Statistica Sinica Preprint No: SS-2023-0385	
Title	Universally Consistent Tests for the Graph of a Gaussian Graphical Model
Manuscript ID	SS-2023-0385
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0385
Complete List of Authors	Thien-Minh Le, Ping-Shou Zhong and Chenlei Leng
Corresponding Authors	Ping-Shou Zhong
E-mails	pszhong@uic.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

Universally Consistent Tests for the Graph of a Gaussian Graphical Model

Thien-Minh Le[1], Ping-Shou Zhong[2], and Chenlei Leng[3]

[1] *University of Tennessee at Chattanooga*

[2] *University of Illinois Chicago*

[3] *University of Warwick*

Abstract:

The Gaussian graphical model is routinely employed to model the joint distribution of multiple random variables. The graph it induces is not only useful for describing the relationship between these variables but also critical for improving statistical estimation precision. In high-dimensional data analysis, despite abundant literature on estimating this graph structure, tests for the adequacy of its specification at a global level are severely underdeveloped. To make progress, this paper proposes novel goodness-of-fit tests that are computationally easy and theoretically tractable. The first contribution of this paper is the development of a new direct plug-in test statistic. We show that its asymptotic distribution under the null follows a Gumbel distribution with a location parameter depending on the underlying true graph structure. The direct test, however, has no power for detecting structures including the truth but not equal. Our second contribution is the development of a novel consistency-empowered test statistic that gains power

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

by, interestingly, amplifying the noise incurred in estimation. The improved test is shown to be universally consistent for all fixed alternatives. Extensive simulation illustrates that the proposed test procedures have the right size under the null, and is powerful under alternatives. As an application, we apply the tests to the analysis of a COVID-19 data set, demonstrating that our test can serve as a valuable tool in choosing a graph structure to improve estimation efficiency.

Key words and phrases: Dependence, Gaussian graphical model, Goodness-of-fit test, Gumbel distribution, High-dimensional data

1. Introduction

The Gaussian graphical model is commonly used for describing the joint distribution of multiple random variables (Lauritzen, 1996). The graph structure induced by this model not only delineates the conditional dependence between these variables, but also is critical for improving estimation precision. In estimating regression parameters in generalized estimating equations (GEE) for example, Zhou and Song (2016) found that incorporating a suitable dependence structure of covariates can improve estimation efficiency, sometimes substantially. In another example, Li and Li (2008) showed that a correctly specified dependence structure is useful to improve estimation efficiency in regularized estimation and variable selection. On the other hand however, a mis-specified dependence structure affects efficiency

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

negatively (Zhou and Song, 2016). Therefore, specifying an appropriate graph is critical for efficiently estimating a parameter of interest.

In practice, the underlying graph structure of a given dataset may be provided by existing studies or prior knowledge or assumed a priori. In genomic studies (Li and Li, 2008; Goeman and Mansmann, 2008), rich biological knowledge is available due to intensive biomedical studies, especially for complex diseases. Existing knowledge and information are publicly available through databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO). The ontology terms of GO are structured as a graph, with terms as nodes and the relations between them as edges. Details on using GO to create graphical structures can be found on the website `https://geneontology.org/docs/ontology-relations/`. Gene pathway information can be converted into graphical structures using R packages such as graphite (Sales et al., 2012). Therefore, a natural question is whether these prior graphs are adequate to describe the data from a statistical perspective. This paper aims to develop novel goodness-of-fit tests to address this challenge, in the context of high-dimensional data in which dimensionality can exceed the sample size.

There is abundant literature focusing on estimating the underlying graph in the Gaussian graphical model. For fixed-dimensional data, Edwards (2000)

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

studied this problem by using a model selection approach that employs stepwise likelihood ratio tests, while Drton and Perlman (2004) developed a multiple testing procedure using partial correlations. For high-dimensional data, a popular approach is to employ a penalized likelihood approach, with a penalty explicitly formulated to encourage the sparsity of the resulting precision matrix that induces the underlying dependence structure. On this, we refer to Yuan and Lin (2007); Friedman et al. (2007); Cai et al. (2011); Liu and Wang (2017); Eftekhari et al. (2021), among many others. On testing the graphical structure itself, there exist methods for testing elements of the graphical structure. For example, Liu (2013) proposed a bias-corrected estimator of the precision matrix and applied it to test individual components of the precision matrix. Similar tests for individual components in a precision matrix are also discussed in Janková and van de Geer (2017); Ren et al. (2015); Ning and Liu (2017). There are also some existing global tests for precision matrices taking limited form, for example, in Xia et al. (2015) and Cheng et al. (2017). However, there is a lack of general global specification tests for precision matrices considered in this paper for testing the entire graph structure.

Our work is also related to a growing body of literature on testing specific covariance structures for high-dimensional data. In this vein, Chen

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

et al. (2010) considered testing sphericity and identity structures, Qiu and Chen (2012) and Wang et al. (2023) developed tests for bandedness structures, Zhong et al. (2017) developed tests for some parameterized covariance structures such as autoregressive and moving average structures, Zheng et al. (2019) considered tests on linear structures, and Guo and Tang (2020) considered specification tests for covariance matrices with nuisance parameters in regression models. These tests are not applicable to test graph structures. Moreover, compared with the above tests which usually involve the estimation of a finite number of nuisance parameters, one significant challenge associated with testing the graph structure in this paper is the need to estimate a high dimensional nuisance parameter.

The main novelty of this paper lies in a new goodness-of-fit test that explores the difference between a graph structure specified under the null and the true underlying graph structure, based on an appropriate maximum norm distance. We overcome the challenge of estimating the high dimensional nuisance parameter by employing a simple and direct plug-in method, thus bypassing the need of choosing tuning parameters employed in many regularization methods in the literature for estimating a graph. Despite its simplicity, the direct plug-in test is not universally consistent. It has a limitation in that it is not consistent whenever the graph under the null

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

encompasses but is not equal to the true graph. To tackle this, we develop a novel consistency-empowered test statistic by amplifying the noise, in the sense that small stochastic noises as a result of estimating zero entries in the graph will be enlarged. This modified test statistic is shown to be universally consistent for testing all types of graphs.

The paper is organized as follows. In Section 2, we introduce basic setting and our proposed test statistic. Section 3 summarizes asymptotic distributions of the proposed test statistic, and the universally consistent of the proposed test. Simulation studies are presented in Section 4. Section 5 provides an application of the proposed methods to a COVID-19 dataset in selecting appropriate graphical structures for improving the estimation efficiency. All technical proofs, additional simulation results, and a detailed procedure for selecting data-driven tuning parameters are provided in the supplementary material.

2. Setting and Proposed Test Statistics

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent and identically distributed realizations of a p-dimensional random vector $\mathbf{X}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}^* = (\sigma^*_{ij})$. The corresponding precision matrix is denoted as $\boldsymbol{\Omega}^* = (\omega^*_{ij}) = \boldsymbol{\Sigma}^{*-1}$. It is known that $\boldsymbol{\Omega}^*$ naturally induces a graph denoted as $\mathcal{G}^* = (\mathcal{V}, \mathcal{E}^*)$, where

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

$\mathcal{V} = \{1, \ldots, p\}$ is the set of nodes and $\mathcal{E}^* = \{(i,j) : \omega_{ij}^* \neq 0\} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges consisting of node pairs whose corresponding entries in $\mathbf{\Omega}^*$ are not zero. The absence of a pair of nodes in $\mathcal{E}^*$ indicates that the corresponding variables are conditionally independent given all the others when $\mathbf{X}$ is normally distributed. (Lauritzen, 1996).

