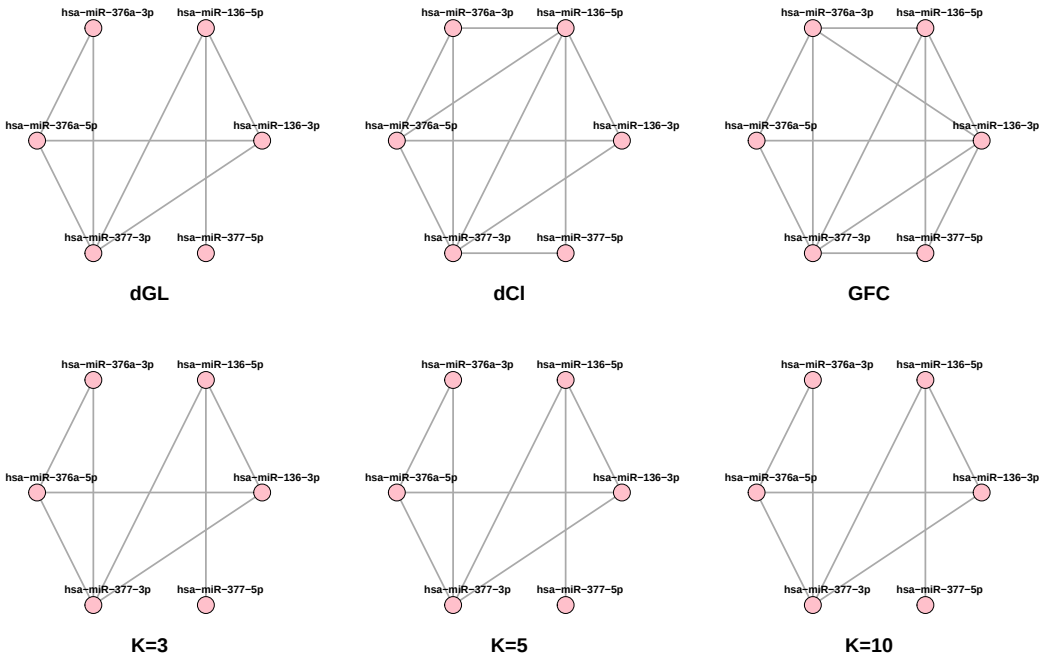


Figure 6: The estimated sub-graph between the genes for the GBM cancer, at the pre-specified level $q = 0.05$.

firms the assertation that in practice, having datasets distributed across multiple machines, in our aggregation framework one does not lose much information and the performance is close to that of the debiased full procedure which would use the entire dataset, but which is unattainable in practice.

Appendix A. Preliminary assumptions

Two regularity assumptions are provided here for the reader's convenience.

Denote by $\mathbf{H}$ the Hessian of the negative log-likelihood function $\ell(\mathbf{\Theta}) \propto \text{trace}(\hat{\mathbf{\Sigma}}\mathbf{\Theta}) - \log\det(\mathbf{\Theta})$, where $\hat{\mathbf{\Sigma}} = \mathbf{X}^\top\mathbf{X}/n$ is the sample covariance matrix. It has been shown that $\mathbf{H} = \mathbf{\Sigma} \otimes \mathbf{\Sigma}$ (see Ravikumar et al., 2011). By definition, $\mathbf{H}$ is a $p^2 \times p^2$ matrix indexed by the pair of elements from the node set, such that $\mathbf{H} = [\mathbf{H}_{(a,b),(c,d)}]$, where $(a,b),(c,d) \in S$. Let $\mathbb{S} = \big\{S \cup \{(1,1),(2,2),\ldots,(p,p)\}\big\}$ with cardinality $\int$, which is equal to $\int = s + p$, and denote its complement set with $\mathbb{S}^c$. The following regularity assumptions (see for example, Ravikumar et al., 2011; Jankova and van de Geer, 2015) are considered for the theoretical guarantees in the estimation of $\mathbf{\Theta}$:

(A1) The irrepresentability condition holds for the true precision matrix $\mathbf{\Theta}$, i.e., there exists $\alpha \in (0,1]$ such that $\max_{e\in\mathbb{S}^c} ||\mathbf{H}_{e\mathbb{S}}\big(\mathbf{H}_{\mathbb{S}\mathbb{S}}\big)^{-1}||_1 \leq 1-\alpha$, where $\mathbf{H}_{\mathbb{S}\mathbb{S}} \in \mathbb{R}^{\int\times\int}$ is a sub-matrix of $\mathbf{H}$ whose rows and columns are indexed by the elements of $\mathbb{S}$. Moreover,

e is a pair $(a,b) \in \mathbb{S}^c$ such that $\mathbf{H}_{e\mathbb{S}}$ is an $\int$-dimensional column vector with elements $\mathbf{H}_{e,(c,d)}$, where $(c,d) \in \mathbb{S}$.

(A2) The eigenvalues of the precision matrix $\mathbf{\Theta}$ are bounded, i.e., there exists a constant $\Lambda \geq 1$ such that $\frac{1}{\Lambda} \leq \Lambda_{\min}(\mathbf{\Theta}) \leq \Lambda_{\max}(\mathbf{\Theta}) \leq \Lambda$, where $\Lambda_{\min}(\mathbf{\Theta})$ and $\Lambda_{\max}(\mathbf{\Theta})$ are the minimum and maximum eigenvalues of $\mathbf{\Theta}$, respectively.

Appendix B. Proofs of the main theorems and lemmas

B.1 Proof of Lemma 1

Consider $T_{ab} = \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \frac{\mathbf{Z}_{ab,l}}{\sigma_{ab}}$, such that $\mathbf{Z}_{ab,l} = \mathbf{\Theta}_a^\top \dot{\mathbf{X}}_l \dot{\mathbf{X}}_l^\top \mathbf{\Theta}_b - \mathbf{\Theta}_{ab}$. It is shown in Jankova and van de Geer (2015) that under the multivariate Gaussian distribution, $\mathbb{V}\mathrm{ar}(\mathbf{Z}_{ab,l}) = \sigma_{ab}^2$, where $\sigma_{ab}^2 = \mathbf{\Theta}_{aa}\mathbf{\Theta}_{bb} + \mathbf{\Theta}_{ab}^2$. As such, on the weakly correlated pair nodes, i.e., on the set $\Gamma^c(\gamma, c_0)$, the conditions of Lemma 6.1 of Liu (2013) are fulfilled, and since $t_p < 2\sqrt{\log(p)}$, for some constant $C > 0$ we have

$$\begin{aligned}\max_{a,b} \sup_{0\leq t\leq t_p} \left| \frac{\mathbb{P}\big(|T_{ab}| > t\big)}{G(t)} - 1 \right| &\leq \max_{a,b} \sup_{0\leq t\leq 2\sqrt{\log(p)}} \left| \frac{\mathbb{P}\big(|\sum_{l=1}^{n} \mathbf{Z}_{ab,l}/\sigma_{ab}| > t\sqrt{n}\big)}{G(t)} - 1 \right| \\ &\leq C\big(\log(p)\big)^{-1-\gamma_1},\end{aligned}$$

where $\gamma_1 = \min\{1, \gamma\}$ and $\gamma > 0$ is defined in (12). To show (14), we have

$$\begin{aligned}\max_{a,b} \big|\tilde{T}_{ab} - T_{ab}\big| &= \max_{a,b} \left| \frac{\sigma_{ab}}{\hat{\sigma}_{ab}} T_{ab} + \frac{\sqrt{n}\mathbf{R}_{ab}}{\hat{\sigma}_{ab}} - T_{ab} \right| \\ &\leq \max_{a,b} \left\{ \big|T_{ab}\big| \times \sigma_{ab} \big| \frac{1}{\hat{\sigma}_{ab}} - \frac{1}{\sigma_{ab}} \big| \right\} + \sqrt{n} \max_{a,b} \frac{\big|\mathbf{R}_{ab}\big|}{\hat{\sigma}_{ab}}. \end{aligned} \tag{20}$$

Using Lemma 2 of Jankova and van de Geer (2015), we have

$$\big|\hat{\sigma}_{ab}^2 - \sigma_{ab}^2\big| = O_p\big(\frac{1}{\alpha}\kappa_{\mathbf{H}}\sqrt{\log(p)/n}\big), \tag{21}$$

where $\kappa_{\mathbf{H}} = |||(\mathbf{H}_{\mathbb{SS}})^{-1}|||_\infty$ and α are defined in Appendix A. As such, it can be shown that the inverse ratio also satisfies

$$\big|(\sigma_{ab}/\hat{\sigma}_{ab}) - 1\big| = O_p\big(\sqrt{\log(p)/n}\big), \tag{22}$$

by letting $1/\alpha = O(1)$ and $\kappa_{\mathbf{H}} = O(1)$. Under assumption (A2), using Lemma 1 of Jankova and van de Geer (2015), it holds that

$$|\mathbf{R}_{ab}| = O_p\bigg(\frac{1}{\alpha^2}\kappa_{\mathbf{H}} \max\big\{d^{3/2}\log(p)/n, \frac{1}{\alpha}\kappa_{\mathbf{H}} d^2 (\log(p)/n)^{3/2}\big\}\bigg), \tag{23}$$

which is $o_p(1/\sqrt{n})$, under the sparsity condition $d^{3/2} = o\big(\sqrt{n}/\log(p)\big)$, and by letting $1/\alpha = O(1)$ and $\kappa_{\mathbf{H}} = O(1)$.

