Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

TEST FOR HIGH DIMENSIONAL COVARIANCE MATRICES

BY YUEFENG HAN AND WEI BIAO WU

University of Chicago

The paper introduces a new test for testing structures of covariances for high dimensional vectors and the data dimension can be much larger than the sample size. Under proper normalization, central and non-central limit theorems are established. The asymptotic theory is attained without imposing any explicit restriction between data dimension and sample size. To facilitate the related statistical inference, we propose the balanced Rademacher weighted differencing scheme, which is also the delete-half jackknife, to approximate the distribution of the proposed test statistics. We also develop a new testing procedure for substructures of precision matrices. The simulation results show that the tests outperform the exiting methods both in terms of size and power. Our test procedure is applied to a colorectal cancer dataset.

1. Introduction. Driven by a diversity of contemporary scientific applications, analysis of high dimensional data has emerged as one of the most important and active areas in statistics. High-dimensional data, where the dimension can be much larger than the sample size, are encountered in genomics, medical imaging, financial economics and others. Knowledge of the covariance structure is essential in the associated statistical inference. For instance, structural assumptions are needed for estimation of high-dimensional covariance matrices, for example the banding method in Wu and Pourahmadi (2009) and Bickel and Levina (2008); tapering in Furrer and Bengtsson (2007) and Cai, Zhang and Zhou (2010); regularizing principal components in Cai, Ma and Wu (2015); factoring in Fan, Fan and Lv (2008) and Fan, Liao and Mincheva (2013). In addition, some researchers considered parametric models of covariance structures, such as autoregressive moving average, compound symmetry and Matérn class covariance function (e.g., see Gneiting, Kleiber and Schlather (2010), Wiesel, Bibi and Globerson (2013) and Pourahmadi (2013)).

AMS 2000 subject classifications: Primary 62F05; secondary 62E17
Keywords and phrases: Gaussian approximation, high-dimensional covariance matrices, covariance function, resampling

1.1. *Testing covariance structure.* Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent and identically distributed (i.i.d.) samples drawn from a p-dimensional distribution with mean μ and covariance matrix $\Sigma = (\sigma_{jk})_{j,k \le p}$. A fundamental problem in the inference of covariance is to test:

$$H_0 : \sigma_{jk} = \sigma_{jk,0} \text{ for all } (j,k) \in \mathcal{S}, \tag{1.1}$$

where $\sigma_{jk,0}$ are pre-specified or from certain parametric families $\sigma_{jk,0}(\theta)$ for some θ, $\mathcal{S}$ is the index set of covariance structure of interest. An incorrectly specified covariance structure could result in inaccurate statistical inference. One motivation of such models comes from spatial statistics and machine learning, where parametric covariance functions are widely used, such as Matérn covariance functions $f(m) = \sigma^2 2^{-\theta}\Gamma(\theta)^{-1}(\sqrt{\theta}m/\rho)^{\theta}K_{\theta}(\sqrt{\theta}m/\rho)$ (Stein (1999)) and the rational quadratic covariance function $f(m) = (1 + m^2/(\theta\sigma^2))^{-\theta/2}$ (Rasmussen and Williams (2006)), where m is the distance, Γ is the gamma function, K_θ is the modified Bessel function of the second kind, and σ^2, ρ and θ are non-negative parameters of the covariance. An important task is to test the validity of such parametric forms.

In the classical fixed dimensional setting, when the data is Gaussian, the conventional likelihood ratio test (LRT) can be used to access the structure of the covariance and it has certain optimality properties; see Anderson (2003) for details. When the dimension p grows with the sample size n, the standard LRT is no longer applicable. There has been a set of high dimensional tests on different covariance structures. Bai et al. (2009) proposed a corrected LRT for the identity hypothesis $H_0 : \Sigma = I$ and demonstrated that the test is valid when X_i are Gaussian and $p/n \to c \in (0,1)$. The result is further extended in Zhang, Peng and Wang (2013) and Zheng, Bai and Yao (2015). Ledoit and Wolf (2002) showed the test in John (1971, 1972) for sphericity with $H_0 : \Sigma = \sigma^2 I$ is consistent even when $p/n \to c$ for a positive constant c. Chen, Zhang and Zhong (2010) proposed tests for sphericity and identity of covariance matrices without normality assumption and without specifying an explicit relationship between p and n. For normally distributed data, Jiang (2004) proposed testing for diagonal Σ by considering the coherence statistic $L_{n,p} = \max_{1\le j<k\le p}|\hat{r}_{jk}|$, where $\hat{r}_{jk}$ is the (j,k)-th sample correlation. Cai and Jiang (2011) extended the test of Jiang (2004) for the bandedness of Σ based on the test statistic $L_{n,p,\kappa} = \max_{|j-k|\ge\kappa}|\hat{r}_{jk}|$ for Gaussian vectors. Xiao and Wu (2013) extended the results on more testing problems, such as stationarity, bandedness and tapering, and allowed non-Gaussianity. Qiu and Chen (2012) proposed a test based on a U-statistic which is an unbiased estimator of $\sum_{|j-k|\ge\kappa}\sigma_{jk}^2$ for testing bandedness. Cai and Ma (2013) studied the optimality of one sample tests for $H_0 : \Sigma = I$.

Li and Chen (2012) considered tests for the equality of covariance matrices. More recently, in regression setting, to access the adequacy of some specified parametric forms of error covariance structures with $H_0 : \Sigma = \Sigma(\boldsymbol{\theta})$ for unknown parameter $\boldsymbol{\theta}$, Zhong et al. (2017) proposed a bias adjusted test based on $\mathrm{tr}\{(\Sigma - \Sigma(\boldsymbol{\theta}))^2\}$ for normally distributed random vectors. He and Chen (2016) proposed a test procedure that focuses on testing along the super-diagonals of the covariance matrix to detect sparse signals and parametric structures. This was further extended to the case of two samples in He and Chen (2018). In many applications, the diagonal elements of the covariance may not be useful in the testing. This motivates us to develop a test to examine the appropriateness of covariance structure specification via the off-diagonals of the covariance matrices.

Define the sample mean $\bar{\boldsymbol{X}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i$ and the sample covariance matrix $\hat{\Sigma} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T = (\hat{\sigma}_{jk})_{j,k\le p}$. We propose a test for the hypothesis H_0 in (1.1) based on an unbiased estimator of the quadratic form $\sum_{(j,k)\in\mathcal{S}}(\sigma_{jk} - \sigma_{jk,0})^2$. We first consider testing for off-diagonal covariance structures. A distributional approximation for the test statistic of Gaussian vectors with same covariance structure is obtained. It is shown that our Gaussian approximation theorem covers the cases where the test statistic does not have a limit Gaussian distribution as $n \to \infty$ and $p \to \infty$. In some cases, after a suitable normalization, the test statistic could have a standard normal distribution as the limiting distribution, but the approximation to a standard normal distribution requires some restrictions on the covariance structure Σ. We provide a sufficient and necessary condition, which extends the sufficient condition for Gaussian data in Cai and Ma (2013). It is also worth noting that the proposed test does not require explicit conditions in the relationship between p and n. The power of the test is also investigated. In order to overcome the difficulty to consistently estimate the fourth moments of $\boldsymbol{X}_i$ and quantify the difference of the c.d.f of the test statistic and that by estimated moments, we propose using the balanced Rademacher weighted differencing scheme, called half-sampling; see also Wu, Lou and Han (2018). Wu (1990) showed that in the one-dimensional case the histogram of the delete-d jackknife with a suitable d, the number of deleted observations, can be consistent in estimating the sampling distribution for linear and certain non-linear statistics (in particular, U-statistics), and is optimal if d is taken to be on the same order as the sample size. We extend his idea and show that the balanced Rademacher weighted differencing scheme (half-sampling approach), which is also the delete-$n/2$ Jackknife, leads to a consistent estimator of the distribution function of the test statistic. The proofs of the validity of the half-sampling approach require a more involved

Gaussian approximation result.

To study the case where $\sigma_{jk,0}$ in (1.1) are from certain parametric families $\sigma_{jk,0}(\theta)$ for some θ, we first estimate the involved parameters, then establish the distributional approximation of the test statistic with estimated parameters and implement the half-sampling procedure accordingly. In particular, the asymptotic mean of the test statistic varies for different parametric forms and different relationship between n and p, which may not vanish due to the bias induced by the estimation of unknown parameters. It is worth noting that our half sampling approach avoids the estimation of the unknown mean of the test statistic, and thus can be easily applied to test parametric covariance functions. The numerical results indicate that our proposed test estimates size accurately. In comparison, the test in Zhong et al. (2017) tends to overestimate the size at low nominal levels.

Besides testing for off-diagonal covariance structures, we also develop a test for sub-matrices. The interest on such a test arises naturally in applications in genomics and other fields, when we are interested in knowing the between pathway associations in genomics where each pathway stands for a group of genes, or studying the relationships between a diverse range of disease phenotypes and genomic markers in PheWAS (see, e.g., Kelley and Ideker (2005)). Asymptotic properties of the test are derived and a half-sampling estimator of the distribution function of the test statistic is studied.

1.2. *Testing precision matrices.* Precision matrix plays a fundamental role in many high dimensional inference problems. It is of significant interest to understand structure or substructure of the precision matrices. For example, under the Gaussian graphical model framework, a submatrix of the precision matrix characterizes the network of two groups, which measures the conditional dependence network structure of two groups of variables. See De la Fuente (2010), Hudson, Reverter and Dalrymple (2009), Ideker and Krogan (2012), Jia et al. (2011), Li, Agarwal and Rajagopalan (2008), Ren et al. (2015), among others. One can also use it to study interactions between two groups that adjust for effects from other variables.

Let $\Omega = \Sigma^{-1} = (\omega_{jk})_{j,k \le p}$ be the precision matrix. Testing the hypothesis $H_0 : \Omega = \Omega_0$ for a given Ω_0 is equivalent to testing $H_0 : \Sigma = \Sigma_0$, which has been well studied under various alternatives. However, in many applications, one aims at studying the group structure of the network, by testing a given substructure of the precision matrix Ω,

$$H_0 : \omega_{jk} = 0 \text{ for all } (j,k) \in \mathcal{S}, \tag{1.2}$$

where $\mathcal{S}$ is an index set. In such cases, it is essential to work on the precision matrix directly, instead of the covariance matrix. Testing procedures on the covariance matrix cannot leverage information on the given substructure of the precision matrix. More importantly, due to the notable difference between conditional and unconditional dependencies, the various procedures for testing the covariance matrix may not be well adapted to testing specific substructure of the precision matrix. To the best of our knowledge, there are no currently available methods with theoretical guarantees to infer about substructure of the precision matrix when the dimension of the substructure can go to infinity. Xia, Cai and Cai (2015) proposed a procedure for testing the differential network by using the maximum entrywise deviation of the precision matrix. Xia, Cai and Cai (2018) considered testing a given submatrix of the precision matrix under a Gaussian graphical model when the dimension of the submatrix is fixed. In our paper, we develop a novel testing procedure for substructures of the precision matrices. The test statistic is based on the Frobenius norm of a substructure estimate of the precision matrix without imposing any structure assumptions. Theoretical properties under sub-Gaussian tails and linear process model are discussed. The testing procedure is easy to implement.

1.3. *Organization of the paper.* The paper is organized as follows. Section 2 introduces the procedure for testing off-diagonal covariance structure and its asymptotic properties of the test statistic and the theoretical properties of the half-sampling estimator. Properties of the test for parametric covariance functions are presented in Sections 3. A new testing procedure for a given substructure of the precision matrix is proposed and its theoretical properties are presented in Section 4. Numerical performance of the tests are given in Section 5. The readers are referred to Appendix (supplementary material) Section A and B for properties of the test for the off-diagonal sub-matrix, and power evaluations, respectively. A real data example is illustrated in Appendix C. Appendix D includes more simulation results. All technical details are relegated to Appendix E.

1.4. *Notation.* Throughout this paper, for a matrix $A = (a_{ij})$ write $|A|_\infty = \max_{i,j} |a_{ij}|$ and the Frobenius norm $|A|_F = (\sum_{ij} a_{ij}^2)^{1/2}$. For a vector $x = (x_1, ..., x_p)^T$, define $|x| = |x|_2 = (x_1^2 + ... + x_p^2)^{1/2}$. Let $\boldsymbol{\xi} = (\xi_1, ..., \xi_p)^T$ be a random vector. Write $\boldsymbol{\xi} \in \mathcal{L}^m$, $m \geq 1$, if the m-norm $\|\boldsymbol{\xi}\|_m := (\mathsf{E}|\boldsymbol{\xi}|^m)^{1/m} < \infty$. For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, write $a_n = O(b_n)$ (resp. $a_n \asymp b_n$) if there exists a constant C such that $|a_n| \leq C|b_n|$ (resp. $1/C \leq a_n/b_n \leq C$) holds for all sufficiently large n, and write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$. Let $\lceil a \rceil = \min\{k \in \mathbb{Z} : k \geq a\}$.

2. Testing Off-diagonal Covariance Structure.

2.1. *Overview.* A natural test statistic for the hypothesis H_0 in (1.1) is based on the quadratic form $\sum_{(j,k)\in\mathcal{S}}(\hat\sigma_{jk}-\sigma_{jk,0})^2$. It is noted that $\sum_{(j,k)\in\mathcal{S}}(\hat\sigma_{jk}-\sigma_{jk,0})^2$ is a biased estimator of $\sum_{(j,k)\in\mathcal{S}}(\sigma_{jk}-\sigma_{jk,0})^2$, since $\mathsf{E}(\hat\sigma_{jk}-\sigma_{jk,0})^2 = \mathrm{var}(\hat\sigma_{jk})+(\sigma_{jk}-\sigma_{jk,0})^2$. Following the spirit of Chen, Zhang and Zhong (2010) and Li and Chen (2012), we propose

$$\mathcal{T}_{\mathcal{S}} = \sum_{(j,k)\in\mathcal{S}} M_{jk}, \tag{2.1}$$

which is an unbiased estimator of $\sum_{(j,k)\in\mathcal{S}}(\sigma_{jk}-\sigma_{jk,0})^2$, where

$$\begin{aligned} M_{jk} =& \frac{1}{P_n^1}\sum^{*}_{i_1,i_2} X_{i_1j}X_{i_1k}X_{i_2j}X_{i_2k} - \frac{2}{P_n^2}\sum^{*}_{i_1,i_2,i_3} X_{i_1j}X_{i_2j}X_{i_2k}X_{i_3k} \\ &- \frac{2}{n}\sigma_{jk,0}\sum_{i_1}^{n} X_{i_1j}X_{i_1k} + \frac{2}{P_n^1}\sigma_{jk,0}\sum^{*}_{i_1,i_2} X_{i_1j}X_{i_2k} + \sigma_{jk,0}^2 \\ &+ \frac{1}{P_n^3}\sum^{*}_{i_1,i_2,i_3,i_4} X_{i_1j}X_{i_2j}X_{i_3k}X_{i_4k} \text{ and } P_n^k := \prod_{j=n-k}^{n} j. \end{aligned} \tag{2.2}$$

Throughout this paper, $\sum^*$ denotes summation over mutually different subscripts shown, for example, $\sum^*_{i_1,i_2,i_3}$ denotes summation over $\{(i_1,i_2,i_3): i_1\neq i_2, i_2\neq i_3, i_1\neq i_3, 1\le i_1,i_2,i_3\le n\}$. Elementary derivations show that $\mathsf{E}M_{jk}=(\sigma_{jk}-\sigma_{jk,0})^2$ for all $1\le j,k\le p$, then $\mathcal{T}_S$ is unbiased for $\sum_{(j,k)\in\mathcal{S}}(\sigma_{jk}-\sigma_{jk,0})^2$. Besides the unbiasedness, $\mathcal{T}_S$ is invariant under the location shift. This means that, without loss of generality, we can assume $\mu=\mathsf{E}\boldsymbol{X}_i=0$ in the rest of the paper. To calculate $\mathcal{T}_S$, it is computationally more efficient to use an equivalent formula given by Himeno and Yamada (2014) which reduces the computational cost from $O(n^4)$ to $O(n)$.

We reject H_0 if $\mathcal{T}_S$ exceeds certain cutoff values. The problem of deriving asymptotic distribution of $\mathcal{T}_S$ is open. In many of earlier papers it is assumed that Σ_0 has special structures such as being diagonal or spheric and/or $\boldsymbol{X}_i$ is Gaussian or has independent entries. Here we shall obtain an asymptotic theory for $\mathcal{T}_S$ for Volterra process model, a generalization of linear process models, which will be specified in this section.

Let us first consider testing the off-diagonal covariance structure:

$$H_{0a}: \sigma_{jk}=\sigma_{jk,0} \text{ for all } (j,k)\in\mathcal{S}_1, \text{ where } \mathcal{S}_1=\{(j,k): 1\le j\neq k\le p\}. \tag{2.3}$$

For $\boldsymbol{X} = (X_1, \dots, X_p)^T$, let $\mathcal{W}(\boldsymbol{X}, \mathcal{S}) := (X_j X_k - \sigma_{jk})_{(j,k)\in\mathcal{S}}$. In particular, let $\hat{T}_n = \mathcal{T}_{\mathcal{S}_1}$ and

$$\mathcal{W}(\boldsymbol{X}, \mathcal{S}_1) = \begin{pmatrix} X_1X_2 - \sigma_{12} \\ \dots \\ X_1X_p - \sigma_{1p} \\ X_2X_1 - \sigma_{12} \\ \dots \\ X_pX_{p-1} - \sigma_{p,p-1} \end{pmatrix} \tag{2.4}$$

be a $p(p-1)$-dimensional vector. Let the random vector $\boldsymbol{X}$ be identically distributed as $\boldsymbol{X}_i$. Denote $\boldsymbol{W} = \mathcal{W}(\boldsymbol{X}, S_1)$, $\boldsymbol{W}_i = \mathcal{W}(\boldsymbol{X}_i, S_1)$ and $\bar{\boldsymbol{W}}_n = \sum_{i=1}^n \boldsymbol{W}_i/n$. Then the covariance matrix $\Gamma = (\gamma_{\alpha,\alpha'})_{\alpha,\alpha'\in\mathcal{S}_1}$ for $\boldsymbol{W}$ is $p(p-1) \times p(p-1)$ with entries

$$\begin{aligned} \gamma_{(j,k),(m,q)} &= \mathsf{E}\left(\left(X_jX_k - \sigma_{jk}\right)\left(X_mX_q - \sigma_{mq}\right)\right) \\ &= \mathsf{E}(X_jX_kX_mX_q) - \sigma_{jk}\sigma_{mq} \\ &= \operatorname{cum}(X_j, X_k, X_m, X_q) + \sigma_{jm}\sigma_{kq} + \sigma_{jq}\sigma_{km}. \end{aligned} \tag{2.5}$$

The square of the Frobenius norm of Γ is

$$|\Gamma|_F^2 = \sum_{\alpha,\alpha'\in\mathcal{S}_1} \gamma_{\alpha\alpha'}^2 := |\mathsf{E}(\boldsymbol{W}\boldsymbol{W}^T)|_F^2.$$

Suppose the following Lyapunov-type condition for $\boldsymbol{W}_i$ is satisfied: there exists a constant K such that, for some $\delta > 0$,

$$(K_\delta^W)^{2+\delta} := \mathsf{E}\left|\frac{\boldsymbol{W}_1^T\boldsymbol{W}_2}{|\Gamma|_F}\right|^{2+\delta} < K < \infty. \tag{2.6}$$

The basic idea of our test procedure is to bound the Kolmogorov distance between the distribution of $n\hat{T}_n/|\Gamma|_F$ and its Gaussian analog under condition (2.6). Under the null hypothesis H_{0a}, we can establish

$$\sup_{t\in\mathbb{R}}\left|\mathsf{P}\left(\frac{n\hat{T}_n}{|\Gamma|_F} \le t\right) - \mathsf{P}\left(\frac{1}{(n-1)|\Gamma|_F}\sum_{i\neq l}^n \boldsymbol{Y}_i^T\boldsymbol{Y}_l \le t\right)\right| \longrightarrow 0,$$

where $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$ are i.i.d. $N(0, \Gamma)$, as the Gaussian analog of $\boldsymbol{W}_i$ in the sense of having the same mean and the same covariance matrix. Then we shall use a half-sampling technique to obtain an asymptotically unbiased and consistent estimator of the cumulative distribution function of $n\hat{T}_n$, since the covariance matrix Γ is unknown and the associated estimation issue can be quite challenging. Rigorous analysis will be carried out afterwards.

2.2. *Asymptotic properties.* To present an asymptotic theory of $\hat{T}_n$, we impose the following conditions:

Assumption 2.1. *$X_{ij} = \mu_j + \sum_{l_1=1}^N b_{j,l_1}\xi_{il_1} + \sum_{l_1<l_2}^N a_{j,l_1l_2}\xi_{il_1}\xi_{il_2} + \cdots + \sum_{l_1<l_2<\ldots<l_d}^N a_{j,l_1l_2\ldots l_d}\xi_{il_1}\xi_{il_2}\ldots\xi_{il_d}$ for all $1 \le j \le p$ where d is a fixed number, $\{\xi_{il}\}_{1\le i\le n, 1\le l\le N}$ are i.i.d. random variables with mean 0, variance 1, $\mathsf{E}\xi_{11}^3 = 0$ and $\mathsf{Var}(\xi_{11}^2) = \nu < \infty$.*

Specifically, for Gaussian vector $\boldsymbol{X}_i$, Assumption 2.1 always holds with $N = p$ and $a_{j,l_1l_2} = 0, \ldots, a_{j,l_1l_2\ldots l_d} = 0$ for all $1 \le l_1 < l_2 < \ldots < l_d \le N$. The requirement of $\xi_{i1}, \ldots, \xi_{iN}$ being i.i.d. and $\mathsf{E}\xi_{11}^3 = 0$ is not essential and is purely for the sake of simpler notion. Differently from Chen, Zhang and Zhong (2010) and Qiu and Chen (2012), we do not assume $N \ge p$.

Furthermore, many papers in testing high dimensional covariance matrices assume linear process model, while we extend to nonlinear process model, *i.e.,* Volterra process model. Linear process is considered in Xu, Zhang and Wu (2014) and Li and Chen (2012). In the study of nonlinear systems, Volterra processes are of fundamental importance; see Schetzen (1980), Rugh (1981), Casti (1985), Priestley (1988) and Bendat (1990), among others. The Volterra process has been widely applied as nonlinear system modeling technique with considerable success, since a wide range of nonlinear process models admit Volterra process. At the technical level, Volterra process involves recursive application of Rosenthal's inequality.

Assumption 2.2. *For some constant $C > 0$,*

$$|\Gamma|_F^2 \ge C \sum_{(j,k)\in\mathcal{S}_1} \sum_{(m,q)\in\mathcal{S}_1} (\sigma_{jm}^2\sigma_{kq}^2 + \sigma_{jm}\sigma_{jq}\sigma_{km}\sigma_{kq}). \tag{2.7}$$

We now discuss Assumption 2.2. Let $Q := \sum_{(j,k)\in\mathcal{S}_1}\sum_{(m,q)\in\mathcal{S}_1} (\sigma_{jm}\sigma_{kq} + \sigma_{jq}\sigma_{km})^2$. Note that from (2.5),

$$\begin{aligned}|\Gamma|_F^2 =& Q + \sum_{(j,k)\in\mathcal{S}_1}\sum_{(m,q)\in\mathcal{S}_1} \big(\mathrm{cum}(X_j, X_k, X_m, X_q)^2 \\ &+ 2\mathrm{cum}(X_j, X_k, X_m, X_q)(\sigma_{jm}\sigma_{kq} + \sigma_{jq}\sigma_{km})\big).\end{aligned}$$

Assume that there exists a constant $c < 1/4$ such that

$$\sum_{(j,k)\in\mathcal{S}_1}\sum_{(m,q)\in\mathcal{S}_1} \mathrm{cum}(X_j, X_k, X_m, X_q)^2 \le cQ. \tag{2.8}$$

Similar conditions are commonly imposed for cumulant analysis; see, e.g., Kalouptsidis and Koukoulas (2005), Xiao and Wu (2013) and Cherif and Fnaiech (2015). Then (2.8) implies Assumption 2.2 by the Cauchy-Schwarz inequality

$$|\Gamma|_F^2 \geq 2(1-2\sqrt{c}) \sum_{(j,k)\in\mathcal{S}_1} \sum_{(m,q)\in\mathcal{S}_1} (\sigma_{jm}^2\sigma_{kq}^2 + \sigma_{jm}\sigma_{jq}\sigma_{km}\sigma_{kq}).$$

Typical examples that satisfy (2.8) include Gaussian vectors whose 4th cumulants are 0 and the linear process models, that is, under Assumption 2.1 with $a_{j,l_1l_2\ldots l_i}=0$ for all $1\leq l_1<l_2<\ldots<l_i\leq N, 2\leq i\leq d, 1\leq j\leq p$; see Lemma E.2 in the supplementary material for details.

The following theorem provides a Berry-Esseen type bound of the asymptotic approximation of $\hat{T}_n$ by a linear combination of χ_1^2 random variables.

THEOREM 2.1. *Suppose Assumptions 2.1 and 2.2 hold and $\|\xi_{11}\|_{4+2\delta}<\infty$ with $0<\delta\leq 1$. Then under the null hypothesis H_{0a} (2.3), we have that*

$$\sup_t \left| P\left(\frac{n\hat{T}_n}{|\Gamma|_F}\leq t\right) - P\left(\sum_{d=1}^{p(p-1)} \frac{\lambda_d}{|\Gamma|_F}(\eta_d-1)\leq t\right)\right| = O(n^{-\delta/(10+4\delta)}), \tag{2.9}$$

where $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_{p^2-p}\geq 0$ are eigenvalues of Γ and η_d, $d\geq 1$, are i.i.d. χ_1^2.

REMARK 2.1. *We conjecture that better rate can be possibly derived by applying the more sophisticated mathematical argument that involves solutions to Stein's equations. Solutions to Stein's equation with normal distribution have a close form which is relatively easy to work with and it can lead to sharp Berry-Esseen bound. Chatterjee (2008)'s new version of Stein's method can be applied to obtain sharp Berry-Esseen bounds of quadratic form for normal approximation. However, it is difficult to work with Stein's equation with distribution being linear combinations of χ_1^2 random variables. A recent breakthrough of Stein's method with distribution being linear combination of χ_1^2 random variables is considered in Arras et al. (2016). Due to its extreme complexity, we are not able to apply it to our problem. The optimal rate of L_2 type Gaussian approximation is still open.* □

Note that $\sum_{d=1}^{p(p-1)}\lambda_d\eta_d$ and $\boldsymbol{Y}^T\boldsymbol{Y}$ have the same distribution, with $\boldsymbol{Y}\sim N(0,\Gamma)$. Under H_{0a}, Theorem 2.1 implies that the asymptotic variance of $n\hat{T}_n$ is $\mathsf{E}(\sum_{d=1}^{p(p-1)}\lambda_d(\eta_d-1))^2 = 2|\Gamma|_F^2$. If the null hypothesis H_{0a} does not hold, a similar argument as Theorem 2.1 implies the following corollary.

Corollary 2.1. *Suppose $\|\xi_{11}\|_{4+2\delta} < \infty$ with $0 < \delta \le 1$. Assume that $\sum_{j\neq k}^{p}(\sigma_{jk} - \sigma_{jk,0})^2/|\Gamma|_F = O(1)$. Under Assumptions 2.1 and 2.2, we have that*

$$\sup_t \left| P\left(\frac{n\hat{T}_n}{|\Gamma|_F} \le t\right) - P\left(\frac{(\boldsymbol{Y} + \sqrt{n}\mu_Y)^T(\boldsymbol{Y} + \sqrt{n}\mu_Y) - \mathrm{tr}(\Gamma)}{|\Gamma|_F} \le t\right)\right| = O(n^{-\delta/(10+4\delta)}). \tag{2.10}$$

where $\boldsymbol{Y} \sim N(0, \Gamma)$ and $\mu_Y = (\sigma_{12} - \sigma_{12,0}, \sigma_{13} - \sigma_{13,0}, ..., \sigma_{p,p-1} - \sigma_{p,p-1,0})^T$. On the other hand, if $\sum_{j\neq k}^{p}(\sigma_{jk} - \sigma_{jk,0})^2/|\Gamma|_F \to \infty$, under Assumptions 2.1 and 2.2, we have that $n\hat{T}_n/|\Gamma|_F \to \infty$ in probability.

Remark 2.2. *The idea of formulating the test statistics for off-diagonal covariance structure can be used for testing $H_0 : \sigma_{jk} = \sigma_{jk,0}$ for all $|j - k| > \kappa$, e.g., the banding structure. With little modification of $\hat{T}_n$, we can construct a test statistic on the super-diagonals $|j - k| > \kappa$. Similar asymptotic properties in Theorem 2.1 and Corollary 2.1 can be obtained.* □

The asymptotic approximation in Theorem 2.1 is attained *without any restriction* on p. In the low dimensional case with $p = O(1)$, which may be viewed as having finite dimension, the Berry-Esseen style theorem as conveyed in Theorem 2.1 and Corollary 2.1 still hold.

By Theorem 2.1, in general, the approximating distribution of $\hat{T}_n$ is a linear combination of χ_1^2. The following corollary concerns a central limit theorem for $\hat{T}_n$.

Corollary 2.2. *Under conditions of Theorem 2.1, the central limit theorem $n\hat{T}_n/|\Gamma|_F \xrightarrow{d} N(0, 2)$ holds if and only if*

$$\rho_\Gamma := \frac{\mathrm{tr}(\Gamma^4)}{\mathrm{tr}^2(\Gamma^2)} \to 0, \ \text{as } p \to \infty. \tag{2.11}$$

Assume $\sum_{(j,k)\in\mathcal{S}_1}\sum_{(m,q)\in\mathcal{S}_1}(\sigma_{jm}^2\sigma_{kq}^2 + \sigma_{jm}\sigma_{jq}\sigma_{km}\sigma_{kq}) \ge K\mathrm{tr}^2(\Sigma^2)$ for some constant $K > 0$. If $\{\boldsymbol{X}_i\}_{i=1}^n$ follows the linear process model, that is, under Assumption 2.1 with $a_{j,l_1l_2...l_i} = 0$ for all $1 \le l_1 < l_2 < ... < l_i \le N, 2 \le i \le d, 1 \le j \le p$, then, (2.11) is equivalent to

$$\rho_\Sigma \to 0, \ \text{as } p \to \infty. \tag{2.12}$$

In other words, condition (2.12) for linear process models is the necessary and sufficient one to achieve the central limit theorem. Condition (2.12) is widely used in the literature of high dimensional hypothesis testing problems; see e.g., Chen, Zhang and Zhong (2010), Li and Chen (2012). This result is consistent with Proposition 3 in Cai and Ma (2013) which deals with tests for high dimensional covariance matrices for Gaussian vectors. They developed the Berry-Esseen bound $(1/n + \rho_\Sigma)^{1/5}$ for a similar test statistic which is asymptotically Gaussian under (2.12). Condition (2.12) is violated, for instance the rational quadratic covariance structure in Example 2.1 below or the simple linear factor model $X_{ij} = F_i + \xi_{ij}$ where $\{F_i\}$ and $\{\xi_{ij}\}$ are i.i.d. mean 0 and variance 1, $\mathrm{tr}(\Sigma^4) \asymp \mathrm{tr}^2(\Sigma^2)$.

EXAMPLE 2.1. *Consider the rational quadratic covariance structure* $\Sigma_0 = \{(\sigma_{jk,0}(\theta))_{p\times p} : \sigma_{jk,0}(\theta) = (1 + \theta_1^{-1}\theta_2^{-2}|j-k|^2)^{-\theta_1/2}$ *and* $0 < \theta_1 < 1/2, \theta_2 > 0\}$. *It can be shown that* $\mathrm{tr}(\Sigma^4) \asymp p^{4-4\theta_1}$ *and* $\mathrm{tr}(\Sigma^2) \asymp p^{2-2\theta_1}$, *leading to* $\rho_\Sigma \not\to 0$, *as* $p \to \infty$. *Then the classical central limit theorem in Corollary 2.2 does not apply, while Theorem 2.1 still holds with a non-Gaussian approximating distribution.* □

2.3. *Estimating the distribution of* $n\hat{T}_n$. In general, by Theorem 2.1, the asymptotic distribution of $n\hat{T}_n$ can be used for testing with estimated critical values via estimation of $\{\lambda_d\}_{d=1}^{p(p-1)}$. It is also called a plug-in resampling procedure based on the sample version of Γ (see Xu, Zhang and Wu (2014)). However, estimation of the eigenvalues of matrix Γ is highly nontrivial, since by (2.5) Γ is a very high $p(p-1) \times p(p-1)$ dimensional matrix. To formulate a computational feasible test procedure, we use a half-sampling approach (also balanced Rademacher weighted differencing scheme) to avoid such estimation problems, and obtain an asymptotically unbiased and consistent estimator of the cumulative distribution function of $n\hat{T}_n$.