While graph $\mathcal{E}^*$ is rarely known, in practice it can be estimated via the penalized likelihood methods discussed in the Introduction or assumed a priori. For the latter, when the dimension p is high, a convenient assumption popular in the literature is that $\mathbf{\Omega}^*$ admits some simple structure such as a banding or a block diagonal structure. We will denote the corresponding graph under the assumption as $\mathcal{E}_0$ and the main aim of this paper is to ascertain whether this assumption is valid. That is, we consider the following hypothesis

$$H_0 : \mathcal{E}^* = \mathcal{E}_0 \quad \text{vs.} \quad H_1 : \mathcal{E}^* \neq \mathcal{E}_0,$$

where in our high-dimensional setup, $\mathcal{E}_0$ usually has a cardinality much smaller than p^2. If $\mathbf{X}$ is normally distributed, the above hypothesis corresponds to a hypothesis for testing the support of the precision matrix $\mathbf{\Omega}^*$, where its non-zero elements are completely unspecified. Specifically, the hypothesis is:

$$H_0^* : \text{supp}(\mathbf{\Omega}^*) = \text{supp}(\mathbf{\Omega}_0) \quad \text{vs.} \quad H_1^* : \text{supp}(\mathbf{\Omega}^*) \neq \text{supp}(\mathbf{\Omega}_0),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

where $\text{supp}(\mathbf{\Omega}^*) = \{(i,j) : \omega_{ij}^* \neq 0\}$ denotes the support of $\mathbf{\Omega}^*$ (i.e., the indices of its non-zero elements), and $\mathbf{\Omega}_0 = (w_{ij,0})$ is a $p \times p$ precision matrix of $\mathbf{X}$ that is compatible with $\mathcal{E}_0$ under the null hypothesis. This compatibility means that if $(i,j) \notin \mathcal{E}_0$, then $w_{ij,0} = 0$. The number of the unknown parameters under the null is allowed to grow with p, which is drastically different from existing tests in the literature for testing a covariance matrix Σ^* with its inverse under the null often specified up to a finite number of unknown parameters (e.g.,Zhong et al. (2017); Zheng et al. (2019))

Our main idea is that if $\mathcal{E}_0$ is correctly specified, $\mathbf{\Omega}_0$ will be equal to $\mathbf{\Omega}^*$; that is, $\mathbf{\Sigma}^*\mathbf{\Omega}_0 - \mathbf{I}_p = \mathbf{0}_p$, where $\mathbf{I}_p$ is the p-dimensional identity matrix and $\mathbf{0}_p$ is a $(p \times p)$-dimensional matrix with entries all being zero. That is, if $\mathcal{E}_0 = \mathcal{E}^*$, we can write the above equation elementwise as

$$\max_{1 \leq i,j \leq p} |\mathbf{e}_j^{\mathrm{T}} \mathbf{\Sigma}^* \mathbf{w}_{i,0} - \mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i| = 0, \tag{2.1}$$

where $\mathbf{\Omega}_0 = (\mathbf{w}_{1,0}, \ldots, \mathbf{w}_{p,0})$ by denoting $\mathbf{w}_{i,0}$ as the i-th column of $\mathbf{\Omega}_0$ and $\mathbf{I}_p = (\mathbf{e}_1, \ldots, \mathbf{e}_p)$ with $\mathbf{e}_i$ being the i-th basis vector. On the other hand, if $\mathcal{E}_0$ is not correctly specified in the sense that $\mathcal{E}_0 \neq \mathcal{E}^*$, the maximum element of $\mathbf{\Sigma}^*\mathbf{\Omega}_0 - \mathbf{I}_p$ may be different from zero.

Thus, to assess whether H_0 (or H_0^*) is true is equivalent to check (2.1). If $\mathbf{\Omega}_0$ and so $\mathbf{w}_{i,0}$ is known in advance, an estimator of $(\mathbf{e}_j^{\mathrm{T}} \mathbf{\Sigma}^* \mathbf{w}_{i,0} - \mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i)^2$ may be obtained by replacing $\mathbf{\Sigma}^*$ by the sample covariance matrix $\boldsymbol{S}_n =$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

$\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{\mathrm{T}}/(n-1)$ with $\bar{\mathbf{X}} = (\bar{X}_1, \cdots, \bar{X}_p)^{\mathrm{T}} = \sum_{i=1}^{n}\mathbf{X}_i/n$. Then, we may use the following statistic D_n to distinguish H_0 and H_1,

$$D_n = \max_{1\le i,j\le p} D_{ij}^2, \quad D_{ij}^2 := (\mathbf{e}_j^{\mathrm{T}}\boldsymbol{S}_n\mathbf{w}_{i,0} - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i)^2/\theta_{ij,0},$$

where $\theta_{ij,0} = \mathrm{var}(\mathbf{e}_j^{\mathrm{T}}\boldsymbol{S}_n\mathbf{w}_{i,0} - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i)$. Note that the statistic D_n depends on the plug-in estimators of $\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_{i,0}$, which can be expressed as

$$\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_{i,0} = \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\mathbf{w}_{i1,0},$$

where $\mathbf{w}_{i1,0}$ represents the nonzero sub-vectors of $\mathbf{w}_{i,0}$, and $\mathbf{B}_{i,0}$ is a $p \times s_i$ matrix with elements equal to either 0 or 1, such that $\mathbf{B}_{i,0}\mathbf{w}_{i1,0} = \mathbf{w}_{i,0}$. It is important to note that $\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}$ and $\mathbf{w}_{i1,0}$ are both vectors of dimension s_i. Therefore, using the sample covariance $\boldsymbol{S}_n$ to estimate these two vectors remains reasonable as long as s_i satisfies condition (C2) below. Assume $\mathbf{X} = \boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\mu}$, where $\boldsymbol{\Gamma}$ is an $m \times p$ matrix and $\mathbf{Z} = (Z_1, \ldots, Z_m)^{\mathrm{T}}$ follows the multivariate model described in Assumption (D1) of the supplemental material (Bai and Saranadasa, 1996; Chen et al., 2010). This model specifies that $E(\mathbf{Z}) = 0$, $\mathrm{var}(\mathbf{Z}) = \mathbf{I}_m$, and $E(Z_i^4) = 3 + \kappa$. This multivariate model generalizes the Gaussian distribution. The leading-order term of $\theta_{ij,0}$ is provided in the following lemma.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

Lemma 1. *Under Assumption (D1) in the supplemental file, we have*

$$\theta_{ij,0} = var(\mathbf{e}_j^{\mathrm{T}} \boldsymbol{S}_n \mathbf{w}_{i,0} - \mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i) = \begin{cases} \omega_{ii}^* \sigma_{jj}^*/n, & \text{for } 1 \le i \ne j \le p \\ (\omega_{ii}^* \sigma_{ii}^* + 1 + \kappa)/n, & \text{for } 1 \le i = j \le p \end{cases}$$

In particular, if the normality assumptions hold, $\kappa = 0$ in the above expression.

However, D_n is not directly applicable because several quantities involved are unknown. Noting that under the null hypothesis, $\boldsymbol{\Omega}_0$ is a sparse matrix, we denote the number of nonzero entries in the jth column of $\boldsymbol{\Omega}_0$ as s_j, where $\max_{1\le j\le p} s_j = o(\sqrt{n})$ is a typical assumption made in estimating high-dimensional precision matrices (Cai et al., 2011; Liu and Wang, 2017). The precision matrix $\boldsymbol{\Omega}_0$ can be estimated in the following column-by-column fashion. Denote $\boldsymbol{\Sigma}_0 = \boldsymbol{\Omega}_0^{-1}$. By definition, $\boldsymbol{\Sigma}_0 \mathbf{w}_{i,0} = \boldsymbol{\Sigma}_0 \mathbf{B}_{i,0} \mathbf{w}_{i1,0} = \mathbf{e}_i$ and then $\mathbf{B}_{i,0}^{\mathrm{T}} \boldsymbol{\Sigma}_0 \mathbf{B}_{i,0} \mathbf{w}_{i1,0} = \mathbf{B}_{i,0}^{\mathrm{T}} \mathbf{e}_i$. Thus, $\mathbf{w}_{i1,0} = (\mathbf{B}_{i,0}^{\mathrm{T}} \boldsymbol{\Sigma}_0 \mathbf{B}_{i,0})^{-1} \mathbf{B}_{i,0}^{\mathrm{T}} \mathbf{e}_i$. Under H_0, $\mathbf{B}_{i,0}^{\mathrm{T}} \mathbf{X}_1, \ldots, \mathbf{B}_{i,0}^{\mathrm{T}} \mathbf{X}_n$ are s_i-dimensional independent and identically distributed random vectors with covariance $\mathbf{B}_{i,0}^{\mathrm{T}} \boldsymbol{\Sigma}_0 \mathbf{B}_{i,0}$. Because s_i are of smaller order of $\sqrt{n}$, $\mathbf{B}_{i,0}^{\mathrm{T}} \boldsymbol{\Sigma}_0 \mathbf{B}_{i,0}$ can be consistently estimated by the sample covariance of $\mathbf{B}_{i,0}^{\mathrm{T}} \mathbf{X}_1, \cdots, \mathbf{B}_{i,0}^{\mathrm{T}} \mathbf{X}_n$ given by $\mathbf{B}_{i,0}^{\mathrm{T}} \boldsymbol{S}_n \mathbf{B}_{i,0}$ under H_0. Then, $\hat{\mathbf{w}}_{i1,0} = (\mathbf{B}_{i,0}^{\mathrm{T}} \boldsymbol{S}_n \mathbf{B}_{i,0})^{-1} \mathbf{B}_{i,0}^{\mathrm{T}} \mathbf{e}_i$ and $\hat{\mathbf{w}}_{i,0} = \mathbf{B}_{i,0} \hat{\mathbf{w}}_{i1,0}$ is a consistent estimator of $\mathbf{w}_{i,0}$ under H_0. By assembling $\hat{\mathbf{w}}_{i,0}$ as $\hat{\boldsymbol{\Omega}}_0$, we have a consistent estimator of $\boldsymbol{\Omega}_0$. The technical detail of the preceding argument can be