Moreover, under assumption (A2), by Lemma 8 of Jankova and van de Geer (2015), it also holds that

$$|T_{ab}| = O_p\big(d\sqrt{\log(p)/n}\big). \tag{24}$$

Combining (22), (23) with (24), and then substituting in (20), under the sparsity condition $d^{3/2} = o\big(\sqrt{n}/(\log(p))^{3/2}\big)$, the $o_p\big(1/\sqrt{\log(p)}\big)$ result follows directly. Note that due to the small cardinality of the highly correlated set, i.e., $\Upsilon = o(p^{\delta\rho})$, $\delta \in (0, 1/2]$, $\rho \in (0, 1)$, relative to the cardinality of the set $\Gamma^c(\gamma, c_0)$, the impact of the highly correlated statistics T_{ab} on the final rates can be neglected following the arguments of Liu (2013) and Li et al. (2022).

B.2 Proof of Theorem 2

Recall that $p_1 = p(p-1)/2$, $G(t) = 2(1 - \Phi(t))$ and let $Q(t) = G(t)/2$. First suppose that t_0 in (11) exists. By (10), we have

$$\begin{aligned}
\mathrm{FDP}_{\mathrm{dir}}(t_0) &\leq \frac{\sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(\tilde{T}_{ab} \geq t_0) + \sum_{(a,b)\in S_{\geq 0}} \mathbb{I}(\tilde{T}_{ab} \leq -t_0)}{R(t_0) \vee 1} \\
&= \frac{\sum_{(a,b)\in S_{\leq 0}} \{\mathbb{I}(\tilde{T}_{ab} \geq t_0) - Q(t_0)\}}{R(t_0) \vee 1} \\
&\quad + \frac{\sum_{(a,b)\in S_{\geq 0}} \{\mathbb{I}(\tilde{T}_{ab} \leq -t_0) - Q(t_0)\} + G(t_0)(p_1 - s/2)}{R(t_0) \vee 1} \\
&\leq \frac{p_1 G(t_0) A_p + G(t_0)(p_1 - s/2)}{R(t_0) \vee 1} \leq q(A_p + 1),
\end{aligned}$$

where using (11) and with $q = \frac{p_1 G(t_0)}{R(t_0)\vee 1}$, the term A_p is defined as

$$\begin{aligned}
A_p &:= \sup_{0\leq t\leq t_p} \left| \frac{\sum_{(a,b)\in S_{\leq 0}} \{\mathbb{I}(\tilde{T}_{ab} \geq t) - Q(t)\} + \sum_{(a,b)\in S_{\geq 0}} \{\mathbb{I}(\tilde{T}_{ab} \leq -t) - Q(t)\}}{p_1 G(t)} \right| \\
&\leq \sup_{0\leq t\leq t_p} \left| \frac{\sum_{(a,b)\in S_{\leq 0}} \{\mathbb{I}(\tilde{T}_{ab} \geq t) - Q(t)\}}{p_1 G(t)} \right| + \sup_{0\leq t\leq t_p} \left| \frac{\sum_{(a,b)\in S_{\geq 0}} \{\mathbb{I}(\tilde{T}_{ab} \leq -t) - Q(t)\}}{p_1 G(t)} \right|.
\end{aligned}$$

As such, to show (15), it is enough to show that

$$\sup_{0\leq t\leq t_p} \left| \frac{\sum_{(a,b)\in S_{\leq 0}} \{\mathbb{I}(\tilde{T}_{ab} \geq t) - Q(t)\}}{p_1 G(t)} \right| \xrightarrow{p} 0, \tag{25}$$

and by similar techniques, one can show that the second term in the upper bound for A_p also tends to zero in probability. Using Lemma 1, the maximal difference between $\tilde{T}_{ab}$ and T_{ab} is $o_p\big(1/\sqrt{\log(p)}\big)$. As such, to show (25), we only need to prove that

$$\sup_{0\leq t\leq t_p} \left| \frac{\sum_{(a,b)\in S_{\leq 0}} \{\mathbb{I}(T_{ab} \geq t) - Q(t)\}}{p_1 G(t)} \right| \xrightarrow{p} 0,$$

and to this end, we need to show that for every $\upsilon > 0$,

$$\sup_{0\leq t\leq t_p} \mathbb{P}\left\{ \left| \frac{\sum_{(a,b)\in S_{\leq 0}} \{\mathbb{I}(T_{ab} \geq t) - Q(t)\}}{p_1 Q(t)} \right| \geq \upsilon \right\} = o(1) \tag{26}$$

and

$$\int_0^{t_p} \mathbb{P}\left\{ \left| \frac{\sum_{(a,b)\in S_{\leq 0}} \{\mathbb{I}(T_{ab} \geq t) - Q(t)\}}{p_1 Q(t)} \right| \geq \upsilon \right\} dt = o(\nu_p), \tag{27}$$

where $\nu_p = 1/\sqrt{\log(p)}$. The reader can refer to Lemma 7.3 of Javanmard and Javadi (2019) and Lemma 6.3 of Liu (2013) for more details on the motivation behind (26) and (27). We show that these two conditions hold in this situation and for the reader's convenience we first briefly explain the intuition behind.

By discretization of the interval $[0, t_p]$ to $0 \leq \tau_1 < \tau_2 < \ldots < \tau_b = t_p$, such that $\tau_j - \tau_{j-1} = \nu_p$ for $1 \leq j \leq b-1$ and $\tau_b - \tau_{b-1} \leq \nu_p$, where $\nu_p = 1/\sqrt{\log(p)}$, we have $b \approx t_p/\nu_p$. For any $t \in [\tau_{j-1}, \tau_j]$, it holds that

$$\begin{aligned} \frac{\sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(T_{ab} \geq \tau_j)}{p_1 Q(\tau_j)} \times \frac{Q(\tau_j)}{Q(\tau_{j-1})} &\leq \sum_{(a,b)\in S_{\leq 0}} \frac{\mathbb{I}(T_{ab} \geq t)}{p_1 Q(t)} \\ &\leq \frac{\sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(T_{ab} \geq \tau_{j-1})}{p_1 Q(\tau_{j-1})} \times \frac{Q(\tau_{j-1})}{Q(\tau_j)}. \end{aligned}$$

As such, it suffices to show that $\max_{0\leq j\leq b} \left| \frac{\sum_{(a,b)\in S_{\leq 0}} \{\mathbb{I}(T_{ab}\geq \tau_j) - Q(\tau_j)\}}{p_1 Q(\tau_j)} \right| \xrightarrow{p} 0$. We have that

$$\begin{aligned} &\mathbb{P}\left(\max_{0\leq j\leq b} \left| \frac{\sum_{(a,b)\in S_{\leq 0}} \{\mathbb{I}(T_{ab} \geq \tau_j) - Q(\tau_j)\}}{p_1 Q(\tau_j)} \right| \geq v \right) \\ &\leq \sum_{j=1}^{b} \mathbb{P}\left(\left| \frac{\sum_{(a,b)\in S_{\leq 0}} \{\mathbb{I}(T_{ab} \geq \tau_j) - Q(\tau_j)\}}{p_1 Q(\tau_j)} \right| \geq v \right) \\ &\leq \frac{1}{\nu_p} \int_0^{t_p} \mathbb{P}\left(\left| \frac{\sum_{(a,b)\in S_{\leq 0}} \left\{\mathbb{I}(T_{ab} \geq t) - Q(t)\right\}}{p_1 Q(t)} \right| \geq v \right) dt \\ &\quad + \sum_{j=b-1}^{b} \mathbb{P}\left(\left| \frac{\sum_{(a,b)\in S_{\leq 0}} \left\{\mathbb{I}(T_{ab} \geq \tau_j) - Q(\tau_j)\right\}}{p_1 Q(\tau_j)} \right| \geq v \right). \end{aligned}$$

As such, it is enough to show (26) and (27).

