The Annals of Statistics
2020, Vol. 48, No. 6, 3565–3588
https://doi.org/10.1214/20-AOS1943

TEST FOR HIGH DIMENSIONAL COVARIANCE MATRICES

BY YUEFENG HAN[1] AND WEI BIAO WU[2]

[1]*Department of Statistics, Rutgers University, yuefeng.han@rutgers.edu*
[2]*Department of Statistics, University of Chicago, wbwu@galton.uchicago.edu*

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 noncentral 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)).

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

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

where $\sigma_{jk,0}$ are prespecified 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 nonnegative parameters of the covariance. An important task is to test the validity of such parametric forms.

Received September 2018; revised January 2020.
MSC2020 subject classifications. Primary 62F05; secondary 62E17.
Key words and phrases. Gaussian approximation, high dimensional covariance matrices, covariance function, resampling.

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), John (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 $\operatorname{tr}\{(\Sigma - \Sigma(\boldsymbol{\theta}))^2\}$ 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 nonlinear 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 submatrices. 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,

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

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 the Appendix (Supplementary Material, Han and Wu (2020)) 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 $x = (x_1, \dots, x_p)^T$, define $|x| = |x|_2 = (x_1^2 + \cdots + x_p^2)^{1/2}$. Let $\boldsymbol{\xi} = (\xi_1, \dots, \xi_p)^T$ be a random vector. Write $\boldsymbol{\xi} \in \mathcal{L}^m$, $m \ge 1$, 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 $|a_n| \le C|b_n|$ (resp., $1/C \le a_n/b_n \le C$) holds for all sufficiently large n, and write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$. Let $\lceil a \rceil = \min\{k \in \mathbb{Z} : k \ge a\}$.

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

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

Throughout this paper, $\sum^*$ denotes summation over mutually different subscripts shown, for example, $\sum^*_{i_1,i_2,i_3}$ denotes summation over $\{(i_1, i_2, i_3) : i_1 \ne i_2, i_2 \ne i_3, i_1 \ne i_3, 1 \le i_1, i_2, i_3 \le n\}$. 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 the 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:

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

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 $\boldsymbol{W} = \mathcal{W}(\boldsymbol{X}, \mathcal{S}_1)$, $\boldsymbol{W}_i = \mathcal{W}(\boldsymbol{X}_i, \mathcal{S}_1)$ 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

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

The square of the Frobenius norm of Γ is

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

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

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

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

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

where $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$ 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.

$$X_{ij} = \mu_j + \sum_{l_1=1}^N b_{j,l_1}\xi_{il_1} + \sum_{l_1<l_2}^N a_{j,l_1l_2}\xi_{il_1}\xi_{il_2} + \cdots + \sum_{l_1<l_2<\cdots<l_d}^N a_{j,l_1l_2\cdots l_d}\xi_{il_1}\xi_{il_2}\cdots\xi_{il_d}$$

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 $a_{j,l_1l_2} = 0, \dots, a_{j,l_1l_2\cdots l_d} = 0$ for all $1 \le l_1 < l_2 < \cdots < l_d \le N$. The requirement of $\xi_{i1}, \dots, \xi_{iN}$ 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, that is, 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 $C > 0$,

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

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, for example, 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

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

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 $a_{j,l_1l_2\cdots l_i} = 0$ for all $1 \le l_1 < l_2 < \cdots < l_i \le N$, $2 \le i \le d$, $1 \le j \le p$; 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 $0 < \delta \le 1$. Then under the null hypothesis H_{0a} (2.3), we have that*

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

where $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{p^2-p} \ge 0$ are eigenvalues of Γ and η_d, $d \ge 1$, 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 a 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*

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

where $\boldsymbol{Y} \sim N(0, \Gamma)$ *and* $\mu_Y = (\sigma_{12} - \sigma_{12,0}, \sigma_{13} - \sigma_{13,0}, \dots, \sigma_{p,p-1} - \sigma_{p,p-1,0})^T$. *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 $|j-k| > \kappa$, for example, the banding structure. With little modification of $\hat{T}_n$, we can construct a test statistic on the superdiagonals $|j-k| > \kappa$. 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*

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

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

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

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 for example, 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, $\operatorname{tr}(\Sigma^4)\asymp \operatorname{tr}^2(\Sigma^2)$.

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 $\operatorname{tr}(\Sigma^4)\asymp p^{4-4\theta_1}$ and $\operatorname{tr}(\Sigma^2)\asymp p^{2-2\theta_1}$, 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 $B\subset\{1,2,\dots,n\}$, $B^c=\{1,\dots,n\}\backslash B$, and $|B|=|B^c|=m=n/2$. Define respectively,

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

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

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

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

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

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 $\{1, 2, \dots, n\}$. Because $\binom{n}{m}$ can be too large, $\tilde{F}(t)$ may be difficult to compute. Instead, a stochastic approximation may be employed. Let $B_1, \dots, B_L$ be i.i.d. uniformly sampled from the class $\mathcal{B} := \{B : B \subset \{1, \dots, n\}, |B| = m\}$. 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)),

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

where $\mathsf{P}^*(\cdot) = \mathsf{P}(\cdot | \boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$ is the conditional probability given the original data $\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$. 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 subsampling 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 $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$ 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 $l = 1, 2, \dots, n$. Then $\boldsymbol{Z}_2, \boldsymbol{Z}_3, \dots, \boldsymbol{Z}_n$ 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 and Wallis (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.