Assume that n is even. Let $B \subset \{1, 2, ..., n\}$, $B^c = \{1, ..., n\}\backslash B$, and $|B| = |B^c| = m = n/2$. Define respectively:

$$J_B(\mathcal{S}_1, \Sigma_0) = \sum_{(j,k)\in\mathcal{S}_1} R_{jk}(B, \sigma_{jk,0}), \tag{2.13}$$

$$C_{B,B^c}(\mathcal{S}_1, \Sigma_0) = \sum_{(j,k)\in\mathcal{S}_1} N_{jk}(B, B^c, \sigma_{jk,0}), \tag{2.14}$$

where recall the notation $\sum^*$ means summation over mutually different subscripts shown, $P_m^k := m(m-1)\cdots(m-k)$, and

(2.15)
$$N_{jk}(B, B^c, \sigma_{jk,0}) = \left(\frac{1}{m}\sum_{i_1\in B} X_{i_1j}X_{i_1k} - \frac{1}{P_m^1}\sum^*_{i_1,i_2\in B} X_{i_1j}X_{i_2k} - \sigma_{jk,0}\right) \cdot \left(\frac{1}{n-m}\sum_{i_3\in B^c} X_{i_3j}X_{i_3k} - \frac{1}{P_{n-m}^1}\sum^*_{i_3,i_4\in B^c} X_{i_3j}X_{i_4k} - \sigma_{jk,0}\right),$$

(2.16)
$$\begin{aligned} R_{j,k}(B,\sigma_{jk,0}) =& \frac{1}{P_m^1}\sum^*_{i_1,i_2\in B} X_{i_1j}X_{i_1k}X_{i_2j}X_{i_2k} - \frac{2}{P_m^2}\sum^*_{i_1,i_2,i_3\in B} X_{i_1j}X_{i_2j}X_{i_2k}X_{i_3k} \\ &+ \frac{1}{P_m^3}\sum^*_{i_1,i_2,i_3,i_4\in B} X_{i_1j}X_{i_2j}X_{i_3k}X_{i_4k} + \sigma_{jk,0}^2 \\ &- \frac{2}{m}\sigma_{jk,0}\sum_{i_1\in B} X_{i_1j}X_{i_1k} + \frac{2}{P_m^1}\sigma_{jk,0}\sum^*_{i_1,i_2\in B} X_{i_1j}X_{i_2k}. \end{aligned}$$

We consider the balanced Rademacher weighted differencing scheme (half-sampling approach). The half sampling estimator is defined as

$$\tilde{F}(t) = \frac{1}{\binom{n}{m}}\sum_{B\in\mathcal{B}} \mathbf{1}_{m(1-m/n)(J_B(\mathcal{S}_1,\Sigma_0)+J_{B^c}(\mathcal{S}_1,\Sigma_0)-2C_{B,B^c}(\mathcal{S}_1,\Sigma_0))\le t}, \tag{2.17}$$

where $\mathcal{B}$ contains all the subsets of size m of $\{1,2,...,n\}$. Because $\binom{n}{m}$ can be too large, $\tilde{F}(t)$ may be difficult to compute. Instead, a stochastic approximation may be employed. Let $B_1,...,B_L$ be i.i.d. uniformly sampled from the class $\mathcal{B} := \{B : B\subset\{1,...,n\}, |B| = m\}$. Assuming $\{\boldsymbol{X}_i\}$ and the sampling process $\{B_l\}$ are independent. The balanced Rademacher weighted differences is defined by $m(1-m/n)(J_{B_l}(\mathcal{S}_1,\Sigma_0)+J_{B_l^c}(\mathcal{S}_1,\Sigma_0)-2C_{B_l,B_l^c}(\mathcal{S}_1,\Sigma_0))$. Following Politis, Romano and Wolf (1999), $\tilde{F}(t)$ can be approximated by

$$\hat{F}_L(t) = \frac{1}{L}\sum_{l=1}^{L} \mathbf{1}_{m(1-m/n)(J_{B_l}(\mathcal{S}_1,\Sigma_0)+J_{B_l^c}(\mathcal{S}_1,\Sigma_0)-2C_{B_l,B_l^c}(\mathcal{S}_1,\Sigma_0))\le t}. \tag{2.18}$$

By the Dvoretzky-Kiefer-Wolfowitz-Massart inequality (cf. Massart (1990)),

$$\mathsf{P}^*\left(\sup_t |\hat{F}_L(t) - \tilde{F}(t)| \ge u\right) \le 2e^{-2Lu^2}, \quad u\ge 0, \tag{2.19}$$

where $\mathsf{P}^*(\cdot) = \mathsf{P}(\cdot|\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ is the conditional probability given the original data $\{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\}$. Hence, the distribution function of $F(t) := \mathsf{P}(n\hat{T}_n \le t)$ can be estimated by $\tilde{F}(t)$ (cf. Theorem 2.2), which is well approximated by $\hat{F}_L(t)$ by choosing $L \ge n$.

Politis, Romano and Wolf (1999) assume that $m/n \to 0$, whereas, motivated by numerical performance (see Example 2.2 below), we build a new half sampling procedure under the case $m = n/2$. In contrast, Xu, Zhang and Wu (2014) considered a sub-sampling procedure with $m = o(n)$. The convergence rate they developed for subsampling is much worse than our Theorem 2.2. In practice, we directly use the stochastic approximation of the half sampling estimator, $\hat{F}_L(t)$, instead of the original half sampling estimator $\tilde{F}(t)$. When the sample size is too small, the total number of possible subsamples can be small, then the method is less reliable. In practice, we recommend the sample size $n \ge 20$ and resampling replications $L \ge 1000$.

Our half sampling procedure is motivated by the Hadamard matrices. For ease of presentation, consider the mean test problem. Assume that $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$ are i.i.d. $N(\mu, \Sigma)$. Let H be an $n \times n$ Hadamard matrix where its first row consists all 1's, all its entries take values 1 or -1, and its rows are mutually orthogonal, so that $HH^T = nI_n$. Let $\boldsymbol{Z}_l = n^{-1/2}\sum_{i=1}^n H_{li}\boldsymbol{Y}_i$ for $l = 1, 2, ..., n$. Then $\boldsymbol{Z}_2$, $\boldsymbol{Z}_3$, ..., $\boldsymbol{Z}_n$ are i.i.d. $N(0, \Sigma)$ and the empirical cumulative distribution function

$$\hat{F}_n(t) = \frac{1}{n-1}\sum_{l=2}^{n} \mathbf{1}_{|\boldsymbol{Z}_l|_2^2 \le t}$$

converges uniformly to $F(t) = \mathsf{P}(n|\bar{\boldsymbol{Y}} - \mu|_2^2 \le t)$. We can reject the null hypothesis $\mu = 0$ at level $\alpha \in (0, 1)$ if $n|\bar{\boldsymbol{Y}}|_2^2 > \hat{t}_{1-\alpha}$, where $\hat{t}_{1-\alpha}$ is the $(1-\alpha)$th sample quantile of $\hat{F}_n(t)$. As an important feature of the latter method, one does not need to estimate the covariance matrix Σ. However, it is highly nontrivial to construct Hadamard matrices; see Hedayat et al. (1978) and Yarlagadda and Hershey (2012). To circumvent the construction problem of Hadamard matrices, we shall obtain asymptotically independent realizations by using balanced Rademacher weighted differencing scheme. See Wu, Lou and Han (2018) for more details.

The example below numerically illustrates the benefits of the half-sampling approach over the usual sub-sampling procedure with $m = o(n)$. Our half sampling approach goes far beyond the theoretical results about sub-sampling approach in Xu, Zhang and Wu (2014). The proofs of the validity of half-sampling approach are highly nontrivial and require a more involved Gaussian approximation result than theirs.

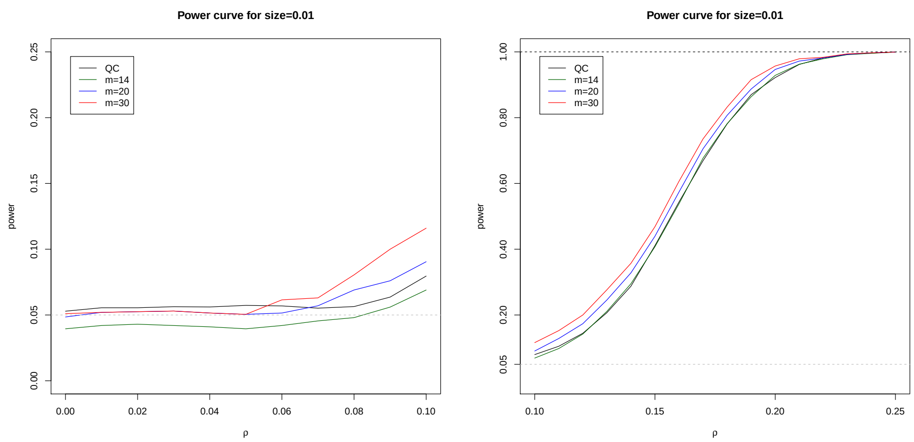

FIG 1. *Power curve of the test given in* Qiu and Chen (2012) *(abbr. QC), the sub-sampling procedures with resampling size* $m = 14, 20$ *and the half sampling procedure with* $m = 30$ *at size* $= 0.05$*. The resampling sizes are 2000.*

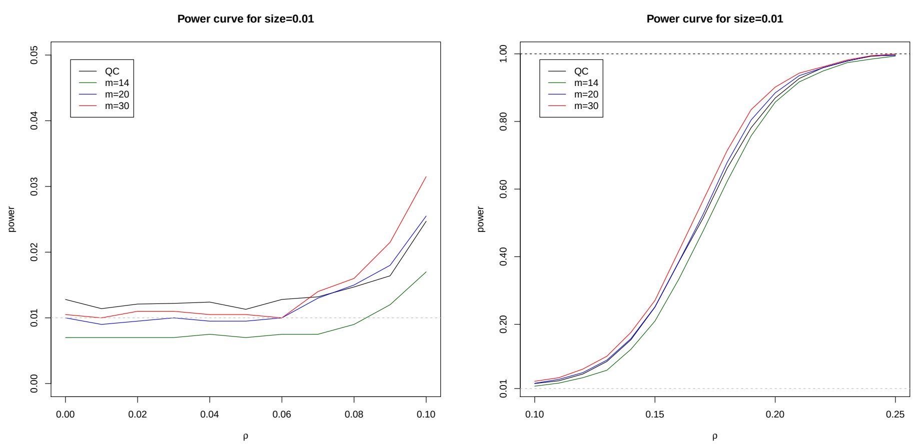

FIG 2. *Power curve of the test given in* Qiu and Chen (2012) *(abbr. QC), the sub-sampling procedures with resampling size* $m = 14, 20$ *and the half sampling procedure with* $m = 30$ *at size* $= 0.01$*. The resampling sizes are 5000.*

EXAMPLE 2.2. *Consider the following model:*

$$X_{ij} = Z_{ij} + \rho\zeta_i, \qquad 1 \le i \le n, 1 \le j \le p,$$

where Z_{ij}'s and ζ_i's are i.i.d $N(0,1)$, and ρ is a parameter. To obtain the power curve, the data set is simulated by setting ρ from 0 (under the null) to 0.25. We set $p = 120$ and $n = 60$. Figures 1 and 2 display the power curve of the test given in Qiu and Chen (2012) (abbr. QC), the sub-sampling procedures with resampling size $m = 14, 20$ and the half sampling procedure with $m = 30$. The empirical size and power of the tests are estimated from 10000 realizations. The result shows that sub-sampling with resampling size $m = 14$ leads to a smaller empirical size than the nominal level, while all the other tests have correct sizes. It can be noted that the half sampling procedure is the best one in both size accuracy and power. In addition, the sub-sampling with $m = 20$ also improves the power over the sub-sampling with $m = 14$ and the QC test. □

Let $y^*_\alpha = \inf\{y : \tilde{F}(y) \ge \alpha\}$ be the α-quantile of half-sampling estimator $\hat{F}(t)$. It can be approximated by $y^*_{L,\alpha} = \inf\{y : \hat{F}_L(y) \ge \alpha\}$. Theorem 2.2 shows convergence property of the half-sampling estimator $\tilde{F}(t)$:

THEOREM 2.2. *Let $F(t) = \mathsf{P}(n\hat{T}_n \le t)$. Suppose Assumptions 2.1 and 2.2 hold, and $\|\xi_{11}\|_{4+2\delta} < \infty$ where $0 < \delta \le 1$. Let $m = \lceil n/2 \rceil$, then under the null hypothesis H_{0a} in (2.3),*

$$\sup_t \mathsf{E}|\tilde{F}(t) - F(t)|^2 = O(n^{-\delta/(10+4\delta)}). \tag{2.20}$$

Based on Theorem 2.2, at a given significance level $0 < \alpha < 1$, we propose the test $\Phi_{a,\alpha} = \mathbf{1}(n\hat{T}_n \ge y^*_{1-\alpha})$. In practice, we use $y^*_{L,1-\alpha}$ instead of $y^*_{1-\alpha}$. The null hypothesis H_{0a} is rejected whenever $\Phi_{a,\alpha} = 1$. Power analysis is discussed in the supplementary materials. In multiple testing problems that are common in genomics, researchers use either normal approximation based method, or the normal quantile transformation of mixture of χ^2_1 distribution; see e.g. Xia, Cai and Cai (2018).

3. Testing Parametric Forms of Covariance Functions. In this section, we aim to test:

$$H_{0a} : \sigma_{jk} = \sigma_{jk,0}(\boldsymbol{\theta}) \text{ for all } (j,k) \in \mathcal{S}_1,\ \mathcal{S}_1 = \{(j,k) : 1 \le j \ne k \le p\}, \tag{3.1}$$

where the unknown parameter $\boldsymbol{\theta} = (\theta_1, ..., \theta_d)^T \subset \mathbb{R}^d$ and d is finite. We estimate $\boldsymbol{\theta}$ by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{j\neq k}^{p} (\hat{\sigma}_{jk} - \sigma_{jk,0}(\boldsymbol{\theta}))^2. \tag{3.2}$$

Assume that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_{\mathsf{P}}(\alpha_{n,p})$, where $\alpha_{n,p}$ is the rate of convergence. For example, it can be verified that $\alpha_{n,p} = (\sqrt{np})^{-1}$ for the sphericity structure $\Sigma_0(\theta) = \theta I_p$, and $\alpha_{n,p} = (\sqrt{n})^{-1}$ for the compound symmetry structure $\Sigma_0(\theta) = I_p + \theta(\mathbf{1}\mathbf{1}^T - I_p)$.

We first introduce some notation. Let θ_j be the j-th ($j = 1, ..., d$) component of the d-dimensional vector $\boldsymbol{\theta}$. Let $V = (v_{m,q})_{1\le m,q\le d}$ with

$$v_{mq} = \sum_{j\neq k}^{p} (\frac{\partial \sigma_{jk,0}(\boldsymbol{\theta})}{\partial \theta_m} \cdot \frac{\partial \sigma_{jk,0}(\boldsymbol{\theta})}{\partial \theta_q}).$$

In addition, let $\Psi = (\Psi_1, ..., \Psi_d)$ and

$$\Psi_m = (\frac{\partial \sigma_{12,0}(\boldsymbol{\theta})}{\partial \theta_m}, \frac{\partial \sigma_{13,0}(\boldsymbol{\theta})}{\partial \theta_m}, ..., \frac{\partial \sigma_{p,p-1,0}(\boldsymbol{\theta})}{\partial \theta_m})^T$$

for $1 \le m \le d$. Moreover, define

$$\Upsilon = \Psi V^{-1} \Psi'.$$

For the process $\boldsymbol{W}_i = \mathcal{W}(\boldsymbol{X}_i, \mathcal{S}_1)$ as $\mathcal{W}(\boldsymbol{X}_i, \mathcal{S}_1)$ defined in (2.4), let

$$\kappa_{\varrho}^{2+\varrho} := \mathsf{E}\left| \frac{\boldsymbol{W}_1^T \Upsilon \boldsymbol{W}_1 - \mathrm{tr}(\Upsilon\Gamma)}{|\Gamma - \Upsilon\Gamma|_F} \right|^{2+\varrho}. \tag{3.3}$$

To facilitate the theoretical analysis, the following technical conditions are considered (see Zhong et al. (2017)).

Assumption 3.1. *Assume that $\tilde{\boldsymbol{\theta}}$ is in a small neighborhood of $\boldsymbol{\theta}$. (i). For any $1 \le m, q \le d$,*

$$\begin{aligned} \sum_{j\neq k}^{p} \frac{\partial^2 \sigma_{jk,0}(\tilde{\boldsymbol{\theta}})}{\partial \theta_m \partial \theta_q} (\sigma_{jk,0}(\boldsymbol{\theta}) - \sigma_{jk}) &= o\left\{ \sum_{j\neq k}^{p} \frac{\partial \sigma_{jk,0}(\boldsymbol{\theta})}{\partial \theta_m} \frac{\partial \sigma_{jk,0}(\boldsymbol{\theta})}{\partial \theta_q} \right\}, \\ \sum_{j\neq k}^{p} \left(\frac{\partial^2 \sigma_{jk,0}(\boldsymbol{\theta})}{\partial \theta_m \partial \theta_q} (\sigma_{jk,0}(\boldsymbol{\theta}) - \sigma_{jk}) \right)^2 &= O\left\{ \sum_{j\neq k}^{p} \left(\frac{\partial \sigma_{jk,0}(\boldsymbol{\theta})}{\partial \theta_m} \frac{\partial \sigma_{jk,0}(\boldsymbol{\theta})}{\partial \theta_q} \right)^2 \right\}. \end{aligned}$$

(ii). For any $1 \le m, q, s \le d$,

$$\begin{aligned}\sum_{j\neq k}^{p}\left(\frac{\partial^3\sigma_{jk,0}(\tilde{\boldsymbol{\theta}})}{\partial\theta_m\partial\theta_q\partial\theta_s}\sigma_{jk}\right)^u &= O\left\{\sum_{j\neq k}^{p}\left(\frac{\partial^2\sigma_{jk,0}(\boldsymbol{\theta})}{\partial\theta_m\partial\theta_q}\sigma_{jk}\right)^u\right\} \text{ for } u=1,2,\\ \sum_{j\neq k}^{p}\left(\frac{\partial^2\sigma_{jk,0}(\boldsymbol{\theta})}{\partial\theta_m\partial\theta_q}\sigma_{jk,0}(\boldsymbol{\theta})\right)^2 &= O\left\{\sum_{j\neq k}^{p}\left(\frac{\partial\sigma_{jk,0}(\boldsymbol{\theta})}{\partial\theta_m}\frac{\partial\sigma_{jk,0}(\boldsymbol{\theta})}{\partial\theta_q}\right)^2\right\}.\end{aligned}$$

Similar to $\hat{T}_n = \mathcal{T}_{\mathcal{S}_1}$ in (2.1), we define $\hat{T}_n(\hat{\boldsymbol{\theta}})$ with $\sigma_{jk,0}$ in (2.2) replaced by $\sigma_{jk,0}(\hat{\boldsymbol{\theta}})$. The asymptotic behavior with estimated parameters is more complicated. The estimated parameters can play a nontrivial role, leading to dichotomous limiting behaviors, c.f. Theorem 3.1. We supplemented the Gaussian approximation results in Xu, Zhang and Wu (2014) with another type of approximating distribution when the bias term is the leading term in the test statistic. The following theorem presents the asymptotic properties of $\hat{T}_n(\hat{\boldsymbol{\theta}})$.

THEOREM 3.1. *Suppose Assumptions 2.1, 2.2 and 3.1 hold and* $\|\xi_{11}\|_{4+2\delta} < \infty$, $\kappa_\varrho < \infty$ *with* $0 < \delta \le 1, \varrho \ge 0$. *(i) If* $\kappa_0/\sqrt{n} \to 0$, *then under the null hypothesis* H_{0a} *in* (3.1),

(3.4)
$$\sup_t\left|P\left(\frac{n\hat{T}_n(\hat{\boldsymbol{\theta}})}{|\Gamma-\Upsilon\Gamma|_F}\le t\right)-P\left(\frac{1}{|\Gamma-\Upsilon\Gamma|_F}(\sum_{d=1}^{p(p-1)}\lambda_d\eta_d-\mathrm{tr}(\Gamma))\le t\right)\right|\to 0.$$

where λ_d *are eigenvalues of* $(I-\Upsilon)^{1/2}\Gamma(I-\Upsilon)^{1/2}$ *and* η_d *are i.i.d.* χ_1^2.
(ii) If $\sqrt{n}/\kappa_0 \to 0$ *and the Lindeberg condition holds,* i.e.,

$$E\left(\left|\frac{\boldsymbol{W}_1^T\Upsilon\boldsymbol{W}_1-\mathrm{tr}(\Upsilon\Gamma)}{\kappa_0|\Gamma-\Upsilon\Gamma|_F}\right|^2\mathbf{1}_{|\boldsymbol{W}_1^T\Upsilon\boldsymbol{W}_1-\mathrm{tr}(\Upsilon\Gamma)|\ge\sqrt{n}\varepsilon\kappa_0|\Gamma-\Upsilon\Gamma|_F}\right)\to 0 \tag{3.5}$$

for any $\varepsilon > 0$, *then under the null hypothesis* H_{0a} (3.1),

$$\sup_t\left|P\left(\frac{\sqrt{n}(n\hat{T}_n(\hat{\boldsymbol{\theta}})+\mathrm{tr}(\Upsilon\Gamma))}{\kappa_0|\Gamma-\Upsilon\Gamma|_F}\le t\right)-\Phi(t)\right|\to 0, \tag{3.6}$$

where Φ *is the standard Gaussian cdf.*

REMARK 3.1. *When* $\kappa_0/\sqrt{n} \to 0$, *Theorem 3.1(i) reveals that the asymptotic mean of* $n\hat{T}_n(\hat{\boldsymbol{\theta}})/|\Gamma-\Upsilon\Gamma|_F$ *is* $(\mathrm{tr}((I-\Upsilon)\Gamma)-\mathrm{tr}(\Gamma))/|\Gamma-\Upsilon\Gamma|_F = -\mathrm{tr}(\Upsilon\Gamma)/|\Gamma-\Upsilon\Gamma|_F$, *which may not converge to* 0 *as* $n, p \to \infty$. □

REMARK 3.2. *As pointed out in Chen and Qin (2010), although the term* $\sum_{i=1}^n X_i' X_i$ *in* $|\bar{X}|_2^2$ *is not useful in testing of the mean, it may impose extra restriction on* p *and* n*. Likewise, our* κ_0 *controls the effect of* $\sum_{i=1} \boldsymbol{W}_i^T \Upsilon \boldsymbol{W}_i$*, which is a bias term induced by the estimation of the unknown parameters. In practice,* $\kappa_0/\sqrt{n} \to 0$ *means that the estimation of* $\boldsymbol{\theta}$ *does not affect the asymptotic behavior of the test statistic. In contrast, if* $\sqrt{n}/\kappa_0 \to 0$*, the estimation of* $\boldsymbol{\theta}$ *incurs leading order effects of the test statistic. Then under proper normalization, we can still achieve asymptotic normality,* i.e., *Theorem 3.1(ii).* □

The test statistic $n\hat{T}_n(\hat{\boldsymbol{\theta}})$ can have two different asymptotic distributions, depending on the magnitudes of κ_0 and $\sqrt{n}$. Note that the asymptotic order of κ_0 is related to the convergence rate of $\hat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}$. We next present several examples to illustrate the asymptotic orders of κ_0 and the corresponding limiting distributions. For notational simplicity, we assume $\boldsymbol{X}_i \sim N(0, \Sigma_0)$ in the examples.

EXAMPLE 3.1. *Consider the compound symmetry covariance structure* $\Sigma_0 = I_p + \theta(\mathbf{1}\mathbf{1}^T - I_p)$ *with* $\theta \in (0,1)$ *and let* $\boldsymbol{X}_i \sim N(0, \Sigma_0)$*. It can be shown that* $\Upsilon = (p(p-1))^{-1}\mathbf{1}_{p(p-1)}\mathbf{1}_{p(p-1)}^T$, $\mathrm{tr}(\Upsilon\Gamma) = 2\theta^2(p-2)(p-3) + 4(\theta^2+\theta)(p-2) + 2(\theta^2+1)$ *and* $\mathrm{tr}(\Gamma - \Upsilon\Gamma)^2 \asymp 4(\theta-\theta^2)^2 p(p-1)(p-2)$*. Then basic calculation shows that* $\kappa_0 \asymp \sqrt{p}$*. Consequently, if* $p/n \to 0$, $(n(n-1))^{-1}\sum_{i\neq l}^n \boldsymbol{W}_i^T \boldsymbol{W}_l$ *is the leader term and we shall apply Theorem 3.1(i); in contrast, if* $n/p \to 0$, $n^{-2}\sum_{i,l}^n \boldsymbol{W}_i^T \Upsilon \boldsymbol{W}_l$ *is the leader term and the Lindeberg condition holds, then we shall apply Theorem 3.1(ii).* □

EXAMPLE 3.2. *Consider the exponential covariance class* $\Sigma_0 = \{(\sigma_{jk,0}(\theta))_{p\times p} : \sigma_{jk,0}(\theta) = \theta_1 \exp(-|j-k|/\theta_2)$ *and* $\theta_1, \theta_2 > 0\}$ *and let* $\boldsymbol{X}_i \sim N(0, \Sigma_0)$*. It can be shown that* $\mathrm{tr}(\Upsilon\Gamma) \asymp 1$, $\mathrm{tr}(\Upsilon\Gamma)^2 \asymp 1$ *and* $\mathrm{tr}(\Gamma - \Upsilon\Gamma)^2 \asymp 1$*. Then* $\kappa_0 \asymp 1$*. Thus,* $\kappa_0/\sqrt{n} \to 0$, $(n(n-1))^{-1}\sum_{i\neq l}^n \boldsymbol{W}_i^T \boldsymbol{W}_l$ *is the leader term and we shall apply Theorem 3.1(i).* □

EXAMPLE 3.3. *Consider the rational quadratic covariance structure* $\Sigma_0 = \{(\sigma_{jk,0}(\theta))_{p\times p} : \sigma_{jk,0}(\theta) = (1 + \theta_1^{-1}\theta_2^{-2}|j-k|^2)^{-\theta_1/2}$ *and* $\theta_1, \theta_2 > 0\}$ *and let* $\boldsymbol{X}_i \sim N(0, \Sigma_0)$*. If* $0 < \theta_1 < 1/2$*, by elementary calculations,* $\mathrm{tr}(\Upsilon\Gamma) \asymp p^{2-2\theta_1}$, $\mathrm{tr}(\Upsilon\Gamma)^2 \asymp p^{4-4\theta_1}$, $\mathrm{tr}(\Gamma^2) \asymp p^{4-4\theta_1}$ *and* $\mathrm{tr}(\Gamma - \Upsilon\Gamma)^2 \asymp p^{4-4\theta_1}$*. Then* $\kappa_0 \asymp 1$*. On the other hand, if* $\theta_1 > 1/2$*, then* $\mathrm{tr}(\Upsilon\Gamma) \asymp p^{3-4\theta_1}\log^2(p) + 1$, $\mathrm{tr}(\Upsilon\Gamma)^2 \asymp p^{6-8\theta_1}\log^4(p) + 1$, $\mathrm{tr}(\Gamma^2) \asymp p^2$ *and* $\mathrm{tr}(\Gamma - \Upsilon\Gamma)^2 \asymp p^2$*. This leads to* $\kappa_0 \asymp p^{2-4\theta_1}\log^2(p) + 1/p$*. Thus, on both cases,* $\kappa_0/\sqrt{n} \to 0$, $(n(n-1))^{-1}\sum_{i\neq l}^n \boldsymbol{W}_i^T \boldsymbol{W}_l$ *is the leader term and we shall apply Theorem 3.1(i).* □

Similar to Section 2.3, we can formulate a half sampling procedure. Let $\hat{\boldsymbol{\theta}}_B$ (resp. $\hat{\boldsymbol{\theta}}_{B^c}$) be the least squares estimator of equation (3.2) via $\{\boldsymbol{X}_i\}_{i\in B}$ (resp. $\{\boldsymbol{X}_i\}_{i\in B^c}$). Define $J_B(\mathcal{S}_1, \hat{\boldsymbol{\theta}})$ and $C_{B,B^c}(\mathcal{S}_1, \hat{\boldsymbol{\theta}})$ with $\sigma_{jk,0}$ in (2.13) and (2.14) replaced by $\sigma_{jk,0}(\hat{\boldsymbol{\theta}}_B)$ and $\sigma_{jk,0}(\hat{\boldsymbol{\theta}}_{B^c})$. Similarly as (2.17) and (2.18), we write the half sampling estimator and its stochastic approximation of the distribution function of $n\hat{T}_n(\hat{\boldsymbol{\theta}})$ as $\tilde{F}_{\hat{\boldsymbol{\theta}}}(t)$ and $\hat{F}_{L,\hat{\boldsymbol{\theta}}}(t)$, respectively. A more detailed version is provided in the appendix.

Thus, we have the following asymptotic property for the half-sampling estimator $\tilde{F}_{\hat{\boldsymbol{\theta}}}(t)$:

THEOREM 3.2. *Write $F_\theta(t) := \mathsf{P}(n\hat{T}_n(\boldsymbol{\theta}) \le t)$. Suppose Assumptions 2.1, 2.2 and 3.1 hold, and $\|\xi_{11}\|_{4+2\delta} < \infty$ where $0 < \delta \le 1$. If $\sqrt{n}/\kappa_0 \to 0$, then assume the Lindeberg condition (3.5) holds. If $m = \lceil n/2 \rceil \to \infty$, then under the null hypothesis H_{0a} in (3.1),*

$$\sup_t |\tilde{F}_{\hat{\boldsymbol{\theta}}}(t) - F_{\hat{\boldsymbol{\theta}}}(t)| \xrightarrow{\mathsf{P}} 0. \tag{3.7}$$

Based on Theorem 3.2, at a given significance level $0 < \alpha < 1$, we propose the test $\Phi_{a,\alpha,\hat{\boldsymbol{\theta}}} = \mathbf{1}(n\hat{T}_n(\hat{\boldsymbol{\theta}}) \ge y^*_{1-\alpha})$, where $y^*_{1-\alpha}$ is the $(1-\alpha)$th quantile of $\tilde{F}_{\hat{\boldsymbol{\theta}}}(t)$. In practice, we use $y^*_{L,1-\alpha} := \inf\{y : \hat{F}_{L,\hat{\boldsymbol{\theta}}}(y) \ge 1-\alpha\}$ instead of $y^*_{1-\alpha}$. The null hypothesis H_{0a} is rejected whenever $\Phi_{a,\alpha,\hat{\boldsymbol{\theta}}} = 1$. Note that our half sampling procedure is valid on both cases in Theorems 3.1. We shall evaluate the numeric performance of the new test method in Section 5. It is also worth noting that our test procedure $\Phi_{a,\alpha,\hat{\boldsymbol{\theta}}}$ can be applied to test general parametric structures, and do not need to estimate the bias induced by estimation of unknown parameters.

4. Testing a Given Substructure of the Precision Matrix. In this section, we consider testing

$$H_{0c} : \omega_{jk} = 0 \text{ for all } (j,k) \in \mathcal{S},$$

where $\mathcal{S}$ is the index set of the precision matrix Ω of interest. Under the Gaussian graphical model framework, a submatrix of the precision matrix characterizes the network of two groups. See De la Fuente (2010), Hudson, Reverter and Dalrymple (2009), Ideker and Krogan (2012), Jia et al. (2011), Li, Agarwal and Rajagopalan (2008), among others. In general, testing substructure of Σ is not directly useful for testing substructure of Ω. So it is essential to work on the precision matrix directly, not the covariance matrix.

A natural approach to test H_{0c} is to first construct estimators of ω_{jk}, and then base the test on the sum of squares of the entries in the index set $\mathcal{S}$.

In the high-dimensional setting, there is no sample precision matrix that one can use to approximate Ω. In this section, we assume $p = o(n)$, then we can use the inverse of sample covariance matrix as an estimate of the precision matrix. That is, $\hat{\Omega} = \hat{\Sigma}^{-1} = (\hat{\omega}_{jk})_{j,k\le p}$. We propose the following test statistic for testing the null hypothesis H_{0c},

$$\hat{G}_n = \sum_{(j,k)\in\mathcal{S}} \hat{\omega}_{jk}^2. \tag{4.1}$$

The method in this paper does not take into account any structural information, which can be useful in analyzing high-dimensional data in situations that such information is not available.

Before studying the null distribution of $\hat{G}_n$, we first introduce the following regularity conditions.

ASSUMPTION 4.1 (Sub-Gaussian). *Suppose ξ_{il}, $1 \le i \le n, 1 \le l \le N$, are i.i.d mean 0 sub-Gaussian random variables with*

$$E \exp\left(t\xi_{il}^2\right) \le K < \infty,$$

for some constant $K > 0$ and $t > 0$.

ASSUMPTION 4.2. *Assume for some constant $K_0 > 0$, $K_0^{-1} \le \lambda_{\min}(\Omega) \le \lambda_{\max}(\Omega) \le K_0$, where $\lambda_{\max}(\Omega)$ and $\lambda_{\min}(\Omega)$ denote the largest and the smallest eigenvalues of Ω, respectively.*

Assumption 4.2 on the eigenvalues is a common assumption in the high dimensional setting, for instance, Xia, Cai and Cai (2015) and Xia, Cai and Cai (2018). Note that this assumption is equivalent to $K_0^{-1} \le \lambda_{\min}(\Sigma) \le \lambda_{\max}(\Sigma) \le K_0$.