Before proving (26), to simplify notation, define $\Psi_{\leq 0}(\rho) = \Psi(\rho) \bigcap S_{\leq 0}$. By dividing the set $S_{\leq 0}$ into two disjoint sets $\Psi_{\leq 0}(\rho)$ and $S_{\leq 0} \setminus \Psi_{\leq 0}(\rho)$, in (26) we have

$$\begin{aligned} \mathbb{P}\left\{ \left| \frac{\sum_{(a,b)\in S_{\leq 0}} \{\mathbb{I}(T_{ab} \geq t) - Q(t)\}}{p_1 Q(t)} \right| \geq \upsilon \right\} &\leq \mathbb{P}\left\{ \left| \sum_{(a,b)\in \Psi_{\leq 0}(\rho)} \{ \frac{\mathbb{I}(T_{ab} \geq t) - Q(t)}{p_1 Q(t)} \} \right| \geq \upsilon/2 \right\} \\ &\quad + \mathbb{P}\left\{ \left| \sum_{(a,b)\in \{S_{\leq 0} \setminus \Psi_{\leq 0}(\rho)\}} \{ \frac{\mathbb{I}(T_{ab} \geq t) - Q(t)}{p_1 Q(t)} \} \right| \geq \upsilon/2 \right\}. \end{aligned} \tag{28}$$

Using Markov's inequality for the first probability on the right hand side of (28), we get

$$\begin{aligned}
&\mathbb{P}\Big\{\Big|\sum_{(a,b)\in\Psi_{\leq 0}(\rho)}\{\frac{\mathbb{I}(T_{ab}\geq t)-Q(t)}{p_1Q(t)}\}\Big|\geq \upsilon/2\Big\}\\
&\leq \frac{2}{\upsilon p_1Q(t)}\mathbb{E}\Big\{\Big|\sum_{(a,b)\in\Psi_{\leq 0}(\rho)}\big\{\mathbb{I}(T_{ab}\geq t)-Q(t)\big\}\Big|\Big\}\\
&\leq \frac{2}{\upsilon p_1Q(t)}\sum_{(a,b)\in\Psi_{\leq 0}(\rho)}\big\{\mathbb{E}(\mathbb{I}(T_{ab}\geq t))+Q(t)\big\}=O(\frac{p^{\delta\rho}}{p_1\sqrt{\log(p)}}),
\end{aligned}\tag{29}$$

where the last equality holds using the uniform convergence of the tail probability of T_{ab} to the tail probability of the normal distribution provided in Lemma 1, and the fact that $\Psi_{\leq 0}(\rho)\subseteq\Psi(\rho)$ with cardinality of $\Psi(\rho)$ satisfying $\psi=O(p^{\delta\rho}/\sqrt{\log(p)})$. Note that the probability in (29) becomes $o(1)$ under $0<\delta\leq 1/2$ and $0<\rho<1$. Using Chebyshev's inequality for the second probability on the right hand side of (28), we have

$$\begin{aligned}
&\mathbb{P}\Big\{\Big|\sum_{(a,b)\in\{S_{\leq 0}\setminus\Psi_{\leq 0}(\rho)\}}\{\frac{\mathbb{I}(T_{ab}\geq t)-Q(t)}{p_1Q(t)}\}\Big|\geq \upsilon/2\Big\}\\
&\leq \frac{4}{p_1^2Q^2(t)\upsilon^2}\mathbb{E}\Big\{\Big|\sum_{(a,b)\in\{S_{\leq 0}\setminus\Psi_{\leq 0}(\rho)\}}\big\{\mathbb{I}(T_{ab}\geq t)-Q(t)\big\}\Big|^2\Big\}.
\end{aligned}\tag{30}$$

Substituting (43) from Lemma 7 into (30) and then together with (29), the $o(1)$ result of (26) follows.

To show (27), by substituting (49) from the proof of Lemma 7 into (30), then together with (29), and finally taking the integral for some constants $C>0$ and $C'>0$, we get

$$\begin{aligned}
&\int_0^{t_p}\mathbb{P}\Big\{\Big|\frac{\sum_{(a,b)\in S_{\leq 0}}\{\mathbb{I}(T_{ab}\geq t)-Q(t)\}}{p_1Q(t)}\Big|\geq\upsilon\Big\}\,dt\\
&\leq 4t_pO\big(\frac{p^{\delta\rho}}{p_1\sqrt{\log(p)}}\big)+\frac{4C'o(p^{2\delta\rho})t_p}{\upsilon^2p_1^2}\exp\big((1-\rho)(3+\rho)t_p^2/4\big)+\frac{4Ct_p}{\upsilon^2\big(\log(p)\big)^{1+\gamma_1}}\\
&=O(\frac{p^{\delta\rho}}{p_1})+\frac{4C'o(p^{2\delta\rho})O(\sqrt{\log(p)})}{\upsilon^2p_1^2}\times\big(\frac{p_1}{\log(p_1)}\big)^{(1-\rho)(3+\rho)/2}+\frac{4CO(\sqrt{\log(p)})}{\upsilon^2\log(p)^{1+\gamma_1}}\\
&=o\big(1/\sqrt{\log(p)}\big),
\end{aligned}$$

where the first equality is deduced by substituting $t_p=\sqrt{2\log(p_1)-2\log(\log(p_1))}$, which is $O(\sqrt{\log(p)})$, and the last equality is deduced since the last term in the third line is the slowest term and it is $o\big(1/\sqrt{\log(p)}\big)$.

Now, suppose that t_0 in (11) does not exist, and consider $t_0=\sqrt{2\log(p_1)}$, our alternative value for the threshold that was proposed in Section 4.1. We have

$$\begin{aligned}
\mathbb{P}\big(\sum_{\substack{(a,b)\in\mathcal{V}\times\mathcal{V}\\ a<b}}\mathbb{I}(\widehat{\text{sign}}_{ab}\neq\text{sign}(\boldsymbol{\Theta}_{ab}))\geq 1\big)&\leq\mathbb{P}\big(\sum_{(a,b)\in S_{\leq 0}}\mathbb{I}(\hat{T}_{ab}\geq t_0)\geq 1\big)\\
&\quad+\mathbb{P}\big(\sum_{(a,b)\in S_{\geq 0}}\mathbb{I}(\hat{T}_{ab}\leq -t_0)\geq 1\big).
\end{aligned}\tag{31}$$

Since on $S_{\leq 0}$, $\hat{T}_{ab} \leq \tilde{T}_{ab}$, for the first probability on the right hand side of (31) we can write

$$\mathbb{P}\big(\sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(\hat{T}_{ab} \geq t_0) \geq 1\big) \leq \mathbb{P}\big(\sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(\tilde{T}_{ab} \geq t_0) \geq 1\big) = \mathbb{P}\big(\bigcup_{(a,b)\in S_{\leq 0}} \{\tilde{T}_{ab} \geq t_0\}\big).$$

Using the law of total probability, by conditioning on the event $\big\{\max_{a,b}|\tilde{T}_{ab} - T_{ab}| \leq \frac{1}{\sqrt{\log(p)}}\big\}$ and its complement set, we get

$$\begin{aligned}
\mathbb{P}\big(\sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(\hat{T}_{ab} \geq t_0) \geq 1\big) &\leq \mathbb{P}\Big(\bigcup_{(a,b)\in S_{\leq 0}} \{\tilde{T}_{ab} \geq t_0\} \,\Big|\, \max_{a,b}|\tilde{T}_{ab} - T_{ab}| \leq \frac{1}{\sqrt{\log(p)}}\Big) \\
&\quad + \mathbb{P}\big(\max_{a,b}|\tilde{T}_{ab} - T_{ab}| > \frac{1}{\sqrt{\log(p)}}\big) \\
&\leq p_1 \max_{(a,b)\in S_{\leq 0}} \mathbb{P}\big(\tilde{T}_{ab} \geq t_0 \,\big|\, \max_{a,b}|\tilde{T}_{ab} - T_{ab}| \leq \frac{1}{\sqrt{\log(p)}}\big) + o(1) \\
&\leq p_1 \max_{(a,b)\in S_{\leq 0}} \mathbb{P}\big(T_{ab} \geq t_0 - \frac{1}{\sqrt{\log(p)}}\big) + o(1), \qquad (32)
\end{aligned}$$