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

found in Le and Zhong (2022). However, the estimated precision matrix $\hat{\mathbf{\Omega}}_0 = (\hat{\mathbf{w}}_{1,0}, \cdots, \hat{\mathbf{w}}_{p,0})$ may not be symmetric or positive definite. To address this, one can use the perturbation method proposed by Liu and Luo (2015) to ensure that $\hat{\mathbf{\Omega}}_0$ is positive definite, and then symmetrize it by averaging it with its transpose: $(\hat{\mathbf{\Omega}}_0 + \hat{\mathbf{\Omega}}_0^T)/2$. We conducted a simulation study to compare our proposed test using this estimator $\hat{\mathbf{\Omega}}_0$ with its symmetrized and positive definite version. The results of this comparison are presented in Table 5 of the supplemental file.

Based on Lemma 1, we can then estimate $\theta_{ij,0}$ as $\hat{\theta}_{ij,0} = \{\hat{\omega}_{ii,0}s_{jj} + (1 + \hat{\kappa})\delta_{ij}\}/n$ where $\hat{\omega}_{ii,0}$ is the (i,i)th element of $\hat{\mathbf{\Omega}}_0$, s_{jj} is the (j,j)th element of matrix $\boldsymbol{S}_n$, $\hat{\kappa} = \sum_{i=1}^n s_{jj}^{-4}(X_{ij} - \bar{X}_i)^4/(np)$, and $\delta_{ij} = 1$ if $i \neq j$ and $\delta_{ij} = 0$ if $i = j$. If $\mathbf{X}$ is normally distributed, we set $\hat{\kappa} = 0$. Replacing the unknown parameters by their estimators, we construct a test statistic $\hat{D}_n$ using the plug-in estimators $\hat{\mathbf{w}}_{i,0}$,

$$\hat{D}_n = \max_{1 \leq i,j \leq p} \hat{D}_{ij}^2,$$

with $\hat{D}_{ij}^2 = (\mathbf{e}_j^{\mathrm{T}} \boldsymbol{S}_n \hat{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i)^2/\hat{\theta}_{ij,0}$.

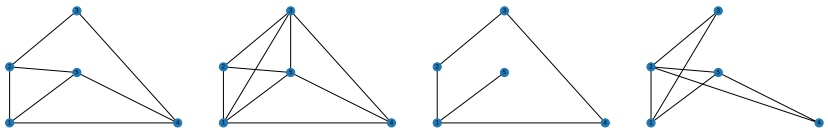
The above test statistic $\hat{D}_n$ is free of tuning and extremely easy to calculate for practical use. These advantages should be compared to those penalized likelihood methods, such as GLASSO (Friedman et al., 2007), for which the choice of tuning parameters is crucial for the performance of the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

resulting estimator. In Section 6.3 of the supplemental file, we compare $\hat{D}_n$ using our proposed estimator $\hat{\mathbf{w}}_{i,0}$ with results based on GLASSO. Our proposed method shows better performance in terms of empirical sizes, power, and computational efficiency. The time complexity of $\hat{D}_n$ with respect to p is $\max(s_0^3 p, p^2)$, where $s_0^3 p$ represents the time complexity to compute all $\hat{D}_{ij}$, and p^2 is the time complexity to compute the maximum operator. See Figure 2 in the simulation study for details on the relationship between estimated computational time and data dimension.

Despite the above advantages, $\hat{D}_n$ does not have much power in rejecting $\mathcal{E}_0$ if $\mathcal{E}^* \subsetneq \mathcal{E}_0$; that is, when $\mathcal{E}^*$ is included in $\mathcal{E}_0$ but they are not equal. For notational convenience, we collect all the included structures in the alternatives H_1 as $H_2 : \mathcal{E}^* \subsetneq \mathcal{E}_0$. Clearly, H_2 is a subset of H_1 and we call the alternatives in H_2 included structures. An example is given in Figure 1 where $\hat{D}_n$ will have no power in rejecting the $\mathcal{E}_0$ in (b), while it does for rejecting the $\mathcal{E}_0$ in (c) or (d). For (b), this is simply because under the null hypothesis that $\mathcal{E} = \mathcal{E}_0$, any reasonable estimator of $\mathbf{\Omega}_0$ of $\mathbf{\Omega}^*$ denoted as $\hat{\mathbf{\Omega}}_0$, including the one discussed above, will asymptotically converge to $\mathbf{\Omega}^*$, making $\mathbf{\Sigma}^* \hat{\mathbf{\Omega}}_0 - \mathbf{I}_p$ very small stochastically.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385



(a) (b) (c) (d)

Figure 1: Different dependence structures: (a) the true graph $\mathcal{E}^*$; (b) $\mathcal{E}_0$ that satisfies $\mathcal{E}^* \subsetneq \mathcal{E}_0$ (included structure); (c) $\mathcal{E}_0$ that satisfies $\mathcal{E}_0 \subsetneq \mathcal{E}^*$; (d) $\mathcal{E}_0$ that is not nested within or outside $\mathcal{E}^*$.

2.1 A novel consistency-empowered test statistic

When $\mathcal{E}^* \subsetneq \mathcal{E}_0$ as illustrated in Figure 1, we know that its compatible estimator $\hat{\mathbf{\Omega}}_0$ will be close to $\mathbf{\Omega}^*$ loosely speaking. Thus, if $(i,j) \in \mathcal{E}_0$ but $(i,j) \notin \mathcal{E}^*$, $\hat{\omega}_{ij,0}$ will be close to zero. For the test to have power, we need to offset the effect of those small estimates. Our idea is to augment those small estimates with a constant that is just large enough for us to reject the null. Thus, in a certain sense, we are amplifying those small noises as a means to empower the consistency of a new test statistic. Of course, how close is close to zero for a small noise should be gauged against its standard error, which motivates the development of the following consistency-empowered test statistic.

Let $\hat{\omega}_{i1,0}^{(j)}$ be the jth component of $\hat{\mathbf{w}}_{i1,0}$ with the associated standard error $\sigma_{i1,0}^{(j)}$, where $\hat{\mathbf{w}}_{i1,0}$ is defined in the previous section. Let $\hat{\sigma}_{i1,0}^{(j)}$ be a

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

consistent estimator of $\sigma_{i1,0}^{(j)}$ which will be defined shortly. Define $\tilde{\mathbf{w}}_{i1,0} = (\tilde{\omega}_{i1,0}^{(1)}, \ldots, \tilde{\omega}_{i1,0}^{(s_i)})^{\mathrm{T}}$ where

$$\tilde{\mathbf{w}}_{i1,0}^{(j)} = \hat{\mathbf{w}}_{i1,0}^{(j)} + \Delta_{i1}^{(j)},$$

with $\Delta_{i1}^{(j)} = C_n I(|\hat{\omega}_{i1,0}^{(j)}|/\hat{\sigma}_{i1,0}^{(j)} \leq \delta_n)$. Here $C_n \neq 0$ and δ_n are tuning parameters which will be discussed in the next section. Clearly, what this procedure does is to add a constant to those elements of $\hat{\mathbf{w}}_{i1,0}$ that are stochastically small. Or put differently, it simply amplifies the noise, as opposed to the usual notion of filtering out noises for better estimation accuracy.