We now introduce some notation. Let $\boldsymbol{W}_i = \mathcal{W}(\boldsymbol{X}_i, S_0)$, where $S_0 = \{(j,k) : 1 \le j,k \le p\}$. Then denote the covariance matrix for $\boldsymbol{W}_i$ as $\Gamma = (\gamma_{\alpha,\alpha'})_{\alpha,\alpha'\in S_0}$. Let $\Lambda = (\Lambda_{(m_1,q_1),(m_2,q_2)})_{1\le m_1,m_2,q_1,q_2\le p}$ with

$$\Lambda_{(m_1,q_1),(m_2,q_2)} = \sum_{j,k\in\mathcal{S}} \omega_{jm_1}\omega_{jm_2}\omega_{kq_1}\omega_{kq_2},$$

where $\mathcal{S}$ is the index set of the precision matrix Ω of interest. Define

$$\tau_{\varrho}^{2+\varrho} := \mathsf{E}\left|\frac{\boldsymbol{W}_1^T\Lambda\boldsymbol{W}_1 - \mathrm{tr}(\Lambda\Gamma)}{|\Lambda\Gamma|_F}\right|^{2+\varrho}. \tag{4.2}$$

The following theorem states the asymptotic properties of $\hat{G}_n$. Let $|\mathcal{S}|$ be the cardinality of $\mathcal{S}$; let $\lambda_1 \ge \ldots \ge \lambda_{p^2} \ge 0$ be eigenvalues of $\Lambda^{1/2}\Gamma\Lambda^{1/2}$ and $f_k = (\sum_{d=1}^{p^2} \lambda_d^k)^{1/k}$, $k > 0$. Then $\mathrm{tr}(\Lambda\Gamma) = f_1$ and $|\Lambda\Gamma|_F = f_2$.

THEOREM 4.1. *Consider the linear process model* $X_{ij} = \sum_{l=1}^{N} b_{j,l}\xi_{il}$, $1 \le j \le p$, *where* ξ_{il} *are i.i.d. and satisfy Assumption 4.1. Suppose that Assumption 4.2 holds and* $\tau_\varrho < \infty$ *with* $0 < \delta \le 1, \varrho \ge 0$. *(i) If* $\tau_0/\sqrt{n} \to 0$ *and* $p^2|\mathcal{S}|f_1/(nf_2^2) \to 0$, *then under the null hypothesis* H_{0c},

$$\sup_t \left| P\left(\frac{n\hat{G}_n - f_1}{f_2} \le t\right) - P\left(\sum_{d=1}^{p(p-1)} \frac{\lambda_d}{f_2}(\eta_d - 1) \le t\right)\right| \to 0. \tag{4.3}$$

where η_d *are i.i.d.* χ_1^2. *(ii) If* $\sqrt{n}/\tau_0 \to 0$, $p^2|\mathcal{S}|f_1/(\tau_0^2 f_2^2) \to 0$, *and the Lindeberg condition holds,* i.e., *for any* $\varepsilon > 0$,

$$E\left(\left|\frac{\boldsymbol{W}_1^T \Lambda \boldsymbol{W}_1 - f_1}{\tau_0 f_2}\right|^2 \mathbf{1}_{|\boldsymbol{W}_1 \Lambda \boldsymbol{W}_1^T - f_1| \ge \sqrt{n}\varepsilon\tau_0 f_2}\right) \to 0,$$

then under the null H_{0c}, *we have the CLT*

$$\frac{\sqrt{n}(n\hat{G}_n - f_1)}{\tau_0 f_2} \Rightarrow N(0,1). \tag{4.4}$$

REMARK 4.1. *Assume* $\boldsymbol{X}_i \sim N(0,\Sigma)$. *Then, under Assumption 4.2, by elementary calculations, we have that* $E\left|\boldsymbol{W}_1^T \Lambda \boldsymbol{W}_1\right|^2 \asymp p^2|\mathcal{S}|^2$, $f_1 \asymp p|\mathcal{S}|$ *and* $f_2^2 \asymp p^2|\mathcal{S}|^2$. *This leads to* $\tau_0 = O(1)$. *Thus, we shall apply Theorem 4.1(i). Meanwhile, the allowed dimension* p *can be as large as* $p = o(n)$. □

The estimation of $\Lambda\Gamma$ is technically challenging, since correlations among the estimates of the entries of ω_{jk} for $(j,k) \in \mathcal{S}$ not only depend on the entries within the submatrix, but also heavily depend on the entries outside of it. To incorporate this dependency structure, we use the half sampling approach in previous sections. Let $B_1, ..., B_L$ be i.i.d. uniformly sampled from the class $\mathcal{B} := \{B : B \subset \{1,...,n\}, |B| = m\}$, where $m = \lceil n/2 \rceil$. Denote the empirical precision matrix estimated by $\{\boldsymbol{X}_i\}_{i\in B}$ (resp. $\{\boldsymbol{X}_i\}_{i\in B^c}$) as $\Omega(B) := (\omega_{jk,B})$ (resp. $\Omega(B^c) := (\omega_{jk,B^c})$). Then we estimate the distribution function of $F_G(t) := \mathsf{P}(n\hat{G}_n \le t)$ by

$$\tilde{F}_G(t) = \frac{1}{\binom{n}{m}} \sum_{B\in\mathcal{B}} \mathbf{1}_{m(1-m/n)(\sum_{(j,k)\in\mathcal{S}}(\omega_{jk,B} - \omega_{jk,B^c})^2) \le t}. \tag{4.5}$$

Similarly as (2.18), define its stochastic approximation $\hat{F}_{L,G}(t)$. Our half-sampling procedure is as follows.

(1) Generate a subset B of size m of $\{1,...,n\}$. Then compute the empirical precision matrix estimation $\Omega(B)$ and $\Omega(B^c)$, and obtain the half-sampling test statistic $m(1-m/n)\sum_{(j,k)\in\mathcal{S}}(\omega_{jk,B} - \omega_{jk,B^c})^2$.

(2) Repeat the above step independently L times ($L > n$) and collect all the corresponding half-sampling test statistics.
(3) Construct half-sampling estimator $\hat{F}_{L,G}(t)$, and calculate the $(1-\alpha)$-quantile of $\hat{F}_{L,G}(t)$: $y^*_{L,1-\alpha} = \inf\{y : \hat{F}_{L,G}(y) \geq 1-\alpha\}$.

The test for H_{0c} is then defined as $\Phi_{c,\alpha} = \mathbf{1}(n\hat{G}_n \geq y^*_{L,1-\alpha})$. We shall reject the null hypothesis H_{0c} at level α, whenever $\Phi_{c,\alpha} = 1$. Besides, p-value can be estimated as $\hat{F}_{L,G}(n\hat{G}_n)$.

5. Simulation Studies. In this section, we shall evaluate the numerical performance of the proposed methods based on the tests $\Phi_{a,\alpha}$, $\Phi_{a,\alpha,\theta}$ and $\Phi_{b,\alpha}$ for two subvectors (c.f. Appendix A). All these testing procedures use the half sampling approach. In practice, we recommend the sample size $n \geq 20$ and resampling replications should be at least 1000. As other resampling methods, the computational cost of our procedure is high. The test $\Phi_{a,\alpha}$ is compared with several other tests, including the test given in Qiu and Chen (2012) which is based on the sum-of-squares type statistics and the test proposed in Chernozhukov, Chetverikov and Kato (2013) which uses Gaussian Multiplier Bootstrap, and is based on the maximum deviation type statistics. These tests are denoted respectively by Qiu-Chen and CCK in the rest of this section. The test $\Phi_{a,\alpha,\theta}$ is compared with a sum-of-squares type statistic given in Zhong et al. (2017), which is denoted as ZLST. For the test $\Phi_{b,\alpha}$, it is compared with CCK only. More simulation results are given in the supplemental material.

We first consider the test for $H_{0a} : \sigma_{jk} = \sigma_{jk,0}$ for all $(j,k) \in \mathcal{S}_1$. To compare with the tests for the banded Σ proposed by Qiu and Chen (2012), we consider the case $\sigma_{jk,0} = 0$ for all $(j,k) \in \mathcal{S}_1$. The following model under the null, $\sigma_{jk} = 0$ for all $(j,k) \in \mathcal{S}_1$, is used to study the size of the tests:

$$X_{ij} = \sqrt{\Delta_j} Z_{ij}, \quad i = 1, ..., n, j = 1, ..., p, \tag{5.1}$$

where $\Delta_j = \sqrt{p} \cdot \text{Unif}(0.5, 2.5)$ for $j = 1, 2$, otherwise, $\Delta_j = \text{Unif}(0.5, 2.5)$ for $j = 3, ..., p$.

To evaluate the power, we generate multivariate random vector $X_i = (X_{i1}, \cdots, X_{ip})$ independently according to the moving average model,

$$X_{ij} = \sqrt{\Delta_j}(Z_{i,j} + 3Z_{i,j+1}), \quad i = 1, ..., n, j = 1, ..., p, \tag{5.2}$$

where three distributions are assigned to the i.i.d. Z_{ij}: (i) standard normal; (ii) centralized Gamma(4,1); and (iii) the student t_5. The last two cases are designed to assess the performance under non-normality and heavy tails.

TABLE 1
Empirical sizes for $H_{0a} : \sigma_{jk} = \sigma_{jk,0}$ for all $j \neq k$ at 5% significance, based on 2000 replications with normal, gamma and student-t innovations in Model (5.1)

p	Proposed Test $\Phi_{a,\alpha}$			Qiu-Chen			CCK		
n	20	50	100	20	50	100	20	50	100
Normal									
32	0.050	0.055	0.044	0.027	0.025	0.024	0.026	0.020	0.022
64	0.048	0.053	0.057	0.026	0.025	0.028	0.023	0.024	0.029
128	0.061	0.052	0.049	0.027	0.026	0.017	0.019	0.025	0.016
256	0.053	0.054	0.053	0.019	0.024	0.034	0.020	0.020	0.025
512	0.061	0.052	0.052	0.028	0.026	0.019	0.029	0.020	0.018
1024	0.055	0.053	0.048	0.017	0.030	0.022	0.020	0.032	0.024
Gamma									
32	0.042	0.048	0.049	0.025	0.034	0.028	0.023	0.024	0.017
64	0.048	0.055	0.048	0.020	0.023	0.017	0.018	0.022	0.021
128	0.054	0.053	0.056	0.021	0.028	0.018	0.020	0.015	0.022
256	0.062	0.051	0.054	0.035	0.025	0.023	0.016	0.019	0.019
512	0.051	0.051	0.049	0.025	0.026	0.022	0.014	0.027	0.018
1024	0.056	0.054	0.050	0.022	0.022	0.020	0.018	0.020	0.017
Student t									
32	0.041	0.049	0.050	0.023	0.024	0.022	0.014	0.029	0.018
64	0.051	0.048	0.050	0.020	0.020	0.021	0.019	0.022	0.022
128	0.053	0.047	0.052	0.017	0.018	0.030	0.014	0.018	0.024
256	0.054	0.053	0.062	0.032	0.025	0.024	0.025	0.022	0.023
512	0.050	0.054	0.044	0.012	0.022	0.019	0.014	0.027	0.028
1024	0.043	0.057	0.054	0.025	0.016	0.024	0.028	0.016	0.017

We choose a set of data dimensions $p = 32, 64, 128, 256, 512, 1024$, while the sample size is $n = 20, 50, 100$, respectively. The nominal significance level for all the tests is set at $\alpha = 0.05$. The empirical size and power of the tests, reported in Tables 1 and 2, are estimated from 2000 replications.

It can be seen from Table 1 that the estimated sizes of our proposed test $\Phi_{a,\alpha}$ are close to the nominal level 0.05 in all the cases. And the size is not sensitive to the dimensionality indicated by its robust performance. This reflects the fact that the null distribution of the test statistic is well approximated by our half-sampling approach. The empirical sizes using Qiu and Chen (2012) (Qiu-Chen) or Chernozhukov, Chetverikov and Kato (2013) (CCK) encounter serious size distortion. The actual sizes are around 0.02 for both tests. This phenomenon is expected as the Qiu-Chen test is constructed based on the asymptotic normality (cf. (2.12)), which is no longer valid for model (5.1) due to the fact that $\mathrm{tr}(\Sigma^4) \asymp \mathrm{tr}^2(\Sigma^2) \asymp p^2$ and $\rho_\Sigma \not\to 0$, and the CCK based test works for sparsity scenario.

The power results in Table 2 show that the proposed test has a much higher power than the other tests in all settings. The results show clearly that the powers of all these test improves with the sample size increases. However, the power of the Qiu-Chen test deteriorates as the dimension p grows. Overall, the new test significantly outperforms the other two tests.

TABLE 2

Empirical powers for $H_{0a} : \sigma_{jk} = \sigma_{jk,0}$ for all $j \neq k$ at 5% significance, based on 2000 replications with normal, gamma and student-t innovations in Model (5.2)

p	Proposed Test $\Phi_{a,\alpha}$			Qiu-Chen			CCK		
n	20	50	100	20	50	100	20	50	100
Normal									
32	0.255	0.590	0.903	0.176	0.507	0.863	0.192	0.539	0.852
64	0.264	0.580	0.890	0.169	0.474	0.837	0.190	0.523	0.855
128	0.266	0.608	0.924	0.164	0.462	0.820	0.194	0.527	0.879
256	0.260	0.596	0.910	0.168	0.444	0.793	0.197	0.525	0.843
512	0.253	0.581	0.892	0.173	0.467	0.786	0.173	0.552	0.858
1024	0.275	0.619	0.912	0.177	0.471	0.817	0.196	0.515	0.840
Gamma									
32	0.250	0.587	0.907	0.176	0.486	0.824	0.182	0.501	0.834
64	0.243	0.579	0.929	0.171	0.469	0.803	0.178	0.504	0.857
128	0.252	0.597	0.896	0.164	0.472	0.793	0.186	0.521	0.834
256	0.263	0.588	0.919	0.168	0.454	0.800	0.192	0.513	0.856
512	0.260	0.593	0.906	0.150	0.446	0.826	0.191	0.488	0.841
1024	0.248	0.602	0.910	0.139	0.481	0.814	0.178	0.498	0.846
Student t									
32	0.263	0.587	0.890	0.173	0.515	0.843	0.161	0.478	0.795
64	0.240	0.573	0.892	0.168	0.480	0.835	0.161	0.481	0.806
128	0.264	0.599	0.913	0.173	0.469	0.791	0.169	0.484	0.783
256	0.248	0.590	0.908	0.167	0.470	0.777	0.170	0.483	0.799
512	0.253	0.584	0.887	0.176	0.455	0.791	0.169	0.479	0.781
1024	0.267	0.606	0.891	0.180	0.466	0.786	0.162	0.483	0.778

Next, we conduct two simulation studies (Example 3.1 and Example 3.3) to evaluate the finite sample performance of the test $\Phi_{a,\alpha,\theta}$ for $H_{0a} : \sigma_{jk} = \sigma_{jk,0}(\boldsymbol{\theta})$ for all $(j,k) \in \mathcal{S}_1$. Data dimension p is chosen to be 60, 120, 240, 480, 720, 960, and the sample size is $n = 60, 120$. The empirical size and power of the tests at the nominal level 0.05 and 0.01 are reported in Tables 3, 4, 5 and 6, based on 2000 replications and 10000 replications, respectively. We also compare our test statistic $\Phi_{a,\alpha,\theta}$ with the ZLST test proposed by Zhong et al. (2017) for Gaussian data.

The null hypothesis for testing compound symmetry covariance structure is

$$H_{0a} : \Sigma_0 = I_p + \theta(\mathbf{1}\mathbf{1}^T - I_p), \quad \theta \in (0,1). \tag{5.3}$$

We generate multivariate random vector X_i according to the following model:

$$X_{ij} = \delta X_{i,j-1} + \sqrt{\theta} f_i + \sqrt{(1-\delta^2)(1-\theta)}\epsilon_{ij}, \quad i = 1,...,n, j = 1,...,p,$$

where X_{i0}, f_i and ϵ_{ij} are i.i.d. and have mean 0, variance 1. We consider three setups for the distribution of X_{i0}, f_i and ϵ_{ij}: (i) standard normal; (ii) standardized Gamma(4,1); and (iii) standardized student t_5. To study the size of the test, we generate the data by setting $\delta = 0$ and $\theta = 0.15$. In contrast, we generate the data by setting $\delta = 0.4$ and $\theta = 0.15$, to access the power of the test.

TABLE 3
Empirical sizes and powers for testing compound symmetry covariance structure in (5.3) at 5% significance, based on 2000 replications with normal, gamma and student-t innovations

p		Normal				Gamma		Student t	
		$\Phi_{a,\alpha,\theta}$		ZLST		$\Phi_{a,\alpha,\theta}$		$\Phi_{a,\alpha,\theta}$	
	n	60	120	60	120	60	120	60	120
size									
60		0.055	0.052	0.054	0.041	0.042	0.054	0.040	0.059
120		0.048	0.053	0.055	0.061	0.041	0.046	0.061	0.046
240		0.053	0.054	0.064	0.046	0.059	0.053	0.049	0.050
480		0.054	0.046	0.056	0.062	0.049	0.056	0.044	0.047
720		0.046	0.047	0.062	0.042	0.046	0.044	0.059	0.048
960		0.047	0.053	0.058	0.063	0.052	0.049	0.053	0.051
power									
60		0.918	1.000	0.863	1.000	0.878	1.000	0.856	1.000
120		0.773	1.000	0.715	0.995	0.749	0.999	0.733	0.992
240		0.606	0.934	0.556	0.915	0.566	0.939	0.558	0.928
480		0.532	0.816	0.452	0.756	0.484	0.768	0.493	0.756
720		0.476	0.696	0.417	0.631	0.455	0.703	0.465	0.687
960		0.433	0.625	0.378	0.585	0.400	0.616	0.404	0.610

Another example is to test the rational quadratic covariance structure

$$H_{0a} : \sigma_{jk,0}(\theta) = (1 + \theta_2 |j-k|^2)^{-\theta_1/2}, \quad \theta_1 > 0, \theta_2 > 0. \tag{5.4}$$

We generate random samples from multivariate model $X_i = \Gamma_X Z_i$, with $\Gamma_X \Gamma_X' = \Sigma_0(\theta)$. The components of $Z_i = (Z_{i1}, ..., Z_{ip})'$ are i.i.d. We consider the following covariance structure $\Sigma_0(\theta)$,

$$\sigma_{jk,0}(\theta) = (1-\delta)(1 + \theta_2 |j-k|^2)^{-\theta_1/2} + \delta \cdot 0.4^{|j-k|}, \quad 1 \le j, k \le p,$$

where $0 \le \delta < 1$ and $\theta_1, \theta_2 > 0$. Similarly, three distributions Z_{ij} are concerned: (i) standard normal; (ii) standardized Gamma(4,1); and (iii) s-tandardized student t_5. To study the size of the test, we generate the data by setting $\delta = 0$, $\theta_1 = 0.4$ and $\theta_2 = 0.4$. In contrast, we generate the data by setting $\delta = 0.4$, $\theta_1 = 0.4$ and $\theta_2 = 0.4$, to evaluate the power of the test.

It can be seen from Tables 3 and 5 that both our test $\Phi_{a,\alpha,\theta}$ and ZLST test control the size very well at the nominal level 0.05, for both examples. The results in Tables 4 and 6 show that the estimated sizes of our new test $\Phi_{a,\alpha,\theta}$ are close to the nominal level 0.01 in all the cases. For compound symmetry covariance structure, the estimated sizes of ZLST test are close to the nominal level 0.01 only when $n = 120$. When $n = 60$, ZLST test leads to

Table 4
Empirical sizes and powers for testing compound symmetry covariance structure in (5.3) at 1% significance, based on 10000 replications with normal, gamma and student-t innovations

p		Normal				Gamma		Student t	
		$\Phi_{a,\alpha,\theta}$		ZLST		$\Phi_{a,\alpha,\theta}$		$\Phi_{a,\alpha,\theta}$	
	n	60	120	60	120	60	120	60	120
size									
60		0.0089	0.0114	0.0127	0.0113	0.0101	0.0079	0.0084	0.0093
120		0.0093	0.0121	0.0126	0.0111	0.0100	0.0104	0.0105	0.0107
240		0.0096	0.0088	0.0137	0.0110	0.0112	0.0096	0.0097	0.0095
480		0.0104	0.0112	0.0153	0.0116	0.0079	0.0116	0.0087	0.0109
720		0.0085	0.0094	0.0161	0.0103	0.0111	0.0105	0.0117	0.0092
960		0.0107	0.0096	0.0174	0.0121	0.0102	0.0103	0.0108	0.0102
power									
60		0.807	1.000	0.779	0.999	0.794	1.000	0.780	1.000
120		0.645	1.000	0.580	0.980	0.628	0.994	0.622	0.989
240		0.455	0.889	0.408	0.845	0.449	0.857	0.436	0.845
480		0.354	0.679	0.305	0.623	0.342	0.667	0.337	0.669
720		0.325	0.558	0.298	0.499	0.309	0.528	0.305	0.536
960		0.282	0.516	0.251	0.460	0.273	0.499	0.261	0.483

Table 5
Empirical sizes and powers for testing rational quadratic covariance structure in (5.4) at 5% significance, based on 2000 replications with normal, gamma and student-t innovations

p		Normal				Gamma		Student t	
		$\Phi_{a,\alpha,\theta}$		ZLST		$\Phi_{a,\alpha,\theta}$		$\Phi_{a,\alpha,\theta}$	
	n	60	120	60	120	60	120	60	120
size									
60		0.042	0.049	0.056	0.040	0.060	0.049	0.053	0.047
120		0.051	0.047	0.045	0.048	0.047	0.053	0.043	0.058
240		0.049	0.053	0.046	0.047	0.043	0.045	0.044	0.054
480		0.049	0.054	0.059	0.045	0.044	0.045	0.048	0.048
720		0.046	0.045	0.056	0.053	0.058	0.047	0.052	0.043
960		0.056	0.051	0.051	0.048	0.050	0.053	0.051	0.047
power									
60		0.226	0.498	0.090	0.311	0.221	0.530	0.228	0.485
120		0.234	0.633	0.099	0.389	0.240	0.610	0.261	0.608
240		0.270	0.717	0.126	0.457	0.311	0.701	0.289	0.691
480		0.339	0.779	0.124	0.498	0.317	0.761	0.348	0.780
720		0.385	0.848	0.135	0.525	0.357	0.809	0.376	0.844
960		0.465	0.903	0.143	0.562	0.431	0.884	0.457	0.923

TABLE 6

Empirical sizes and powers for testing rational quadratic covariance structure in (5.4) at 1% significance, based on 10000 replications with normal, gamma and student-t innovations

p	Normal				Gamma		Student t	
	$\Phi_{a,\alpha,\theta}$		ZLST		$\Phi_{a,\alpha,\theta}$		$\Phi_{a,\alpha,\theta}$	
n	60	120	60	120	60	120	60	120
size								
60	0.0111	0.0113	0.0190	0.0231	0.0087	0.0117	0.0079	0.0125
120	0.0088	0.0104	0.0196	0.0184	0.0089	0.0120	0.0104	0.0088
240	0.0111	0.0097	0.0176	0.0170	0.0107	0.0103	0.0086	0.0084
480	0.0106	0.0114	0.0177	0.0161	0.0097	0.0101	0.0098	0.0102
720	0.0096	0.0097	0.0169	0.0141	0.0113	0.0110	0.0095	0.0097
960	0.0102	0.0093	0.0171	0.0168	0.0105	0.0096	0.0099	0.0096
power								
60	0.082	0.256	0.028	0.137	0.093	0.267	0.072	0.248
120	0.096	0.369	0.032	0.177	0.100	0.355	0.095	0.318
240	0.138	0.428	0.036	0.230	0.135	0.425	0.129	0.421
480	0.182	0.466	0.039	0.260	0.164	0.465	0.173	0.446
720	0.232	0.507	0.046	0.276	0.205	0.498	0.218	0.499
960	0.302	0.556	0.051	0.298	0.264	0.545	0.281	0.549

an inflatted size at the nominal level 0.01. For rational quadratic covariance structure, ZLST test suffers from the size distortion at the nominal level 0.01, the actual sizes are around 0.02. This reflects that our proposed method has more accurate small tail probabilities than ZLST test.

The power results show that the proposed test has a higher power than ZLST test in all settings, especially for rational quadratic covariance structure. It can be seen in Tables 3 and 4 that the estimated powers of both tests tend to decrease when the dimension p increases. However, for the rational quadratic covariance structure in Tables 5 and 6, the estimated powers rise as the dimension p increases. Overall, for both examples, the new test $\Phi_{a,\alpha,\theta}$ significantly outperforms ZLST test.

We then conduct simulations to evaluate the performance of the test for $H_{0b} : \Sigma_{12} = \Sigma_{12,0}$, where $\Sigma_{12,0}$ is pre-assigned. We partition equally the entire random vector X_i into two subvectors of $p_1 = p/2$ and $p_2 = p - p_1$. Without loss of generality, we shall always take $\Sigma_{12,0} = \mathbf{0}$ in the simulations. Factor models for X_{ij} are considered. In the size evaluation, the following linear factor model is considered:

$$X_{ij} = \begin{cases} b_{j1}^T f_{i1} + \epsilon_{ij}, & 1 \le j \le p_1, \\ b_{j2}^T f_{i2} + \epsilon_{ij}, & p_1 + 1 \le j \le p, \end{cases} \tag{5.5}$$

where b_{j1}, b_{j2} are vectors of factor loadings, f_{i1}, f_{i2} is a 2×1 vector of common factors and ϵ_{ij} is the error term, f_{i1}, f_{i2} and ϵ_{ij} are independent. All elements of b_{j1} and b_{j2}, $j = 1, ..., p$, are chosen from Unif$(0.5, 2.5)$.

In the simulation for the power, we generate the sample from the following factor model.

$$X_{ij} = \begin{cases} b_{j1}^T f_{i1} + \rho f_{i3} + \epsilon_{ij}, & 1 \le j \le p_1, \\ b_{j2}^T f_{i2} + \rho f_{i3} + \epsilon_{ij}, & p_1 + 1 \le j \le p, \end{cases} \tag{5.6}$$

where f_{i3} is a 1×1 common factor and f_{i1}, f_{i2}, f_{i3} and ϵ_{ij} are independent. In this study, ρ is chosen to be 1.5. Same distributions are considered for i.i.d sequences f_{i1},f_{i2}, f_{i3} and $(\epsilon_{ij})_{j=1}^p$. The sample sizes are taken to be $n = 20, 50, 100$, while the dimension p varies over the values $32, 64, 128, 256, 512, 1024$. The simulation results for the second test are reported in Tables 7 and 8, based on 2000 replications.

TABLE 7
Empirical sizes for $H_{0b} : \Sigma_{12} = \mathbf{0}$ at 5% significance, based on 2000 replications with normal, gamma and student-t innovations in Model (5.5)

p	Proposed Test $\Phi_{b,\alpha}$			CCK		
n	20	50	100	20	50	100
Normal						
32	0.056	0.048	0.049	0.011	0.020	0.027
64	0.045	0.057	0.043	0.012	0.015	0.018
128	0.053	0.052	0.063	0.012	0.020	0.021
256	0.054	0.059	0.049	0.009	0.012	0.023
512	0.062	0.053	0.057	0.008	0.018	0.019
1024	0.055	0.049	0.055	0.004	0.014	0.019
Gamma						
32	0.058	0.055	0.060	0.007	0.018	0.026
64	0.055	0.052	0.054	0.006	0.015	0.025
128	0.052	0.046	0.044	0.008	0.015	0.020
256	0.046	0.054	0.046	0.007	0.013	0.019
512	0.059	0.055	0.050	0.003	0.013	0.017
1024	0.053	0.045	0.049	0.003	0.012	0.016
Student t						
32	0.052	0.054	0.044	0.015	0.013	0.014
64	0.057	0.051	0.051	0.011	0.013	0.016
128	0.054	0.045	0.048	0.012	0.013	0.018
256	0.051	0.045	0.048	0.009	0.010	0.017
512	0.055	0.046	0.048	0.003	0.006	0.010
1024	0.060	0.047	0.054	0.001	0.004	0.008

Table 7 reports the empirical sizes of the proposed test $\Phi_{b,\alpha}$ (c.f. Appendix A) and the CCK test for the factor model at the 5% significance level. For each choice of p and n, it can be seen that the estimated sizes are reasonably close to the nominal level 0.05 for the proposed test, whereas the sizes of the CCK test tend to be smaller than the nominal level. It is observed that the empirical sizes of the CCK test decreases with p, but increases with n.

Table 8, which compares the powers, shows that the new test $\Phi_{b,\alpha}$ uniformly and significantly outperforms the CCK test over all choices of n and

TABLE 8
Empirical powers for $H_{0b} : \Sigma_{12} = \mathbf{0}$ at 5% significance, based on 2000 replications with normal, gamma and student-t innovations in Model (5.6)

p	Proposed Test $\Phi_{b,\alpha}$			CCK		
n	20	50	100	20	50	100
Normal						
32	0.263	0.624	0.923	0.075	0.279	0.764
64	0.274	0.608	0.916	0.060	0.257	0.595
128	0.256	0.619	0.910	0.049	0.266	0.573
256	0.263	0.621	0.916	0.045	0.234	0.553
512	0.273	0.616	0.902	0.034	0.238	0.522
1024	0.270	0.637	0.910	0.022	0.225	0.501
Gamma						
32	0.252	0.627	0.893	0.059	0.247	0.661
64	0.259	0.630	0.898	0.045	0.226	0.567
128	0.240	0.633	0.883	0.037	0.201	0.509
256	0.265	0.627	0.907	0.022	0.178	0.508
512	0.248	0.611	0.901	0.022	0.174	0.482
1024	0.256	0.628	0.918	0.016	0.133	0.402
Student t						
32	0.258	0.610	0.864	0.053	0.268	0.658
64	0.248	0.619	0.873	0.038	0.226	0.517
128	0.244	0.634	0.876	0.022	0.169	0.493
256	0.267	0.611	0.870	0.016	0.140	0.415
512	0.249	0.626	0.859	0.010	0.106	0.353
1024	0.266	0.605	0.886	0.003	0.071	0.289

p. We also observed that the powers of the CCK test improves with the sample size, but deteriorates as the dimension p increases in our simulation setting.

Acknowledgement. We are grateful to the referees, an Associate Editor and the Editor for their many helpful comments. We would also like to thank Dr. Danna Zhang for helpful comments and discussions.

References.

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis. Wiley Series in Probability and Statistics.* Wiley, New York.

ARRAS, B., MIJOULE, G., POLY, G. and SWAN, Y. (2016). A new approach to the Stein-Tikhomirov method: with applications to the second Wiener chaos and Dickman convergence. *arXiv preprint arXiv:1605.06819.*

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840.

BAUDOT, A., DE LA TORRE, V. and VALENCIA, A. (2010). Mutated genes, pathways and processes in tumours. *EMBO reports* **11** 805–810.

BENDAT, J. S. (1990). *Nonlinear system analysis and identification from random data.* Wiley-Interscience.

BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604.

Burkholder, D. L. (1988). Sharp inequalities for martingales and stochastic integrals. *Astérisque* **157** 75–94.

Cai, T. T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525.

Cai, T. T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388.

Cai, T. T., Ma, Z. and Wu, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815.

Cai, T. T., Zhang, C.-H. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144.

Casti, J. L. (1985). *Nonlinear system theory* **175**. Academic Press.

Chatterjee, S. (2008). A new method of normal approximation. *Ann. Probab.* **36** 1584–1610.

Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835.

Chen, S. X., Zhang, L.-X. and Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819.

Cherif, I. and Fnaiech, F. (2015). Nonlinear System Identification with a Real–Coded Genetic Algorithm (RCGA). *Int. J. Appl. Math. Comput. Sci.* **25** 863–875.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819.

Colussi, D., Brandi, G., Bazzoli, F. and Ricciardiello, L. (2013). Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention. *International journal of molecular sciences* **14** 16365–16385.

De la Fuente, A. (2010). From 'differential expression' to 'differential networking'–identification of dysfunctional regulatory networks in diseases. *Trends in genetics* **26** 326–333.

Demmel, J. W. (1997). *Applied numerical linear algebra* **56**. Siam.

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

Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 603–680.

Furrer, R. and Bengtsson, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivariate Anal.* **98** 227–255.

Gneiting, T., Kleiber, W. and Schlather, M. (2010). Matérn Cross-Covariance Functions for Multivariate Random Fields. *J. Amer. Statist. Assoc.* **105** 1167-1177.

He, J. and Chen, S. X. (2016). Testing super-diagonal structure in high dimensional covariance matrices. *J. Econometrics* **194** 283–297.

He, J. and Chen, S. X. (2018). High-dimensional two-sample covariance matrix testing via super-diagonals. *Statist. Sinica* **28** 2671–2696.

Hedayat, A., Wallis, W. D. et al. (1978). Hadamard matrices and their applications. *Ann. Statist.* **6** 1184–1238.

Himeno, T. and Yamada, T. (2014). Estimations for some functions of covariance matrix in high dimension under non-normality and its applications. *J. Multivariate Anal.* **130** 27–44.

Hudson, N. J., Reverter, A. and Dalrymple, B. P. (2009). A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS computational biology* **5** e1000382.

Ideker, T. and Krogan, N. J. (2012). Differential network biology. *Molecular systems biology* **8** 565.

Jia, P., Kao, C.-F., Kuo, P.-H. and Zhao, Z. (2011). A comprehensive network and pathway analysis of candidate genes in major depressive disorder. *BMC systems biology* **5** S12.

Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880.

John, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127.

John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59** 169–173.

Kalouptsidis, N. and Koukoulas, P. (2005). Blind identification of Volterra-Hammerstein systems. *IEEE Trans. Signal Process.* **53** 2777–2787.

Kelley, R. and Ideker, T. (2005). Systematic interpretation of genetic interactions using protein networks. *Nature biotechnology* **23** 561–566.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102.

Li, Y., Agarwal, P. and Rajagopalan, D. (2008). A global pathway crosstalk network. *Bioinformatics* **24** 1442–1447.

Li, J. and Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940.

Massart, P. (1990). The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality. *Ann. Probab.* **18** 1269–1283.