where the $o(1)$ result is deduced using the first result of Lemma 1. Using the Cramér result from Lemma 1, we can approximate $\mathbb{P}\big(T_{ab} \geq t_0 - \frac{1}{\sqrt{\log(p)}}\big)$ with $Q\big(t_0 - \frac{1}{\sqrt{\log(p)}}\big)$. Then using the upper bound in Lemma 7.1 from Javanmard and Javadi (2019), which implies that $Q(t) < \phi(t)/t$, for all $t \geq 0$, with $\phi(t) = \exp(-t^2/2)/\sqrt{2\pi}$ as the density of the standard normal distribution at point t and π as the mathematical constant, and substituting $t_0 = \sqrt{2\log(p_1)}$ into (32), we get

$$\begin{aligned}
&\mathbb{P}\big(\sum_{(a,b)\in S_{\leq 0}} \mathbb{I}(\hat{T}_{ab} \geq t_0) \geq 1\big) \\
&\leq \frac{p_1}{\sqrt{2\pi}\big(\sqrt{2\log(p_1)} - \frac{1}{\sqrt{\log(p)}}\big)} \exp\big(-(\sqrt{2\log(p_1)} - \frac{1}{\sqrt{\log(p)}})^2/2\big) + o(1) \\
&= \frac{1}{\sqrt{2\pi}\big(\sqrt{2\log(p_1)} - \frac{1}{\sqrt{\log(p)}}\big)} \exp\Big(\sqrt{\frac{2\log(p_1)}{\log(p)}} - \frac{1}{2\log(p)}\Big) + o(1),
\end{aligned}$$

which tends to zero with increasing p. By similar techniques, we can show that the second probability on the right hand side of (31) also tends to zero with increasing p, and the proof follows directly.

B.3 Proof of Theorem 3

Since the denominator of the threshold defined in (11) is data-dependent and different from the denominator of the directional power in (2), which does not depend on data, we first upper bound the threshold t_0 by a data-independent threshold $t_\star$ defined as

$$t_\star = \Phi^{-1}\big(1 - \frac{qs}{2p_1}(1 - o(1))\big).$$

By the same technique as in the proof of Lemma E.12 from Cui et al. (2021), it can be shown that $t_0 \leq t_\star$ with high probability.

For every $\epsilon_1 > 0$ and $\epsilon_2 > 0$, define the event

$$D(\epsilon_1, \epsilon_2) := \big\{ \max_{a,b} |\tilde{T}_{ab} - T_{ab}| \leq \epsilon_1, \quad \max_{a,b} |\hat{\sigma}_{ab} - \sigma_{ab}| \leq \epsilon_2 \big\}.$$

Let $S_{>0} := \{1 \leq a < b \leq p : \mathbf{\Theta}_{ab} > 0\}$ and $S_{<0} := \{1 \leq a < b \leq p : \mathbf{\Theta}_{ab} < 0\}$. Then $S = S_{>0} \bigcup S_{<0}$ and we have

$$\begin{aligned}
\text{Power}_{\text{dir}} &= \frac{1}{s} \sum_{(a,b)\in S_{>0}} \mathbb{P}(\hat{T}_{ab} \geq t_0) + \frac{1}{s} \sum_{(a,b)\in S_{<0}} \mathbb{P}(\hat{T}_{ab} \leq -t_0) \\
&\geq \frac{1}{s} \sum_{(a,b)\in S_{>0}} \mathbb{P}(\hat{T}_{ab} \geq t_\star) + \frac{1}{s} \sum_{(a,b)\in S_{<0}} \mathbb{P}(\hat{T}_{ab} \leq -t_\star) \\
&\geq \frac{1}{s} \sum_{(a,b)\in S_{>0}} \mathbb{P}\big(\{\hat{T}_{ab} \geq t_\star\} \bigcap D\big) + \frac{1}{s} \sum_{(a,b)\in S_{<0}} \mathbb{P}\big(\{\hat{T}_{ab} \leq -t_\star\} \bigcap D\big) - \mathbb{P}(D^c) \\
&= \frac{1}{s} \sum_{(a,b)\in S_{>0}} \mathbb{P}\bigg(\big\{\tilde{T}_{ab} \geq t_\star - \frac{\sqrt{n}\mathbf{\Theta}_{ab}}{\hat{\sigma}_{ab}}\big\} \bigcap D\bigg) - \mathbb{P}(D^c) \\
&\quad + \frac{1}{s} \sum_{(a,b)\in S_{<0}} \mathbb{P}\bigg(\big\{\tilde{T}_{ab} \leq -t_\star - \frac{\sqrt{n}\mathbf{\Theta}_{ab}}{\hat{\sigma}_{ab}}\big\} \bigcap D\bigg) \\
&\geq \frac{1}{s} \sum_{(a,b)\in S_{>0}} \mathbb{P}\bigg(T_{ab} \geq t_\star + \epsilon_1 - \frac{\sqrt{n}\mathbf{\Theta}_{ab}}{\sigma_{ab} + \epsilon_2}\bigg) - \mathbb{P}(D^c) \\
&\quad + \frac{1}{s} \sum_{(a,b)\in S_{<0}} \mathbb{P}\bigg(T_{ab} \leq -t_\star - \epsilon_1 - \frac{\sqrt{n}\mathbf{\Theta}_{ab}}{\sigma_{ab} + \epsilon_2}\bigg) \\
&= \frac{1}{s} \sum_{(a,b)\in S_{>0}} \mathbb{P}\bigg(T_{ab} \geq t_\star + \epsilon_1 - \frac{\sqrt{n}|\mathbf{\Theta}_{ab}|}{\sigma_{ab} + \epsilon_2}\bigg) - \mathbb{P}(D^c) \\
&\quad + \frac{1}{s} \sum_{(a,b)\in S_{<0}} \mathbb{P}\bigg(T_{ab} \leq -(t_\star + \epsilon_1 - \frac{\sqrt{n}|\mathbf{\Theta}_{ab}|}{\sigma_{ab} + \epsilon_2})\bigg),
\end{aligned} \tag{33}$$

where the second equality is deduced due to the fact that $\hat{T}_{ab} = \tilde{T}_{ab} + \frac{\sqrt{n}\mathbf{\Theta}_{ab}}{\hat{\sigma}_{ab}}$. Using Lemma 7.1 of Javanmard and Javadi (2019), we have

$$G(\sqrt{2\log(p_1/s)}) \leq \frac{1}{\sqrt{\pi \log(p_1/s)}} \exp\{-\log(p_1/s)\} = \frac{s}{p_1\sqrt{\pi \log(p_1/s)}}. \tag{34}$$

It can be shown that $t_\star < \sqrt{2\log(p_1/s)}$. Assuming otherwise, i.e., $t_\star \geq \sqrt{2\log(p_1/s)}$, we have $G(t_\star) \leq G(\sqrt{2\log(p_1/s)})$, because $G(\cdot)$ is a decreasing function. Combined with (34), we get

$$\frac{qs}{p_1}(1 - o(1)) = G(t_\star) \leq \frac{s}{p_1\sqrt{\pi \log(p_1/s)}},$$

which is a contradiction, since $s \leq pd$ and $d = o\big(\frac{n^{1/3}}{\log(p)}\big)$. As a result $\frac{s}{p_1} = o\big(\frac{n^{1/3}}{(p-1)\log(p)}\big)$ which tends to zero when $p > n^{1/3}$. Therefore, $t_\star < \sqrt{2\log(p_1/s)}$. By assuming $|\boldsymbol{\Theta}_{ab}| > \sqrt{\frac{2\sigma_{ab}^2 \log(p_1/s)}{n}}$, we have

$$\frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}} > \sqrt{2\log(p_1/s)} > t_\star > \Phi^{-1}(1 - \frac{qs}{2p_1}).$$

As such, $\Phi\big(\Phi^{-1}(1-\frac{qs}{2p_1}) - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\big) < 1/2$ and the denominator of (16) is lower bounded by a constant. Thus, under the conditions of Lemma 1 and Lemma 2 of Jankova and van de Geer (2015), we have

$$\frac{\mathbb{P}(D^c)}{\frac{1}{s}\sum_{(a,b)\in S}\left(1 - \Phi\big(\Phi^{-1}(1-\frac{qs}{2p_1}) - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\big)\right)} \to 0,$$

as $n \to \infty$. As (33) holds for all positive small values ϵ_1 and ϵ_2, we can write