Recall the definition of $\mathbf{B}_{i,0}$ in the last section. Let $\tilde{\mathbf{w}}_{i,0} = \mathbf{B}_{i,0}\tilde{\mathbf{w}}_{i1,0}$ and $\Delta_i = \mathbf{B}_{i,0}\Delta_{i1}$ where $\Delta_{i1} = (\Delta_{i1}^{(1)}, \ldots, \Delta_{i1}^{(s_i)})^{\mathrm{T}}$. Our proposed consistency-empowered test statistic is then

$$\tilde{D}_n = \max_{1 \leq i,j \leq p} (\mathbf{e}_j^{\mathrm{T}} \boldsymbol{S}_n \tilde{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i)^2 / \hat{\theta}_{ij,0} = \max_{1 \leq i,j \leq p} \tilde{D}_{ij}^2,$$

where $\tilde{D}_{ij}^2 = (\mathbf{e}_j^{\mathrm{T}} \boldsymbol{S}_n \tilde{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i)^2 / \hat{\theta}_{ij,0}$. The consistency-empowered estimator $\tilde{\mathbf{w}}_{i1,0}^{(j)}$ aims to ensure that the non-zero components $\mathbf{w}_{i1,0}^{(j)}$ are indeed estimated by some non-zero estimators. Interestingly, the form of the consistency-empowered estimator $\tilde{\mathbf{w}}_{i1,0}^{(j)}$ is an opposite of the conventional threshold tests (Fan, 1996), where small components under some threshold are set to zeros. The proposed test statistic $\tilde{D}_n$ is also different from the power-enhanced test statistic proposed by Fan et al. (2015) which includes a combination of a test

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

statistic that has an asymptotically correct size and a power enhancement component. Our proposed $\tilde{D}_n$ is not a combination of two components. We modify the estimator $\hat{\mathbf{w}}_{i1,0}^{(j)}$ to ensure the consistency of the proposed test for all types of null hypotheses. Having said this, we point out that the test statistic $\hat{D}_n$ without consistency empowerment is powerless to test those null hypotheses in which the true graph is nested within the graph of the null, regardless of the sample size.

We now discuss a consistent estimator of $\sigma_{i1,0}^{(j)}$ required in $\Delta_{i1}^{(j)}$. Any non-zero element $\hat{\omega}_{i1,0}^{(j)}$ of $\hat{\mathbf{w}}_{i1,0}$ corresponds to $\hat{\omega}_{ik,0}$ $(k = 1, \ldots, p)$, (i,k)-th component of $\hat{\mathbf{\Omega}}_0$, such that $\hat{\omega}_{i1,0}^{(j)} = \hat{\omega}_{ik,0}$. Le and Zhong (2022) established that $\hat{\omega}_{i1,0}^{(j)}$ is asymptotically normal, in the sense that

$$\sqrt{n}(\hat{\omega}_{i1,0}^{(j)} - \omega_{i1,0}^{(j)}) = \sqrt{n}(\hat{\omega}_{ik,0} - \omega_{ik,0}) \to N(0, h_{ik}) \tag{2.2}$$

in distribution, where $h_{ik} = \omega_{ii}^* \omega_{kk}^* + \omega_{ik}^*$. Thus, a consistent estimator of $\sigma_{i1,0}^{(j)}$ is simply $\hat{\sigma}_{i1,0}^{(j)} = \sqrt{(\hat{\omega}_{ii,0}\hat{\omega}_{kk,0} + \hat{\omega}_{ik,0}^2)}/\sqrt{n}$.

3. Main Results

In this section, we study the asymptotic distributions of $\hat{D}_n$ and the consistency-empowered test statistic $\tilde{D}_n$ under the general distributional assumptions (D1) and (D2) specified in the supplemental material, as $p \to \infty$ and $n \to \infty$. Although the asymptotic results in Theorems 1-4 are applicable

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

to non-Gaussian random variables $\mathbf{X}$, our discussion following each theorem will primarily focus on Gaussian graphical models.. We assume the following regularity conditions.

(C1) There exist constants $C_1, C_2 > 0$ such that $||\mathbf{\Sigma}^*||_1 \leq C_1$ and $||\mathbf{\Omega}^*||_1 \leq C_2$, where $||\mathbf{M}||_1 = \max_{1\leq j\leq n} \sum_{i=1}^m |M_{ij}|$ for any $m \times n$ matrix $\mathbf{M} = (M_{ij})$;

(C2) $s_0\sqrt{(\log p/n)} = o(1)$, where $s_0 = \max_{1\leq j\leq p} s_j$.

These two conditions are commonly employed in the literature (e.g., Zhou et al. (2011); Liu and Luo (2015)). Many commonly assumed precision matrix structures such as banded and factor models satisfy Condition (C1). See the supplementary materials for details.

A node is called *isolated* if it does not connect with any other nodes. That is, node i in $\mathcal{E}^*$ or the support for the i-th variable in supp($\mathbf{\Omega}^*$) is isolated if and only if $\omega_{ij}^* = 0$ for all $j \neq i$. We have the following results on the asymptotic distribution of $\hat{D}_n$.

Theorem 1. *Under conditions (C1)-(C2) and Assumptions (D1) and (D2) in the supplemental file, if supp*($\mathbf{\Omega}^*$) *has k isolated variables where* $\lim_{p\to\infty} k/p = \beta, 0 \leq \beta < 1$, *then under the null* H_0^*,

$$pr\{\hat{D}_n - 4\log p + \log(\log p) \leq t\} \to \exp\{-\exp(-t/2)/\sqrt{(2\gamma\pi)}\}$$

where $\gamma = (1 - \beta^2/2)^{-2}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

Interestingly, the asymptotic distribution of $\hat{D}_n$ depends on k, the number of isolated nodes in $\mathcal{E}^*$. From Theorem 1, if the number of isolated nodes k is of a smaller order of the number of variables p as $k = o(p)$, then $\hat{D}_n$ converges to the following Gumbel distribution

$$\text{pr}\{\hat{D}_n - 4\log p + \log(\log p) \le t\} \to \exp\{-\exp(-t/2)/\sqrt{(2\pi)}\}. \qquad (3.1)$$

Example 1 and Example 2 below further illustrate Theorem 1.

Example 1. Assume that $\mathcal{E}^*$ has a Toeplitz structure $\mathcal{E} = \{(i,j), |i-j| \le s_0\}$. The number of the isolated node in $\mathcal{E}^*$ is 0 and hence $\gamma = 1$. Then under the null $H_0 : \mathcal{E}_0 = \mathcal{E}$, we have $\text{pr}\{\hat{D}_n - 4\log p + \log(\log p) \le t\} \to \exp\{-\exp(-t/2)/\sqrt{(2\pi)}\}$. The limiting distribution of $\hat{D}_n$ is Gumbel$(-\log 2\pi, 2)$.

Example 2. Assume that $\mathcal{E}^*$ follows a factor model structure with $\mathbf{\Omega}^* = \mathbf{I}_p + \boldsymbol{u}_1\boldsymbol{u}_1^T$, where $\boldsymbol{u}_1 = (1,1,1,0,\ldots,0) \in \mathbb{R}^p$. The limiting distribution of $\hat{D}_n$ satisfies $\text{pr}\{\hat{D}_n - 4\log p + \log(\log p) \le t\} \to \exp\{-\exp(-t/2)/\sqrt{(8\pi)}\}$. In this case, the limiting distribution of $\hat{D}_n$ is Gumbel$(-\log 8\pi, 2)$.

As discussed in Section 2.1, $\hat{D}_n$ cannot detect the alternative hypothesis under which $\mathcal{E}^* \subsetneq \mathcal{E}_0$ or $\text{supp}(\mathbf{\Omega}^*) \subsetneq \text{supp}(\mathbf{\Omega}_0)$, that is, when $\mathcal{E}_0 \neq \mathcal{E}^*$ but $\mathcal{E}_0$ includes the true network structure $\mathcal{E}^*$ or $\text{supp}(\mathbf{\Omega}^*) \neq \text{supp}(\mathbf{\Omega}_0)$ but $\text{supp}(\mathbf{\Omega}_0)$ includes $\text{supp}(\mathbf{\Omega}^*)$. In fact, for testing $H_0 : \mathcal{E}^* = \mathcal{E}_0$ vs $H_2 : \mathcal{E}^* \subsetneq \mathcal{E}_0$, which is equivalent to $H_0^* : \text{supp}(\mathbf{\Omega}^*) = \text{supp}(\mathbf{\Omega}_0)$ vs $H_2^* : \text{supp}(\mathbf{\Omega}^*) \subsetneq \text{supp}(\mathbf{\Omega}_0)$,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

if H_2 is true or H_2^* is true, the the test statistic $\hat{D}_n$ converges to the same distribution as the test statistic under H_0 or H_0^*. The following Theorem states this fact.