Onatski, A., Moreira, M. J. and Hallin, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231.

Onatski, A., Moreira, M. J. and Hallin, M. (2014). Signal detection in high dimension: The multispiked case. *Ann. Statist.* **42** 225–254.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling. Springer series in statistics*. Springer-Verlag, New York.

Pourahmadi, M. (2013). *High-Dimensional Covariance Estimation*. John Wiley & Sons, Inc.

Priestley, M. B. (1988). *Non-linear and non-stationary time series analysis*. Academic press.

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

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian Processes for Machine Learning*. MIT press Cambridge.

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

Rio, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.* **22** 146–163.

Rugh, W. J. (1981). *Nonlinear system theory*. Johns Hopkins University Press Baltimore.

Sabates-Bellver, J., Van der Flier, L. G., de Palo, M., Cattaneo, E., Maake, C., Rehrauer, H., Laczko, E., Kurowski, M. A., Bujnicki, J. M., Menigatti, M. et al. (2007). Transcriptome profile of human colorectal adenomas. *Molecular cancer research* **5** 1263–1275.

Schetzen, M. (1980). *The Volterra and Wiener theories of nonlinear systems*. Wiley.

Stein, M. L. (1999). *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York.

Vershynin, R. (2012). How close is the sample covariance matrix to the actual covariance matrix? *J. Theoret. Probab.* **25** 655–686.

Wiesel, A., Bibi, O. and Globerson, A. (2013). Time Varying Autoregressive Moving Average Models for Covariance Estimation. *IEEE Trans. Signal Process.* **61** 2791–2801.

Wu, C. F. J. (1990). On the Asymptotic Properties of the Jackknife Histogram. *Ann. Statist.* **18** 1438–1452.

Wu, W. B., Lou, Z. and Han, Y. (2018). Hypothesis Testing for High-Dimensional Data. In *Handbook of Big Data Analytics* (W. K. Härdle, H. H.-S. Lu and X. Shen, eds.) Springer International Publishing, Springer.

Wu, W. B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica* **19** 1755–1768.

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

Xia, Y., Cai, T. and Cai, T. T. (2018). Multiple Testing of Submatrices of a Precision Matrix with Applications to Identification of Between Pathway Interactions. *J. Amer. Statist. Assoc.* **113** 328–339.

Xiao, H. and Wu, W. B. (2013). Asymptotic theory for maximum deviations of sample covariance matrix estimates. *Stochastic Process. Appl.* **123** 2899–2920.

Xu, M., Zhang, D. and Wu, W. B. (2014). L^2 Asymptotics for High-Dimensional Data. *arXiv preprint arXiv:1405.7244*.

Yarlagadda, R. K. and Hershey, J. E. (2012). *Hadamard matrix analysis and synthesis: with applications to communications and signal/image processing.* Springer Science & Business Media.

Zhang, R., Peng, L. and Wang, R. (2013). Tests for covariance matrix with fixed or divergent dimension. *Ann. Statist.* **41** 2075–2096.

Zheng, S., Bai, Z. and Yao, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591.

Zhong, P.-S., Lan, W., Song, P. X. K. and Tsai, C.-L. (2017). Tests for Covariance Structures with High-dimensional Repeated Measurements. *Ann. Statist.* **45** 1185–1213.

SUPPLEMENTARY MATERIAL TO "TEST OF HIGH DIMENSIONAL COVARIANCE STRUCTURES"

By Yuefeng Han and Wei Biao Wu

University of Chicago

In this supplementary material, we shall provide testing covariance between two subvectors, power analysis, real data analysis, additional simulations, the proofs of main results in the paper and some lemmas that are useful in proofs of the paper.

The readers are referred to Appendix A for properties of the test for the off-diagonal sub-matrix. Power evaluations are presented in Appendix B. A real data example is illustrated in Appendix C. Appendix D includes more simulation results. All technical details are relegated to Appendix E.

We first formally define $J_B(\mathcal{S}_1, \hat{\boldsymbol{\theta}})$ and $C_{B,B^c}(\mathcal{S}_1, \hat{\boldsymbol{\theta}})$ in Section 3 of the paper. Let $B \subset \{1, 2, ..., n\}, B^c = \{1, ..., n\} \backslash B$, and $|B| = |B^c| = m = n/2$. Define respectively:

$$\begin{aligned} J_B(\mathcal{S}_1, \hat{\boldsymbol{\theta}}) &= \sum_{(j,k)\in\mathcal{S}_1} R_{jk}(B, \hat{\boldsymbol{\theta}}), \\ C_{B,B^c}(\mathcal{S}_1, \hat{\boldsymbol{\theta}}) &= \sum_{(j,k)\in\mathcal{S}_1} N_{jk}(B, B^c, \hat{\boldsymbol{\theta}}), \end{aligned}$$

and

$$\begin{aligned} R_{j,k}(B, \hat{\boldsymbol{\theta}}) = & \frac{1}{m(m-1)} \sum^{*}_{i_1,i_2\in B} X_{i_1j}X_{i_1k}X_{i_2j}X_{i_2k} - \frac{2}{m(m-1)(m-2)} \sum^{*}_{i_1,i_2,i_3\in B} X_{i_1j}X_{i_2j}X_{i_2k}X_{i_3k} \\ & + \frac{1}{m(m-1)(m-2)(m-3)} \sum^{*}_{i_1,i_2,i_3,i_4\in B} X_{i_1j}X_{i_2j}X_{i_3k}X_{i_4k} \\ & + \sigma_{jk,0}(\hat{\boldsymbol{\theta}}_B)^2 - \frac{2}{n}\sigma_{jk,0}(\hat{\boldsymbol{\theta}}_B) \sum^{n}_{i_1\in B} X_{i_1j}X_{i_1k} + \frac{2}{n(n-1)}\sigma_{jk,0}(\hat{\boldsymbol{\theta}}_B) \sum^{*}_{i_1,i_2\in B} X_{i_1j}X_{i_2k}. \end{aligned}$$

$$\begin{aligned} N_{jk}(B, B^c, \hat{\boldsymbol{\theta}}) = & \left(\frac{1}{m} \sum_{i_1\in B} X_{i_1j}X_{i_1k} - \frac{1}{m(m-1)} \sum^{*}_{i_1,i_2\in B} X_{i_1j}X_{i_2k} - \sigma_{jk,0}(\hat{\boldsymbol{\theta}}_B) \right) \\ & \cdot \left(\frac{1}{n-m} \sum_{i_3\in B^c} X_{i_3j}X_{i_3k} - \frac{1}{(n-m)(n-m-1)} \sum^{*}_{i_3,i_4\in B^c} X_{i_3j}X_{i_4k} - \sigma_{jk,0}(\hat{\boldsymbol{\theta}}_{B^c}) \right), \end{aligned}$$

where $\hat{\boldsymbol{\theta}}_B$ (resp. $\hat{\boldsymbol{\theta}}_{B^c}$) is the least squares estimator of equation (3.1) via $\{\boldsymbol{X}_i\}_{i\in B}$ (resp. $\{\boldsymbol{X}_i\}_{i\in B^c}$).

APPENDIX A: TESTING COVARIANCE BETWEEN TWO SUBVECTORS

Consider partition of data vector $\boldsymbol{X}$ into two subvectors of dimension p_1 and p_2, *i.e.*,

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_{(1)} \\ \boldsymbol{X}_{(2)}) \end{pmatrix},$$

and the partition of Σ by

$$\begin{pmatrix} \boldsymbol{X}_{(1)} \\ \boldsymbol{X}_{(2)}) \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

In this section, we intend to test $H_{0b} : \Sigma_{12} = \Sigma_{12,0}$ vs $H_{1b} : \Sigma_{12} \neq \Sigma_{12,0}$, where $\Sigma_{12,0}$ is pre-assigned. Recall (2.2) for M_{jk}. With the same considerations as we proposed for the estimator $\hat{T}_n$, it can be checked that an unbiased estimator of $\mathrm{tr}(\Sigma_{12} - \Sigma_{12,0})^2$ is $\hat{Q}_n = \mathcal{T}_{\mathcal{S}_2}$ with

$$\hat{Q}_n = \sum_{(j,k)\in\mathcal{S}_2} M_{jk}, \text{ where } \mathcal{S}_2 = \{(j,k) : 1 \le j \le p_1, p_1 + 1 \le k \le p\}. \tag{A.1}$$

Testing for subvectors was also considered in Li and Chen (2012) in the case of the two samples. They only established asymptotic normality under a restrictive condition on the covariances, however, our approximating distribution is more general.

Denote $\boldsymbol{U}_i = \mathcal{W}(\boldsymbol{X}_i, \mathcal{S}_2)$ and $\bar{\boldsymbol{U}}_n = \sum_{i=1}^n \boldsymbol{U}_i/n$. Then the covariance matrix $\Xi = (\gamma_{\alpha,\alpha'})_{\alpha,\alpha'\in\mathcal{S}_2}$ for $\boldsymbol{U}_i$ is $p_1p_2 \times p_1p_2$ with entries $\gamma_{\alpha,\alpha'}$ given in (2.5). The square of Frobenius norm of Ξ is

$$|\Xi|_F^2 = \sum_{\alpha,\alpha'\in\mathcal{S}_2} \gamma_{\alpha\alpha'}^2 := |\mathsf{E}(\boldsymbol{U}\boldsymbol{U}^T)|_F^2.$$

Then, a strategy similar to the previous section is to derive asymptotic distribution of $\hat{Q}_n$, construct a half-sampling estimator of the empirical distribution of $n\hat{Q}_n$, and use it to develop a test procedure.

Assumption A.1. *For some constant $C > 0$,*

$$|\Xi|_F^2 \ge C\left(\mathrm{tr}(\Sigma_{11}^2)\mathrm{tr}(\Sigma_{22}^2) + \mathrm{tr}^2(\Sigma_{12}\Sigma_{21})\right). \tag{A.2}$$

This condition is similar to Assumption 2.2 in the paper.

Then $\hat{Q}_n$ can be approximated by a linear combination of χ_1^2 random variables, given in the following theorem.

THEOREM A.1. *Under Assumption 2.1 and A.1, suppose* $\|\xi_{11}\|_{4+2\delta} < \infty$ *where* $0 < \delta \leq 1$, *if* H_{0b} *holds, then*

$$\sup_t \left| P\left(\frac{n\hat{Q}_n}{|\Xi|_F} \leq t\right) - P\left(\sum_{d=1}^{p_1p_2} \frac{\theta_d}{|\Xi|_F}(\eta_d - 1) \leq t\right)\right| = O(n^{-\delta/(10+4\delta)}), \tag{A.3}$$

where $\theta_1 \geq ... \geq \theta_{p_1p_2} \geq 0$ *are eigenvalues of* Ξ *and* η_d *are i.i.d.* χ_1^2.

Similar to the analysis on $\hat{T}_n$, the approximating distribution of $\hat{Q}_n$ under alternatives can be established in the following corollaries.

COROLLARY A.1. *Suppose* $\|\xi_{11}\|_{4+2\delta} < \infty$ *with* $0 < \delta \leq 1$. *Assume that* $\mathrm{tr}(\Sigma_{12} - \Sigma_{12,0})^2/|\Xi|_F = O(1)$. *Under Assumption 2.1 and A.1, we have that*

$$\sup_t \left| P\left(\frac{n\hat{Q}_n}{|\Xi|_F} \leq t\right) - P\left(\frac{(\boldsymbol{Z} + \sqrt{n}\mu_Z)^T(\boldsymbol{Z} + \sqrt{n}\mu_Z) - \mathrm{tr}(\Xi)}{|\Xi|_F} \leq t\right)\right| = O(n^{-\delta/(10+4\delta)}), \tag{A.4}$$

where $\boldsymbol{Z} \sim N(0, \Xi)$ *and* $\mu_Z = (\sigma_{1,p_1+1} - \sigma_{1,p_1+1,0}, \sigma_{1,p_1+2} - \sigma_{1,p_1+2,0}, ..., \sigma_{p_1,p_1+p_2} - \sigma_{p_1,p_1+p_2,0})^T$.

On the other hand, if $\mathrm{tr}(\Sigma_{12} - \Sigma_{12,0})^2/|\Xi|_F \to \infty$, *under Assumptions 2.1 and A.1, we have that* $n\hat{Q}_n/|\Xi|_F \to \infty$ *in probability.*

The asymptotic normality of $\hat{Q}_n$ is summarized in the following corollary, whose proof is trivial through the argument in Corollary 2.2.

COROLLARY A.2. *Let* $\theta_1 \geq ... \geq \theta_{p_1p_2} \geq 0$ *be eigenvalues of* Ξ. *Under conditions of Theorem A.1, the classical central limit theorem*

$$\frac{n\hat{Q}_n}{|\Xi|_F} \xrightarrow{d} N(0, 2)$$

holds if and only if

$$\rho_\Xi \to 0, \text{ as } p \to \infty. \tag{A.5}$$

If $\boldsymbol{X}_i$ *follows the linear process model, that is, under Assumption 2.1 with* $a_{j,l_1l_2...l_i} = 0$ *for all* $1 \leq l_1 < l_2 < ... < l_i \leq N, 2 \leq i \leq d, 1 \leq j \leq p$, *then, (A.5) is equivalent to* $\mathrm{tr}(\Sigma_{11}^4)\mathrm{tr}(\Sigma_{22}^4) + \mathrm{tr}^2((\Sigma_{12}\Sigma_{21})^2) = o(\mathrm{tr}^2(\Sigma_{11}^2)\mathrm{tr}^2(\Sigma_{22}^2))$; *or equivalently,* $\mathrm{tr}(\Sigma_{11}^3)\mathrm{tr}(\Sigma_{22}^3) + \mathrm{tr}((\Sigma_{12}\Sigma_{21})^3) = o(\mathrm{tr}^{3/2}(\Sigma_{11}^2)\mathrm{tr}^{3/2}(\Sigma_{22}^2))$; *or equivalently,* $\theta_1/|\Xi|_F \to 0$.

It is noted that the condition $\mathrm{tr}(\Sigma_{11}^4)\mathrm{tr}(\Sigma_{22}^4)+\mathrm{tr}^2((\Sigma_{12}\Sigma_{21})^2) = o(\mathrm{tr}^2(\Sigma_{11}^2)\mathrm{tr}^2(\Sigma_{22}^2))$ is milder than the one $\mathrm{tr}(\Sigma^4) = o(\mathrm{tr}^2(\Sigma^2))$ for linear process models; c.f. Li and Chen (2012). Since $\mathrm{tr}(\Sigma^2) \asymp \mathrm{tr}(\Sigma_{11}^2 + \Sigma_{22}^2)$ and $\mathrm{tr}(\Sigma^4) \asymp \mathrm{tr}(\Sigma_{11}^4 + \Sigma_{22}^4 + (\Sigma_{12}\Sigma_{21})^2)$, $\mathrm{tr}(\Sigma^4) = o(\mathrm{tr}^2(\Sigma^2))$ always implies the one in Corollary A.2. The condition $\mathrm{tr}(\Sigma_{11}^4)\mathrm{tr}(\Sigma_{22}^4) + \mathrm{tr}^2((\Sigma_{12}\Sigma_{21})^2) = o(\mathrm{tr}^2(\Sigma_{11}^2)\mathrm{tr}^2(\Sigma_{22}^2))$ may be violated in high dimensional data, for instance the linear factor model mentioned in Section 2.2.

To formulate a test procedure, we use the half-sampling approach to construct an unbiased and consistent estimator of the cumulative distribution function of $n\hat{Q}_n$. Consider a subset $B \subset \{1, 2, ..., n\}$ of size $m = \lceil n/2 \rceil$. Define $J_B(\mathcal{S}_2, \Sigma_{12,0})$, $C_{B,B^c}(\mathcal{S}_2, \Sigma_{12,0})$:

$$J_B(\mathcal{S}_2, \Sigma_{12,0}) = \sum_{(j,k)\in\mathcal{S}_2} R_{jk}(B, \sigma_{jk,0}), \tag{A.6}$$

$$C_{B,B^c}(\mathcal{S}_2, \Sigma_{12,0}) = \sum_{(j,k)\in\mathcal{S}_2} N_{jk}(B, B^c, \sigma_{jk,0}), \tag{A.7}$$

where $R_{jk}(B, \sigma_{jk,0})$ and $N_{jk}(B, B^c, \sigma_{jk,0})$ are defined in (2.13) and (2.14) respectively.

The half-sampling procedure samples L subsets of size $m = \lceil n/2 \rceil$ without replacement from the original n data points, uniformly at random. Let index sets $B_1, B_2, ..., B_L \subset \mathcal{B}$, with $\mathcal{B} := \{B : B \subset \{1, ..., n\}, |B| = m\}$. The empirical distribution of $\mathsf{P}(n\hat{Q}_n \le t)$ is estimated by $\tilde{F}_Q(t)$, where

$$\tilde{F}_Q(t) = \frac{1}{\binom{n}{m}} \sum_{B\in\mathcal{B}} \mathbf{1}_{m(1-m/n)(J_{B_l}(\mathcal{S}_2,\Sigma_{12,0})+J_{B_l^c}(\mathcal{S}_2,\Sigma_{12,0})-2C_{B_l,B_l^c}(\mathcal{S}_2,\Sigma_{12,0}))\le t}. \tag{A.8}$$

Similarly as (2.16), define its stochastic approximation $\hat{F}_{L,Q}(t)$. By the Dvoretzky-Kiefer-Wolfowitz-Massart inequality,

$$\mathsf{P}^*\left(\sup_t |\hat{F}_{L,Q}(t) - \tilde{F}_Q(t)| \ge u\right) \le 2e^{-2Lu^2}. \tag{A.9}$$

Define the α-quantile of half-sampling estimator $\hat{F}_Q(t)$ as follows:

$$g_\alpha^* = \inf\left\{g : \tilde{F}_Q(g) \ge \alpha\right\}, \tag{A.10}$$

which can be approximated by $g_{L,\alpha}^* = \inf\left\{g : \hat{F}_{L,Q}(g) \ge \alpha\right\}$.

THEOREM A.2. *Under Assumption 2.1 and A.1, suppose* $\|\xi_{1l}\|_{4+2\delta} < \infty$ *where* $0 < \delta \le 1$. *Let* $m = \lceil n/2 \rceil$, *then under the null hypothesis* H_{0b},

$$\sup_t \mathsf{E}|\hat{F}_Q(t) - \mathsf{P}(n\hat{Q}_n \le t)|^2 = O(n^{-\delta/(10+4\delta)}). \tag{A.11}$$

Based on the results of Theorems A.1 and A.2, at a given significance level $0 < \alpha < 1$, asymptotically α-level test can be defined as $\Phi_{b,\alpha}$ by

$$\Phi_{b,\alpha} = \mathbf{1}(n\hat{Q}_n \geq g^*_{1-\alpha})$$

where $g^*_{1-\alpha}$ is the $(1-\alpha)$th quantile of $\hat{F}_Q(t)$. In practice, we use $g^*_{L,1-\alpha}$ instead of $g^*_{1-\alpha}$. The null hypothesis H_{0b} is rejected whenever $\Phi_{b,\alpha} = 1$. Power analysis is discussed in Section B.

APPENDIX B: POWER ANALYSIS

B.1. Power analysis for testing off-diagonal covariance structure. We now turn our attention to the power analysis of $\hat{T}_n$. Let $\beta_n(\Sigma, \alpha) = \mathsf{P}(n\hat{T}_n \geq y^*_{1-\alpha}|H_{1a})$ be the power of the test under the alternative hypothesis $H_{1a} : \sigma_{jk} \neq \sigma_{jk,0}$ for some $(j,k) \in \mathcal{S}_1$, where $y^*_{1-\alpha}$ is the $(1-\alpha)$th quantile of $\tilde{F}(t)$, $0 < \alpha < 1$. Let $\boldsymbol{Y} \sim N(0, \Gamma)$ and $\mu_Y = (\sigma_{12} - \sigma_{12,0}, \sigma_{13} - \sigma_{13,0}, ..., \sigma_{p,p-1} - \sigma_{p,p-1,0})^T$. Denote $\tilde{y}_{1-\alpha} = y^*_{1-\alpha}/|\Gamma|_F$. From Theorem 2.1, we can obtain that $\mathsf{P}(n\hat{T}_n \geq y^*_{1-\alpha}) = \alpha + o(1)$. Recall Theorem 2.1 implies that $n\hat{T}_n/|\Gamma|_F$ can be approximated by $V := \sum_{d=1}^{p(p-1)} \lambda_d(\eta_d - 1)/|\Gamma|_F$ with $\mathsf{E}V^2 = 2$. Then $\mathsf{P}(|V| \geq 2\alpha^{-1/2}) \leq \alpha \mathsf{E}V^2/4 = \alpha/2$ and $\mathsf{P}(|V| \geq 2(1-\alpha)^{-1/2}) \leq (1-\alpha)/2$. Hence $\mathsf{P}(V \geq 2\alpha^{-1/2}) \leq \alpha/2$ and $\mathsf{P}(V \geq -2(1-\alpha)^{-1/2}) \geq (1+\alpha)/2$. Thus, for n large enough, we have $-2(1-\alpha)^{-1/2} < \tilde{y}_{1-\alpha} < 2\alpha^{-1/2}$. Thus $\tilde{y}_{1-\alpha} = O(1)$. From Corollary 2.1, when $\sum_{j\neq k}^{p}(\sigma_{jk} - \sigma_{jk,0})^2/|\Gamma|_F = O(1)$, we note that

$$\beta_n(\Sigma, \alpha) = \mathsf{P}\left(\frac{\boldsymbol{Y}^T\boldsymbol{Y} - \mathrm{tr}(\Gamma)}{|\Gamma|_F} \geq \tilde{y}_{1-\alpha} - \frac{n\sum_{j\neq k}(\sigma_{jk} - \sigma_{jk,0})^2}{|\Gamma|_F} - \frac{2\sqrt{n}\mu_Y^T\boldsymbol{Y}}{|\Gamma|_F}\right) + o(1) \tag{B.1}$$

Elementary calculation shows that

$$\frac{\sqrt{n}\mu_Y^T\boldsymbol{Y}}{|\Gamma|_F} = O_{\mathsf{P}}\left(\frac{\sqrt{n\mu_Y^T\Gamma\mu_Y}}{|\Gamma|_F}\right). \tag{B.2}$$

Note that

$$\mu_Y^T\Gamma\mu_Y \leq \rho(\Gamma)\sum_{j\neq k}(\sigma_{jk} - \sigma_{jk,0})^2,$$

where $\rho(\Gamma)$ is the spectral norm of Γ.

Thus, (B.1) and (B.2) indicate that the signal to noise ratio

$$SNR(\Sigma) = \sum_{j\neq k}(\sigma_{jk} - \sigma_{jk,0})^2/|\Gamma|_F$$

is instrumental in determining the power of the test. Furthermore,

$$\begin{aligned}\beta_n(\Sigma,\alpha) &\le \mathsf{P}\left(\frac{\boldsymbol{Y}^T\boldsymbol{Y} - \mathrm{tr}(\Gamma) + n\mu_Y^T\mu_Y + 2\sqrt{\boldsymbol{Y}^T\boldsymbol{Y}^T}\sqrt{\mu_Y^T\mu_Y}}{|\Gamma|_F} \ge \tilde{y}_{1-\alpha}\right) + o(1)\\ &\le \mathsf{P}\left(\frac{\boldsymbol{Y}^T\boldsymbol{Y} - \mathrm{tr}(\Gamma)}{|\Gamma|_F} \ge \frac{\tilde{y}_{1-\alpha}}{2} - \frac{nSNR(\Sigma)}{2} - \sqrt{nSNR(\Sigma)}\right) + o(1).\end{aligned}$$

Similarly,

$$\begin{aligned}\beta_n(\Sigma,\alpha) &\ge \mathsf{P}\left(\frac{\boldsymbol{Y}^T\boldsymbol{Y} - \mathrm{tr}(\Gamma) + n\mu_Y^T\mu_Y - 2\sqrt{\boldsymbol{Y}^T\boldsymbol{Y}^T}\sqrt{\mu_Y^T\mu_Y}}{|\Gamma|_F} \ge \tilde{y}_{1-\alpha}\right) + o(1)\\ &\ge \mathsf{P}\left(\frac{\boldsymbol{Y}^T\boldsymbol{Y} - \mathrm{tr}(\Gamma)}{|\Gamma|_F} \ge \tilde{y}_{1-\alpha} - nSNR(\Sigma) + 2\sqrt{nSNR(\Sigma)}\right) + o(1).\end{aligned}$$

Then under the alternative H_{1a}, when $\sum_{j\neq k}^p(\sigma_{jk} - \sigma_{jk,0})^2/|\Gamma|_F = O(1)$, the asymptotic power is bounded above and below by

$$\begin{aligned}\mathsf{P}&\left(\frac{\boldsymbol{Y}^T\boldsymbol{Y} - \mathrm{tr}(\Gamma)}{|\Gamma|_F} \ge \tilde{y}_{1-\alpha} - nSNR(\Sigma) + 2\sqrt{nSNR(\Sigma)}\right) + o(1) \le \beta_n(\Sigma,\alpha)\\ &\le \mathsf{P}\left(\frac{\boldsymbol{Y}^T\boldsymbol{Y} - \mathrm{tr}(\Gamma)}{|\Gamma|_F} \ge \frac{\tilde{y}_{1-\alpha}}{2} - \frac{nSNR(\Sigma)}{2} - \sqrt{nSNR(\Sigma)}\right) + o(1).\end{aligned}$$

By Corollary 2.1, when $\sum_{j\neq k}^p(\sigma_{jk} - \sigma_{jk,0})^2/|\Gamma|_F \to \infty$, $\beta_n(\Sigma,\alpha) \to 1$.

If the difference between σ_{jk} and $\sigma_{jk,0}$ for $j \neq k$ is not too small so that $|\Gamma|_F = O(n\sum_{j\neq k}(\sigma_{jk} - \sigma_{jk,0})^2)$, the test will be powerful. When p is fixed, this condition trivially holds while $n \to \infty$. For high dimensional data, if $n\sum_{j\neq k}(\sigma_{jk} - \sigma_{jk,0})^2/|\Gamma|_F \to \infty$, the power will converge to 1. If $n\sum_{j\neq k}(\sigma_{jk} - \sigma_{jk,0})^2/|\Gamma|_F \to 0$, the test cannot distinguish H_{0a} from H_{1a}, *i.e.* $\beta_n(\Sigma,\alpha) \to \alpha$. Furthermore, to better appreciate the power of the test, let

$$\vartheta_0 = \sqrt{\sum_{j\neq k}(\sigma_{jk} - \sigma_{jk,0})^2/(p(p-1))}$$

be the average signal strength. Then the test has a nontrivial power if ϑ_0 is at least the order of $\sqrt{|\Gamma|_F/(np^2)}$, while the order is $n^{-1/2}$ for fixed dimension situations $p = O(1)$.

Next, we consider cases involving sparse and faint signals. Let $\Sigma = I + \upsilon\upsilon'$, where $\upsilon = (\delta, ..., \delta, 0, ..., 0)'$ with the first s elements to be δ and otherwise 0. Since the signals are sparse and faint, we assume the signal strength $\delta = o(1)$ and sparse level $s = o(p)$. If $\boldsymbol{X}_i$ follows the linear process model, then basic calculation shows that $|\Gamma|_F \asymp |\Sigma|_F^2 \asymp p + s^2\delta^4$, and $\sum_{j\neq k}(\sigma_{jk} - \sigma_{jk,0}) = s^2\delta^2$. That is, $SNR(\Sigma) \asymp s^2\delta^2/p$. Hence, if $ns^2\delta^2/p \to \infty$, or equivalently $s\delta \gg \sqrt{p/n}$, the power converges to 1. As the minimum rate of detectable signals is usually $(\log(p)/n)^{1/2}$, $s\delta \gg \sqrt{p/n}$ implies $s \gg \sqrt{p}$. Then, the number of non-zero covariances is at a higher order than p. The test could be powerless if the number of signal is smaller than p. This is due to the natural of the L_2 type statistics. We thank the reviewer for pointing this out.

It is also worth noting that our test statistics based on quadratic forms are designed for general alternative hypotheses without imposing any structure assumptions on covariances. If we are interested in specific alternatives such as the spiked covariance structures for Gaussian data, one can apply for example Onatski, Moreira and Hallin (2013, 2014).

B.2. Power analysis for testing covariance between two subvectors. Let $\beta_n(\Sigma_{12}, \alpha) = \mathsf{P}(n\hat{Q}_n \geq g^*_{1-\alpha}|H_{1b})$ be the power of the test under the alternative hypothesis $H_{1b} : \sigma_{jk} \neq \sigma_{jk,0}$ for some $(j, k) \in \mathcal{S}_2$, where $g^*_{1-\alpha}$ is the $(1-\alpha)$th quantile of $\hat{F}_Q(t)$, $0 < \alpha < 1$. Let $\boldsymbol{Z} \sim N(0, \Xi)$ and $\mu_Z = (\sigma_{1p_1+1} - \sigma_{1p_1+1,0}, \sigma_{1p_1+2} - \sigma_{1p_1+2,0}, ..., \sigma_{p_1p_1+p_2} - \sigma_{p_1p_1+p_2,0})^T$. Denote $\tilde{g}_{1-\alpha} = g^*_{1-\alpha}/|\Xi|_F$, then $\tilde{g}_{1-\alpha} = O(1)$. From Corollary A.1, when $\mathrm{tr}(\Sigma_{12} - \Sigma_{12,0})^2/|\Xi|_F = O(1)$, we note that

(B.3)
$$\beta_n(\Sigma_{12}, \alpha) = \mathsf{P}\left(\frac{\boldsymbol{Z}^T\boldsymbol{Z} + 2\sqrt{n}\mu_Z^T\boldsymbol{Z} - \mathrm{tr}(\Xi)}{|\Xi|_F} \geq \tilde{g}_{1-\alpha} - \frac{n\mathrm{tr}\{(\Sigma_{12} - \Sigma_{12,0})(\Sigma_{12} - \Sigma_{12,0})^T\}}{|\Xi|_F}\right) + o(1)$$

Similar to Section B.1,

$$SNR(\Sigma_{12}) = \mathrm{tr}\{(\Sigma_{12} - \Sigma_{12,0})(\Sigma_{12} - \Sigma_{12,0})^T\}/|\Xi|_F$$

is key quantity in determining the power of the test. When when $\mathrm{tr}(\Sigma_{12} - \Sigma_{12,0})^2/|\Xi|_F = O(1)$, it can be shown that

$$\mathsf{P}\left(\frac{\boldsymbol{Z}^T\boldsymbol{Z} - \mathrm{tr}(\Xi)}{|\Xi|_F} \geq \tilde{g}_{1-\alpha} - nSNR(\Sigma_{12}) + 2\sqrt{nSNR(\Sigma_{12})}\right) + o(1) \leq \beta_n(\Sigma_{12}, \alpha)$$
$$\leq \mathsf{P}\left(\frac{\boldsymbol{Z}^T\boldsymbol{Z} - \mathrm{tr}(\Xi)}{|\Xi|_F} \geq \frac{\tilde{g}_{1-\alpha}}{2} - \frac{nSNR(\Sigma_{12})}{2} - \sqrt{nSNR(\Sigma_{12})}\right) + o(1).$$

When $\mathrm{tr}(\Sigma_{12} - \Sigma_{12,0})^2/|\Xi|_F \to \infty$, by Corollary A.1, $\beta_n(\Sigma_{12}, \alpha) \to 1$.

If $\Delta := n\text{tr}\{(\Sigma_{12} - \Sigma_{12,0})(\Sigma_{12} - \Sigma_{12,0})^T\}/\sqrt{\text{tr}(\Sigma_{11}^2)\text{tr}(\Sigma_{22}^2)} \to \infty$, the power will converge to 1. If $\Delta \to 0$, the test cannot distinguish H_{0b} from H_{1b}, *i.e.* $\beta_n(\Sigma_{12}, \alpha) \to \alpha$. A further analysis on the power can be made, which is similar to Section B.1. Details are omitted.

APPENDIX C: REAL DATA ANALYSIS

We now apply our testing procedures to the analysis of a colorectal cancers dataset (Sabates-Bellver et al. (2007)), preprocessed from NCBI's Gene Expression Omnibus, accessible through GEO Series accession number GSE8671 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE8671). This study consists of 32 subjects with colorectal adenomas. The transcriptomes (RNA) of 32 adenomatous polyps (tumor group) and segment-matched samples of normal colorectal mucosa (normal group) from the same individuals are measured. There are 54,675 genes in this microarry data. What we are interested in is to test the existence of associations between two subvectors, which can be useful for identifying sets of genes which are significantly correlated.