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{\text{Power}_{\text{dir}}}{\frac{1}{s}\sum_{(a,b)\in S}\left(1 - \Phi\big(\Phi^{-1}(1-\frac{qs}{2p_1}) - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\big)\right)} \\
&\geq \liminf_{n\to\infty} \frac{1}{\frac{1}{s}\sum_{(a,b)\in S}\left(1 - \Phi\big(\Phi^{-1}(1-\frac{qs}{2p_1}) - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\big)\right)}\Bigg\{ - \mathbb{P}(D^c) \\
&\quad + \frac{1}{s}\sum_{(a,b)\in S_{>0}} \mathbb{P}\left(T_{ab} \geq t_\star - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\right) + \frac{1}{s}\sum_{(a,b)\in S_{<0}} \mathbb{P}\left(T_{ab} \leq -\big(t_\star - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\big)\right)\Bigg\} \\
&= \liminf_{n\to\infty} \frac{1}{\frac{1}{s}\sum_{(a,b)\in S}\left(1 - \Phi\big(\Phi^{-1}(1-\frac{qs}{2p_1}) - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\big)\right)}\Bigg\{ - \mathbb{P}(D^c) \\
&\quad + \frac{1}{s}\sum_{(a,b)\in S_{>0}} \left(1 - \Phi\big(t_\star - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\big)\right) + \frac{1}{s}\sum_{(a,b)\in S_{<0}} \Phi\left(-\big(t_\star - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\big)\right)\Bigg\} \\
&= \liminf_{n\to\infty} \frac{1}{\frac{1}{s}\sum_{(a,b)\in S}\left(1 - \Phi\big(\Phi^{-1}(1-\frac{qs}{2p_1}) - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\big)\right)} \times \frac{1}{s}\sum_{(a,b)\in S}\left(1 - \Phi\big(t_\star - \frac{\sqrt{n}|\boldsymbol{\Theta}_{ab}|}{\sigma_{ab}}\big)\right) \\
&= 1,
\end{aligned}$$

where the first equality is deduced using the Cramér result from Lemma 1, and the last equality is deduced by substituting $t_\star = \Phi^{-1}\big(1 - \frac{qs}{2p_1}(1 - o(1))\big)$.

Note that $1 - \Phi\big(\Phi^{-1}(1-\frac{qs}{2p_1}) - u\big)$ is an increasing function of u. As such, by considering $\boldsymbol{\Theta}_{\min} = \min_{(a,b)\in S}|\boldsymbol{\Theta}_{ab}|$, we get

$$\liminf_{n\to\infty} \frac{\text{Power}_{\text{dir}}}{\frac{1}{s}\sum_{(a,b)\in S}\left(1 - \Phi\big(\Phi^{-1}(1-\frac{qs}{2p_1}) - \frac{\sqrt{n}\boldsymbol{\Theta}_{\min}}{\sigma_{ab}}\big)\right)} \geq 1.$$

Thus, to show $\text{Power}_{\text{dir}} \to 1$, it is enough to show that

$$\frac{1}{s}\sum_{(a,b)\in S}\Big(1-\Phi\big(\Phi^{-1}(1-\frac{qs}{2p_1})-\frac{\sqrt{n}\boldsymbol{\Theta}_{\min}}{\sigma_{ab}}\big)\Big)\to 1,$$

or, equivalently, it is enough to show that $\frac{\sqrt{n}\boldsymbol{\Theta}_{\min}}{\sigma_{ab}}-\Phi^{-1}(1-\frac{qs}{2p_1})\to\infty$. Using Lemma 7.1 of Javanmard and Javadi (2019), we have that $G(\sqrt{2\log(\frac{2p_1}{qs})})\leq\frac{qs}{2p_1\sqrt{\pi\log(\frac{2p_1}{qs})}}<\frac{qs}{p_1}$, from which we deduce that $\sqrt{2\log(\frac{2p_1}{qs})}\geq\Phi^{-1}(1-\frac{qs}{2p_1})$. As such,

$$\frac{\sqrt{n}\boldsymbol{\Theta}_{\min}}{\sigma_{ab}}-\Phi^{-1}(1-\frac{qs}{2p_1})\geq\frac{\sqrt{n}\boldsymbol{\Theta}_{\min}}{\sigma_{ab}}-\sqrt{2\log(\frac{2p_1}{qs})}.$$

Assuming $\sqrt{n}\boldsymbol{\Theta}_{\min}-\sqrt{2\max_{(a,b)\in S}\sigma_{ab}^2\log(\frac{2p_1}{qs})}\to\infty$, it can be easily shown that $\frac{\sqrt{n}\boldsymbol{\Theta}_{\min}}{\sigma_{ab}}-\Phi^{-1}(1-\frac{qs}{2p_1})\to\infty$, and the proof is completed.

B.4 Proof of Lemma 4

By defining the random variable

$$U_{ab}=\sum_{k=1}^{K}\sum_{l=1}^{n_k}\frac{\sigma_{ab}}{\sqrt{n}\hat{\sigma}^2_{ab,k}}\big(\boldsymbol{\Theta}_a^\top\dot{\mathbf{X}}_{l,k}\dot{\mathbf{X}}_{l,k}^\top\boldsymbol{\Theta}_b-\boldsymbol{\Theta}_{ab}\big),$$

we get $\max_{a,b}|\tilde{U}_{ab}-T_{ab}|\leq\max_{a,b}|\tilde{U}_{ab}-U_{ab}|+\max_{a,b}|U_{ab}-T_{ab}|$. First, to show the negligibility of the term $\max_{a,b}|U_{ab}-T_{ab}|$, we have

$$\max_{a,b}|U_{ab}-T_{ab}|\leq\max_{a,b}\Big\{\sum_{k=1}^{K}\frac{\sigma_{ab}}{\sqrt{n}}\big|\frac{1}{\hat{\sigma}^2_{ab,k}}-\frac{1}{\sigma^2_{ab}}\big|\times\Big|\sum_{l=1}^{n_k}(\boldsymbol{\Theta}_a^\top\dot{\mathbf{X}}_{l,k}\dot{\mathbf{X}}_{l,k}^\top\boldsymbol{\Theta}_b-\boldsymbol{\Theta}_{ab})\Big|\Big\}. \tag{35}$$

Using (21) and the k-th sample, it can be shown that for the consistent estimator $\hat{\sigma}^2_{ab,k}$, we have

$$\big|\frac{1}{\hat{\sigma}^2_{ab,k}}-\frac{1}{\sigma^2_{ab}}\big|=O_p\big(\sqrt{\log(p)/n_k}\big),\quad k=1,\ldots,K, \tag{36}$$

under the conditions $1/\alpha=O(1)$ and $\kappa_{\mathbf{H}}=O(1)$. Moreover, under assumption (A2) and using Lemma 8 of Jankova and van de Geer (2015), it holds that

$$\big|\sum_{l=1}^{n_k}(\boldsymbol{\Theta}_a^\top\dot{\mathbf{X}}_{l,k}\dot{\mathbf{X}}_{l,k}^\top\boldsymbol{\Theta}_b-\boldsymbol{\Theta}_{ab})\big|=O_p\big(d\sqrt{n_k\log(p)}\big). \tag{37}$$

Substituting (36) and (37) into (35), we get

$$\max_{a,b}|U_{ab}-T_{ab}|=O_p\big(Kd\log(p)/\sqrt{n}\big). \tag{38}$$

To bound the term $\max_{a,b}|\tilde{U}_{ab} - U_{ab}|$, we have

$$\max_{a,b}|\tilde{U}_{ab} - U_{ab}| \leq \max_{a,b}\left|\mathbf{R}_{ab,\dagger}\right| + \max_{a,b}\left\{\left|\sqrt{\frac{\sum_{k=1}^{K} n_k/\sigma_{ab}^2}{\sum_{k=1}^{K} n_k/\hat{\sigma}_{ab,k}^2}} - 1\right| \times \left|\sum_{k=1}^{K}\sum_{l=1}^{n_k} \frac{\sigma_{ab}}{\sqrt{n}\hat{\sigma}_{ab,k}^2}(\boldsymbol{\Theta}_a^\top \dot{\mathbf{X}}_{l,k}\dot{\mathbf{X}}_{l,k}^\top \boldsymbol{\Theta}_b - \boldsymbol{\Theta}_{ab})\right|\right\}. \tag{39}$$

It has been shown in Nezakati and Pircalabelu (2023b) that

$$|\mathbf{R}_{ab,\dagger}| = O_p\big(\frac{K}{\sqrt{n}}\max\{d^{3/2}\log(p), d^2(\log(p))^{3/2}/\sqrt{n_\dagger}\}\big). \tag{40}$$