Theorem 2. *Under the alternative hypothesis* $H_2^* : supp(\mathbf{\Omega}^*) \subsetneq supp(\mathbf{\Omega}_0)$ *when* $supp(\mathbf{\Omega}_0)$ *satisfies the sparsity assumption (C2) and Assumptions (D1) and (D2) in the supplemental file, the test statistic* $\hat{D}_n$ *converges to the same limiting distribution specified in Theorem 1.*

The results in Theorem 2 imply that the test based on $\hat{D}_n$ is not consistent for alternatives defined by $H_2 : \mathcal{E}^* \subsetneq \mathcal{E}_0$. We now show that the modified test $\tilde{D}_n$ is universally consistent for all types of alternatives when the tuning parameters C_n and δ_n involved in its definition are chosen appropriately. A strategy is to choose these two parameters such that $\tilde{D}_n$ and $\hat{D}_n$ have the same asymptotic distribution under H_0, while the test based on $\tilde{D}_n$ can reject network structures satisfying $\mathcal{E}^* = \mathcal{E}_0$ with probability one when $\mathcal{E}^* \subsetneq \mathcal{E}_0$. Under H_0 where $\mathcal{E} = \mathcal{E}_0$, $\omega_{i1,0}^{(j)}$ $(j = 1, \ldots, s_i)$ are all non-zeros. However, if $\mathcal{E}^* \subset \mathcal{E}_0$ but $\mathcal{E}^* \neq \mathcal{E}_0$, $\omega_{i1,0}^{(j)}$ $(j = 1, \ldots, s_i)$ are supposed to be non-zeros because of the specification of $\mathbf{\Omega}_0$, but some of them will be estimated as zeros (in asymptotic sense). Based on the asymptotic normality in (2.2), we have $\hat{\omega}_{i1,0}^{(j)} = \omega_{i1,0}^{(j)} + O_p(1/\sqrt{n})$. This result holds when the true value of $\omega_{i1,0}^{(j)}$ is zero or non-zero. If $\omega_{i1,0}^{(j)} \neq 0$, then $\hat{\omega}_{i1,0}^{(j)}/\sigma_{i1,0}^{(j)} = O_p(\sqrt{n})$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

where $\sigma_{i1,0}^{(j)}$ is defined in equation (2.2). If $\omega_{i1,0}^{(j)} = 0$, then $\hat{\omega}_{i1,0}^{(j)}/\sigma_{i1,0}^{(j)} = O_p(1)$. Based on these observations, we may choose $C_n = C\sqrt{\log(p)}$ for some $C > 0$ and $\delta_n = \sqrt{\log(n)}$ so that $\tilde{D}_n$ and $\hat{D}_n$ have the same asymptotic distribution under H_0. The details can be found in the proof of the following theorem.

Theorem 3. *Under conditions (C1)-(C2), and Assumptions (D1) and (D2) in the supplemental file, if supp($\mathbf{\Omega}^*$) has k isolated variables where $\lim_{p\to\infty} k/p = \beta$ for some $0 \le \beta < 1$, $\delta_n = \sqrt{\log(n)}$ and $C_n = C\sqrt{\log(p)}$ for some constant $C > 0$, then under the null H_0^*,*

$$pr\{\tilde{D}_n - 4\log p + \log(\log p) \le t\} \to \exp\{-\exp(-t/2)/\sqrt{(2\gamma\pi)}\},$$

where $\gamma = (1 - \beta^2/2)^{-2}$.

Based on Theorem 3, we can use the same cutoff as that used for $\hat{D}_n$ to construct the test. If the tuning parameters δ_n and C_n are selected at the levels specified by Theorem 3, the test $\tilde{D}_n$ can maintain the type I error asymptotically. Furthermore, the following Theorem shows that $\tilde{D}_n$ is universally consistent for all types of fixed alternatives. In particular, it rejects any network structure in H_2 satisfying $\mathcal{E}_0 \supsetneq \mathcal{E}^*$ with probability one.

Theorem 4. *If we choose $\delta_n = \sqrt{\log(n)}$ and $C_n = C\sqrt{\log(p)}$ for some $C > \max_{i,j} 4(\omega_{ii}^*\sigma_{jj}^* + 1)/(\sigma_{ii}^*\sigma_{jj}^* + 2\sigma_{ij}^{*2})$, under conditions (C1)-(C2) and Assumptions (D1) and (D2) in the supplemental file, the consistency-empowered*

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

test based on $\tilde{D}_n$ is universally consistent for all types of fixed alternatives.

By comparing Theorems 2 and 4, we observe that the modified test based on $\tilde{D}_n$ is more powerful than the test based on $\hat{D}_n$ in the sense that $\tilde{D}_n$ is consistent for alternatives in H_2 but $\hat{D}_n$ is not. More importantly, the results in Theorem 4 formally establish that the proposed test $\tilde{D}_n$ is universally consistent for all types of fixed alternatives. Theorem 4 provides us some guidelines on the choice of C. The magnitude of C could be chosen by $\max_{i,j} 4(\hat{\omega}_{ii}\hat{\sigma}_{jj}+1)/(\hat{\sigma}_{ii}\hat{\sigma}_{jj}+2\hat{\sigma}_{ij}^2)$ where $\hat{\omega}_{ii}$ and $\hat{\sigma}_{ij}$ are, respectively, estimators of ω_{ii}^* and σ_{ii}^*. We propose a data-driven procedure for selecting the tuning parameters δ_n and C_n. Due to space constraints, the detailed procedure and a small simulation study are provided in Section 7 of the supplementary material.

4. Simulation Studies

4.1 Numerical performance of the test statistic $\hat{D}_n$

We perform numerical study to evaluate the finite sample performance of the proposed test statistic $\hat{D}_n$ in terms of its size and power properties. We generate n i.i.d. multivariate normally distributed p-dimensional random vectors with mean vector 0 and covariance matrix $\mathbf{\Sigma}^*$ with its corresponding graph $\mathcal{E}^*$ admitting a banded structure such that $\mathcal{E}^* = \{(i,j) : |i-j| < s_0\}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

In Section 6.1 of the supplemental file, we present simulation studies to assess the robustness of the proposed methods with respect to the normality assumption. Our findings indicate that the proposed tests perform reasonably well under non-Gaussian conditions.

Let $\mathbf{\Omega}^* = \mathbf{\Sigma}^{*-1} = (\omega^*_{ij})_{p\times p}$ be the precision matrix. Because different precision matrices can correspond to the same underlying graph, we specify two precision matrices to examine the performance of the proposed test statistic. For the first precision matrix, we set it as banded with its non-zero components decaying at an exponential rate away from its diagonals. More specifically, we set $\omega^*_{ij} = 0.6^{-|i-j|}$ for $|i-j| < s_0$ and $\omega^*_{ij} = 0$ otherwise. For the other precision matrix, we again set it as banded with its non-zero components decaying at the polynomial rate, that is, $\omega^*_{ij} = (1+|i-j|)^{-2}$ for $|i-j| < s_0$ and $\omega^*_{ij} = 0$ otherwise. We consider two different sparsity levels as $s_0 = 4$ or 6. To evaluate the performance of the proposed tests under various scenarios for $\mathbf{\Omega}^*$, we also conduct simulation studies on data with both a sparse and random underlying precision matrix, as well as a precision matrix with small signals. These studies are detailed in Section 6.2 of the supplemental material.

To evaluate the empirical size and power of the proposed test, we consider various specifications of $\mathcal{E}_0$, the structure specified in the null hypothesis. For

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

evaluating the empirical size, we consider $\mathcal{E}_0 = \mathcal{E}^*$. To evaluate the power of the proposed test, we consider the following four different specifications of $\mathcal{E}_0$.

1) (Isolated structure) Set $\mathcal{E}_0 = \mathcal{E}_{0,1} = \{(i,j) : i = j\}$. All the nodes are isolated.

2) (Nested structure) Set $\mathcal{E}_0 = \mathcal{E}_{0,2} = \{(i,j) : |i-j| < 3\}$. This structure is nested in the true network structure $\mathcal{E}^*$.