We consider genetic pathways of this colorectal cancers dataset. Abnormal regulation of gene pathways is the key causative factor in colorectal cancers. According to the molecular signature database, we refer to the colorectal cancer pathway as the targeted pathway of colorectal cancer. Among the 54,675 genes, 119 are mapped to this pathway. There are many pathways related to colorectal cancer, including several major signaling pathways. Assembled based on existing literature (see Baudot, De La Torre and Valencia (2010); Colussi et al. (2013)), we consider the WNT signaling pathway (263 genes), MAPK signaling pathway (475 genes), p53 signaling pathway (121 genes), mTOR signaling pathway (90 genes), GnRH signaling pathway (192 genes), Adipocytokine signaling pathway (117 genes) and Type I diabetes mellitus pathway (84 genes). Note that many of the pathways share genes while our method requires group indices to be non-overlapping since two overlapped groups are obviously dependent of each other. To remove the influence of such trivial dependence, we shall test whether the colorectal cancer pathway is correlated with these common pathways after removing overlapping genes. Let $X_i^{(1)}, ..., X_i^{(8)}$ be the expression levels of individual i from the tumor group for the colorectal cancer pathway, WNT signaling pathway, MAPK signaling pathway, p53 signaling pathway, mTOR signaling pathway, GnRH signaling pathway, Adipocytokine signaling pathway and Type I diabetes mellitus pathway, respectively. The null hypotheses are $H_{01}^T : \text{cov}(X_i^{(1)}, X_i^{(2)}) = \mathbf{0}_{68\times212}$, $H_{02}^T : \text{cov}(X_i^{(1)}, X_i^{(3)}) = \mathbf{0}_{74\times430}$, $H_{03}^T : \text{cov}(X_i^{(1)}, X_i^{(4)}) = \mathbf{0}_{109\times111}$, $H_{04}^T : \text{cov}(X_i^{(1)}, X_i^{(5)}) = \mathbf{0}_{94\times65}$,

H_{05}^T : $\mathrm{cov}(X_i^{(1)}, X_i^{(6)}) = \mathbf{0}_{102\times175}$, H_{06}^T : $\mathrm{cov}(X_i^{(1)}, X_i^{(7)}) = \mathbf{0}_{107\times105}$, H_{07}^T : $\mathrm{cov}(X_i^{(1)}, X_i^{(8)}) = \mathbf{0}_{119\times84}$. Similar null hypothesis $H_{01}^N, ..., H_{07}^N$ can be formulated for the normal group. Our proposed method using half sampling approach ($\Phi_{b,\alpha}$) is compared with the test proposed in Chernozhukov, Chetverikov and Kato (2013) which uses Gaussian Multiplier Bootstrap (denoted by C-CK), and a test method given in Qiu and Chen (2012) (abbr. QC)). The results are summarized in the following table.

TABLE 9

Estimated p-values of tests for covariances between pathway "colorectal cancer" and other different pathways, base on $N = 10^5$ half-sampling implementations

	tumor group			normal group		
pathway	$\Phi_{b,\alpha}$	QC	CCK	$\Phi_{b,\alpha}$	QC	CCK
WNT	0.00001	1.65×10^{-8}	0.38612	0.11247	0.04677	0.66285
MAPK	0.00002	4.44×10^{-16}	0.39525	0.00000	6.81×10^{-10}	0.42463
p53	0.00011	7.48×10^{-7}	0.34558	0.00018	5.79×10^{-7}	0.72479
mTOR	0.16414	0.00691	0.68261	0.00116	0.00016	0.46266
GnRH	0.00008	6.93×10^{-11}	0.31194	0.00005	3.30×10^{-8}	0.17098
Adipocytokine	0.00042	1.10×10^{-9}	0.12459	0.00240	1.81×10^{-5}	0.14529
Type I diabetes	0.02457	0.01303	0.02527	0.08692	0.01929	0.81330

It can be seen from Table 9 that the CCK test is not able to reject any null hypotheses at 5% level. All the p-values obtained by the proposed test and QC test are very small, and have similar magnitudes. Using the proposed test $\Phi_{b,\alpha}$ and QC test, H_{01}^T is rejected at 5% level, suggesting that there is a substantial correlation between the colorectal cancer pathway and WNT signaling pathway. However, for the normal group, H_{01}^N is not rejected by the proposed test, since it gives a p-value of 0.11247, while rejected by QC test with p-vlaue 0.04677. In contrast, using the proposed test, H_{04}^T for tumor group is not rejected at 5% level while H_{04}^N for normal group is rejected. The proposed test also suggests that, at 0.1% level, both H_{04}^T and H_{04}^N are rejected. Using QC test, both H_{04}^T and H_{04}^N are rejected at 5% level, and only H_{04}^N is rejected at 0.1% level. Moreover, at 0.1% level, the proposed test suggests that H_{06}^N for normal group is rejected.

APPENDIX D: ADDITIONAL SIMULATION RESULTS

In this section we present additional simulation results comparing the numerical performance of the proposed tests with that of other tests, particularly in the setting $\mathrm{tr}(\Sigma^4)/\mathrm{tr}^2(\Sigma^2) \to 0$ and the nominal level $\alpha = 0.01$ case.

D.1. Additional simulation results for the test $\Phi_{a,\alpha}$. We now consider the model (5.1) in the paper to see the size behavior, that is,

$$X_{ij} = \sqrt{\Delta_j} Z_{ij}, \quad i = 1, ..., n, j = 1, ..., p, \tag{D.1}$$

and model (5.2) in the paper for power analysis,

$$X_{ij} = \sqrt{\Delta_j}(Z_{i,j} + 3Z_{i,j+1}), \quad i = 1, ..., n, j = 1, ..., p, \tag{D.2}$$

at the nominal level $\alpha = 0.01$. Let $\Delta_j = \sqrt{p} \cdot \text{Unif}(0.5, 2.5)$ for $j = 1, 2$, and, $\Delta_j = \text{Unif}(0.5, 2.5)$ for $j = 3, ..., p$. Three distributions are assigned to the i.i.d. Z_{ij}: (i) standard normal; (ii) centralized Gamma(4,1); and (iii) the student t_5. The empirical size and power of the proposed test $\Phi_{a,\alpha}$ for H_{0a}, reported in Tables 10, are estimated from 10000 replications. The resampling implementations are 5000.

Table 10

Empirical sizes and powers for $H_{0a} : \sigma_{jk} = \sigma_{jk,0}$ for all $j \neq k$ at 1% significance level, based on 10000 replications with normal, gamma and student-t innovations in Models (D.1) and (D.2)

p	size for $\Phi_{a,\alpha}$			power for $\Phi_{a,\alpha}$		
n	20	50	100	20	50	100
Normal						
32	0.0126	0.0110	0.0112	0.142	0.470	0.822
64	0.0103	0.0095	0.0092	0.154	0.418	0.799
128	0.0138	0.0125	0.0087	0.143	0.431	0.761
256	0.0069	0.0118	0.0116	0.160	0.412	0.787
512	0.0130	0.0116	0.0100	0.143	0.400	0.751
1024	0.0123	0.0091	0.0104	0.140	0.394	0.787
Gamma						
32	0.0130	0.0114	0.0102	0.145	0.420	0.813
64	0.0124	0.0093	0.0101	0.167	0.444	0.797
128	0.0138	0.0097	0.0120	0.145	0.398	0.781
256	0.0117	0.0111	0.0096	0.170	0.442	0.765
512	0.0084	0.0103	0.0115	0.162	0.402	0.761
1024	0.0138	0.0088	0.0124	0.164	0.379	0.794
Student t						
32	0.0120	0.0105	0.0088	0.138	0.437	0.803
64	0.0091	0.0096	0.0102	0.141	0.400	0.754
128	0.0104	0.0113	0.0105	0.151	0.439	0.775
256	0.0103	0.0112	0.0123	0.156	0.407	0.747
512	0.0101	0.0102	0.0094	0.148	0.413	0.741
1024	0.0113	0.0087	0.0117	0.139	0.421	0.716

In Table 10, the results indicate that our test control the size well when $n = 50, 100$. We can observe notable fluctuation of empirical size when sample size $n = 20$. Under the alternative hypothesis, the power rises as the sample size increases.

Then, we consider the setting $\mathrm{tr}(\Sigma^4)/\mathrm{tr}^2(\Sigma^2) \to 0$. We still consider the same model (D.1) for the size analysis,

$$X_{ij} = \sqrt{\Delta_j} Z_{ij}, \quad i = 1, ..., n, j = 1, ..., p, \tag{D.3}$$

and the model

$$X_{ij} = \sqrt{\Delta_j}(Z_{i,j} + 0.2Z_{i,j+1}), \quad i = 1, ..., n, j = 1, ..., p, \tag{D.4}$$

for power analysis. Here let $\Delta_j = \mathrm{Unif}(0.5, 2.5)$ for $j = 1, ..., p$. We present the simulation results for H_{0a} in Tables 11, 12, 13 and 14. The resampling implementations are 5000.

TABLE 11

Empirical sizes for $H_{0a} : \sigma_{jk} = \sigma_{jk,0}$ for all $j \neq k$ at 5% significance level, based on 2000 replications with normal, gamma and student-t innovations in Model (D.3)

p	Proposed Test $\Phi_{a,\alpha}$			Qiu-Chen		
n	20	50	100	20	50	100
Normal						
32	0.048	0.045	0.050	0.072	0.048	0.049
64	0.045	0.056	0.055	0.063	0.057	0.056
128	0.052	0.055	0.048	0.060	0.056	0.059
256	0.060	0.045	0.047	0.070	0.056	0.053
512	0.061	0.045	0.048	0.057	0.053	0.046
1024	0.055	0.049	0.057	0.067	0.051	0.051
Gamma						
32	0.041	0.046	0.049	0.052	0.048	0.043
64	0.050	0.051	0.051	0.067	0.049	0.053
128	0.048	0.045	0.048	0.069	0.057	0.059
256	0.051	0.041	0.043	0.062	0.052	0.041
512	0.045	0.053	0.046	0.066	0.055	0.051
1024	0.042	0.054	0.045	0.061	0.056	0.042
Student t						
32	0.041	0.058	0.050	0.055	0.049	0.044
64	0.056	0.055	0.044	0.064	0.055	0.059
128	0.043	0.048	0.043	0.063	0.050	0.053
256	0.043	0.047	0.053	0.069	0.055	0.056
512	0.052	0.051	0.059	0.060	0.055	0.050
1024	0.049	0.052	0.055	0.067	0.057	0.052

The results in Table 11 show that, at the nominal level 5%, both our test and Qiu-Chen test control the size very well when $n = 50, 100$. When sample size $n = 20$, our proposed test still control the size very well, but the empirical size of Qiu-Chen test is a bit larger than the nominal level 5%. It can be seen from Table 13 that the estimated sizes of our proposed test are close to the nominal level 1%, when $n = 50, 100$. When $n = 20$, our proposed test encounters some fluctuations of empirical size. In contrast, Qiu-Chen test only controls the size well when $n = 100$. When $n = 50$, Qiu-Chen test tends to be a bit larger than the nominal level 1%. With small sample size $n = 20$,

TABLE 12
Empirical powers for $H_{0a} : \sigma_{jk} = \sigma_{jk,0}$ for all $j \neq k$ at 5% significance level, based on 2000 replications with normal, gamma and student-t innovations in Model (D.4)

p	Proposed Test $\Phi_{a,\alpha}$			Qiu-Chen		
n	20	50	100	20	50	100
Normal						
32	0.132	0.398	0.884	0.090	0.323	0.818
64	0.154	0.401	0.894	0.113	0.359	0.818
128	0.148	0.411	0.901	0.118	0.342	0.854
256	0.154	0.414	0.912	0.133	0.380	0.846
512	0.152	0.402	0.909	0.123	0.332	0.862
1024	0.162	0.425	0.921	0.145	0.402	0.856
Gamma						
32	0.166	0.371	0.888	0.111	0.323	0.802
64	0.162	0.391	0.899	0.134	0.361	0.817
128	0.154	0.405	0.912	0.136	0.351	0.869
256	0.182	0.415	0.928	0.116	0.360	0.883
512	0.169	0.428	0.927	0.140	0.353	0.885
1024	0.175	0.423	0.922	0.129	0.386	0.856
Student t						
32	0.158	0.379	0.880	0.137	0.334	0.833
64	0.188	0.396	0.898	0.134	0.330	0.840
128	0.174	0.416	0.900	0.138	0.366	0.881
256	0.170	0.433	0.915	0.131	0.347	0.873
512	0.162	0.413	0.924	0.125	0.356	0.861
1024	0.176	0.415	0.934	0.135	0.368	0.883

TABLE 13
Empirical sizes for $H_{0a} : \sigma_{jk} = \sigma_{jk,0}$ for all $j \neq k$ at 1% significance level, based on 10000 replications with normal, gamma and student-t innovations in Model (D.3)

p	Proposed Test $\Phi_{a,\alpha}$			Qiu-Chen		
n	20	50	100	20	50	100
Normal						
32	0.0121	0.0088	0.0095	0.0161	0.0139	0.0101
64	0.0082	0.0094	0.0087	0.0172	0.0111	0.0113
128	0.0134	0.0114	0.0093	0.0168	0.0122	0.0108
256	0.0135	0.0089	0.0112	0.0170	0.0140	0.0114
512	0.0126	0.0087	0.0104	0.0196	0.0119	0.0114
1024	0.0124	0.0094	0.0105	0.0175	0.0137	0.0107
Gamma						
32	0.0131	0.0103	0.0084	0.0162	0.0126	0.0086
64	0.0110	0.0095	0.0084	0.0166	0.0136	0.0102
128	0.0124	0.0114	0.0103	0.0165	0.0107	0.0090
256	0.0138	0.0086	0.0097	0.0154	0.0123	0.0107
512	0.0123	0.0085	0.0107	0.0170	0.0108	0.0097
1024	0.0084	0.0104	0.0105	0.0157	0.0115	0.0103
Student t						
32	0.0130	0.0115	0.0110	0.0128	0.0085	0.0100
64	0.0104	0.0101	0.0095	0.0175	0.0110	0.0086
128	0.0083	0.0081	0.0105	0.0166	0.0102	0.0107
256	0.0135	0.0103	0.0104	0.0167	0.0126	0.0118
512	0.0124	0.0111	0.0113	0.0172	0.0124	0.0102
1024	0.0137	0.0113	0.0097	0.0181	0.0115	0.0130

TABLE 14
Empirical powers for $H_{0a}: \sigma_{jk} = \sigma_{jk,0}$ for all $j \neq k$ at 1% significance level, based on 10000 replications with normal, gamma and student-t innovations in Model (D.4)

p	Proposed Test $\Phi_{a,\alpha}$			Qiu-Chen		
n	20	50	100	20	50	100
Normal						
32	0.077	0.202	0.807	0.051	0.180	0.681
64	0.070	0.210	0.781	0.056	0.195	0.765
128	0.077	0.237	0.776	0.055	0.184	0.726
256	0.078	0.218	0.780	0.054	0.185	0.700
512	0.071	0.231	0.793	0.055	0.191	0.731
1024	0.083	0.258	0.809	0.055	0.197	0.713
Gamma						
32	0.083	0.204	0.701	0.045	0.159	0.679
64	0.078	0.191	0.734	0.057	0.185	0.709
128	0.080	0.223	0.747	0.056	0.199	0.695
256	0.072	0.234	0.783	0.052	0.191	0.729
512	0.081	0.239	0.772	0.054	0.196	0.711
1024	0.078	0.239	0.794	0.052	0.184	0.709
Student t						
32	0.078	0.204	0.695	0.050	0.189	0.641
64	0.074	0.213	0.739	0.050	0.197	0.675
128	0.072	0.210	0.768	0.048	0.191	0.717
256	0.069	0.224	0.747	0.052	0.193	0.706
512	0.072	0.238	0.786	0.053	0.196	0.710
1024	0.074	0.224	0.776	0.054	0.190	0.714

Qiu-Chen test has size distortion. From the results of Tables 12 and 14, the proposed test has a higher power than Qiu-Chen test in our simulation settings. Overall, these numerical results corroborates that our proposed test outperforms Qiu-Chen test under the setting $\operatorname{tr}(\Sigma^4)/\operatorname{tr}^2(\Sigma^2) \to 0$.

D.2. Additional simulation results for the test $\boldsymbol{\Phi_{b,\alpha}}$. We now consider the following model (D.5) for size,

$$X_{ij} = \begin{cases} b_{j1}^T f_{i1} + \epsilon_{ij}, & 1 \leq j \leq p_1, \\ b_{j2}^T f_{i2} + \epsilon_{ij}, & p_1 + 1 \leq j \leq p, \end{cases} \tag{D.5}$$

and model (D.6) for power analysis,

$$X_{ij} = \begin{cases} b_{j1}^T f_{i1} + \rho f_{i3} + \epsilon_{ij}, & 1 \leq j \leq p_1, \\ b_{j2}^T f_{i2} + \rho f_{i3} + \epsilon_{ij}, & p_1 + 1 \leq j \leq p, \end{cases} \tag{D.6}$$

at the nominal level $\alpha = 0.01$. All elements of factor loadings b_{j1} and b_{j2}, $j = 1, ..., p$, are chosen from Unif$(0.5, 2.5)$. Let f_{i1}, f_{i2} be 2×1 vectors of common factors, and f_{i3} be a 1×1 common factor. Besides, f_{i1}, f_{i2}, f_{i3} and ϵ_{ij} are independent. Same distributions are considered for i.i.d sequences f_{i1}, f_{i2}, f_{i3} and $(\epsilon_{ij})_{j=1}^p$: (i) standard normal; (ii) centralized Gamma(4,1);

and (iii) the student t_5. The empirical size and power of the tests for H_{0b} at the nominal level 0.01, reported in Table 15, are estimated from 10000 replications. The resampling implementations are 5000.

TABLE 15
Empirical sizes and powers for $H_{0b} : \Sigma_{12} = \mathbf{0}$ at 1% significance level, based on 10000 replications with normal, gamma and student-t innovations in Models (D.5) and (D.6)

p	size for $\Phi_{b,\alpha}$			power for $\Phi_{b,\alpha}$		
n	20	50	100	20	50	100
Normal						
32	0.0134	0.0125	0.0106	0.102	0.365	0.714
64	0.0094	0.0105	0.0088	0.120	0.396	0.783
128	0.0100	0.0079	0.0110	0.137	0.382	0.786
256	0.0095	0.0126	0.0110	0.122	0.406	0.797
512	0.0134	0.0110	0.0115	0.147	0.396	0.796
1024	0.0074	0.0103	0.0100	0.134	0.386	0.802
Gamma						
32	0.0130	0.0123	0.0103	0.108	0.357	0.740
64	0.0110	0.0097	0.0110	0.095	0.390	0.776
128	0.0139	0.0085	0.0121	0.126	0.389	0.738
256	0.0086	0.0115	0.0079	0.146	0.399	0.766
512	0.0126	0.0103	0.0088	0.121	0.374	0.798
1024	0.0110	0.0116	0.0105	0.121	0.383	0.777
Student t						
32	0.0111	0.0119	0.0092	0.127	0.387	0.648
64	0.0128	0.0100	0.0113	0.140	0.362	0.649
128	0.0147	0.0105	0.0095	0.118	0.351	0.712
256	0.0123	0.0108	0.0100	0.131	0.398	0.676
512	0.0095	0.0114	0.0102	0.105	0.381	0.693
1024	0.0136	0.0117	0.0099	0.121	0.390	0.733

In Table 15, the results indicate that our test control the size well when $n = 50, 100$. Similarly, we observe notable fluctuation of empirical size when sample size $n = 20$. Under the alternative hypothesis, the power rises as the sample size increases.

APPENDIX E: PROOFS OF MAIN RESULTS IN THE PAPER

Throughout the proof, assume without loss of generality that $\boldsymbol{\mu} = \mathbf{0}$. Denote by C a constant that is independent of n and p and its value may change from place to place.

LEMMA E.1. *Considering $\{\boldsymbol{X}_i\}_{i=1}^n$ with Assumption 2.1, we have*

$$E|\boldsymbol{W}_1^T\boldsymbol{W}_2|^{2+\delta} \le K_\delta^W|\Gamma|_F^{2+\delta}, \tag{E.1}$$

$$E|\boldsymbol{U}_1^T\boldsymbol{U}_2|^{2+\delta} \le K_\delta^U|\Xi|_F^{2+\delta}, \tag{E.2}$$

where K_δ^W and K_δ^U are bounded constants, only depending on δ, ν and $\|\xi_{11}\|_{4+2\delta}$.

Proof. For the convenience of presentation, we assume $d = 2$ in Assumption 2.1. If $d > 2$, the argument shown as follows still can be applied to prove the Lemma with more tedious calculations. Recall $\mathsf{E}\left(\xi_{11}^3\right) = 0$, $\mathsf{Var}\left(\xi_{11}^2\right) = \nu > 0$, $X_j = \sum_{l_1<l_2}^N a_{j,l_1l_2}\xi_{1l_1}\xi_{1l_2} + \sum_{l_1=1}^N b_{j,l_1}\xi_{1l_1}$. Then $\sigma_{jk} = \sum_{l_1<l_2}^N a_{j,l_1l_2}a_{k,l_1l_2} + \sum_{l_1=1}^N b_{j,l_1}b_{k,l_1}$. We rewrite $X_jX_k - \sigma_{jk}$ as follows,

$$\begin{aligned}
X_jX_k - \sigma_{jk} &= \left(\sum_{l_1<l_2}^N a_{j,l_1l_2}\xi_{1l_1}\xi_{1l_2} + \sum_{l_1=1}^N b_{j,l_1}\xi_{1l_1}\right)\left(\sum_{l_1<l_2}^N a_{k,l_1l_2}\xi_{1l_1}\xi_{1l_2} + \sum_{l_1=1}^N b_{k,l_1}\xi_{1l_1}\right) - \sigma_{jk}\\
&= \sum_{l_1<l_2<l_3<l_4} c_{(j,k),l_1l_2l_3l_4,(1)}\xi_{1l_1}\xi_{1l_2}\xi_{1l_3}\xi_{1l_4}\\
&\quad + \sum_{l_1<l_2<l_3} c_{(j,k),l_1l_2l_3,(1)}(\xi_{1l_1}^2 - 1)\xi_{1l_2}\xi_{1l_3}\\
&\quad + \sum_{l_1<l_2<l_3} c_{(j,k),l_1l_2l_3,(2)}\xi_{1l_1}(\xi_{1l_2}^2 - 1)\xi_{1l_3}\\
&\quad + \sum_{l_1<l_2<l_3} c_{(j,k),l_1l_2l_3,(3)}\xi_{1l_1}\xi_{1l_2}(\xi_{1l_3}^2 - 1)\\
&\quad + \sum_{l_1<l_2<l_3} c_{(j,k),l_1l_2l_3,(4)}\xi_{1l_1}\xi_{1l_2}\xi_{1l_3}\\
&\quad + \sum_{l_1<l_2} c_{(j,k),l_1l_2,(1)}(\xi_{1l_1}^2 - 1)(\xi_{1l_2}^2 - 1)\\
&\quad + \sum_{l_1<l_2} c_{(j,k),l_1l_2,(2)}(\xi_{1l_1}^2 - 1)\xi_{1l_2}\\
&\quad + \sum_{l_1<l_2} c_{(j,k),l_1l_2,(3)}\xi_{1l_1}(\xi_{1l_2}^2 - 1)\\
&\quad + \sum_{l_1<l_2} c_{(j,k),l_1l_2,(4)}\xi_{1l_1}\xi_{1l_2}\\
&\quad + \sum_{l_1} c_{(j,k),l_1,(1)}(\xi_{1l_1}^2 - 1) + \sum_{l_1} c_{(j,k),l_1,(2)}\xi_{1l_1},
\end{aligned} \tag{E.3}$$

where

$$
\begin{aligned}
c_{(j,k),l_1l_2l_3l_4,(1)} &= a_{j,l_1l_2}a_{k,l_3l_4} + a_{j,l_1l_3}a_{k,l_2l_4} + a_{j,l_1l_4}a_{k,l_2l_3} + a_{j,l_3l_4}a_{k,l_1l_2} + a_{j,l_2l_4}a_{k,l_1l_3} + a_{j,l_2l_3}a_{k,l_1l_4},\\
c_{(j,k),l_1l_2l_3,(1)} &= a_{j,l_1l_2}a_{k,l_1l_3} + a_{j,l_1l_3}a_{k,l_1l_2},\\
c_{(j,k),l_1l_2l_3,(2)} &= a_{j,l_1l_2}a_{k,l_2l_3} + a_{j,l_2l_3}a_{k,l_1l_2},\\
c_{(j,k),l_1l_2l_3,(3)} &= a_{j,l_1l_3}a_{k,l_2l_3} + a_{j,l_2l_3}a_{k,l_1l_3},\\
c_{(j,k),l_1l_2l_3,(4)} &= b_{j,l_1}a_{k,l_2l_3} + b_{j,l_2}a_{k,l_1l_3} + b_{j,l_3}a_{k,l_1l_2} + a_{j,l_1l_2}b_{k,l_3} + a_{j,l_1l_3}b_{k,l_2} + a_{k,l_2l_3}b_{j,l_1},\\
c_{(j,k),l_1l_2,(1)} &= a_{j,l_1l_2}a_{k,l_1l_2},\\
c_{(j,k),l_1l_2,(2)} &= a_{j,l_1l_2}b_{k,l_1} + b_{j,l_1}a_{k,l_1l_2},\\
c_{(j,k),l_1l_2,(3)} &= a_{j,l_1l_2}b_{k,l_2} + b_{j,l_2}a_{k,l_1l_2},\\
c_{(j,k),l_1l_2,(4)} &= \sum_{l_3>l_2}(a_{j,l_1l_3}a_{k,l_2l_3} + a_{j,l_2l_3}a_{k,l_1l_3}) + \sum_{l_1<l_3<l_2}(a_{j,l_1l_3}a_{k,l_3l_2} + a_{j,l_3l_2}a_{k,l_1l_3})\\
&\quad + \sum_{l_3<l_1}(a_{j,l_3l_1}a_{k,l_3l_2} + a_{j,l_3l_2}a_{k,l_3l_1}) + b_{j,l_1}b_{k,l_2} + b_{j,l_2}b_{k,l_1},\\
c_{(j,k),l_1,(1)} &= \sum_{l_2>l_1} a_{j,l_1l_2}a_{k,l_1l_2} + \sum_{l_2<l_1} a_{j,l_2l_1}a_{k,l_2l_1} + b_{j,l_1}b_{k,l_1},\\
c_{(j,k),l_1,(2)} &= \sum_{l_2>l_1}(a_{j,l_1l_2}b_{k,l_1} + b_{j,l_1}a_{k,l_1,l_2}) + \sum_{l_2<l_1}(a_{j,l_2l_1}b_{k,l_1} + b_{j,l_1}a_{k,l_2l_1}).
\end{aligned}
$$

Note that each term in the above equation (E.3) of $X_jX_k - \sigma_{jk}$ is uncorrelated, for instance, $\mathsf{E}((\xi_{1l_1}^2 - 1)(\xi_{1l_2}^2 - 1)\cdot(\xi_{1l_1}^2 - 1)) = 0$. Let η be a vector expanded by $\xi_{1l_1}\xi_{1l_2}\xi_{1l_3}\xi_{1l_4}$, $(\xi_{1l_1}^2-1)\xi_{1l_2}\xi_{1l_3}$, $\xi_{1l_1}(\xi_{1l_2}^2-1)\xi_{1l_3}$, $\xi_{1l_1}\xi_{1l_2}(\xi_{1l_3}^2-1)$, $\xi_{1l_1}\xi_{1l_2}\xi_{1l_3}$, $(\xi_{1l_1}^2 - 1)(\xi_{1l_2}^2 - 1)$, $(\xi_{1l_1}^2 - 1)\xi_{1l_2}$, $\xi_{1l_1}(\xi_{1l_2}^2 - 1)$, $\xi_{1l_1}^2\xi_{1l_2}^2$, $(\xi_{1l_1}^2 - 1)$ and ξ_{1l_1} in alphabetical order. For

$$\alpha = (l_1, l_2, l_3, l_4, n)$$

with n being $1, 2, 3, 4$, define $L_{(j,k),\alpha} := c_{(j,k),\alpha}$. That is, $L_{(j,k),(l_1,l_2,l_3,l_4,1)} := c_{(j,k),l_1l_2l_3l_4,(1)}$, $L_{(j,k),(l_1,l_2,l_3,\cdot,3)} := c_{(j,k),l_1l_2l_3,(3)}$, $L_{(j,k),(l_1,\cdot,\cdot,\cdot,1)} := c_{(j,k),l_1,(1)}$, etc. Then, we can write

$$X_jX_k - \sigma_{jk} := \sum_{\alpha} L_{(j,k),\alpha}\eta_\alpha := L_{(j,k),\cdot}^T\eta \tag{E.4}$$

and

$$\boldsymbol{W} = L\eta.$$

We also have

$$\Gamma = \mathsf{E}(\boldsymbol{W}\boldsymbol{W}^T) = L\mathsf{E}(\eta\eta^T)L^T,$$

where $\mathsf{E}(\eta\eta^T)$ is a diagonal matrix with its elements being 1, ν or ν^2.

Denote $D = L^T L := (d_{\alpha,\beta})$. Simple calculation shows that

$$d_{\alpha,\beta} = \sum_{(j,k)\in\mathcal{S}_1} c_{(j,k),\alpha} c_{(j,k),\beta}.$$

Since

$$|\Gamma|_F^2 = \mathrm{tr}((L\mathsf{E}(\eta\eta^T)L^T))^2 = \mathrm{tr}((\mathsf{E}(\eta\eta^T)L^T L))^2,$$

we can show that

$$|\Gamma|_F^2 \le \max\{1,\nu^4\}\sum_{\alpha,\beta} d_{\alpha,\beta}^2.$$

Denote $h = 2+\delta$. Let η and η^* be i.i.d. Followed from (E.3), we have, by elementary calculations that

$$\begin{aligned}
&\|\boldsymbol{W}_1^T\boldsymbol{W}_2\|_h^2 = \|(L\eta)^T L\eta^*\|_h^2 \\
&\le C_\delta \| \sum_{(j,k)\in\mathcal{S}_1}\sum_{l_1<l_2<l_3<l_4} c_{(j,k),l_1l_2l_3l_4,(1)}\xi_{1l_1}\xi_{1l_2}\xi_{1l_3}\xi_{1l_4}\sum_{l_5<l_6<l_7<l_8} c_{(j,k),l_5l_6l_7l_8,(1)}\xi_{1l_5}^*\xi_{1l_6}^*\xi_{1l_7}^*\xi_{1l_8}^*\|_h^2 \\
&\quad + C_\delta \| \sum_{(j,k)\in\mathcal{S}_1}\sum_{l_1<l_2<l_3} c_{(j,k),l_1l_2l_3,(1)}(\xi_{1l_1}^2-1)\xi_{1l_2}\xi_{1l_3}\sum_{l_4<l_5<l_6} c_{(j,k),l_4l_5l_6,(1)}(\xi_{1l_4}^2-1)\xi_{1l_5}\xi_{1l_6}\|_h^2 \\
&\quad + \cdots \\
&\quad + C_\delta \| \sum_{(j,k)\in\mathcal{S}_1}\sum_{l_1<l_2<l_3<l_4} c_{(j,k),l_1l_2l_3l_4,(1)}\xi_{1l_1}\xi_{1l_2}\xi_{1l_3}\xi_{1l_4}\sum_{l_5<l_6<l_7} c_{(j,k),l_5l_6l_7,(1)}(\xi_{1l_5}^2-1)\xi_{1l_6}\xi_{1l_7}\|_h^2 \\
&\quad + \cdots \\
&:= C_\delta(R_1+R_2+\cdots+R_{55}),
\end{aligned}$$

for some bounded positive constant C_δ. Applying Lemma F.1 (Burkholder's inequality) to each step, we have

$$\begin{aligned}
&R_1 \le (h-1)\sum_{l_4>l_3}\|\xi_{1l_4}\|_h^2\|\sum_{l_1<l_2<l_3}\sum_{l_5<l_6<l_7<l_8} d_{(l_1,l_2,l_3,l_4,1),(l_5,l_6,l_7,l_8,1)}\xi_{1l_1}\xi_{1l_2}\xi_{1l_3}\xi_{1l_5}^*\xi_{1l_6}^*\xi_{1l_7}^*\xi_{1l_8}^*\|_h^2 \\
&\le (h-1)^2\sum_{l_4>l_3>l_2}\|\xi_{1l_4}\|_h^2\|\xi_{1l_3}\|_h^2\|\sum_{l_1<l_2}\sum_{l_5<l_6<l_7<l_8} d_{(l_1,l_2,l_3,l_4,1),(l_5,l_6,l_7,l_8,1)}\xi_{1l_1}\xi_{1l_2}\xi_{1l_5}^*\xi_{1l_6}^*\xi_{1l_7}^*\xi_{1l_8}^*\|_h^2 \\
&\ldots \\
&\le (h-1)^8\sum_{l_1<l_2<l_3<l_4}\sum_{l_5<l_6<l_7<l_8} d_{(l_1,l_2,l_3,l_4,1),(l_5,l_6,l_7,l_8,1)}^2\|\xi_{11}\|_h^{16}
\end{aligned}$$

Denote

$$\bar{K}_\delta = (\max\{(h-1)^2\|\xi_{11}\|_h^4,(h-1)\|\xi_{11}^2-1\|_h^2,1\})^4.$$

Adopting similar arguments to $R_2, ..., R_{55}$, we can obtain

$$\|\boldsymbol{W}_1^T \boldsymbol{W}_2\|_h^2 \le C_\delta \bar{K}_\delta \sum_{\alpha,\beta} d_{\alpha,\beta}^2.$$

Then (E.1) follows by setting

$$K_\delta^W := \left(\frac{C_\delta \bar{K}_\delta}{\min\{1, \nu^4\}} \right)^{(2+\delta)/2}.$$

Clearly, $K_\delta^W < \infty$.