To bound $\max_{a,b}\left|\sqrt{\frac{\sum_{k=1}^{K} n_k/\sigma_{ab}^2}{\sum_{k=1}^{K} n_k/\hat{\sigma}_{ab,k}^2}} - 1\right|$, we have

$$\begin{aligned}\left|\frac{\sum_{k=1}^{K} n_k/\hat{\sigma}_{ab,k}^2}{\sum_{k=1}^{K} n_k/\sigma_{ab}^2} - 1\right| &= \sigma_{ab}^2\left|\sum_{k=1}^{K}\frac{n_k}{n}\big(\frac{1}{\hat{\sigma}_{ab,k}^2} - \frac{1}{\sigma_{ab}^2}\big)\right| \\ &= \sum_{k=1}^{K}\frac{n_k}{n}O_p\big(\sqrt{\log(p)/n_k}\big) = O_p\big(K\sqrt{\log(p)/n}\big),\end{aligned}$$

where the second equality is deduced by using (36). Applying the upper bound indicated, we have as well that

$$\max_{a,b}\left|\sqrt{\frac{\sum_{k=1}^{K} n_k/\sigma_{ab}^2}{\sum_{k=1}^{K} n_k/\hat{\sigma}_{ab,k}^2}} - 1\right| = O_p\big(K\sqrt{\log(p)/n}\big). \tag{41}$$

Moreover, to bound $\max_{a,b}\left|\sum_{k=1}^{K}\sum_{l=1}^{n_k}\frac{\sigma_{ab}}{\sqrt{n}\hat{\sigma}_{ab,k}^2}(\boldsymbol{\Theta}_a^\top \dot{\mathbf{X}}_{l,k}\dot{\mathbf{X}}_{l,k}^\top \boldsymbol{\Theta}_b - \boldsymbol{\Theta}_{ab})\right|$, using the triangle inequality, we have

$$\begin{aligned}\max_{a,b}\left|\sum_{k=1}^{K}\sum_{l=1}^{n_k}\frac{\sigma_{ab}}{\sqrt{n}\hat{\sigma}_{ab,k}^2}(\boldsymbol{\Theta}_a^\top \dot{\mathbf{X}}_{l,k}\dot{\mathbf{X}}_{l,k}^\top \boldsymbol{\Theta}_b - \boldsymbol{\Theta}_{ab})\right| &\leq \left|\sum_{k=1}^{K}\sum_{l=1}^{n_k}\frac{1}{\sqrt{n}\sigma_{ab}}(\boldsymbol{\Theta}_a^\top \dot{\mathbf{X}}_{l,k}\dot{\mathbf{X}}_{l,k}^\top \boldsymbol{\Theta}_b - \boldsymbol{\Theta}_{ab})\right| \\ &+ \left|\sum_{k=1}^{K}\sum_{l=1}^{n_k}\frac{\sigma_{ab}}{\sqrt{n}}\big(\frac{1}{\hat{\sigma}_{ab,k}^2} - \frac{1}{\sigma_{ab}^2}\big)\big(\boldsymbol{\Theta}_a^\top \dot{\mathbf{X}}_{l,k}\dot{\mathbf{X}}_{l,k}^\top \boldsymbol{\Theta}_b - \boldsymbol{\Theta}_{ab}\big)\right|.\end{aligned}$$

Using the bounds from (36) and (37), we get

$$\begin{aligned}\max_{a,b}\left|\sum_{k=1}^{K}\sum_{l=1}^{n_k}\frac{\sigma_{ab}}{\sqrt{n}\hat{\sigma}_{ab,k}^2}(\boldsymbol{\Theta}_a^\top \dot{\mathbf{X}}_{l,k}\dot{\mathbf{X}}_{l,k}^\top \boldsymbol{\Theta}_b - \boldsymbol{\Theta}_{ab})\right| &= O_p\big(Kd\sqrt{\log(p)}\big) + O_p\big(Kd\log(p)/\sqrt{n}\big) \\ &= O_p\big(Kd\sqrt{\log(p)}\big).\end{aligned} \tag{42}$$

Substituting (40), (41) and (42) into (39), we have

$$\max_{a,b}|\tilde{U}_{ab}-U_{ab}| = O_p\big(\frac{K}{\sqrt{n}}\max\{d^{3/2}\log(p), d^2(\log(p))^{3/2}/\sqrt{n_\dagger}\}\big) + O_p\big(K^2 d\log(p)/\sqrt{n}\big).$$

Combining this with (38), we get

$$\begin{aligned}\max_{a,b}|\tilde{U}_{ab}-T_{ab}| = {} & O_p\big(\frac{K}{\sqrt{n}}\max\{d^{3/2}\log(p), d^2(\log(p))^{3/2}/\sqrt{n_\dagger}\}\big) + O_p\big(K^2 d\log(p)/\sqrt{n}\big) \\ & + O_p\big(Kd\log(p)/\sqrt{n}\big).\end{aligned}$$

By considering $K = O\big(n^{1/4}/(d\log(p))\big)$ and $d^{3/2} = o\big(\sqrt{n_\dagger}/(\log(p))^{3/2}\big)$, we have

$$\max_{a,b}|\tilde{U}_{ab}-T_{ab}| = o_p\big(\frac{n_\dagger^{1/6}}{n^{1/4}(\log(p))^{1/2}}\big) + O_p(\frac{1}{n^{1/4}}) + O_p(\frac{1}{d\log(p)}),$$

and the result follows directly.

B.5 Proof of Theorem 5

The proof of this theorem deploys a similar technique as that of Theorem 2, and we omit it. However, when the given t_0' does not exist and we let $t_0' = \sqrt{2\log(p_1)}$, we need to condition on the event $\big\{\max_{a,b}|\tilde{U}_{ab}-T_{ab}| \le \max\{\frac{n_\dagger^{1/6}}{n^{1/4}\sqrt{\log(p)}}, \frac{1}{n^{1/3}}, \frac{1}{\log(p)}\}\big\}$ and its complement set. Then the result follows directly.

Appendix C. A technical lemma and its proof

Lemma 7 Let $p \le n^r$, for some $r > 0$ and suppose that assumptions (A1)-(A2) in Appendix A hold. Consider the random variable T_{ab} defined in (13). Under the sparsity condition $d^{3/2} = o(\sqrt{n}/(\log(p))^{3/2})$, and considering cardinalities $\Upsilon = o(p^{\delta\rho})$, $\rho \in (0,1)$, $\delta \in (0, 1/2]$ and $\psi = O(p^{\delta\rho}/\sqrt{\log(p)})$, for the sets $\Gamma(\gamma, c_0)$ and $\Psi(\rho)$, respectively, it holds that

$$\mathbb{E}\bigg\{\bigg|\frac{\sum_{(a,b)\in\{S_{\le 0}\setminus\Psi_{\le 0}(\rho)\}}\big\{\mathbb{I}(T_{ab}\ge t) - Q(t)\big\}}{p_1 Q(t)}\bigg|^2\bigg\} = o(1). \tag{43}$$

Proof To show (43), we have

$$\begin{aligned}&\mathbb{E}\bigg\{\bigg|\frac{\sum_{(a,b)\in\{S_{\le 0}\setminus\Psi_{\le 0}(\rho)\}}\big\{\mathbb{I}(T_{ab}\ge t) - Q(t)\big\}}{p_1 Q(t)}\bigg|^2\bigg\} \\ &= \frac{1}{p_1^2}\sum_{(a,b),(c,d)\in\{S_{\le 0}\setminus\Psi_{\le 0}(\rho)\}}\bigg\{\frac{\mathbb{P}\big(T_{ab}\ge t, T_{cd}\ge t\big) - Q^2(t)}{Q^2(t)}\bigg\},\end{aligned}$$

where the equality is deduced first by squaring the terms inside the expectation and then using the Cramér result on the tail probability of T_{ab} from Lemma 1. To bound the above expectation, we consider two disjoint sets $\Gamma(\gamma, c_0)$ (highly correlated pair nodes) and $\Gamma^c(\gamma, c_0)$ (weakly correlated pair nodes). To simplify the notation, define

$$\Gamma_{\le 0}(\gamma, c_0) = \Gamma(\gamma, c_0)\bigcap S_{\le 0}, \quad \Gamma^c_{\le 0}(\gamma, c_0) = \Gamma^c(\gamma, c_0)\bigcap S_{\le 0}.$$