3) (1-diff: structure with edges to node 1 different) Set $\mathcal{E}_0 = \mathcal{E}_1^* \cup \mathcal{E}_{0,3}$, where $\mathcal{E}_1^* = \mathcal{E}^*$ on the set of edges $\{(i,j), i,j \neq 1\}$, and $\mathcal{E}_{0,3} = \{(1,3),(1,7),(1,8),(1,9)\}$ are edges connected with node 1.

4) (2-diff: structure with edges to 2 nodes different) Set $\mathcal{E}_0 = \mathcal{E}_2^* \cup \mathcal{E}_{0,4}$, where $\mathcal{E}_2^* = \mathcal{E}^*$ on the set of edges $\{(i,j), i,j \neq 1,2\}$, and $\mathcal{E}_{0,4} = \{(1,3),(1,7),(1,8),(1,9),(2,4),(2,9),(2,12)\}$ are edges connected to nodes 1 and 2.

To understand the effect of sample size and data dimension, we choose two different sample sizes $n = 300$ and $n = 1000$. For each sample size, data dimension is changed by setting p/n at three different values $0.5, 1$, and 2. Because the true precision matrix $\mathbf{\Omega}^*$ specified above satisfies Conditions (C1)-(C2), we applied the results in Theorem 1. More specifically, we rejected the hypothesis if the test statistic values $\hat{D}_n$ are greater than the $4\log p - \log(\log p) + \text{Gumbel}_{.95}(-\log 2\pi, 2)$, where $\text{Gumbel}_{.95}(-\log 2\pi, 2)$ is

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

the 95 % quantile value of the Gumbel distribution with location parameter $-\log 2\pi$ and scale parameter 2. Simulation results are reported based on 500 simulation replications.

Table 1 reports the empirical sizes and power of the proposed test statistic $\hat{D}_n$ for testing different structures $\mathcal{E}_0$ specified in the above (1)-(4). It can be seen that our proposed test controls type I error rate well at the nominal level of 0.05 under various settings. The proposed test statistic is consistent as the power of the tests are one in many scenarios. Based on the pattern of empirical power, we see that the power of the proposed test $\hat{D}_n$ increases as n increases or p decreases. Table 1 also shows that sparsity level has some impact on the power of the test statistic, the increasing of s_0 leads to a decreasing power.

Table 2 summarizes the empirical size and power of the proposed test $\hat{D}_n$ under the polynomial rate decay structure. We see that its patterns are similar to that in Table 1 where the empirical power increases as sample size increases, and decreases as p or s_0 increases. We observe that in this case the power is not as high as those in Table 1 where the precision matrix decays at an exponential rate in Table 1. This is also something expected because the signals of the precision matrix in Table 2 is weaker than the signals in the previous example. For example, when $s_0 = 4$, the non-zeros in the first

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

Table 1: Type I error and power of the proposed test statistic $\hat{D}_n$ under different alternatives when the precision matrix has banded structure and decays at an exponential rate.

			Empirical	Power of the Test $\hat{D}_n$			
s_0	n	p/n	Size	Isolated	Nested	1-diff	2-diff
4	300	0.5	0.034	1.000	1.000	1.000	1.000
		1	0.042	1.000	1.000	1.000	1.000
		2	0.030	1.000	1.000	1.000	1.000
	1000	0.5	0.038	1.000	1.000	1.000	1.000
		1	0.032	1.000	1.000	1.000	1.000
		2	0.044	1.000	1.000	1.000	1.000
6	300	0.5	0.026	1.000	0.824	1.000	1.000
		1	0.032	1.000	0.766	1.000	1.000
		2	0.028	1.000	0.648	1.000	1.000
	1000	0.5	0.032	1.000	1.000	1.000	1.000
		1	0.046	1.000	1.000	1.000	1.000
		2	0.036	1.000	1.000	1.000	1.000

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

Table 2: Type I error rate and power of the proposed test statistic $\hat{D}_n$ under different alternatives where the precision matrix has banded structure and decays at a polynomial rate.

			Empirical	Power of the Test $\hat{D}_n$			
s_0	n	p/n	Size	Isolated	Nested	1-diff	2-diff
4	300	0.5	0.022	1.000	0.034	0.348	0.468
		1	0.034	1.000	0.038	0.226	0.336
		2	0.024	1.000	0.024	0.170	0.248
	1000	0.5	0.034	1.000	0.274	0.998	1.000
		1	0.046	1.000	0.210	0.994	1.000
		2	0.038	1.000	0.204	0.988	1.000
6	300	0.5	0.032	1.000	0.034	0.320	0.474
		1	0.030	1.000	0.028	0.248	0.328
		2	0.024	1.000	0.024	0.164	0.224
	1000	0.5	0.046	1.000	0.130	0.992	1.000
		1	0.032	1.000	0.106	0.994	1.000
		2	0.024	1.000	0.084	0.994	1.000

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

column of polynomial decayed precision matrix $\mathbf{\Omega}^*$ is $(1, 0.6, 0.36, 0.216)^{\mathrm{T}}$ while the first column non-zeros are $(1, 0.25, 0.11, 0.06)^{\mathrm{T}}$ in the exponentially decayed $\mathbf{\Omega}^*$.

4.2 Numerical comparison of $\hat{D}_n$ and $\tilde{D}_n$

In this simulation study, we evaluate the finite sample performance of $\hat{D}_n$ and $\tilde{D}_n$ in terms of empirical sizes and powers in detecting the nested network and the included network structures. Similar to Simulation Settings I, we generate n IID multivariate normally distributed p-dimensional random vectors with mean vector 0 and covariance matrix $\mathbf{\Sigma}^*$. The corresponding precision matrix is $\mathbf{\Omega}^* = \mathbf{\Sigma}^{*-1} = (\omega^*_{ij})_{p\times p}$ where $\omega^*_{ij} = 0.6^{-|i-j|}$ for $|i-j| < s_0$ and $\omega^*_{ij} = 0$ otherwise. For evaluating the empirical sizes, we consider $\mathcal{E}_0 = \mathcal{E}^*$. We consider the following two specified structures hypotheses in the simulation

5) (Nested structure) Set $\mathcal{E}_0 = \mathcal{E}_{0,5} = \{(i,j) : |i-j| < s_0 - 1\}$. The structure $\mathcal{E}_0$ is nested in the true structure $\mathcal{E}^*$.

6) (Included structure) Set $\mathcal{E}_0 = \mathcal{E}_{0,6} = \{(i,j) : |i-j| < s_0 + 1\}$. The structure $\mathcal{E}_0$ includes in the true structure $\mathcal{E}^*$.

Table 3 summarizes the empirical sizes and powers of the tests based on $\hat{D}_n$ and $\tilde{D}_n$. Table 3 demonstrates both tests have the similar power

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

in rejecting the nested structure and control the type 1 error rate. The modified test statistics version $\tilde{D}_n$ has the ability to reject the pre-specified network structures that include the true network structure, while the test statistic $\hat{D}_n$ loses power for included networks. We chose $\delta_n = \sqrt{\log(n)}$ and $C_n = 0.05$ for the test statistics $\tilde{D}_n$ in our simulation studies. Tables 6 and 7 in the supplemental file demonstrate that the proposed test is robust to different choices of C_n; the empirical size and power remain consistent across various values of C_n when δ_n is fixed. Additionally, the proposed tests exhibit similar empirical sizes and power for $k = 2, 4$ across all choices of C_n. This simulation study demonstrates that the proposed tests perform reasonably well with the suggested order of $\delta_n = \sqrt{\log n}$.

To illustrate the computational complexity of $\tilde{D}_n$, Figure 2 presents the average running time in seconds for $\tilde{D}_n$ versus the data dimension, with p on the x-axis and the square root of the running time on the y-axis. We observe a linear relationship between p and the square root of the running time, indicating that the computational time is approximately of the order p^2 with respect to the data dimension.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

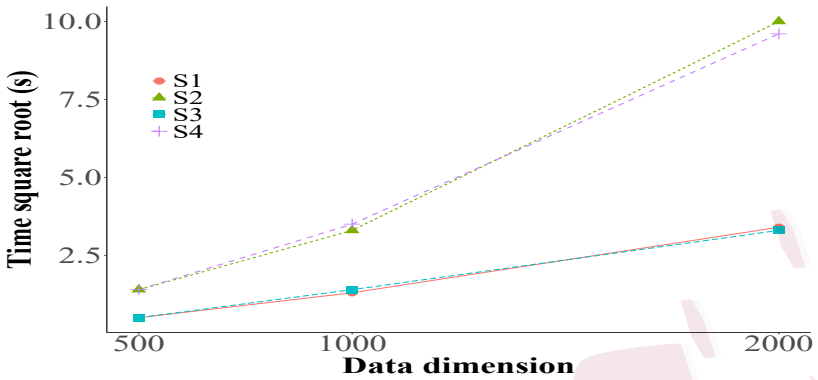


Figure 2: The square root of the running time for the test statistic $\tilde{D}_n$ versus the data dimension for different combinations of n, p, and s_0 is shown for the following scenarios: S_1: $(n, s_0) = (500, 4)$; S_2: $(n, s_0) = (1000, 4)$; S_3: $(n, s_0) = (500, 6)$; and S_4: $(n, s_0) = (1000, 6)$.