Following the same arguments, we can show (E.2). □

Proof of Theorem 2.1. For the convenience of presentation, we assume $d = 2$ in Assumption 2.1. If $d > 2$, the argument shown as follows still can be applied to prove the theorem with more tedious calculations. Denote

$$\tilde{T}_n = \sum_{j \neq k}^{p} \left(\frac{1}{n(n-1)} \sum_{i_1,i_2}^{*} X_{i_1 j} X_{i_1 k} X_{i_2 j} X_{i_2 k} + \sigma_{jk,0}^2 - \frac{2}{n} \sigma_{jk,0} \sum_{i_1}^{n} X_{i_1 j} X_{i_1 k} \right).$$

Write $\hat{T}_n - \tilde{T}_n = -R_1 + R_2$, where

$$R_1 = \frac{2}{n(n-1)(n-2)} \sum_{j \neq k}^{p} \sum_{i_1,i_2,i_3}^{*} X_{i_1 j} (X_{i_2 j} X_{i_2 k} - \sigma_{jk,0}) X_{i_3 k},$$

$$R_2 = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{j \neq k}^{p} \sum_{i_1,i_2,i_3,i_4}^{*} X_{i_1 j} X_{i_2 j} X_{i_3 k} X_{i_4 k}.$$

Note that $\mathsf{E} R_1 = \mathsf{E} R_2 = 0$. By the independence between X_i, we have

$$\begin{aligned}
\mathsf{E} R_2^2 &= \sum_{\substack{j \neq k \\ m \neq q}} \sum_{i_1,i_2,i_3,i_4}^{*} \sum_{i_5,i_6,i_7,i_8}^{*} \frac{\mathsf{E}(X_{i_1 j} X_{i_2 j} X_{i_3 k} X_{i_4 k} X_{i_5 m} X_{i_6 m} X_{i_7 q} X_{i_8 q})}{n^2 (n-1)^2 (n-3)^2 (n-4)^2} \\
&= \frac{8}{n(n-1)(n-3)(n-4)} \sum_{j \neq k} \sum_{m \neq q} (\sigma_{jm}^2 \sigma_{kq}^2 + 2 \sigma_{jm} \sigma_{jq} \sigma_{kq} \sigma_{km}) \\
&\le \frac{8}{n(n-1)(n-3)(n-4)} \cdot C |\Gamma|_F^2.
\end{aligned}$$

Under H_{0a}, we can decompose $\mathsf{E} R_1^2$ as

$$\mathsf{E} R_1^2 = \frac{4}{n(n-1)(n-2)} \sum_{i=1}^{6} R_{1,i},$$

where

$$
\begin{aligned}
R_{1,1} &= \sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\sigma_{jm}\sigma_{kq}\left(\mathsf{E}(X_{ij}X_{ik}X_{im}X_{iq})-\sigma_{jk}\sigma_{mq}\right),\\
R_{1,2} &= \sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\sigma_{jq}\sigma_{km}\left(\mathsf{E}(X_{ij}X_{ik}X_{im}X_{iq})-\sigma_{jk}\sigma_{mq}\right),\\
R_{1,3} &= \sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\sigma_{jm}\mathsf{E}(X_{ij}X_{ik}X_{iq})\mathsf{E}(X_{ik}X_{im}X_{iq}),\\
R_{1,4} &= \sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\sigma_{jq}\mathsf{E}(X_{ij}X_{ik}X_{im})\mathsf{E}(X_{ik}X_{im}X_{iq}),\\
R_{1,5} &= \sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\sigma_{km}\mathsf{E}(X_{ij}X_{im}X_{iq})\mathsf{E}(X_{ij}X_{ik}X_{iq}),\\
R_{1,6} &= \sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\sigma_{kq}\mathsf{E}(X_{ij}X_{ik}X_{im})\mathsf{E}(X_{ij}X_{im}X_{iq}).
\end{aligned}
$$

Since $\mathsf{E}\left(\xi_{11}^3\right)=0$, elementary calculation shows that $R_{1,3}=R_{1,4}=R_{1,5}=R_{1,6}=0$. By (2.7),

$$
\begin{aligned}
&R_{1,1}+R_{1,2}=\sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\left(\mathsf{E}(X_{ij}X_{ik}X_{im}X_{iq})-\sigma_{jk}\sigma_{mq}\right)\left(\sigma_{jm}\sigma_{kq}+\sigma_{jq}\sigma_{km}\right)\\
&\leq\sqrt{\sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\left(\sigma_{jm}\sigma_{kq}+\sigma_{jq}\sigma_{km}\right)^2}\sqrt{\sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\left(\mathsf{E}(X_{ij}X_{ik}X_{im}X_{iq})-\sigma_{jk}\sigma_{mq}\right)^2}\\
&=\sqrt{2\sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\left(\sigma_{jm}^2\sigma_{kq}^2+\sigma_{jm}\sigma_{jq}\sigma_{km}\sigma_{kq}\right)}\sqrt{\sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\left(\mathsf{E}(X_{ij}X_{ik}-\sigma_{jk})(X_{im}X_{iq}-\sigma_{mq})\right)^2}\\
&\leq C_{\nu,1}|\Gamma|_F\cdot\sqrt{\sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\left(\mathsf{E}(X_{ij}X_{ik}-\sigma_{jk})(X_{im}X_{iq}-\sigma_{mq})\right)^2}.
\end{aligned}
$$

By (E.4), $X_{ij}X_{ik}-\sigma_{jk}=L_{(j,k),\cdot}^T\eta$. Recall in the proof of Lemma E.1, $\mathsf{E}(\eta\eta^T)$ is a diagonal matrix with its elements being $1,\nu,\nu^2$, and $D=L^TL$. Then, we have

$$\sum_{j\neq k}^{p}\sum_{m\neq q}^{p}(\mathsf{E}(X_{ij}X_{ik}-\sigma_{jk})(X_{im}X_{iq}-\sigma_{mq}))^2=\sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\left(L_{(j,k),\cdot}^T\mathsf{E}(\eta\eta^T)L_{(m,q),\cdot}\right)^2$$
$$\leq \max\{1,\nu^4\}\sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\operatorname{tr}\left(L_{(m,q),\cdot}L_{(j,k),\cdot}^T\right)^2$$
$$\leq \max\{1,\nu^4\}\sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\operatorname{tr}\left(L_{(j,k),\cdot}^T L_{(m,q),\cdot}\right)^2$$
$$\leq \max\{1,\nu^4\}\operatorname{tr}(D^2)$$
$$\leq C_{\nu,2}|\Gamma|_F^2.$$

Thus, we can obtain,

$$R_{1,1}+R_{1,2} = \sum_{j\neq k}^{p}\sum_{m\neq q}^{p}\left(\mathsf{E}(X_{ij}X_{ik}X_{im}X_{iq})-\sigma_{jk}\sigma_{mq}\right)\left(\sigma_{jm}\sigma_{kq}+\sigma_{jq}\sigma_{km}\right)$$
$$\leq C_{\nu,3}|\Gamma|_F^2. \tag{E.5}$$

So $\mathsf{E}R_1^2 \leq C(n(n-1)(n-2))^{-1}|\Gamma|_F^2$. Hence, $R_1/|\Gamma|_F = O_{\mathsf{P}}(n^{-3/2})$ and $R_2/|\Gamma|_F = O_{\mathsf{P}}(n^{-2})$.

Under Assumption 2.1, by Lemma E.1, we have (E.1). Adopting Lemma F.4 in the Supplementary Material, under H_{0a}, we obtain

$$\sup_t\left|\mathsf{P}\left(\frac{n\tilde{T}_n}{|\Gamma|_F}\leq t\right)-\mathsf{P}\left(\sum_{d=1}^{p(p-1)}\frac{\lambda_d}{|\Gamma|_F}(\eta_d-1)\leq t\right)\right|=O(n^{-\delta/(10+4\delta)}). \tag{E.6}$$

Then Theorem 2.1 follows by (E.6), triangle inequality and Lemma F.2 in the Supplementary Material. □

Proof of Corollary 2.1. When $\sum_{j\neq k}^{p}(\sigma_{jk}-\sigma_{jk,0})^2/|\Gamma|_F = O(1)$, it can be proved using similar arguments in the proof of Theorem 2.1 and Lemma F.3.

When $\sum_{j\neq k}^{p}(\sigma_{jk}-\sigma_{jk,0})^2/|\Gamma|_F \to \infty$, similar to the proof of Theorem 2.1, we can show

$$\frac{n\hat{T}_n}{|\Gamma|_F}=\frac{n\tilde{T}_n}{|\Gamma|_F}(1+o_{\mathsf{P}}(1)),$$

where

$$\tilde{T}_n = \frac{1}{n(n-1)} \sum_{i_1,i_2}^{*} \sum_{j\neq k}^{p} (X_{i_1j}X_{i_1k} - \sigma_{jk})(X_{i_2j}X_{i_2k} - \sigma_{jk}) + \sum_{j\neq k}^{p} (\sigma_{jk,0} - \sigma_{jk})^2 - \frac{2}{n} \sum_{i_1}^{n} \sum_{j\neq k}^{p} (\sigma_{jk,0} - \sigma_{jk})(X_{i_1j}X_{i_1k} - \sigma_{jk}).$$

By Theorem 2.1, since $\sum_{j\neq k}^{p}(\sigma_{jk} - \sigma_{jk,0})^2/|\Gamma|_F \to \infty$, it is clear that

$$\frac{n\tilde{T}_n}{|\Gamma|_F} \to \infty \text{ in probability.}$$

Corollary 2.1 then follows. □

Before proving Corollary 2.2, we need the following Lemma.

LEMMA E.2. *Assume that* $\{\boldsymbol{X}_i\}_{i=1}^{n}$ *follows a linear process, that is, under Assumption 2.1 with* $a_{j,l_1l_2\ldots l_i} = 0$ *for all* $1 \leq l_1 < l_2 < \ldots < l_i \leq N, 2 \leq i \leq d, 1 \leq j \leq p$. *Then, we have*

(E.7)

$$|\Gamma|_F^2 \geq \min\left\{\frac{\nu^2}{2}, 2\right\} \sum_{(j,k)\in\mathcal{S}_1} \sum_{(m,q)\in\mathcal{S}_1} \left(\sigma_{jm}^2\sigma_{kq}^2 + \sigma_{jm}\sigma_{jq}\sigma_{km}\sigma_{kq}\right),$$

(E.8)

$$|\Gamma|_F^2 \leq \max\left\{\frac{\nu^2}{2}, \frac{(\nu-2)^2}{2} + 2\right\} \sum_{(j,k)\in\mathcal{S}_1} \sum_{(m,q)\in\mathcal{S}_1} \left(\sigma_{jm}^2\sigma_{kq}^2 + \sigma_{jm}\sigma_{jq}\sigma_{km}\sigma_{kq}\right).$$

Note that by Cauchy-Schwarz inequality

$$\sum_{(j,k)\in\mathcal{S}_1} \sum_{(m,q)\in\mathcal{S}_1} (\sigma_{jm}^2\sigma_{kq}^2 + \sigma_{jm}\sigma_{jq}\sigma_{km}\sigma_{kq}) \leq 2\mathrm{tr}^2(\Sigma^2).$$

Since $\sum_{(j,k)\in\mathcal{S}_1}\sum_{(m,q)\in\mathcal{S}_1}(\sigma_{jm}^2\sigma_{kq}^2 + \sigma_{jm}\sigma_{jq}\sigma_{km}\sigma_{kq}) \geq K\mathrm{tr}^2(\Sigma^2)$ for some constant $K > 0$, Lemma *E.2* shows that $|\Gamma|_F^2 \asymp |\Sigma|_F^4$.

PROOF OF LEMMA E.2. Let $\mathcal{B} = \{(i,l), 1 \leq i \leq l \leq N\}$ and $\omega = (\omega_\beta)_{\beta\in\mathcal{B}} \in \mathbb{R}^{N(N+1)/2}$, where $\omega_\beta = \xi_{1i}\xi_{1l}$ for $\beta = (i,l)$, that is

$$\boldsymbol{\omega} = (\varrho_1, \xi_{11}\xi_{12}, \ldots, \xi_{11}\xi_{1N}, \varrho_2, \xi_{12}\xi_{13}, \ldots, \varrho_N)^T, \text{ where } \varrho_i = \xi_{1i}^2 - 1.$$

Let V_W be the covariance matrix of $\boldsymbol{\omega}$. Then $V_W = \text{diag}(\{v_{\beta,\beta}\}_{\beta\in\mathcal{B}})$, where for $\beta = (i,l)$, $v_{\beta,\beta} = \mathsf{Var}(\xi_{1i}^2) = \nu$ if $l = i$ and $v_{\beta,\beta} = 1$ if $l \neq i$. Also define $G = (g_{\alpha,\beta})_{\alpha\in\mathcal{I},\beta\in\mathcal{B}} \in \mathbb{R}^{p(p-1)\times[N(N+1)/2]}$, where for $\alpha = (j,k)$, $\beta = (i,l)$,

$$g_{\alpha,\beta} = \begin{cases} b_{ji}b_{ki}, & \text{if } l = i; \\ b_{ji}b_{kl} + b_{jl}b_{ki}, & \text{if } l > i. \end{cases}$$

Note that $(X_j - \mu_j)(X_k - \mu_k) = \mathbf{g}_\alpha^T\boldsymbol{\omega}$, where $\mathbf{g}_\alpha^T$ is the α'th row of G. Then $W = G\boldsymbol{\omega}$ and

$$\mathsf{E}\left(WW^T\right) = \left(\gamma_{\alpha,\alpha'}\right)_{\alpha,\alpha'\in\mathcal{I}},$$

where for $\alpha = (j,k)$, $\alpha' = (m,q)$,

$$\begin{aligned}
\gamma_{\alpha,\alpha'} &= \mathsf{Cov}\left(X_jX_k, X_mX_q\right) = \mathbf{g}_\alpha^T V_W \mathbf{g}_{\alpha'} \\
&= \nu\sum_i b_{ji}b_{ki}b_{mi}b_{qi} + \sum_{i<l}\left(b_{ji}b_{kl} + b_{jl}b_{ki}\right)\left(b_{qi}b_{ml} + b_{mi}b_{ql}\right) \\
&= (\nu - 2)\sum_i b_{ji}b_{ki}b_{mi}b_{qi} + \sigma_{jm}\sigma_{kq} + \sigma_{jq}\sigma_{km}.
\end{aligned}$$

Note that $\Sigma = BB^T$. Let

$$\begin{aligned}
L_0 &= \sum_{\substack{j\neq k\\ m\neq q}}\left(\sigma_{jm}\sigma_{kq} + \sigma_{jq}\sigma_{km}\right)^2, \\
L_1 &= \sum_{\substack{j\neq k\\ m\neq q}}\left(\sum_i b_{ji}b_{ki}b_{mi}b_{qi}\right)^2 = \sum_{\substack{j\neq k\\ m\neq q}}\sum_{i,l} b_{ji}b_{ki}b_{mi}b_{qi}b_{jl}b_{kl}b_{ml}b_{ql}, \\
L_2 &= \sum_{\substack{j\neq k\\ m\neq q}}\left(\sigma_{jm}\sigma_{kq} + \sigma_{jq}\sigma_{km}\right)\sum_i b_{ji}b_{ki}b_{mi}b_{qi} \\
&= \sum_{\substack{j\neq k\\ m\neq q}}\sum_{i,l,l'} b_{ji}b_{ki}b_{mi}b_{qi}\left(b_{jl}b_{ml}b_{kl'}b_{ql'} + b_{jl}b_{ql}b_{kl'}b_{ml'}\right).
\end{aligned}$$

Since $(\sum_{j\neq k} b_{ji}b_{jl}b_{ki}b_{kl'} + \sum_{j\neq k} b_{ji}b_{jl'}b_{ki}b_{kl})^2 \geq 0$, we can show that $L_2 \geq$

$2L_1$. Thus

$$
\begin{aligned}
|\Gamma|_F^2 &= \sum_{\alpha,\alpha'\in\mathcal{I}} \gamma_{\alpha,\alpha'}^2 \\
&= \sum_{\substack{j\neq k\\ m\neq q}} \Big[(\nu-2)\sum_i b_{ji}b_{ki}b_{mi}b_{qi} + \sigma_{jm}\sigma_{kq} + \sigma_{jq}\sigma_{km}\Big]^2 \\
&= \sum_{\substack{j\neq k\\ m\neq q}} \left(\sigma_{jq}^2\sigma_{km}^2 + \sigma_{jm}^2\sigma_{kq}^2 + 2\sigma_{jq}\sigma_{qk}\sigma_{km}\sigma_{mj} + (\nu-2)^2\left(\sum_i b_{ji}b_{ki}b_{mi}b_{qi}\right)^2\right. \\
&\qquad \left. +2(\nu-2)\left(\sigma_{jm}\sigma_{kq}+\sigma_{jq}\sigma_{km}\right)\sum_i b_{ji}b_{ki}b_{mi}b_{qi}\right) \\
&= L_1(\nu-2)^2 + 2L_2(\nu-2) + L_0.
\end{aligned}
$$

Clearly $|\Gamma|_F^2 \geq L_0$ if $\nu \geq 2$. It is easy to see that $4L_1 - 4L_2 + L_0 \geq 0$, so $L_0 \geq 4L_2 - 4L_1 \geq 4L_1$. If $0 < \nu < 2$, then the quantity

$$
|\Gamma|_F^2 - \frac{L_0\nu^2}{4} = \left(L_1 - \frac{L_0}{4}\right)\nu^2 + 2(L_2 - 2L_1)\nu + L_0 + 4L_1 - 4L_2
$$

is larger than the minimum of its value at $\nu = 0$ and $\nu = 2$, which are both nonnegative. Therefore,

$$
|\Gamma|_F^2 \geq \nu^2 L_0/4 = \frac{\nu^2}{2}\sum_{\substack{j\neq k\\ m\neq q}}\left(\sigma_{jm}^2\sigma_{kq}^2 + \sigma_{jm}\sigma_{jq}\sigma_{km}\sigma_{kq}\right)
$$

for any $\nu \in (0,2)$.
Similarly, we can show that

$$
|\Gamma|_F^2 \leq \max\left\{\frac{\nu^2}{2}, \frac{(\nu-2)^2}{2} + 2\right\}\sum_{\substack{j\neq k\\ m\neq q}}\left(\sigma_{jm}^2\sigma_{kq}^2 + \sigma_{jm}\sigma_{jq}\sigma_{km}\sigma_{kq}\right).
$$

□

PROOF OF COROLLARY 2.2. Note that $\rho_\Sigma = o(1)$ is equivalent to $\mathrm{tr}(\Sigma^3) = o(\mathrm{tr}^{3/2}(\Sigma^2))$. By the Lindeberg Central Limit Theorem, the *necessary and sufficient condition* for

$$
\sum_{d=1}^{p(p-1)} \frac{\lambda_d}{|\Gamma|_F}(\eta_d - 1) \xrightarrow{d} N(0,2)
$$

is $\lambda_1/|\Gamma|_F \to 0$. Since $\sum_{(j,k)\in\mathcal{S}_1}\sum_{(m,q)\in\mathcal{S}_1}(\sigma_{jm}^2\sigma_{kq}^2+\sigma_{jm}\sigma_{jq}\sigma_{km}\sigma_{kq}) \geq K\mathrm{tr}^2(\Sigma^2)$ for some constant $K>0$, by Lemma E.2, $|\Gamma|_F^2 \asymp |\Sigma|_F^4$. Corollary 2.2 then follows. □

Proof of Theorem 2.2. We first prove the theorem under the null hypothesis H_{0a}. Under the alternative hypothesis, a similar argument can be implied. For the convenience of presentation, we assume $d=2$ in Assumption 2.1. If $d>2$, the argument shown as follows still can be applied to prove the theorem with more tedious calculations.
Denote $F(t)=\mathsf{P}(n\hat{T}_n \leq t)$. To simplify the notion, write $J_{B_l} := J_{B_l}(\mathcal{S}_1,\Sigma_0)$, $J_{B_l^c} := J_{B_l^c}(\mathcal{S}_1,\Sigma_0)$, $C_{B_l,B_l^c} := C_{B_l,B_l^c}(\mathcal{S}_1,\Sigma_0)$. For sets $B_l, B_{l'} \in \{1,2,...,n\}$ with $|B_l|=|B_{l'}|=m=n/2$, denote

$$\begin{aligned}
V_{B_l} &= \frac{1}{m}\sum_{j\neq k}^{p}\Bigg(\frac{1}{m-1}\sum_{i_1,i_2\in B_l}^{*}(X_{i_1j}X_{i_1k}-\sigma_{jk,0})(X_{i_2j}X_{i_2k}-\sigma_{jk,0})\\
&\quad+\frac{1}{m-1}\sum_{i_1,i_2\in B_l^c}^{*}(X_{i_1j}X_{i_1k}-\sigma_{jk,0})(X_{i_2j}X_{i_2k}-\sigma_{jk,0})\\
&\quad-\frac{2}{m}\sum_{i_1\in B_l,i_2\in B_l^c}(X_{i_1j}X_{i_1k}-\sigma_{jk,0})(X_{i_2j}X_{i_2k}-\sigma_{jk,0})\Bigg),\\
V_{B_l}^o &= \frac{1}{m}\left(\frac{1}{m-1}\sum_{i_1,i_2\in B_l}^{*}Y_{i_1}^TY_{i_2}+\frac{1}{m-1}\sum_{i_1,i_2\in B_l^c}^{*}Y_{i_1}^TY_{i_2}-\frac{2}{m}\sum_{i_1\in B_l}\sum_{i_2\in B_l^c}Y_{i_1}^TY_{i_2}\right),
\end{aligned}$$

where Y_i i.i.d $N(0,\Gamma)$. Similarly, define $V_{B_{l'}}$ and $V_{B_{l'}}^o$. By elementary manipulations,

$$\frac{m(J_{B_l}+J_{B_l^c}-2C_{B_l,B_l^c})}{2|\Gamma|_F}=\frac{mV_{B_l}}{2|\Gamma|_F}+\frac{mR_l}{2|\Gamma|_F},$$

where we decompose R_l as

$$R_l=-2R_{l,1}-2R_{l,2}+R_{l,3}+R_{l,4}+2R_{l,5}+2R_{l,6}-2R_{l,7}$$

and

$$
\begin{aligned}
R_{l,1} &= \frac{1}{m(m-1)(m-2)} \sum_{i_1,i_2,i_3 \in B_l}^{*} \sum_{j \neq k}^{p} X_{i_1 j}(X_{i_2 j}X_{i_2 k} - \sigma_{jk})X_{i_3 k},\\
R_{l,2} &= \frac{1}{m(m-1)(m-2)} \sum_{i_1,i_2,i_3 \in B_l^c}^{*} \sum_{j \neq k}^{p} X_{i_1 j}(X_{i_2 j}X_{i_2 k} - \sigma_{jk})X_{i_3 k},\\
R_{l,3} &= \frac{1}{m(m-1)(m-2)(m-3)} \sum_{i_1,i_2,i_3,i_4 \in B_l}^{*} \sum_{j \neq k}^{p} X_{i_1 j}X_{i_2 j}X_{i_3 k}X_{i_4 k},\\
R_{l,4} &= \frac{1}{m(m-1)(m-2)(m-3)} \sum_{i_1,i_2,i_3,i_4 \in B_l^c}^{*} \sum_{j \neq k}^{p} X_{i_1 j}X_{i_2 j}X_{i_3 k}X_{i_4 k},\\
R_{l,5} &= \frac{1}{m^2(m-1)} \sum_{i_1,i_2 \in B_l}^{*} \sum_{i_3 \in B_l^c} \sum_{j \neq k}^{p} X_{i_1 j}X_{i_2 k}(X_{i_3 j}X_{i_3 k} - \sigma_{jk}),\\
R_{l,6} &= \frac{1}{m^2(m-1)} \sum_{i_1,i_2 \in B_l^c}^{*} \sum_{i_3 \in B_l} \sum_{j \neq k}^{p} X_{i_1 j}X_{i_2 k}(X_{i_3 j}X_{i_3 k} - \sigma_{jk}),\\
R_{l,7} &= \frac{1}{m^2(m-1)^2} \sum_{i_1,i_2 \in B_l}^{*} \sum_{i_3,i_4 \in B_l^c}^{*} \sum_{j \neq k}^{p} X_{i_1 j}X_{i_2 k}X_{i_3 j}X_{i_4 k}.
\end{aligned}
$$

Similar to R_1 and R_2 in the proof of Theorem 2.1, we can obtain $R_{l,i}/|\Gamma|_F = O_{\mathsf{P}}(m^{-3/2})$ for $i = 1, 2$ and $R_{l,i}/|\Gamma|_F = O_{\mathsf{P}}(m^{-2})$ for $i = 3, 4$. Applying Lemma E.1, under the null H_{0a}, employing (E.5),

$$
\begin{aligned}
\mathsf{E}R_{l,5}^2 &= \sum_{j \neq k} \sum_{m \neq q} \frac{1}{m(m-1)}(\sigma_{jm}\sigma_{kq} + \sigma_{jq}\sigma_{km}) \cdot \frac{1}{m}(\mathsf{E}(X_{ij}X_{ik}X_{im}X_{iq}) - \sigma_{jk}\sigma_{kq})\\
&\leq \frac{1}{m^2(m-1)} \cdot C_{\nu,1}|\Gamma|_F^2.\\
\mathsf{E}R_{l,7}^2 &= \sum_{j \neq k} \sum_{m \neq q} \frac{1}{m^2(m-1)^2}(\sigma_{jm}\sigma_{kq} + \sigma_{jq}\sigma_{km})^2 \leq \frac{1}{2C_\nu} \cdot \frac{1}{m^2(m-1)^2}|\Gamma|_F^2.
\end{aligned}
$$

So $R_{l,i}/|\Gamma|_F = O_{\mathsf{P}}(m^{-3/2})$ for $i = 5, 6$, $R_{l,7}/|\Gamma|_F = O_{\mathsf{P}}(m^{-2})$.

For any $\varepsilon \geq 0$, we know by the triangle inequality that

$$\mathsf{P}\left(\frac{mV_{B_l}}{2|\Gamma|_F} \leq t - \varepsilon\right) - \mathsf{P}\left(\frac{|mR_l|}{2|\Gamma|_F} \geq \varepsilon\right) \leq \mathsf{P}\left(\frac{m(J_{B_l} + J_{B_l^c} - 2C_{B_l,B_l^c})}{2|\Gamma|_F} \leq t\right) \leq \mathsf{P}\left(\frac{mV_{B_l}}{2|\Gamma|_F} \leq t + \varepsilon\right) + \mathsf{P}\left(\frac{|mR_l|}{2|\Gamma|_F} \geq \varepsilon\right). \tag{E.9}$$

By Lemma F.2 in the Supplementary Material, we obtain

$$\mathsf{P}\left(\frac{mV_{B_l}}{2|\Gamma|_F} \leq t\right) - \sqrt{\varepsilon} \cdot \sqrt{8\pi} - K \cdot \frac{1}{m\varepsilon^2} \leq \mathsf{P}\left(\frac{m(J_{B_l} + J_{B_l^c} - 2C_{B_l,B_l^c})}{2|\Gamma|_F} \leq t\right) \leq \mathsf{P}\left(\frac{mV_{B_l}}{2|\Gamma|_F} \leq t\right) + \sqrt{\varepsilon} \cdot \sqrt{8\pi} + K \cdot \frac{1}{m\varepsilon^2}.$$

Taking $\varepsilon = n^{-2/5}$,

$$\sup_t \left| \mathsf{P}\left(\frac{m(J_{B_l} + J_{B_l^c} - 2C_{B_l,B_l^c})}{2|\Gamma|_F} \leq t\right) - \mathsf{P}\left(\frac{mV_{B_l}}{2|\Gamma|_F} \leq t\right) \right| = O(n^{-1/5}). \tag{E.10}$$

Adopting Lemma F.4, Corollary F.3 in the Supplementary Material and (E.10), for all $B_l, B_{l'} \in \mathcal{B}$, we have

$$\sup_t \left| \mathsf{P}\left(\frac{m}{2}(J_{B_l} + J_{B_l^c} - 2C_{B_l,B_l^c}) \leq t\right) - \mathsf{P}\left(\frac{1}{n-1}\sum_{i_1,i_2}^{*} Y_{i_1}^T Y_{i_2} \leq t\right) \right| = O(n^{-\delta/(10+4\delta)}), \tag{E.11}$$

$$\sup_t \left| \mathsf{P}\left(n\hat{T}_n \leq t\right) - \mathsf{P}\left(\frac{1}{n-1}\sum_{i_1,i_2}^{*} Y_{i_1}^T Y_{i_2} \leq t\right) \right| = O(n^{-\delta/(10+4\delta)}). \tag{E.12}$$

Consequently,

$$\sup_t \left| \mathsf{P}\left(\frac{m}{2}(J_{B_l} + J_{B_l^c} - 2C_{B_l,B_l^c}) \leq t\right) - F(t) \right| = O(n^{-\delta/(10+4\delta)}).$$

Let $\mathcal{B}$ be the class of all the $\binom{n}{m}$ subsets of size m of $\{1, 2, ..., n\}$ and

$$\tilde{F}(t) = \frac{1}{\binom{n}{m}} \sum_{B_l \in \mathcal{B}} \mathbf{1}_{m(1-m/n)(J_{B_l} + J_{B_l^c} - 2C_{B_l,B_l^c}) \leq t}.$$

For $B_l, B_{l'} \in \mathcal{B}$, let $I_1 = B_l \cap B_{l'}$, $I_2 = B_l \cap B_{l'}^c$, $I_3 = B_l^c \cap B_{l'}$, $I_4 = B_l^c \cap B_{l'}^c$ and

$$d(B_l, B_{l'}) = \max\left\{\left||I_1| - \frac{n}{4}\right|, \left||I_2| - \frac{n}{4}\right|, \left||I_3| - \frac{n}{4}\right|, \left||I_4| - \frac{n}{4}\right|\right\}.$$

Referring to (E.17) and (E.18), the proportion of pairs $(B_l, B_{l'})$ such that $d(B_l, B_{l'}) > n^{1/2}\log n$ is very small. Now we shall show that for $B_l, B_{l'} \in \mathcal{B}$ with $d(B_l, B_{l'}) \le n^{1/2}\log n$,

$$\begin{aligned}
&\sup_t \Big|\mathsf{P}\left(\frac{m}{2}(J_{B_l} + J_{B_l^c} - 2C_{B_l, B_l^c}) \le t, \frac{m}{2}(J_{B_{l'}} + J_{B_{l'}^c} - 2C_{B_{l'}, B_{l'}^c}) \le t\right) - \\
&\qquad \mathsf{P}\left(\frac{m}{2}(J_{B_l} + J_{B_l^c} - 2C_{B_l, B_l^c}) \le t\right)\mathsf{P}\left(\frac{m}{2}(J_{B_{l'}} + J_{B_{l'}^c} - 2C_{B_{l'}, B_{l'}^c}) \le t\right)\Big| \\
&\quad = O(n^{-\delta/(10+4\delta)}).
\end{aligned} \tag{E.13}$$

For convenience, denote $\bar{Y}_{B_l} = m^{-1}\sum_{i\in B_l} Y_i$ and $\bar{Y}_{B_l^c}, \bar{Y}_{B_{l'}}, \bar{Y}_{B_{l'}^c}$ are similarly defined. Let $\vartheta = (4|I_1| - n)/n$. Then $|I_1| = |I_4| = n/4 + n\vartheta/4$, $|I_2| = |I_3| = n/4 - n\vartheta/4$ and $\vartheta = 4d(B_l, B_{l'})/n \le 4n^{-1/2}\log n$. Define

$$\begin{aligned}
\tilde{Y}_{B_{l'}} &= \bar{Y}_{B_{l'}} - \bar{Y}_{B_{l'}^c} - \vartheta\bar{Y}_{B_l} + \vartheta\bar{Y}_{B_l^c}, \\
\tilde{V}^o_{B_l} &= \frac{\frac{m}{2}(\bar{Y}_{B_l} - \bar{Y}_{B_l^c})^T(\bar{Y}_{B_l} - \bar{Y}_{B_l^c}) - \mathrm{tr}(\Gamma)}{|\Gamma|_F}, \\
\tilde{V}^o_{B_{l'}} &= \frac{\frac{m}{2}(\bar{Y}_{B_{l'}} - \bar{Y}_{B_{l'}^c})^T(\bar{Y}_{B_{l'}} - \bar{Y}_{B_{l'}^c}) - \mathrm{tr}(\Gamma)}{|\Gamma|_F}, \\
\breve{V}^o_{B_{l'}} &= \frac{\frac{m}{2}\tilde{Y}^T_{B_{l'}}\tilde{Y}_{B_{l'}} - (1+\vartheta^2)\mathrm{tr}(\Gamma)}{|\Gamma|_F}.
\end{aligned}$$

A simple calculation shows that

$$\mathsf{Cov}(\bar{Y}_{B_l} - \bar{Y}_{B_l^c}, \tilde{Y}_{B_{l'}}) = 0,$$

which due to Gaussianity, implies that $\tilde{Y}_{B_{l'}}$ is independent of $\bar{Y}_{B_l} - \bar{Y}_{B_l^c}$. Then for any $\varepsilon > 0$, we have

$$\begin{aligned}
\mathsf{P}\left(\frac{mV^o_{B_l}}{2|\Gamma|_F} \le t, \frac{mV^o_{B_{l'}}}{2|\Gamma|_F} \le t\right) &= \mathsf{P}\left(\tilde{V}^o_{B_l} \le t, \tilde{V}^o_{B_{l'}} \le t\right) \\
&\le \mathsf{P}\left(\tilde{V}^o_{B_l} \le t, \breve{V}^o_{B_{l'}} \le t + \varepsilon\right) + \mathsf{P}\left(|\tilde{V}^o_{B_{l'}} - \breve{V}^o_{B_{l'}}| > \varepsilon\right) \\
&= \mathsf{P}\left(\tilde{V}^o_{B_l} \le t\right)\mathsf{P}\left(\breve{V}^o_{B_{l'}} \le t + \varepsilon\right) + \mathsf{P}\left(|\tilde{V}^o_{B_{l'}} - \breve{V}^o_{B_{l'}}| > \varepsilon\right) \\
&\le \mathsf{P}\left(\tilde{V}^o_{B_l} \le t\right)\mathsf{P}\left(\tilde{V}^o_{B_{l'}} \le t + 2\varepsilon\right) + 2\mathsf{P}\left(|\tilde{V}^o_{B_{l'}} - \breve{V}^o_{B_{l'}}| > \varepsilon\right). \qquad \text{(E.14)}
\end{aligned}$$