Note that the set $S_{\leq 0} \setminus \Psi_{\leq 0}(\rho)$ is equal to the union of two disjoint sets $\Gamma_{\leq 0}(\gamma, c_0) \setminus \Psi_{\leq 0}(\rho)$ and $\Gamma^c_{\leq 0}(\gamma, c_0) \setminus \Psi_{\leq 0}(\rho)$. As such,

$$\begin{aligned}
&\mathbb{E}\Big\{\Big|\frac{\sum_{(a,b)\in\{S_{\leq 0}\setminus\Psi_{\leq 0}(\rho)\}}\big\{\mathbb{I}(T_{ab}\geq t)-Q(t)\big\}}{p_1 Q(t)}\Big|^2\Big\}\\
&\leq \frac{1}{p_1^2}\sum_{(a,b),(c,d)\in\{\Gamma_{\leq 0}(\gamma,c_0)\setminus\Psi_{\leq 0}(\rho)\}}\Big\{\frac{\mathbb{P}(T_{ab}\geq t, T_{cd}\geq t)}{Q^2(t)}\Big\}\\
&\quad+\frac{1}{p_1^2}\sum_{(a,b),(c,d)\in\{\Gamma^c_{\leq 0}(\gamma,c_0)\setminus\Psi_{\leq 0}(\rho)\}}\Big\{\frac{\mathbb{P}(T_{ab}\geq t, T_{cd}\geq t)}{Q^2(t)}-1\Big\}. \qquad (44)
\end{aligned}$$

Recall that

$$\mathbf{Z}_{ab,l}=\boldsymbol{\Theta}_a^\top\dot{\mathbf{X}}_l\dot{\mathbf{X}}_l^\top\boldsymbol{\Theta}_b-\boldsymbol{\Theta}_{ab}=\mathbf{e}_a^\top\boldsymbol{\Theta}\dot{\mathbf{X}}_l\dot{\mathbf{X}}_l^\top\boldsymbol{\Theta}\mathbf{e}_b-\boldsymbol{\Theta}_{ab}=\mathbf{Z}_l^a\mathbf{Z}_l^b-\boldsymbol{\Theta}_{ab},$$

where $\mathbf{Z}_l^a$ is the a-th element of $\mathbf{Z}_l:=\boldsymbol{\Theta}\dot{\mathbf{X}}_l\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Theta})$ and $\mathbf{e}_a$ is a unit column vector with 1 at the a-th position and zero everywhere else. Thus, $\mathbb{E}(\mathbf{Z}_l^a\mathbf{Z}_l^b)=\mathbb{C}\mathrm{ov}(\mathbf{Z}_l^a,\mathbf{Z}_l^b)=\boldsymbol{\Theta}_{ab}$ and then $\mathbb{E}(\mathbf{Z}_{ab,l})=\mathbb{E}(\mathbf{Z}_l^a\mathbf{Z}_l^b)-\boldsymbol{\Theta}_{ab}=0$. Moreover,

$$\mathbb{E}(\mathbf{Z}_l^a\mathbf{Z}_l^b\mathbf{Z}_l^c\mathbf{Z}_l^d)=\boldsymbol{\Theta}_{ab}\boldsymbol{\Theta}_{cd}+\boldsymbol{\Theta}_{ac}\boldsymbol{\Theta}_{bd}+\boldsymbol{\Theta}_{ad}\boldsymbol{\Theta}_{bc},$$

and as such, we get

$$\mathbb{C}\mathrm{ov}\Big(\frac{\mathbf{Z}_{ab,l}}{\sigma_{ab}},\frac{\mathbf{Z}_{cd,l}}{\sigma_{cd}}\Big)=\frac{\boldsymbol{\Theta}_{ac}\boldsymbol{\Theta}_{bd}+\boldsymbol{\Theta}_{ad}\boldsymbol{\Theta}_{bc}}{\sqrt{(\boldsymbol{\Theta}_{aa}\boldsymbol{\Theta}_{bb}+\boldsymbol{\Theta}_{ab}^2)(\boldsymbol{\Theta}_{cc}\boldsymbol{\Theta}_{dd}+\boldsymbol{\Theta}_{cd}^2)}}.$$

On $\Gamma_{\leq 0}(\gamma, c_0) \setminus \Psi_{\leq 0}(\rho)$, we have $|\boldsymbol{\Theta}_{ab}|\leq\frac{1}{\sqrt{2}}\big(\frac{1-\rho}{1+\rho}\big)\sqrt{\boldsymbol{\Theta}_{aa}\boldsymbol{\Theta}_{bb}}$, and so,

$$\begin{aligned}
\Big|\mathbb{C}\mathrm{ov}\Big(\frac{\mathbf{Z}_{ab,l}}{\sigma_{ab}},\frac{\mathbf{Z}_{cd,l}}{\sigma_{cd}}\Big)\Big|&\leq\Big(\frac{1-\rho}{1+\rho}\Big)^2\Bigg(\frac{\boldsymbol{\Theta}_{aa}\boldsymbol{\Theta}_{bb}\boldsymbol{\Theta}_{cc}\boldsymbol{\Theta}_{dd}}{\boldsymbol{\Theta}_{aa}\boldsymbol{\Theta}_{bb}\boldsymbol{\Theta}_{cc}\boldsymbol{\Theta}_{dd}+\boldsymbol{\Theta}_{aa}\boldsymbol{\Theta}_{bb}\boldsymbol{\Theta}_{cd}^2+\boldsymbol{\Theta}_{cc}\boldsymbol{\Theta}_{dd}\boldsymbol{\Theta}_{ab}^2+\boldsymbol{\Theta}_{ab}^2\boldsymbol{\Theta}_{cd}^2}\Bigg)^{1/2}\\
&\leq\Big(\frac{1-\rho}{1+\rho}\Big)^2.
\end{aligned}$$

Considering the above bound on the covariance of $\mathbf{Z}_{ab,l}/\sigma_{ab}$ and $\mathbf{Z}_{cd,l}/\sigma_{cd}$ in Lemma 6.2 of Liu (2013), we can write on $\Gamma_{\leq 0}(\gamma, c_0) \setminus \Psi_{\leq 0}(\rho)$,

$$\begin{aligned}
\mathbb{P}\big(T_{ab}\geq t, T_{cd}\geq t\big)&=\mathbb{P}\Big(\sum_{l=1}^n\frac{\mathbf{Z}_{ab,l}}{\sigma_{ab}}\geq t\sqrt{n},\sum_{l=1}^n\frac{\mathbf{Z}_{cd,l}}{\sigma_{cd}}\geq t\sqrt{n}\Big)\\
&\leq C'(1+t)^{-2}\exp\Big(-t^2/\big(1+(\frac{1-\rho}{1+\rho})^2\big)\Big)\\
&\leq C'(1+t)^{-2}\exp\big(-(1+\rho)^2t^2/4\big),
\end{aligned}$$

for some constant $C'>0$. As such, on $\Gamma_{\leq 0}(\gamma, c_0) \setminus \Psi_{\leq 0}(\rho)$,

$$\frac{1}{p_1^2}\sum_{(a,b),(c,d)\in\{\Gamma_{\leq 0}(\gamma,c_0)\setminus\Psi_{\leq 0}(\rho)\}}\frac{\mathbb{P}\big(T_{ab}\geq t, T_{cd}\geq t\big)}{Q^2(t)}\leq\frac{C'\Upsilon^2}{p_1^2Q^2(t)}(1+t)^{-2}\exp\big(-(1+\rho)^2t^2/4\big) \qquad (45)$$

$$\leq\frac{C'\Upsilon^2}{p_1^2}\exp\big((1-\rho)(3+\rho)t^2/4\big), \qquad (46)$$

where recall that $\Upsilon = \#\Gamma(\gamma, c_0)$. The last inequality follows using the lower bound of Lemma 7.1 of Javanmard and Javadi (2019), that is $Q(t) > \frac{t}{t^2+1}\phi(t)$, for all $t \geq 0$, with $\phi(t) = \exp(-t^2/2)/\sqrt{2\pi}$. Using this lower bound on $Q(t)$, we get

$$\frac{1}{Q^2(t)(t+1)^2} < \frac{2\pi(t^2+1)^2}{t^2(t+1)^2}\exp(t^2) \leq 2\pi\exp(t^2).$$

The result in (46) is deduced by substituting this upper bound in (45).