5. Real Data Analysis

We illustrate the use of the proposed test statistics for identifying the structure of a graphical model by applying them to a correlated data analysis. Towards this, we examined a COVID-19 dataset provided by The New York Times (The New York Times, 2021) that is publicly available on `https://github.com/nytimes/covid-19-data`. The data set includes daily confirmed COVID-19 cases observed over 51 states of the U.S. from January 1, 2021 to December 31, 2021. We aggregated the data on a weekly basis such that the data contain 52 weekly confirmed cases in thousands

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

Table 3: Type 1 error and empirical power of the test statistics $\hat{D}_n$ and $\tilde{D}_n$ for both nested and included structures

			$\hat{D}_n$				$\tilde{D}_n$			
				Power		Running		Power		Running
s_0	n	p/n	Size	Nested	Included	Time	Size	Nested	Included	Time
4	500	0.50	0.020	1.000	0.030	0.16	0.020	1.000	0.970	0.17
		1.00	0.050	1.000	0.050	1.08	0.050	1.000	0.990	1.08
		2.00	0.050	1.000	0.040	6.92	0.050	1.000	0.990	6.96
	1000	0.50	0.030	1.000	0.040	1.18	0.030	1.000	1.000	1.19
		1.00	0.030	1.000	0.020	6.62	0.030	1.000	1.000	6.64
		2.00	0.010	1.000	0.010	59.62	0.010	1.000	1.000	59.63
6	500	0.50	0.000	0.110	0.000	0.17	0.010	0.160	0.520	0.17
		1.00	0.030	0.060	0.030	1.10	0.030	0.100	0.690	1.10
		2.00	0.010	0.070	0.000	6.62	0.020	0.090	0.590	6.64
	1000	0.50	0.050	0.670	0.050	1.23	0.040	0.690	1.000	1.24
		1.00	0.050	0.570	0.040	7.15	0.050	0.580	1.000	7.17
		2.00	0.010	0.450	0.010	55.15	0.020	0.460	1.000	55.17

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

from 51 states, which is denoted as a matrix of size $51\times$ 52. Our interest was to understand how the numbers of COVID cases depend on geographical locations. Towards this, we coded three dummy variables according to whether a state is in the North East, West, Mid West, or the South. The following linear regression was postulated

$$E(y_{ij}|x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}, (i = 1, \ldots, 51; j = 1, \ldots, 52), \quad (5.1)$$

where x_{i1} is an indicator variable whether the state is in the North East, x_{i2} is for the Mid West, and x_{i3} is for the West. Denote $\boldsymbol{Y_i} = (y_{i1}, \ldots, y_{i52})^{\mathrm{T}}$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^{\mathrm{T}}$, $\boldsymbol{X_i} = \mathbf{1} \otimes (1, x_{i1}, x_{i2}, x_{i3})$, where $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}$ is a 52×1 matrix, $\otimes$ is the Kronecker product, and $\boldsymbol{X_i}$ is a 52×4 matrix. Since the components of $\boldsymbol{Y_i}$ are correlated, we applied the method of generalized estimation equations (Liang & Zeger, 1986) for estimating $\boldsymbol{\beta}$ by incorporating the correlation structure of $\boldsymbol{Y_i}$. That is, we estimate $\boldsymbol{\beta}$ by solving

$$\sum_{i=1}^{51} \boldsymbol{X_i}^{\mathrm{T}} \boldsymbol{V}^{-1} (\boldsymbol{Y_i} - \boldsymbol{X_i}\boldsymbol{\beta}) = 0, \quad (5.2)$$

where $\boldsymbol{V}$ is the so-called working covariance matrix. It is known that correct specification of $\boldsymbol{V}$ improves the estimation efficiency of the resulting estimator.

To choose an appropriate graph corresponding to $\boldsymbol{\Omega} = \boldsymbol{V}^{-1}$, first, we estimate the underlying graph $\mathcal{E}^*$ of $\mathbf{Y}$ using the TIGER approach

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

(Liu and Wang, 2017) and the GLASSO method (Friedman et al., 2019). The heatmaps of these estimated graphs are provided in Section 9 of the supplemental file. Both methods suggest that either a banded structure or a block diagonal structure may be reasonable for this dataset. To this end, we formally conducted the proposed tests to determine if one of these specified graphical structures fits the data well.

a) (Isolated structure) Set $\mathcal{E}_0 = \mathcal{E}_{0,1} = \{(i,j), i = j\}$.

b) (Banded structure with bandwidth 3, denoted as Band(3)) Set $\mathcal{E}_0 = \mathcal{E}_{0,2} = \{(i,j), |i-j| < 3\}$.

c) (Diagonal blocks structure, denoted as Block(4)) Set $\mathcal{E}_0 = \mathcal{E}_{0,3}$, where $\mathcal{E}_{0,3} = \cup_{k=1}^{13}\{(i,j), 4(k-1)+1 \le i,j \le 4(k-1)+4\}$.

We applied our proposed methods to test the above hypothetical structures. Since the number of isolated nodes of the true structure is unknown, we chose $\gamma = 1$ for the limiting distribution in Theorem 1 so that our test is conservative because we only reject the null hypothesis if our test statistic value is large enough. The test statistic values and its corresponding p-values (in parentheses) for testing Isolated, Band(3), and Block(4) are, respectively, 49.57(< 0.0001), 17.61(0.08), and 16.32(0.14). Therefore we reject the null hypothesis that the true structure is the Isolated structure with 95% confidence. However, we cannot reject the null hypothesis that

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

the true structure is the Band(3) network or the Block(4) network at the 95% confidence level.

Table 4: Estimated coefficients parameters under three different pre-specified structures, standard errors of the estimated parameters in the parentheses, and * denotes p-value less than 0.05.

Coefficients	Isolated	Band(3)	Block(4)
β_1	-0.78 (1.59)	0.64 (0.73)	-0.32 (0.25)
β_2	-0.68 (1.30)	1.01 (0.80)	-0.01 (0.32)
β_3	-0.85 (1.47)	1.91 (0.66)*	0.23 (0.24)

We then used these three pre-specified structures to obtain the estimating coefficients for the model (5.1). Table 4 reports the estimated results for model (5.1), including the estimated coefficients, their standard errors, and their statistical significance (p-value less than 0.05). When using the pre-specified Band(3) structure, there are significant differences in COVID cases between the West and the South. But, under the Isolated structure and Block(4) structure, all the coefficients β_1, β_2 and β_3 are not significant which indicates that there is no significant difference in COVID cases among the four regions in the U.S. We also notice that, the estimated standard errors

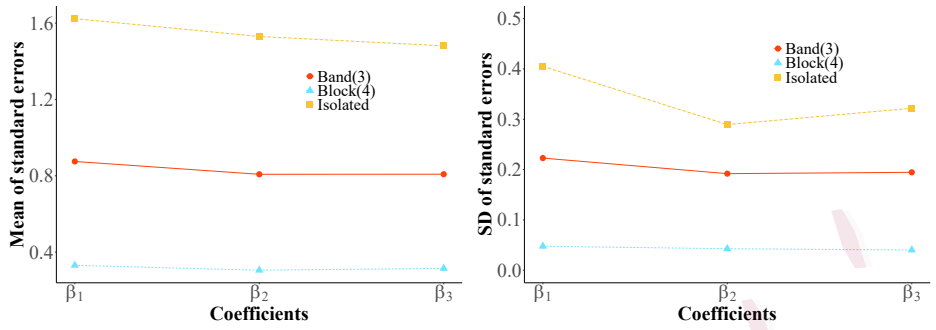
Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

under the Band(3) and Block(4) pre-specified network are much smaller the Isolated structure. These results are consistent with our proposed test statistics since they suggest that the Block(4) and Band(3) structure fit the data well, but not the Isolated structure. This suggests that the proposed test statistics can be used as a powerful tool to identify a good pre-specified structure for further analysis.