For the second term,

$$\tilde{V}^o_{B_{l'}} - \breve{V}^o_{B_{l'}} = \frac{m\vartheta(\bar{Y}_{B_l} - \bar{Y}_{B_l^c})^T(\bar{Y}_{B_{l'}} - \bar{Y}_{B_{l'}^c})}{|\Gamma|_F} - \frac{\frac{m}{2}\vartheta^2(\bar{Y}_{B_l} - \bar{Y}_{B_l^c})^T(\bar{Y}_{B_l} - \bar{Y}_{B_l^c}) - \vartheta^2\mathrm{tr}(\Gamma)}{|\Gamma|_F}.$$

Observe that

$$\frac{m\vartheta(\bar{Y}_{B_l} - \bar{Y}_{B_l^c})^T(\bar{Y}_{B_{l'}} - \bar{Y}_{B_{l'}^c})}{|\Gamma|_F} = \vartheta O_{\mathsf{P}}(1) = O_{\mathsf{P}}(\frac{\log n}{\sqrt{n}}),$$
$$\frac{\frac{m}{2}\vartheta^2(\bar{Y}_{B_l} - \bar{Y}_{B_l^c})^T(\bar{Y}_{B_l} - \bar{Y}_{B_l^c}) - \vartheta^2\mathrm{tr}(\Gamma)}{|\Gamma|_F} = \vartheta^2 O_{\mathsf{P}}(1) = O_{\mathsf{P}}(\frac{(\log n)^2}{n}).$$

Employing Lemma F.6 in the Supplementary Material and (E.14), taking $\varepsilon = n^{-2/5}$, a similar argument implies that

$$\text{(E.15)}\quad \sup_t \left| \mathsf{P}\left(\frac{mV^o_{B_l}}{2|\Gamma|_F} \le t, \frac{mV^o_{B_{l'}}}{2|\Gamma|_F} \le t\right) - \mathsf{P}\left(\frac{mV^o_{B_l}}{2|\Gamma|_F} \le t\right)\mathsf{P}\left(\frac{mV^o_{B_{l'}}}{2|\Gamma|_F} \le t\right)\right| = O(n^{-1/5}).$$

Recall Lemma F.6 shows that for any $B_l, B_{l'} \in \mathcal{B}$, if $d(B_l, B_{l'}) \le n^{1/2}\log n$,

$$\sup_t \left| \mathsf{P}\left(\frac{m}{2}V_{B_l} \le t, \frac{m}{2}V_{B_{l'}} \le t\right) - \mathsf{P}\left(\frac{m}{2}V^o_{B_l} \le t, \frac{m}{2}V^o_{B_{l'}} \le t\right)\right| = O(n^{-\delta/(10+4\delta)}).$$

Thus, applying Lemma F.6, (E.10) and (E.15), we can obtain

$$\begin{aligned}
&\sup_t \Big| \mathsf{P}\left(\frac{m}{2}(J_{B_l} + J_{B_l^c} - 2C_{B_l,B_l^c}) \le t, \frac{m}{2}(J_{B_{l'}} + J_{B_{l'}^c} - 2C_{B_{l'},B_{l'}^c}) \le t\right) - \\
&\qquad \mathsf{P}\left(\frac{m}{2}(J_{B_l} + J_{B_l^c} - 2C_{B_l,B_l^c}) \le t\right)\mathsf{P}\left(\frac{m}{2}(J_{B_{l'}} + J_{B_{l'}^c} - 2C_{B_{l'},B_{l'}^c}) \le t\right)\Big| \\
&= \sup_t \left| \mathsf{P}\left(\frac{m}{2}V_{B_l} \le t, \frac{m}{2}V_{B_{l'}} \le t\right) - \mathsf{P}^2\left(\frac{m}{2}V_{B_l} \le t\right)\right| + O(n^{-1/5}) \\
&\le \sup_t \left| \mathsf{P}\left(\frac{m}{2}V_{B_l} \le t, \frac{m}{2}V_{B_{l'}} \le t\right) - \mathsf{P}\left(\frac{m}{2}V^o_{B_l} \le t, \frac{m}{2}V^o_{B_{l'}} \le t\right)\right| \\
&\qquad + \sup_t \left| \mathsf{P}^2\left(\frac{m}{2}V_{B_l} \le t\right) - \mathsf{P}^2\left(\frac{m}{2}V^o_{B_l} \le t\right)\right| + O(n^{-1/5})
\end{aligned}$$

$$\text{(E.16)}\quad = O(n^{-\delta/(10+4\delta)}).$$

Let $B_0 = \{1, 2, ..., m\}$. Write $H_{B_l} = \mathbf{1}_{\frac{m}{2}(J_{B_l}(\mathcal{S}_1) + J_{B_l^c}(\mathcal{S}_1) - 2C_{B_l,B_l^c}(\mathcal{S}_1)) \le t}$. Then

$$\mathsf{E}|\tilde{F}(t) - F(t)|^2 = \frac{1}{\binom{n}{m}^2}\sum_{B_l,B_{l'}\in\mathcal{B}} \mathsf{Cov}\left(H_{B_l}, H_{B_{l'}}\right) =: I_1 + I_2,$$

where I_1 (resp. I_2) represents the sum with pairs $B_l, B_{l'}$ with $d(B_l, B_{l'}) \leq n^{1/2}\log n$ (resp. $d(B_l, B_{l'}) > n^{1/2}\log n$). Note that

$$\begin{aligned} I_2 &= \frac{1}{\binom{n}{m}} \sum_{B\in\mathcal{B}, d(B_0,B)>n^{1/2}\log n} \mathsf{Cov}(H_{B_l}, H_{B_{l'}}) \\ &\leq \frac{1}{\binom{n}{m}} \#\{B \in \mathcal{B}, d(B_0, B) > n^{1/2}\log n\} \\ &= \mathsf{P}\left(d(B_0, \mathcal{J}(\pi_1, \pi_2 ..., \pi_n)) > n^{1/2}\log n | \pi_1 + ... + \pi_n = m\right), \end{aligned} \tag{E.17}$$

where $\pi_1, ..., \pi_n$ are i.i.d. Bernoulli(1/2) with values 0 or 1, and $\mathcal{J}(\pi_1, \pi_2 ..., \pi_n) \subset \{1, \ldots, n\}$ is an index set such that, if $\pi_i = 1$, then $i \in \{1, ..., n\}$ is chosen. By the Hoeffding inequality,

$$\begin{aligned} I_2 &\leq \mathsf{P}\left(|\pi_1 + ... + \pi_m - m/2| > n^{1/2}\log n\right) / \mathsf{P}\left(\pi_1 + ... + \pi_n = m\right) \\ &\leq \frac{2\exp\left(-4\log^2(n)\right)}{\binom{n}{m}\cdot\frac{1}{2^n}} \leq 2\sqrt{n}\exp\left(-4\log^2(n)\right) =: \rho_n \end{aligned} \tag{E.18}$$

By (E.16), $I_1 \leq (1-\rho_n)\cdot O(n^{-\delta/(10+4\delta)}) = O(n^{-\delta/(10+4\delta)})$. Thus, we obtain

$$\sup_t \mathsf{E}|\tilde{F}(t) - F(t)|^2 = O(n^{-\delta/(10+4\delta)}).$$

□

REMARK E.1. *By the Glivenko-Cantelli argument, Lemma F.2 and Lemma F.4, we also have the uniform version*

$$\sup_t |\tilde{F}(t) - F(t)| \xrightarrow{P} 0.$$

PROOF OF THEOREM 3.1. Under the assumption 3.1, applying Taylor's expansion, similar to Zhong et al. (2017), we can show $\hat{T}_n(\hat{\boldsymbol{\theta}}) = \hat{T}_n(\boldsymbol{\theta})(1 + o_{\mathsf{P}}(1))$.
By carrying out the same route as it in the proof of Theorem 2.1, we have

$$\hat{T}_n(\boldsymbol{\theta}) = \left[\frac{1}{n(n-1)}\sum_{i\neq l}^{n} \boldsymbol{W}_i^T\boldsymbol{W}_l - \frac{1}{n^2}\sum_{i,l}^{n}\boldsymbol{W}_i^T\Upsilon\boldsymbol{W}_l\right](1 + o_{\mathsf{P}}(1)).$$

Following the same arguments as those in the proofs of Lemma E.1, it can be shown that K_δ^Υ defined as follows is bounded,

$$(K_\delta^\Upsilon)^{2+\delta} := \mathsf{E}\left|\frac{\boldsymbol{W}_1^T(I-\Upsilon)\boldsymbol{W}_2}{|\Gamma - \Upsilon\Gamma|_F}\right|^{2+\delta}.$$

Then Theorem 3.1(i) follows by employing Lemma F.5 in the Supplementary Material. Since

$$\mathsf{E}\left[\left(\frac{1}{n(n-1)}\sum_{i\neq l}^{n}\boldsymbol{W}_i^T\boldsymbol{W}_l-\frac{1}{n^2}\sum_{i\neq l}^{n}\boldsymbol{W}_i^T\Upsilon\boldsymbol{W}_l\right)^2\right]=\left|\Gamma-\frac{n-1}{n}\Upsilon\Gamma\right|_F^2$$

and $\sqrt{n}/\kappa_0\to 0$, Theorem 3.1(ii) follows by Lindeberg Central Limit Theorem. □

Proofs of Theorem 3.2. It can be carried out following the same routes as those in the proofs of Theorems 2.2 and 3.1. □

Proof of Theorem A.1. For the convenience of presentation, we assume $d=2$ in Assumption 2.1. If $d>2$, the argument shown as follows still can be applied to prove the Lemma with more tedious calculations. Denote

$$\tilde{Q}_n = \sum_{j=1}^{p_1}\sum_{k=p_1+1}^{p_1+p_2}\left(\frac{1}{n(n-1)}\sum_{i_1,i_2}^{*}X_{i_1j}X_{i_1k}X_{i_2j}X_{i_2k}+\sigma_{jk,0}^2-\frac{2}{n}\sigma_{jk,0}\sum_{i_1}^{n}X_{i_1j}X_{i_1k}\right).$$

Rewrite $\hat{Q}_n-\tilde{Q}_n=-R_1+R_2$, where

$$R_1 = \frac{2}{n(n-1)(n-2)}\sum_{j=1}^{p_1}\sum_{k=p_1+1}^{p_1+p_2}\sum_{i_1,i_2,i_3}^{*}X_{i_1j}(X_{i_2j}X_{i_2k}-\sigma_{jk,0})X_{i_3k},$$

$$R_2 = \frac{1}{n(n-1)(n-2)(n-3)}\sum_{j=1}^{p_1}\sum_{k=p_1+1}^{p_1+p_2}\sum_{i_1,i_2,i_3,i_4}^{*}X_{i_1j}X_{i_2j}X_{i_3k}X_{i_4k}.$$

Note that $\mathsf{E}\xi_{11}^3=0$. Recall that $\boldsymbol{X}_i=B\boldsymbol{\xi}_i$, using the same notation in the proof of Lemma E.1, it is straightforward to show that

$$\mathsf{E}(X_{ij}X_{ik}X_{im}X_{iq})-\sigma_{jk}\sigma_{mq} = \mathsf{E}\left(L_{(j,k),\cdot}^T\eta\cdot L_{(m,q),\cdot}^T\eta\right),$$

$$\mathsf{E}(X_{ij}X_{ik}X_{im}) = 0.$$

To this end, under H_{0a},

$$\mathsf{E}R_1^2=\frac{4}{n(n-1)(n-2)}\sum_{j,m=1}^{p_1}\sum_{k,q=p_1+1}^{p_1+p_2}(\sigma_{jm}\sigma_{kq}+\sigma_{jq}\sigma_{km})\left(L_{(j,k),\cdot}^T\cdot\mathsf{E}(\eta\eta^T)\cdot L_{(m,q),\cdot}\right).$$

Using similar arguments in the proof of (E.5),

$$\mathsf{E}R_1^2\le\frac{1}{n(n-1)(n-2)}\cdot C|\Xi|_F^2.$$

By carrying out similar procedures, by the Cauchy-Schwarz inequality, we can get

$$\begin{aligned}\mathsf{E}R_2^2 &= \frac{4}{n(n-1)(n-3)(n-4)}\left(\mathrm{tr}(\Sigma_{11}^2)\mathrm{tr}(\Sigma_{22}^2)+\mathrm{tr}^2(\Sigma_{12}\Sigma_{21})+4\mathrm{tr}(\Sigma_{11}\Sigma_{12}\Sigma_{22}\Sigma_{21})\right)\\ &\le \frac{1}{n(n-1)(n-3)(n-4)}\cdot C\left(\mathrm{tr}(\Sigma_{11}^2)\mathrm{tr}(\Sigma_{22}^2)+\mathrm{tr}^2(\Sigma_{12}\Sigma_{21})\right).\end{aligned}$$

Hence, $R_1/|\Xi|_F = O_{\mathsf{P}}(n^{-3/2})$ and $R_2/|\Xi|_F = O_{\mathsf{P}}(n^{-2})$.
Adopting Lemma E.1 and F.4, under H_{0b}, we obtain
(E.19)

$$\sup_t \left|\mathsf{P}\left(\frac{n\tilde{Q}_n}{|\Xi|_F}\le t\right)-\mathsf{P}\left(\sum_{d=1}^{p_1p_2}\frac{\theta_d}{|\Xi|_F}(\eta_d-1)\le t\right)\right| = O(n^{-\delta/(10+4\delta)}).$$

Then Theorem A.1 follows by (E.19), triangle inequality and Lemma F.2. □

To prove Corollary A.2, we need the following Lemma.

LEMMA E.3. *Assume that* $\{\boldsymbol{X}_i\}_{i=1}^n$ *follows a linear process, that is, under Assumption 2.1 with* $a_{j,l_1l_2\ldots l_i}=0$ *for all* $1\le l_1<l_2<\ldots<l_i\le N, 2\le i\le d, 1\le j\le p$. *Then, we have*

$$|\Xi|_F^2 \ge \min\left\{\frac{\nu^2}{4},1\right\}\left(\mathrm{tr}(\Sigma_{11}^2)\mathrm{tr}(\Sigma_{22}^2)+\mathrm{tr}^2(\Sigma_{12}\Sigma_{21}))\right), \tag{E.20}$$

$$|\Xi|_F^2 \le \max\left\{\frac{\nu^2}{2},\frac{(\nu-2)^2}{2}+2\right\}\left(\mathrm{tr}(\Sigma_{11}^2)\mathrm{tr}(\Sigma_{22}^2)+\mathrm{tr}^2(\Sigma_{12}\Sigma_{21}))\right). \tag{E.21}$$

PROOF OF LEMMA E.3. Similar to Lemma E.2, define $\alpha=(j,k)$, $\alpha'=(m,q)$ for $1\le j,m\le p_1$, $p_1+1\le k,q\le p_1+p_2=p$, we obtain

$$\begin{aligned}\gamma_{\alpha,\alpha'} &= \mathsf{Cov}\left(X_jX_k,X_mX_q\right)\\ &= (\nu-2)\sum_i b_{ji}b_{ki}b_{mi}b_{qi}+\sigma_{jm}\sigma_{kq}+\sigma_{jq}\sigma_{km},\\ |\Xi|_F^2 &= L_1(\nu-2)^2+2L_2(\nu-2)+L_0.\end{aligned}$$

Let $C = B_{(1)}^T B_{(1)}$ and $D = B_{(2)}^T B_{(2)}$, then

$$\begin{aligned}
L_0 &= \sum_{1\le j,m\le p_1} \sum_{p_1+1\le k,q\le p_1+p_2} (\sigma_{jm}\sigma_{kq} + \sigma_{jq}\sigma_{km})^2 \\
&= \mathrm{tr}(\Sigma_{11}^2)\mathrm{tr}(\Sigma_{22}^2) + (\mathrm{tr}(\Sigma_{12}\Sigma_{21}))^2 + 2\mathrm{tr}(\Sigma_{11}\Sigma_{12}\Sigma_{22}\Sigma_{21}), \\
L_1 &= \sum_{1\le j,m\le p_1} \sum_{p_1+1\le k,q\le p_1+p_2} \left(\sum_i a_{ji}a_{ki}a_{mi}a_{qi}\right)^2 = \sum_{il} c_{il}^2 d_{il}^2, \\
L_2 &= \sum_{1\le j,m\le p_1} \sum_{p_1+1\le k,q\le p_1+p_2} (\sigma_{jm}\sigma_{kq} + \sigma_{jq}\sigma_{km}) \sum_i b_{ji}b_{ki}b_{mi}b_{qi} \\
&= \sum_i \left(\sum_l c_{il}d_{il}\right)^2 + \sum_i \left(\sum_l c_{il}^2 \sum_l d_{il}^2\right).
\end{aligned}$$

Note that $\mathrm{tr}(\Sigma_{11}\Sigma_{12}\Sigma_{22}\Sigma_{21}) = \mathrm{tr}(B_{(1)}^T B_{(1)} B_{(2)}^T B_{(2)})^2 > 0$.
Since $(c_{il}d_{il'} + c_{il'}d_{il})^2 \ge 0$, $\sum_i \sum_{l\ne l'} c_{il}d_{il}c_{il'}d_{il'} + \sum_i \sum_{l\ne l'} c_{il'}^2 d_{il}^2 \ge 0$. We obtain $L_2 \ge 2L_1$. Thus, by carrying out the same route as it in Lemma E.2, we can show

$$\begin{aligned}
|\Xi|_F^2 &\ge \min\left\{\frac{\nu^2}{4}, 1\right\} \left(\mathrm{tr}(\Sigma_{11}^2)\mathrm{tr}(\Sigma_{22}^2) + \mathrm{tr}^2(\Sigma_{12}\Sigma_{21}) + 2\mathrm{tr}(\Sigma_{11}\Sigma_{12}\Sigma_{22}\Sigma_{21})\right) \\
&> \min\left\{\frac{\nu^2}{4}, 1\right\} \left(\mathrm{tr}(\Sigma_{11}^2)\mathrm{tr}(\Sigma_{22}^2) + \mathrm{tr}^2(\Sigma_{12}\Sigma_{21}))\right).
\end{aligned}$$

Similarly,

$$|\Xi|_F^2 \le \max\left\{\frac{\nu^2}{2}, \frac{(\nu-2)^2}{2} + 2\right\} \left(\mathrm{tr}(\Sigma_{11}^2)\mathrm{tr}(\Sigma_{22}^2) + \mathrm{tr}^2(\Sigma_{12}\Sigma_{21}))\right).$$

□

PROOF OF THEOREM A.2. It can be carried out following the same routes as those in the proofs of Theorems 2.2 and A.1. □

PROOF OF THEOREM 4.1. Note that $\hat{\Omega}\hat{\Sigma} = I$ and $\Omega\Sigma = I$. Then,

$$\Omega\Sigma + (\hat{\Omega} - \Omega)\Sigma + \Omega(\hat{\Sigma} - \Sigma) + (\hat{\Omega} - \Omega)(\hat{\Sigma} - \Sigma) = I.$$

It follows that

$$\hat{\Omega} - \Omega = -\Omega(\hat{\Sigma} - \Sigma)\Omega - (\hat{\Omega} - \Omega)(\hat{\Sigma} - \Sigma)\Omega.$$

Let $\Omega_{j\cdot}$ be the $j-$th row of Ω and $\Omega_{\cdot k}$ be the $k-$th column of Ω. Then,

$$\hat{\omega}_{jk} - \omega_{jk} = -\Omega_{j\cdot}(\hat{\Sigma} - \Sigma)\Omega_{\cdot k} - (\hat{\Omega}_{j\cdot} - \Omega_{j\cdot})(\hat{\Sigma} - \Sigma)\Omega_{\cdot k}.$$

Basic calculation shows that

$$\begin{aligned} R := \sum_{j,k\in\mathcal{S}} (\hat{\omega}_{jk} - \omega_{jk})^2 &= \sum_{j,k\in\mathcal{S}} (\Omega_{j\cdot}(\hat{\Sigma} - \Sigma)\Omega_{\cdot k})^2 + \sum_{j,k\in\mathcal{S}} ((\hat{\Omega}_{j\cdot} - \Omega_{j\cdot})(\hat{\Sigma} - \Sigma)\Omega_{\cdot k})^2 \\ &\quad +2\sum_{j,k\in\mathcal{S}} \Omega_{j\cdot}(\hat{\Sigma} - \Sigma)\Omega_{\cdot k}(\hat{\Omega}_{j\cdot} - \Omega_{j\cdot})(\hat{\Sigma} - \Sigma)\Omega_{\cdot k} \\ &:= R_1 + R_2 + 2R_3. \end{aligned}$$

By the Cauchy-Schwarz inequality, $|R_3| \le \sqrt{R_1 R_2}$, then we have

$$|\sqrt{R} - \sqrt{R_1}| \le \sqrt{R_2}.$$

For the index set $\mathcal{S} \subset \{(j,k) : 1 \le j,k \le p\}$, we write

$$R_1 = \sum_{(j,k)\in\mathcal{S}} \left(-\sum_{m,q=1}^{p} \omega_{jm}\omega_{kq}(\hat{\sigma}_{mq} - \sigma_{mq}) \right)^2 := \frac{1}{n^2}\sum_{i,l=1}^{n} \boldsymbol{W}_i \Lambda \boldsymbol{W}_l,$$

where $\Lambda = (\Lambda_{(m_1,q_1),(m_2,q_2)})_{1\le m_1,m_2,q_1,q_2\le p}$ with

$$\Lambda_{(m_1,q_1),(m_2,q_2)} = \sum_{j,k\in\mathcal{S}} \omega_{jm_1}\omega_{jm_2}\omega_{kq_1}\omega_{kq_2}.$$

By Assumption 4.2, for $1 \le j \le p$, $0 < K_0^{-1} \le |\Omega_{j\cdot}|_2 \le K_0$ and $\lambda_{\max}(\Sigma) \le K_0$. Let $\|A\|_2$ be the spectral norm of matrix A. Employing Proposition 2.1 in Vershynin (2012), we have $\|\hat{\Sigma} - \Sigma\|_2 = O_{\mathsf{P}}(\sqrt{p/n})$ under Assumption 4.1. Since $p/n \to 0$, we have $\|\Omega\|_2\|\hat{\Sigma} - \Sigma\|_2 \to 0$ in probability. Under $\|\Omega\|_2\|\hat{\Sigma} - \Sigma\|_2 < 1$, by Demmel (1997), for any vector u,

$$\frac{|(\hat{\Omega} - \Omega)u|_2}{|\Omega u|_2} \le \|[I + \Sigma^{-1}(\hat{\Sigma} - \Sigma)]^{-1}\|_2\|\Sigma^{-1}\|_2\|\hat{\Sigma} - \Sigma\|_2 \le \frac{\|\Omega\|_2\|\hat{\Sigma} - \Sigma\|_2}{1 - \|\Omega\|_2\|\hat{\Sigma} - \Sigma\|_2},$$

which implies

$$\|\hat{\Omega} - \Omega\|_2 \le \frac{\|\Omega\|_2^2\|\hat{\Sigma} - \Sigma\|_2}{1 - \|\Omega\|_2\|\hat{\Sigma} - \Sigma\|_2}.$$

Thus, $\|\hat{\Omega} - \Omega\|_2 = O_{\mathsf{P}}(\sqrt{p/n})$ and $|\hat{\Omega}_{j\cdot} - \Omega_{j\cdot}|_2 = O_{\mathsf{P}}(\sqrt{p/n})$.

By the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
R_2 = \sum_{j,k\in\mathcal{S}} ((\hat{\Omega}_{j\cdot} - \Omega_{j\cdot})(\hat{\Sigma} - \Sigma)\Omega_{\cdot k})^2 &\le \sum_{j,k\in\mathcal{S}} |(\hat{\Omega}_{j\cdot} - \Omega_{j\cdot})(\hat{\Sigma} - \Sigma)|_2^2 \cdot |\Omega_{\cdot k}|_2^2 \\
&\le K_0^2 \sum_{j,k\in\mathcal{S}} |\hat{\Omega}_{j\cdot} - \Omega_{j\cdot}|_2^2 \cdot \|\hat{\Sigma} - \Sigma\|_2^2 \\
&\le K_0^2 |\mathcal{S}| \cdot \|\hat{\Omega} - \Omega\|_2^2 \cdot \|\hat{\Sigma} - \Sigma\|_2^2 \\
&= O_P(|\mathcal{S}| p^2 n^{-2}). \qquad \text{(E.22)}
\end{aligned}
$$

Recall that $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_{p(p-1)} \ge 0$ are eigenvalues of $\Lambda^{1/2}\Gamma\Lambda^{1/2}$, $f_1 = \mathrm{tr}(\Lambda\Gamma)$, $f_2 = (\mathrm{tr}(\Lambda\Gamma)^2)^{1/2}$. Note that $f_1 \ge f_2$. Denote $V := \sum_{d=1}^{p(p-1)} \lambda_d(\eta_d - 1)/f_2$, where η_d are i.i.d. χ_1^2. We first consider case (i). By Corollary F.1,

$$
\sup_t \left| \mathsf{P}\left(\frac{nR_1 - f_1}{f_2} \le t \right) - \mathsf{P}(V \le t) \right| \longrightarrow 0.
$$

Elementary calculation shows that

$$
\begin{aligned}
\mathsf{P}\left(\frac{nR - f_1}{f_2} \le t \right) &\le \mathsf{P}\left(\frac{n(\sqrt{R_1} + \sqrt{R_2})^2 - f_1}{f_2} \le t \right) \\
&= \mathsf{P}\left(\sqrt{nR_1} \le \sqrt{tf_2 + f_1} - \sqrt{nR_2} \right) \\
&= \mathsf{P}\left(\frac{nR_1 - f_1}{f_2} \le \frac{(\sqrt{tf_2 + f_1} - \sqrt{nR_2})^2 - f_1}{f_2} \right).
\end{aligned}
$$

Similarly,

$$
\mathsf{P}\left(\frac{nR - f_1}{f_2} \le t \right) \ge \mathsf{P}\left(\frac{nR_1 - f_1}{f_2} \le \frac{(\sqrt{tf_2 + f_1} + \sqrt{nR_2})^2 - f_1}{f_2} \right).
$$

By (E.22), $nR_2/f_2 + nR_2 f_1/f_2^2 \to 0$ in probability. Note that when $t \to -\infty$ (resp. $t \to \infty$), $\mathsf{P}(V \le t) \to 0$ (resp. $\mathsf{P}(V \le t) \to 1$), then (4.3) holds.

Theorem 4.1(ii) follows by the same routes.

□

APPENDIX F: LEMMAS FOR GAUSSIAN APPROXIMATION

In this section, we present the following lemmas, which are used in the proofs of the paper.

LEMMA F.1 (Burkholder (1988), Rio (2009)). *Let $d > 1$ and $d' = \min\{d, 2\}$; let $D_t, 1 \le t \le n$, be martingale differences, and $D_t \in \mathcal{L}^d$ for* *every t. Write $M_n = \sum_{t=1}^n D_t$. Then*

$$\|M_n\|_d^{d'} \leq C_d^{d'} \sum_{t=1}^n \|D_t\|_d^{d'}, \tag{F.1}$$

where $C_d = (d-1)^{-1}$ if $1 < d \leq 2$ and $C_d = \sqrt{d-1}$ if $d > 2$.

LEMMA F.2. *Let $|a_1| \geq |a_2| \geq ... \geq |a_p| \geq 0$ be such that $\sum_{i=1}^p a_i^2 = 1$; let η_i be i.i.d. χ^2 random variables. Then for all $h > 0$,*

$$\sup_t P(t \leq a_1\eta_1 + \cdots + a_p\eta_p \leq t+h) \leq h^{1/2}\sqrt{4/\pi}. \tag{F.2}$$

REMARK F.1. *In the setting of Lemma F.2, if $|a_2| \geq c$ for some constant $c > 0$, then by elementary calculations the density of $a_1\eta_1 + a_2\eta_2$ is uniformly bounded. So the left hand side of (F.2) has bound $O(h)$ if either $|a_2| \geq c$ for some constant $c > 0$ or $|a_1| \leq 1/2$.*

PROOF. It follows from Lemma 6.2 in Xu, Zhang and Wu (2014). For the sake of completeness, we provide their proofs.
Write $V = \sum_{j=1}^p a_j\eta_j$. Assume $|a_1| \leq 1/2$. Then its characteristic function $\phi_V(s) = \mathsf{E}\exp(\sqrt{-1}sV)$, $s \in \mathbb{R}$, satisfies

$$\begin{aligned} |\phi_V(s)| &= \left|\prod_{j=1}^p (1 - 2\sqrt{-1}a_j s)^{-1/2}\right| \\ &= \prod_{j=1}^p (1 + 4a_j^2 s^2)^{-1/4} \\ &\leq (1 + 4s^2 + 8b_4 s^4 + 32/3 b_6 s^6)^{-1/4}, \end{aligned} \tag{F.3}$$

where $b_4 = \sum_{j\neq k} a_j^2 a_k^2 = 1 - \sum_{k=1}^p a_k^4 \geq 1 - a_1^2 \geq 3/4$ and

$$\begin{aligned} b_6 &= 1 - 3\sum_{j\neq k} a_j^4 a_k^2 - \sum_j a_j^6 \\ &\geq 1 - 3\sum_j a_j^4 \left(\sum_{k\neq j} a_k^2 + a_j^2\right) \geq 1 - 3a_1^2 \geq 1/4. \end{aligned}$$

By the inversion formula and (F.3), the density function $f_V(\cdot)$ of V satisfies

$$f_V(v) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\sqrt{-1}vs}\phi_V(s)ds \leq \frac{1}{2\pi}\int_{-\infty}^{\infty} |\phi_V(s)|ds < 1$$

Now we shall deal with the case that $|a_1| > 1/2$. Note that for all $w > 0$, $\sup_u \mathsf{P}(u \le \eta_1 \le u + w) \le w^{1/2}\sqrt{2/\pi}$. Then $\sup_t \mathsf{P}(t \le V \le t+h) \le (2h)^{1/2}\sqrt{2/\pi}$. Combining with the case $|a_1| \le 1/2$, we obtain the upper bound $\max(h^{1/2}\sqrt{4/\pi}, h)$. Note that (F.2) trivially holds if $h \ge 1$. □

LEMMA F.3. *Let Y_i be i.i.d $N(\mu, \Sigma)$ and $\delta > 0$, then*

$$\begin{aligned} \mathsf{E}\left|\frac{Y_1^T Y_1 - \mathrm{tr}(\Sigma)}{\|\Sigma\|_F}\right|^{2+\delta} &\le \nu_\delta^{2+\delta}, \\ \mathsf{E}\left|\frac{Y_1^T Y_2}{\|\Sigma\|_F}\right|^{2+\delta} &\le d_\delta^{2+\delta}, \end{aligned}$$

where ξ is standard normal distribution, $\nu_\delta = 6(2+\delta)\|\xi^2\|_{2+\delta}(1+\mu^T\mu/\|\Sigma\|_F)$ and $d_\delta = 6(1+\delta)\|\xi\|_{2+\delta}^2(1+\mu^T\mu/\|\Sigma\|_F)$.

PROOF. It can be carried out following the same routes as those in the proofs of Lemma E.1 in the manuscript. □

Assume $W_1, ..., W_n$ are i.i.d. with mean 0 and covariance matrix Σ. Let $\lambda_1 \ge ... \ge \lambda_p$ be the eigenvalues of Σ. Let $Z_1, ..., Z_n$ be i.i.d. $N(0, \Sigma)$. Define the Lyapunov-type condition,

$$\mathsf{E}\left|\frac{W_1^T W_2}{\|\Sigma\|_F}\right|^{2+\delta} := (K_\delta)^{2+\delta}, \tag{C1}$$

with K_δ bounded.

LEMMA F.4. *Assume that (C1) holds with $0 < \delta \le 1$. Then,*

$$\sup_t |\mathsf{P}(R_n \le t) - \mathsf{P}(R_n^\diamond \le t)| = O(n^{-\delta/(10+4\delta)})$$

and

$$\sup_t |\mathsf{P}(R_n \le t) - \mathsf{P}(R_n^* \le t)| = O(n^{-\delta/(10+4\delta)}),$$

where

$$R_n = \frac{\frac{1}{n-1}\sum_{i\ne l}^n W_i^T W_l}{\|\Sigma\|_F}, R_n^\diamond = \frac{\frac{1}{n-1}\sum_{i\ne l}^n Z_i^T Z_l}{\|\Sigma\|_F}, \text{ and } R_n^* = \frac{Z_1^T Z_1 - \mathrm{tr}(\Sigma)}{\|\Sigma\|_F}.$$

PROOF. Let $h(x) = (1 - \min(1, \max(x, 0))^4)^4$ and $h_{\phi,t}(u) = h(\phi(u - t))$, $\phi > 0$. Then it is easy to show

$$\begin{aligned}
&h_* = \sup_x \{|h'(x)| + |h''(x)| + |h'''(x)|\} < \infty,\\
&\sup_{u,t} |h'_{\phi,t}(u)| \le h_*\phi, \quad \sup_{u,t} |h''_{\phi,t}(u)| \le h_*\phi^2, \quad \sup_{u,t} |h'''_{\phi,t}(u)| \le h_*\phi^3,\\
&\mathbf{1}_{u\le t} \le h_{\phi,t}(u) \le \mathbf{1}_{u \le t+\phi^{-1}}.
\end{aligned}$$

Then, $\mathsf{P}(R_n \le t) \le \mathsf{E}h_{\phi,t}(R_n)$.