Now, to bound the second term on the right hand side of (44), we have

$$\frac{\mathbb{P}(T_{ab} \geq t, T_{cd} \geq t)}{Q^2(t)} - 1 \leq \left|\frac{\mathbb{P}\big(|T_{ab}| \geq t, |T_{cd}| \geq t\big)}{Q^2(t)} - 1\right|.$$

As such,

$$\begin{aligned}
&\frac{1}{p_1^2}\sum_{(a,b),(c,d)\in\{\Gamma^c_{\leq 0}(\gamma,c_0)\setminus\Psi_{\leq 0}(\rho)\}}\left\{\frac{\mathbb{P}(T_{ab} \geq t, T_{cd} \geq t)}{Q^2(t)} - 1\right\} \\
&\quad\leq \frac{1}{p_1^2}\sum_{(a,b),(c,d)\in\{\Gamma^c_{\leq 0}(\gamma,c_0)\setminus\Psi_{\leq 0}(\rho)\}}\left|\frac{\mathbb{P}\big(|T_{ab}| \geq t, |T_{cd}| \geq t\big)}{Q^2(t)} - 1\right| \\
&\quad\leq \frac{1}{p_1^2}\sum_{(a,b),(c,d)\in\Gamma^c_{\leq 0}(\gamma,c_0)}\left|\frac{\mathbb{P}\big(|T_{ab}| \geq t, |T_{cd}| \geq t\big)}{Q^2(t)} - 1\right|. \qquad (47)
\end{aligned}$$

Consider $\mathcal{Z}_l = (\frac{\mathbf{Z}_{ab,l}}{\sigma_{ab}}, \frac{\mathbf{Z}_{cd,l}}{\sigma_{cd}})^\top$, $l = 1, \ldots, n$, on $\Gamma^c_{\leq 0}(\gamma, c_0)$, then

$$\big|\big|\mathbb{C}\mathrm{ov}(\mathcal{Z}_l) - \mathbf{I}_2\big|\big|_F = \sqrt{2}\frac{|\mathbf{\Theta}_{ac}\mathbf{\Theta}_{bd} + \mathbf{\Theta}_{ad}\mathbf{\Theta}_{bc}|}{\sigma_{ab}\sigma_{cd}} \leq c_3(\log(p))^{-2-\gamma},$$

for every $\gamma > 0$, and some constant $c_3 > 0$. The last inequality holds because on $\Gamma^c_{\leq 0}(\gamma, c_0)$, we have $|\mathbf{\Theta}^0_{ab}| \leq c_0(\log(p))^{-1-\gamma/2}$. By Lemma 6.1 of Liu (2013),

$$\sup_{0\leq t\leq 2\sqrt{\log(p)}}\left|\frac{\mathbb{P}\big(|T_{ab}| \geq t, |T_{cd}| \geq t\big)}{Q^2(t)} - 1\right| \leq C\big(\log(p)\big)^{-1-\gamma_1},$$

for some constant $C > 0$, where $\gamma_1 = \min\{\gamma, 1/2\}$. Substituting this bound into (47), we get

$$\begin{aligned}
\frac{1}{p_1^2}\sum_{(a,b),(c,d)\in\{\Gamma^c_{\leq 0}(\gamma,c_0)\setminus\Psi_{\leq 0}(\rho)\}}\left\{\frac{\mathbb{P}(T_{ab} \geq t, T_{cd} \geq t)}{Q^2(t)} - 1\right\} &\leq \frac{C(\#\Gamma^c_{\leq 0}(\gamma, c_0))^2}{p_1^2} \\
&\quad\times \big(\log(p)\big)^{-1-\gamma_1}, \qquad (48)
\end{aligned}$$

where $\#\Gamma^c_{\leq 0}(\gamma, c_0)$ is the cardinality of the set $\Gamma^c_{\leq 0}(\gamma, c_0)$. By substituting (46) and (48) into (44), we have

$$\begin{aligned}
&\mathbb{E}\Big\{\Big|\frac{\sum_{(a,b)\in\{S_{\leq 0}\setminus\Psi_{\leq 0}(\rho)\}}\big\{\mathbb{I}(T_{ab}\geq t)-Q(t)\big\}}{p_1 Q(t)}\Big|^2\Big\}\\
&\leq \frac{C' o(p^{2\delta\rho})}{p_1^2}\exp\big((1-\rho)(3+\rho)t_p^2/4\big)+C\big(\log(p)\big)^{-1-\gamma_1} \qquad (49)\\
&= \frac{C' o(p^{2\delta\rho})}{p_1^2}\times\big(\frac{p_1}{\log(p_1)}\big)^{(1-\rho)(3+\rho)/2}+C\big(\log(p)\big)^{-1-\gamma_1}=o(1),
\end{aligned}$$

where (49) is deduced using the fact that $\exp\big((1-\rho)(3+\rho)t^2/4\big)\leq\exp\big((1-\rho)(3+\rho)t_p^2/4\big)$ for all $t\leq t_p$, and the last equality follows by substituting $t_p=\sqrt{2\log(p_1)-2\log(\log(p_1))}$. ■

References

R. Akbani, K. C. Akdemir, B. A. Aksoy, M. Albert, A. Ally, S. B. Amin, H. Arachchi, A. Arora, J. T. Auman, B. Ayala, et al. Genomic classification of cutaneous melanoma. *Cell*, 161(7):1681–1696, 2015.

R. F. Barber and E. J. Candès. A knockoff filter for high-dimensional selective inference. 47(5):2504–2537, 2019.

H. Battey, J. Fan, H. Liu, J. Lu, and Z. Zhu. Distributed testing and estimation under sparse high dimensional models. *The Annals of Statistics*, 46(3):1352–1382, 2018.

T. Cai, W. Liu, and X. Luo. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

T. Cai, W. Liu, and H. Zhou. Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *The Annals of Statistics*, 44(2):455–488, 2016.

Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. *Nature*, 543(7645):378–384, 2017.

A. Cardoso-Cachopo. Improving methods for single-label text categorization. PhD Thesis, Instituto Superior Tecnico, Universidade Tecnica de Lisboa, 2007.

X. Chen and M.-g. Xie. A split-and-conquer approach for analysis of extraordinarily large data. *Statistica Sinica*, 24(4):1655–1684, 2014.

C. Cui, J. Jia, Y. Xiao, and H. Zhang. Directional FDR control for sub-Gaussian sparse glms. *arXiv preprint arXiv:2105.00393*, 2021.

S. T. Dumais. Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments, & Computers*, 23(2):229–236, 1991.

J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics*, 9(3):432–441, 2008.

A. Gelman and F. Tuerlinckx. Type s error rates for classical and Bayesian single and multiple comparison procedures. *Computational statistics*, 15(3):373–390, 2000.

J. Guo, E. Levina, G. Michailidis, and J. Zhu. Joint estimation of multiple graphical models. *Biometrika*, 98(1):1–15, 2011.

C. J. Hsieh, M. A. Sustik, I. S. Dhillon, and P. Ravikumar. Quic: quadratic approximation for sparse inverse covariance estimation. *The Journal of Machine Learning Research*, 15 (1):2911–2947, 2014.

J. Jankova and S. van de Geer. Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, 9(1):1205–1229, 2015.

A. Javanmard and H. Javadi. False discovery rate control via debiased lasso. *Electronic Journal of Statistics*, 13(1):1212–1253, 2019.

J. Jia, K. Rohe, and B. Yu. The lasso under Poisson-like heteroscedasticity. *Statistica Sinica*, 23(1):99–118, 2013.

M. I. Jordan, J. D. Lee, and Y. Yang. Communication-efficient distributed statistical inference. *Journal of the American Statistical Association*, 114(526):668–681, 2018.

S. Li, T. T. Cai, and H. Li. Transfer learning in large-scale Gaussian graphical models with false discovery rate control. *Journal of the American Statistical Association*, 00(0):1–13, 2022. In press.

J. Liu, T. Lichtenberg, K. A. Hoadley, L. M. Poisson, A. J. Lazar, A. D. Cherniack, A. J. Kovatich, C. C. Benz, D. A. Levine, A. V. Lee, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*, 173(2):400–416, 2018.

W. Liu. Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics*, 41(6):2948–2978, 2013.

E. Nezakati and E. Pircalabelu. Estimation and inference in sparse multivariate regression and conditional Gaussian graphical models under an unbalanced distributed setting. pages 1–50, 2023a. Technical report.

E. Nezakati and E. Pircalabelu. Unbalanced distributed estimation and inference for the precision matrix in Gaussian graphical models. *Statistics and Computing*, 33:1–14, 2023b.

B. Ng, G. Varoquaux, J. B. Poline, and B. Thirion. A novel sparse group Gaussian graphical model for functional connectivity estimation. In *Information Processing in Medical Imaging: 23rd International Conference, IPMI 2013, Asilomar, CA, USA, June 28–July 3, 2013. Proceedings 23*, pages 256–267. Springer, 2013.

P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980, 2011.

J. Schäfer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1):1–30, 2005.

L. Tang, L. Zhou, and P. X. K. Song. Distributed simultaneous inference in generalized linear models via confidence distribution. *Journal of Multivariate Analysis*, 176:1–13, 2020.

J. Wang. Joint estimation of sparse multivariate regression and conditional graphical models. *Statistica Sinica*, 25(3):831–851, 2015.

Y. Xia, T. Cai, and T. T. Cai. Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika*, 102(2):247–266, 2015.

M. Yuan and Y. Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

P. Zhao and B. Yu. On model selection consistency of lasso. *The Journal of Machine Learning Research*, 7:2541–2563, 2006.