Finally, we applied a bootstrap method to further evaluate the standard errors of coefficient estimators, and compared the efficiency gain in terms of the standard errors when using different pre-specified network structures. More specifically, we subsampled 40 states from 51 states without replacement for 100 times. At each time we used subsampled data in the GEE equation (5.2) to estimate the coefficients. Note that here to increase the stability of the procedure, we reuse the estimated precision matrix $\boldsymbol{V}^{-1}$ based on the data from all 51 states. At the i-th replication, we denote the corresponding standard errors of each coefficient by $(\mathrm{Sd}_{1,i}, \mathrm{Sd}_{2,i}, \mathrm{Sd}_{3,i})$. To evaluate the variability of standard errors from the subsampling process, we then calculate the corresponding means and standard deviations of $(\mathrm{Sd}_{1,i}, \mathrm{Sd}_{2,i}, \mathrm{Sd}_{3,i})$, for $i = 1, \cdots, 100$ as $\mathrm{AVE}_j = \sum_{i=1}^{100} \mathrm{Sd}_{j,i}/100, \mathrm{SD}_j = \sqrt{\left\{ \sum_{i=1}^{100} (\mathrm{Sd}_{j,i} - \mathrm{AVE}_j)^2/100 \right\}}$ for $j = 1, 2, 3$.

Figure 3 shows the mean and standard deviation of the coefficients

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385



(a) (b)

Figure 3: Average (a) and standard error (b) of absolute prediction errors of three pre-specified graphic structures: Isolated, Band(3), and Block(4).

obtained from the above subsampling procedure. Both panels 2(a) and 2(b) of the figure demonstrate that the standard error obtained from the pre-specified Block(4) structure is the smallest, followed by Band(3), and Isolated network. The result again agrees with our test statistics obtained and confirms that choosing a good pre-specified graphical structure is essential for efficiency gain when using the GEE method. Therefore, the proposed test statistics can help to select a reliable pre-specified graphic structure.

6. Discussion

In this paper, we have developed new methods for testing pre-specified graphs that are available as prior information in some applications. However, in

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

situations where a pre-specified graph is not available, there may be interest in determining whether a certain family of graphical structures, indexed by certain parameters, is appropriate for the observed data. To address this, we have extended our method to assess the following goodness-of-fit hypothesis: $H_0 : \mathcal{E}^* \in \mathcal{E}_0(\gamma)$ vs. $H_1 : \mathcal{E}^* \notin \mathcal{E}_0(\gamma)$, where $\mathcal{E}_0(\gamma)$ represents a family of graphical structures indexed by parameters γ, and γ is unknown. For example, $\mathcal{E}_0(\gamma)$ could represent a banded structure with an unknown bandwidth γ. In this framework, it is not necessary to specify a single graph structure; instead, one only needs to specify a family of graphical structures. This extension is discussed in Section 8 of the supplemental file.

Supplementary Material

All technical proofs, examples of structures satisfying condition (C1), additional simulation results, a data-driven tuning parameter selection procedure, and further details on the real data analysis are provided in the supplementary materials. The source codes and real data sets are included in a Github link: `https://github.com/leminhthien2011/UniversalConsistentTestforGraphsofGGM`.

Acknowledgments

The research was partially supported by NSF grants DMS-1462156 and FRG-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

2152070. The authors thank Editor Prof. Judy Wang, the Associate Editor, and the two referees for their constructive comments, which significantly improved the paper.

References

Bai, Z. and H. Saranadasa (1996). Effect of high dimension: By an example of a two sample problem. *Statistica Sinica 6*(2), 311–329.

Cai, T., W. Liu, and X. Luo (2011). A constrained l1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association 106*(494), 594–607.

Chen, S. X., L.-X. Zhang, and P.-S. Zhong (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association 105*(490), 810–819.

Cheng, G., Z. Zhang, and B. Zhang (2017). Test for bandedness of high-dimensional precision matrices. *Journal of Nonparametric Statistics 29*(4), 884–902.

Drton, M. and M. D. Perlman (2004). Model selection for gaussian concentration graphs. *Biometrika 91*(3), 591–602.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

Edwards, D. (2000). *Introduction to Graphical Modelling*. New York: Springer.

Eftekhari, A., D. Pasadakis, M. Bollhöfer, S. Scheidegger, and O. Schenk (2021). Block-enhanced precision matrix estimation for large-scale datasets. *Journal of Computational Science 53*, 2975–3026.

Fan, J. (1996). Test of significance based on wavelet thresholding and neyman's truncation. *Journal of the American Statistical Association 91*(434), 674–688.

Fan, J., Y. Liao, and J. Yao (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica 83*(4), 1497–1541.

Friedman, J., T. Hastie, and R. Tibshirani (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics 9*(3), 432–441.

Friedman, J., T. Hastie, and R. Tibshirani (2019). *glasso: Graphical Lasso: Estimation of Gaussian Graphical Models*. R package version 1.11.

Goeman, J. J. and U. Mansmann (2008, 01). Multiple testing on the directed acyclic graph of gene ontology. *Bioinformatics 24*(4), 537–544.

Guo, X. and C. Y. Tang (2020). Specification tests for covariance structures in high-dimensional statistical models. *Biometrika 108*(2), 335–351.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

Janková, J. and S. van de Geer (2017). Honest confidence regions and optimality in high-dimensional precision matrix estimation. *Test 26*(26), 143–162.

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

Le, T.-M. and P.-S. Zhong (2022). High-dimensional precision matrix estimation with a known graphical structure. *Stat 11*(1), e424.

Li, C. and H. Li (2008). Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics 24*(9), 1175–1182.

Liu, H. and L. Wang (2017). TIGER: A tuning-insensitive approach for optimally estimating Gaussian graphical models. *Electronic Journal of Statistics 11*(1), 241–294.

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

Liu, W. and X. Luo (2015). Fast and adaptive sparse precision matrix estimation in high dimensions. *Journal of Multivariate Analysis 135*, 153–162.

Ning, Y. and H. Liu (2017). A general theory of hypothesis tests and

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

confidence regions for sparse high dimensional models. *The Annals of Statistics 45*(1), 158–195.

Qiu, Y. and S. X. Chen (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *The Annals of Statistics 40*(3), 1285–1314.

Ren, Z., T. Sun, C.-H. Zhang, and H. H. Zhou (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *The Annals of Statistics 43*(3), 991–1026.

Sales, G., E. Calura, D. Cavalieri, and C. Romualdi (2012). graphite - a bioconductor package to convert pathway topology to gene network. *BMC Bioinformatics 13*(20).

The New York Times (2021). Coronavirus (covid-19) data in the united states. `https://github.com/nytimes/covid-19-data`.

Wang, X., G. Xu, and S. Zheng (2023). Adaptive tests for bandedness of high-dimensional covariance matrices. *Statistica Sinica 33*, 1673–1696.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0385

Yuan, M. and Y. Lin (2007). Model selection and estimation gaussian graphical model. *Biometrika 94*, 19–35.

Zheng, S., Z. Chen, H. Cui, and R. Li (2019). Hypothesis testing on linear structures of high-dimensional covariance matrix. *The Annals of Statistics 47*(6), 3300–3334.

Zhong, P.-S., W. Lan, P. X. K. Song, and C.-L. Tsai (2017). Tests for covariance structures with high-dimensional repeated measurements. *The Annals of Statistics 45*(3), 1185–1213.

Zhou, S., P. Rütimann, M. Xu, and P. Bühlmann (2011). High-dimensional covariance estimation based on gaussian graphical models. *Journal of Machine Learning Research 12*, 2975–3026.

Zhou, Y. and P. X.-K. Song (2016). Regression analysis of networked data. *Biometrika 103*(2), 287–301.

[1] Department of Mathematics - University of Tennessee at Chattanooga, Tennessee, USA. E-mail: thien-le@utc.edu

[2]Department of Mathematics, Statistics and Computer Science - University of Illinois Chicago, Illinois, USA. E-mail: pszhong@uic.edu

[3]Department of Statistics - University of Warwick, Coventry, UK. E-mail: c.leng@warwick.ac.uk