We first show that

$$\sup_t |\mathsf{E}h_{\phi,t}(R_n) - \mathsf{E}h_{\phi,t}(R_n^\diamond)| \le CL_\delta(n, \phi), \tag{F.4}$$

where

$$L_\delta(n,\phi) = \left\{ \frac{\mathsf{E}(W_1 \Sigma W_1)^{1+\delta/2}}{n^{\delta/2}\|\Sigma\|_F^{2+\delta}} + \frac{K_\delta^{2+\delta}}{n^{1+\delta/2}} \right\} \phi^{2+\delta}.$$

Let $\Gamma_i = \sum_{l=1}^{i-1} W_l + \sum_{l=i+1}^n Z_l$ and

$$\begin{aligned}
H_i &= \frac{\Gamma_i^T\Gamma_i - \sum_{l=1}^{i-1} W_l^T W_l - \sum_{l=i+1}^n Z_l^T Z_l}{(n-1)\|\Sigma\|_F},\\
J_i &= \frac{2\Gamma_i^T W_i}{(n-1)\|\Sigma\|_F},\\
M_i &= \frac{2\Gamma_i^T Z_i}{(n-1)\|\Sigma\|_F}.
\end{aligned}$$

Note H_i and Γ_i are independent of W_i and Z_i.

By taylor expansion, we have

$$h_{\phi,t}(H_i + J_i) - h_{\phi,t}(H_i + M_i) = I + II + III,$$

where

$$\begin{aligned}
I &= h'_{\phi,t}(H_i)(J_i - M_i),\\
II &= \frac{1}{2} h''_{\phi,t}(H_i)(J_i^2 - M_i^2).
\end{aligned}$$

Then,

$$\begin{aligned}
\mathsf{E}I &= \mathsf{E}\{\mathsf{E}(h'_{\phi,t}(H_i)(J_i - M_i)|W_i, Z_i)\}\\
&= \frac{2}{(n-1)\|\Sigma\|_F} \mathsf{E}\left[(W_i^T - Z_i^T)\mathsf{E}(h'_{\phi,t}(\Gamma_i)H_i)\right] = 0,
\end{aligned}$$

$$\begin{aligned}\mathsf{E}II &= \frac{1}{2}\mathsf{E}\{\mathsf{E}[h''_{\phi,t}(H_i)(J_i^2 - M_i^2)|W_1, ..., W_{i-1}, Z_{i+1}, ..., Z_n]\} \\ &= \frac{2}{(n-1)^2\|\Sigma\|_F^2}\mathsf{E}\left[h''_{\phi,t}(H_i)\mathsf{E}(\Gamma_i^T W_i W_i^T \Gamma_i - \Gamma_i Z_i Z_i^T \Gamma_i | W_1, ..., W_{i-1}, Z_{i+1}, ..., Z_n)\right] \\ &= 0.\end{aligned}$$

Since $0 \le h(x) \le 1$ for all x and $|h'''_{\phi,t}(u)| \le h_*\phi^3$, we have

$$\begin{aligned}\mathsf{E}III &\le \mathsf{E}\min\left\{h_*\phi^2(|J_i|^2 + |M_i|^2), h_*\phi^3(|J_i|^3 + |M_i|^3)\right\} \\ &\le C\phi^{2+\delta}(\mathsf{E}|J_i|^{2+\delta} + \mathsf{E}|M_i|^{2+\delta}).\end{aligned}$$

Denote $q = 2 + \delta$. For a fixed vector $y \in \mathbb{R}^p$, $Z_n^T y \sim N(0, y^T\Sigma y)$. Then $\mathsf{E}|Z_n^T y|^q = c_q(y^T\Sigma y)^{q/2}$.
By Rosenthal's inequality,

$$\mathsf{E}|\Gamma_i y|_q^q \le c_q\left(i\|W_1^T y\|_q^q + (n-i)\|Z_n^T y\|_q^q + n^{q/2}(y^T\Sigma y)^{q/2}\right).$$

Thus,

$$\|\Gamma_i W_i\|_q^q \le c_q\left(n\|W_1^T W_2\|_q^q + n^{q/2}\mathsf{E}(W_1^T\Sigma W_1)^{q/2}\right).$$

Hence,

$$\mathsf{E}|J_i|^q \le C\frac{n\|W_1^T W_2\|_q^q + n^{q/2}\mathsf{E}(W_1^T\Sigma W_1)^{q/2}}{n^q\|\Sigma\|_F^q}.$$

Similarly, $\|\Gamma_i Z_i\|_q^q \le c_q\left(n\mathsf{E}(W_1^T\Sigma W_1)^{q/2} + n^{q/2}\|\Sigma\|_F^q\right)$. So

$$\mathsf{E}|M_i|^q \le C\frac{n\mathsf{E}(W_1^T\Sigma W_1)^{q/2}}{n^q\|\Sigma\|_F^q} + n^{-q/2}.$$

Observe that

$$h_{\phi,t}(R_n) - h_{\phi,t}(R_n^\diamond) = \sum_{i=1}^n [h_{\phi,t}(H_i + J_i) - h_{\phi,t}(H_i + M_i)].$$

By Hölder inequality, since $\mathsf{E}(W_1^T\Sigma W_1)^{q/2} \ge (\mathsf{E}W_1^T\Sigma W_1)^{q/2} = \|\Sigma\|_F^q$,

$$\begin{aligned}\sup|\mathsf{E}h_{\phi,t}(R_n) - \mathsf{E}h_{\phi,t}(R_n^\diamond)| &\le C\phi^q\left\{\frac{K_\delta^q}{n^{q-2}} + \frac{\mathsf{E}(W_1^T\Sigma W_1)^{q/2}}{n^{q/2-1}\|\Sigma\|_F^q} + \frac{1}{n^{q/2-1}}\right\} \\ &\le C\phi^q\left\{\frac{K_\delta^q}{n^{q-2}} + \frac{\mathsf{E}(W_1^T\Sigma W_1)^{q/2}}{n^{q/2-1}\|\Sigma\|_F^q}\right\}.\end{aligned}$$

Thus,

$$\mathsf{P}(R_n \le t) \le \mathsf{E}h_{\phi,t}(R_n) \le \mathsf{E}h_{\phi,t}(R_n^\diamond) + CL_\delta(n,\phi) \le \mathsf{P}(R_n^\diamond \le t + \phi^{-1}) + CL_\delta(n,\phi).$$

Similarly, we can get

$$\mathsf{P}(R_n \le t) \ge \mathsf{P}(R_n^\diamond \le t - \phi^{-1}) - CL_\delta(n,\phi).$$

Let $\eta_1, ..., \eta_p$ be i.i.d. χ^2_1, $\zeta_1, ..., \zeta_p$ be i.i.d. χ^2_{n-1} and they are mutually independent. Recall that $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_p \ge 0$. Observe that

$$R_n^\diamond = \frac{\sum_{i \ne l} Z_i^T Z_l}{(n-1)\|\Sigma\|_F} =_D \sum_{j=1}^p \frac{\lambda_j}{\|\Sigma\|_F}\left(\eta_j - \frac{\zeta_j}{n-1}\right) =_D R_n^* - R_\triangle,$$

where

$$R_\triangle = \frac{1}{(n-1)\|\Sigma\|_F}\sum_{j=1}^p \lambda_j(\zeta_j - (n-1)).$$

Note that $\mathsf{E}R^2_\triangle = 2(n-1)^{-1}$. By Lemma F.2 and the Markov and triangle inequalities,

$$\mathsf{P}(R_n^\diamond \le t) \le \mathsf{P}(R_n^* \le t - \varepsilon) + \mathsf{P}(|R_\triangle| \ge \varepsilon) \le \mathsf{P}(R_n^* \le t) + \sqrt{\varepsilon}\sqrt{4\pi} + \frac{2}{(n-1)\varepsilon^2}.$$

Similarly, we have

$$\mathsf{P}(R_n^\diamond \le t) \ge \mathsf{P}(R_n^* \le t) - \sqrt{\varepsilon}\sqrt{4\pi} - \frac{2}{(n-1)\varepsilon^2}.$$

Taking $\varepsilon = n^{-2/5}$,

$$|\mathsf{P}(R_n^\diamond \le t) - \mathsf{P}(R_n^* \le t)| \le 3(n-1)^{-1/5}.$$

Applying Lemma F.2, we obtain

$$\mathsf{P}(R_n \le t) \le \mathsf{E}h_{\phi,t}(R_n) \le \mathsf{E}h_{\phi,t}(R_n^\diamond) + CL_\delta(n,\phi) \le \mathsf{P}(R_n^\diamond \le t + \phi^{-1}) + CL_\delta(n,\phi).$$

$$\sup_t |\mathsf{P}(R_n \le t) - \mathsf{P}(R_n^\diamond \le t)| = O(L_\delta(n,\phi) + \phi^{-1/2} + n^{-1/5})$$

and

$$\sup_t |\mathsf{P}(R_n \le t) - \mathsf{P}(R_n^* \le t)| = O(L_\delta(n,\phi) + \phi^{-1/2} + n^{-1/5}).$$

By Jensen's inequality,

$$\mathsf{E}(W_1^T \Sigma W_1)^{q/2} \le \mathsf{E}|W_1^T W_2|^{q/2} = K_\delta^q \|\Sigma\|_F^q.$$

Then we can choose $\phi \asymp n^{(q-2)/(1+2q)} = n^{\delta/(5+2\delta)}$ and the corresponding convergence rate is $O(n^{-\delta/(10+4\delta)})$. □

For symmetric matrix A, define the Lyapunov-type condition,

$$\text{(C2)}\qquad \begin{aligned} \mathsf{E}\left|\frac{W_1^T(I-A)W_2}{\|\Sigma - A\Sigma\|_F}\right|^{2+\delta} &= (K_\delta^A)^{2+\delta}, \\ \mathsf{E}\left|\frac{W_1^T A W_1 - \operatorname{tr}(A\Sigma)}{\|\Sigma - A\Sigma\|_F}\right|^{2+\varrho} &= (\kappa_\varrho)^{2+\varrho}, \\ \mathsf{E}\left|\frac{Z_1^T A Z_1 - \operatorname{tr}(A\Sigma)}{\|\Sigma - A\Sigma\|_F}\right|^{2+\varrho} &= (c_\varrho)^{2+\varrho}. \end{aligned}$$

Lemma F.5. *Assume that (C2) holds with $0 < \delta \le 1$, $\varrho \ge 0$ and K_δ^A bounded. Then,*

$$\sup_t |P(Q_n \le t) - P(Q_n^\diamond \le t)| = O(\phi^{-1/2})$$

where

$$\phi^2\{\frac{1}{\sqrt{n}}\kappa_0 + \frac{1}{n}\kappa_0^2\} + \phi^{2+\delta}\{\frac{1}{n^{\delta/2}} + \frac{1}{n^{1+\delta}}\} = \phi^{-1/2},$$

$$\begin{aligned} Q_n &= \frac{1}{\|\Sigma - A\Sigma\|_F}\left\{\frac{1}{n-1}\sum_{i\ne l}^n W_i^T W_l - \frac{1}{n^2}\sum_{i,l}^n W_i A W_l - \operatorname{tr}(\Sigma)\right\}, \\ Q_n^\diamond &= \frac{1}{\|\Sigma - A\Sigma\|_F}\left\{\frac{1}{n-1}\sum_{i\ne l}^n Z_i^T Z_l - \frac{1}{n^2}\sum_{i,l}^n Z_i A Z_l - \operatorname{tr}(\Sigma)\right\}. \end{aligned}$$

Then the convergence rate is $n^{-\delta/(10+4\delta)} + \kappa_0^{2/5} n^{-1/5}$, which goes to 0 if and only if $\kappa_0/\sqrt{n} \to 0$.

Proof. Firstly, following the same procedure in the proof of Lemma F.4, we need to show

$$\text{(F.5)}\qquad \sup_t |\mathsf{E}h_{\phi,t}(Q_n) - \mathsf{E}h_{\phi,t}(Q_n^\diamond)| \le C L_\delta^\dagger(n,\phi),$$

where

$$L_\delta^\dagger(n,\phi) = \phi^2\{-\frac{4}{\sqrt{n}}\kappa_0 + \frac{1}{n}(\kappa_0^2 + c_0^2)\} + \phi^{2+\delta}\frac{(K_\delta^A)^{2+\delta}}{n^{\delta/2}} + \frac{1}{n^{1+\varrho}}\phi^{2+\varrho}(\kappa_\varrho^{2+\varrho} + c_\varrho^{2+\varrho}).$$

Let $\Gamma_i = \sum_{l=1}^{i-1} W_l + \sum_{l=i+1}^{n} Z_l$ and

$$
\begin{aligned}
H_i &= \frac{1}{\|\Sigma - A\Sigma\|_F}\left[\frac{1}{n-1}\Gamma_i^T(I - \frac{n-1}{n}A)\Gamma_i - \frac{1}{n-1}\sum_{l=1}^{i-1} W_l^T W_l \right. \\
&\left. -\frac{1}{n-1}\sum_{l=i+1}^{n} Z_l^T Z_l - \mathrm{tr}(\Sigma) - \frac{1}{n}\mathrm{tr}(A\Sigma)\right], \\
J_i &= \frac{\frac{2}{n-1}\Gamma_i^T(I - \frac{n-1}{n}A)W_i - \frac{1}{n}W_i^T A W_i + \frac{1}{n}\mathrm{tr}(A\Sigma)}{\|\Sigma - A\Sigma\|_F}, \\
M_i &= \frac{\frac{2}{n-1}\Gamma_i^T(I - \frac{n-1}{n}A)Z_i - \frac{1}{n}Z_i^T A Z_i + \frac{1}{n}\mathrm{tr}(A\Sigma)}{\|\Sigma - A\Sigma\|_F}.
\end{aligned}
$$

Note H_i and Γ_i are independent of W_i and Z_i.

By taylor expansion, we have

$$h_{\phi,t}(H_i + J_i) - h_{\phi,t}(H_i + M_i) = I + II + III,$$

where

$$
\begin{aligned}
I &= h'_{\phi,t}(H_i)(J_i - M_i), \\
II &= \frac{1}{2}h''_{\phi,t}(H_i)(J_i^2 - M_i^2).
\end{aligned}
$$

It is easy to show that $\mathsf{E}I = 0$.
Since $\mathsf{E}(W_i W_i^T|\Gamma_i) = \mathsf{E}(Z_i Z_i^T|\Gamma_i)$ and $\mathsf{E}W_1^T A W_1 = \mathsf{E}Z_1^T A Z_1 = \mathrm{tr}(A\Sigma)$, we have

$$
\begin{aligned}
&\mathsf{E}II \le c\phi^2 \mathsf{E}\{\mathsf{E}[J_i^2 - M_i^2|W_1, ..., W_{i-1}, Z_{i+1}, ..., Z_n]\} \\
&\le \frac{c\phi^2}{\|\Sigma - A\Sigma\|_F^2}\mathsf{E}\left\{\frac{4}{n(n-1)}\mathsf{E}\left[\Gamma_i(I - \frac{n-1}{n}A)Z_i(Z_i A Z_i - \mathrm{tr}(A\Sigma))|W_1, ..., W_{i-1}, Z_{i+1}, ..., Z_n\right]\right. \\
&\quad - \frac{4}{n(n-1)}\mathsf{E}\left[\Gamma_i(I - \frac{n-1}{n}A)W_i(W_i A W_i - \mathrm{tr}(A\Sigma))|W_1, ..., W_{i-1}, Z_{i+1}, ..., Z_n\right] \\
&\quad \left. + \frac{1}{n^2}\mathsf{E}\left[(W_i^T A W_i - \mathrm{tr}(A\Sigma))^2 - (Z_i^T A Z_i - \mathrm{tr}(A\Sigma))^2|W_1, ..., W_{i-1}, Z_{i+1}, ..., Z_n\right]\right\}.
\end{aligned}
$$

Note $\mathsf{E}\left[\Gamma_i(I - \frac{n-1}{n}A)Z_i(Z_i A Z_i - \mathrm{tr}(A\Sigma))|W_1, ..., W_{i-1}, Z_{i+1}, ..., Z_n\right] = 0$.
By Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\mathsf{E}\left|\Gamma_i(I - \frac{n-1}{n}A)W_i(W_i A W_i - \mathrm{tr}(A\Sigma))\right| \\
&\le \|\Gamma_i(I - \frac{n-1}{n}A)W_i\|_2 \, \|W_i A W_i - \mathrm{tr}(A\Sigma)\|_2 \\
&= \sqrt{n-1}\|(I - \frac{n-1}{n}A)\Sigma\|_F \, \|W_i A W_i - \mathrm{tr}(A\Sigma)\|_2 .
\end{aligned}
$$

Thus, by Lyapunov-type condition (C2),

$$\mathsf{E}II \leq \frac{c\phi^2}{\|\Sigma - A\Sigma\|_F^2}\left\{-\frac{4}{n^{3/2}}\kappa_0\|(I - \frac{n-1}{n}A)\Sigma\|_F\|(I-A)\Sigma\|_F + \frac{1}{n^2}(\kappa_0^2 + c_0^2)\|(I-A)\Sigma\|_F^2\right\}.$$

Employing similar derivations, for sufficient large n,

$$\mathsf{E}III \leq c\left\{\frac{(\kappa_\varrho^{2+\varrho} + c_\varrho^{2+\varrho})}{n^{2+\varrho}}\phi^{2+\varrho} + \frac{(K_\delta^A)^{2+\delta}}{n^{1+\delta/2}}\phi_{2+\delta}\right\}.$$

Then, by basic calculation, we can show (F.5). Note c_ϱ is a constant. The rest of the proof follows from the same procedure in the proof of Lemma F.4. □

Define the Lyapunov-type condition,

$$\mathsf{E}\left|\frac{W_1^T A W_2}{\|A\Sigma\|_F}\right|^{2+\delta} = (K_\delta^B)^{2+\delta},$$

$$\text{(C3)} \qquad \mathsf{E}\left|\frac{W_1^T A W_1 - \mathrm{tr}(A\Sigma)}{\|A\Sigma\|_F}\right|^{2+\varrho} = (\tau_\varrho)^{2+\varrho},$$

COROLLARY F.1. *Assume that (C3) holds with $0 < \delta \leq 1$, $\varrho \geq 0$ and K_δ^B bounded. Then,*

$$\sup_t\left|P\left(\frac{1}{n}\sum_{i,l=1}^n \frac{W_i^T A W_l - \mathrm{tr}(A\Sigma)}{\|A\Sigma\|_F} \leq t\right) - P\left(\frac{1}{n}\sum_{i,l=1}^n \frac{Z_i^T A Z_l - \mathrm{tr}(A\Sigma)}{\|A\Sigma\|_F} \leq t\right)\right| = O(\phi^{-1/2})$$

where

$$\phi^2\{\frac{1}{\sqrt{n}}\tau_0 + \frac{1}{n}\tau_0^2\} + \phi^{2+\delta}\{\frac{1}{n^{\delta/2}} + \frac{1}{n^{1+\delta}}\} = \phi^{-1/2}.$$

Then the convergence rate is $n^{-\delta/(10+4\delta)} + \tau_0^{2/5}n^{-1/5}$, which goes to 0 if and only if $\tau_0/\sqrt{n} \to 0$.

LEMMA F.6. *Let $B_1, B_2 \in \{1, 2, ..., n\}$ and $|B_1| = |B_2| = m = n/2$. For*

$k = 1, 2,$ *denote*

$$\begin{aligned} R_{n,k} &= \frac{1}{2\|\Sigma\|_F}\left\{\frac{1}{m-1}\sum_{i\neq l\in B_k} W_i^T W_l + \frac{1}{m-1}\sum_{i'\neq l'\in B_k^c} W_{i'}^T W_{l'} \right. \\ &\qquad \left. -\frac{2}{m}\sum_{i\in B_k, i'\in B_k^c} W_i^T W_{i'}\right\}, \\ R^{\diamond}_{n,k} &= \frac{1}{2\|\Sigma\|_F}\left\{\frac{1}{m-1}\sum_{i\neq l\in B_k} Z_i^T Z_l + \frac{1}{m-1}\sum_{i'\neq l'\in B_k^c} Z_{i'}^T Z_{l'} \right. \\ &\qquad \left. -\frac{2}{m}\sum_{i\in B_k, i'\in B_k^c} Z_i^T Z_{i'}\right\}. \end{aligned}$$

Let $I_1 = B_1 \cap B_2$, $I_2 = B_1 \cap B_2^c$, $I_3 = B_1^c \cap B_2$ *and* $I_4 = B_1^c \cap B_2^c$. *Define*

$$d(B_1, B_2) = \max\left\{\left||I_1| - \frac{n}{4}\right|, \left||I_2| - \frac{n}{4}\right|, \left||I_3| - \frac{n}{4}\right|, \left||I_4| - \frac{n}{4}\right|\right\}.$$

Assume that (C1) *holds with* $0 < \delta \le 1$ *and* $d(B_1, B_2) \le n^{1/2}\log n$. *Then,*

$$\sup_t |P(R_{n,1} \le t, R_{n,2} \le t) - P(R^{\diamond}_{n,1} \le t, R^{\diamond}_{n,2} \le t)| = O(n^{-\delta/(10+4\delta)})$$

PROOF. By simple calculation, we have

$$\begin{aligned} \sum_{i\neq l\in B_1} W_i^T W_l &= \sum_{i\neq l\in I_1} W_i^T W_l + \sum_{i\neq l\in I_2} W_i^T W_l + 2\sum_{i\in I_1, l\in I_2} W_i^T W_l, \\ \sum_{i'\neq l'\in B_1^c} W_{i'}^T W_{l'} &= \sum_{i'\neq l'\in I_4} W_{i'}^T W_{l'} + \sum_{i'\neq l'\in I_3} W_{i'}^T W_{l'} + 2\sum_{i'\in I_4, l'\in I_3} W_{i'}^T W_{l'}, \\ \sum_{i\neq l\in B_2} W_i^T W_l &= \sum_{i\neq l\in I_1} W_i^T W_l + \sum_{i\neq l\in I_3} W_i^T W_l + 2\sum_{i\in I_1, l\in I_3} W_i^T W_l, \\ \sum_{i'\neq l'\in B_2^c} W_{i'}^T W_{l'} &= \sum_{i'\neq l'\in I_4} W_{i'}^T W_{l'} + \sum_{i'\neq l'\in I_2} W_{i'}^T W_{l'} + 2\sum_{i'\in I_4, l'\in I_2} W_{i'}^T W_{l'}, \\ \sum_{i\in B_1, i'\in B_1^c} W_i^T W_{i'} &= \sum_{i\in I_1} W_i^T \sum_{i'\in I_4} W_{i'} + \sum_{i\in I_2} W_i^T \sum_{i'\in I_4} W_{i'} + \sum_{i\in I_1} W_i^T \sum_{i'\in I_3} W_{i'} \\ &\quad + \sum_{i\in I_2} W_i^T \sum_{i'\in I_3} W_{i'}, \\ \sum_{i\in B_2, i'\in B_2^c} W_i^T W_{i'} &= \sum_{i\in I_1} W_i^T \sum_{i'\in I_4} W_{i'} + \sum_{i\in I_3} W_i^T \sum_{i'\in I_4} W_{i'} + \sum_{i\in I_1} W_i^T \sum_{i'\in I_2} W_{i'} \\ &\quad + \sum_{i\in I_3} W_i^T \sum_{i'\in I_2} W_{i'}. \end{aligned}$$

Let

$$\Psi_1 = \frac{1}{2(m-1)|\Gamma|_F^2}\left(\sum_{i\neq l\in I_1} W_i^T W_l + \sum_{i'\neq l'\in I_4} W_{i'}^T W_{l'} + \sum_{i\neq l\in I_2} W_i^T W_l + \sum_{i'\neq l'\in I_3} W_{i'}^T W_{l'}\right.$$
$$\left.-2\sum_{i\in I_1} W_i^T \sum_{i'\in I_4} W_{i'} - 2\sum_{l\in I_2} W_l^T \sum_{l'\in I_3} W_{l'}\right),$$

$$\Psi_2 = \frac{1}{2m|\Gamma|_F^2}\left(\sum_{i\in I_1} W_i^T \sum_{l\in I_2} W_l + \sum_{i\in I_3} W_i^T \sum_{l\in I_4} W_l - \sum_{i\in I_1} W_i^T \sum_{l\in I_3} W_l - \sum_{i\in I_2} W_i^T \sum_{l\in I_4} W_l\right),$$

$$\Psi_1^\diamond = \frac{1}{2(m-1)|\Gamma|_F^2}\left(\sum_{i\neq l\in I_1} Y_i^T Y_l + \sum_{i'\neq l'\in I_4} Y_{i'}^T Y_{l'} + \sum_{i\neq l\in I_2} Y_i^T Y_l + \sum_{i'\neq l'\in I_3} Y_{i'}^T Y_{l'}\right.$$
$$\left.-2\sum_{i\in I_1} Y_i^T \sum_{i'\in I_4} Y_{i'} - 2\sum_{l\in I_2} Y_l^T \sum_{l'\in I_3} Y_{l'}\right),$$

$$\Psi_2^\diamond = \frac{1}{2m|\Gamma|_F^2}\left(\sum_{i\in I_1} Y_i^T \sum_{l\in I_2} Y_l + \sum_{i\in I_3} Y_i^T \sum_{l\in I_4} Y_l - \sum_{i\in I_1} Y_i^T \sum_{l\in I_3} Y_l - \sum_{i\in I_2} Y_i^T \sum_{l\in I_4} Y_l\right).$$

Then, we have

$$R_{n,1} = \Psi_1 + 2\Psi_2 + \frac{2}{m(m-1)|\Gamma|_F^2}\left(\sum_{i\in I_1, l\in I_4} W_i^T W_l + \sum_{i\in I_2, l\in I_3} W_i^T W_l\right.$$
$$\left.+\sum_{i\in I_1, l\in I_2} W_i^T W_l + \sum_{i\in I_3, l\in I_4} W_i^T W_l\right),$$

$$R_{n,2} = \Psi_1 - 2\Psi_2 + \frac{2}{m(m-1)|\Gamma|_F^2}\left(\sum_{i\in I_1, l\in I_4} W_i^T W_l + \sum_{i\in I_2, l\in I_3} W_i^T W_l\right.$$
$$\left.+\sum_{i\in I_1, l\in I_3} W_i^T W_l + \sum_{i\in I_2, l\in I_4} W_i^T W_l\right).$$

Simple calculation shows that, for $j\neq k$ and $j,k=1,2,3,4$,

$$\frac{1}{m(m-1)|\Gamma|_F}\sum_{i\in I_j, l\in I_k} W_i^T W_l = O_{\mathsf{P}}(\frac{1}{m}).$$

By Lemma F.2 and triangle inequality, applying similar argument in the proof of (E.10), we obtain,

$$\sup_t |\mathsf{P}(R_{n,1} \le t, R_{n,2} \le t) - \mathsf{P}(\Psi_1 + 2|\Psi_2| \le t)| = O(n^{-1/5}).$$

We approximate $|x|$ by the function

$$p(x) = \begin{cases} -\frac{1}{16}(5x^8 - 21x^6 + 35x^4 - 35x^2) & |x| \le 1 \\ |x| & \text{o.w..} \end{cases}$$

Then we have that for some constant p_*,

$$sup_x\{|p'(x)| + |p''(x)| + |p'''(x)|\} = p_* < \infty.$$

Let $p_\phi(x) = \phi^{-1}p(\phi x)$, then

$$|x| - \phi^{-1} \le p_\phi(x) \le |x| \le p_\phi(x) + \phi^{-1}.$$

Recall $h_{\phi,t}(\cdot)$ in the proof of Lemma F.4. Define $g_{\phi,t}(x,y) = h_{\phi,t}(x+2p_\phi(y))$. If follows that

$$\mathbf{1}_{x+2|y|\le t} \le g_{\phi,t}(x,y) \le \mathbf{1}_{x+2|y|\le t+3\phi^{-1}}.$$

By simple calculation, we can show that for some constant g_*,

$$\sup_{x,y,t}\left|\frac{\partial g_{\phi,t}(x,y)}{\partial x}\right| \le g_*\phi, \qquad \sup_{x,y,t}\left|\frac{\partial g_{\phi,t}(x,y)}{\partial y}\right| \le g_*\phi,$$
$$\sup_{x,y,t}\left|\frac{\partial^2 g_{\phi,t}(x,y)}{\partial x^2}\right| \le g_*\phi^2, \qquad \sup_{x,y,t}\left|\frac{\partial^2 g_{\phi,t}(x,y)}{\partial x\partial y}\right| \le g_*\phi^2,$$
$$\sup_{x,y,t}\left|\frac{\partial^2 g_{\phi,t}(x,y)}{\partial y^2}\right| \le g_*\phi^2, \qquad \sup_{x,y,t}\left|\frac{\partial^3 g_{\phi,t}(x,y)}{\partial x^3}\right| \le g_*\phi^3,$$
$$\sup_{x,y,t}\left|\frac{\partial^3 g_{\phi,t}(x,y)}{\partial x^2\partial y}\right| \le g_*\phi^3, \qquad \sup_{x,y,t}\left|\frac{\partial^3 g_{\phi,t}(x,y)}{\partial x\partial y^2}\right| \le g_*\phi^3,$$
$$\sup_{x,y,t}\left|\frac{\partial^3 g_{\phi,t}(x,y)}{\partial y^3}\right| \le g_*\phi^3.$$

Note two dimensional taylor expansion,

$$
\begin{aligned}
g_{\phi,t}(x,y) &= g_{\phi,t}(x_0,y_0) + \frac{\partial g_{\phi,t}(x_0,y_0)}{\partial x}(x-x_0) + \frac{\partial g_{\phi,t}(x_0,y_0)}{\partial y}(y-y_0) \\
&+ \frac{\partial^2 g_{\phi,t}(x_0,y_0)}{\partial x^2}\frac{(x-x_0)^2}{2} + 2\frac{\partial^2 g_{\phi,t}(x_0,y_0)}{\partial x \partial y}\frac{(x-x_0)(y-y_0)}{2} \\
&+ \frac{\partial^2 g_{\phi,t}(x_0,y_0)}{\partial y^2}\frac{(y-y_0)^2}{2} + \frac{\partial^3 g_{\phi,t}(x_*,y_*)}{\partial x^3}\frac{(x-x_0)^3}{3!} \\
&+ 3\frac{\partial^3 g_{\phi,t}(x_*,y_*)}{\partial x^2 \partial y}\frac{(x-x_0)^2(y-y_0)}{3!} + 3\frac{\partial^3 g_{\phi,t}(x_*,y_*)}{\partial x \partial y^2}\frac{(x-x_0)(y-y_0)}{3!} \\
&+ \frac{\partial^3 g_{\phi,t}(x_*,y_*)}{\partial y^3}\frac{(y-y_0)^3}{3!}.
\end{aligned}
$$

By expanding $g_{\phi,t}(\Psi_1,\Psi_2) - g_{\phi,t}(\Psi_1^\diamond,\Psi_2^\diamond)$ similarly as in Lemma F.4, we can prove, there exist a constant $C > 0$,

$$\sup_t |\mathsf{E} g_{\phi,t}(\Psi_1,\Psi_2) - \mathsf{E} g_{\phi,t}(\Psi_1^\diamond,\Psi_2^\diamond)| \le C L_\delta(n,\phi),$$

where $L_\delta(n,\phi)$ is the same as the one in Lemma F.4.
Thus,

$$
\begin{aligned}
\mathsf{P}(R_{n,1} \le t, R_{n,2} \le t) &\le \mathsf{E} g_{\phi,t}(\Psi_1,\Psi_2) + Cn^{-1/5} \le \mathsf{E} g_{\phi,t}(\Psi_1^\diamond,\Psi_2^\diamond) + CL_\delta(n,\phi) + Cn^{-1/5} \\
&\le \mathsf{P}(R_{n,1}^\diamond \le t + 3\phi^{-1}, R_{n,2} \le t + 3\phi^{-1}) + CL_\delta(n,\phi) + Cn^{-1/5}.
\end{aligned}
$$

Similarly, we can get

$$\mathsf{P}(R_{n,1} \le t, R_{n,2} \le t) \ge \mathsf{P}(R_{n,1}^\diamond \le t - 3\phi^{-1}, R_{n,2} \le t - 3\phi^{-1}) - CL_\delta(n,\phi) - Cn^{-1/5}.$$

Applying Lemma F.2, we obtain

$$\sup_t |\mathsf{P}(R_{n,1} \le t, R_{n,2} \le t) - \mathsf{P}(R_{n,1}^\diamond \le t, R_{n,2}^\diamond \le t)| = O(L_\delta(n,\phi) + \phi^{-1/2} + n^{-1/5}).$$

Then we can choose $\phi \asymp n^{\delta/(5+2\delta)}$ and the corresponding convergence rate is $O(n^{-\delta/(10+4\delta)})$. □

COROLLARY F.2. *Assume that* (C1) *holds with* $0 < \delta \le 1$. *Then,*

$$\sup_t |P(R_{n,1} \le t) - P(R_n^\diamond \le t)| = O(n^{-\delta/(10+4\delta)}).$$

DEPARTMENT OF STATISTICS
GEORGE HERBERT JONES LABORATORY
5747 S. ELLIS AVENUE
CHICAGO, IL 60637
E-MAIL: yhan2@galton.uchicago.edu
wbwu@galton.uchicago.edu