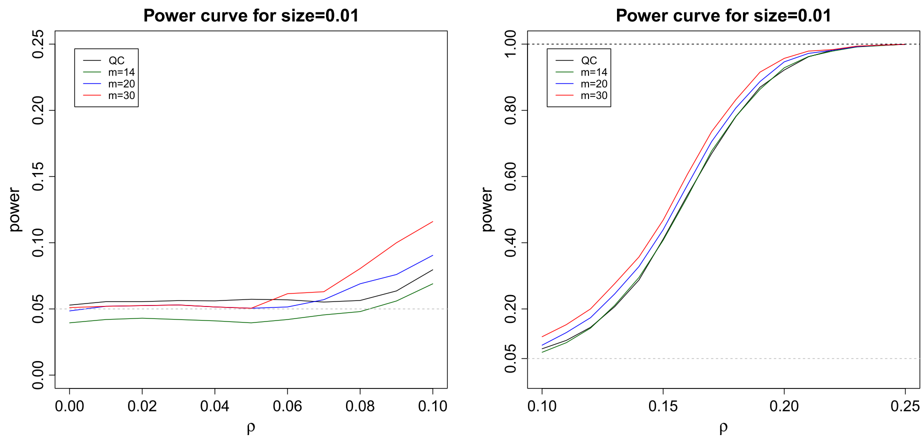

FIG. 1. *Power curve of the test given in* Qiu and Chen (2012) (*abbr. QC*), *the subsampling procedures with resampling size* $m = 14, 20$ *and the half-sampling procedure with* $m = 30$ *at size* $= 0.05$. *The resampling sizes are* 2000.

The example below numerically illustrates the benefits of the half-sampling approach over the usual subsampling procedure with $m = o(n)$. Our half- sampling approach goes far beyond the theoretical results about subsampling 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.

EXAMPLE 2.2. Consider the following model:

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

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 subsampling 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 subsampling 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

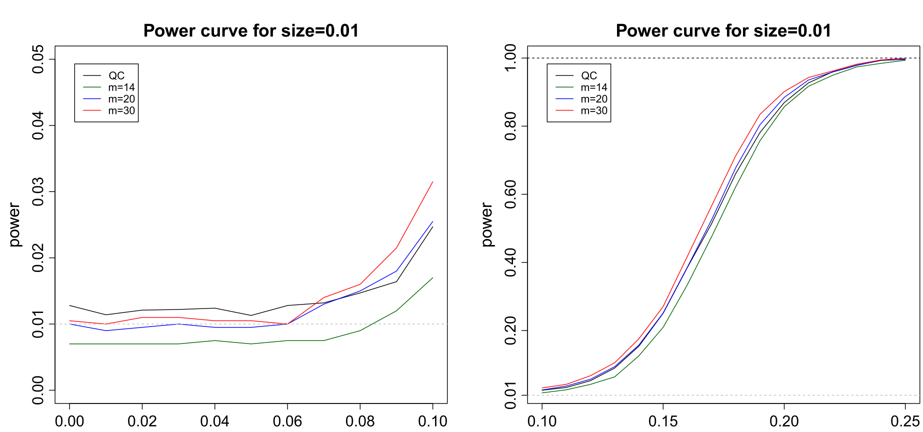

FIG. 2. *Power curve of the test given in* Qiu and Chen (2012) (*abbr. QC*), *the subsampling procedures with resampling size* $m = 14, 20$ *and the half-sampling procedure with* $m = 30$ *at size* $= 0.01$. *The resampling sizes are* 5000.

noted that the half-sampling procedure is the best one in both size accuracy and power. In addition, the subsampling with $m = 20$ also improves the power over the subsampling with $m = 14$ and the QC test.

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

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

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

Based on Theorem 2.2, at a given significance level $0 < \alpha < 1$, we propose the test $\Phi_{a,\alpha} = \mathbf{1}(n\hat{T}_n \geq y_{1-\alpha}^*)$. 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 Material. 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_1^2 distribution; see, for example, Xia, Cai and Cai (2018).

3. Testing parametric forms of covariance functions. In this section, we aim to test:

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

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

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

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 jth $(j = 1, \ldots, d)$ component of the d-dimensional vector $\boldsymbol{\theta}$. Let $V = (v_{m,q})_{1 \leq m,q \leq d}$ with

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

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

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

for $1 \leq m \leq d$. 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

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

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 $1 \le m, q \le d$,

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

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

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

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

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

Similar to $\hat{T}_n = \mathcal{T}_{S_1}$ 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; cf. 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),

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

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*, that is,

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

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

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

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 $(\operatorname{tr}((I - \Upsilon)\Gamma) - \operatorname{tr}(\Gamma))/|\Gamma - \Upsilon\Gamma|_F = -\operatorname{tr}(\Upsilon\Gamma)/|\Gamma - \Upsilon\Gamma|_F$, 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, *that is*, 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 $\Upsilon = (p(p-1))^{-1}\mathbf{1}_{p(p-1)}\mathbf{1}_{p(p-1)}^T$, $\operatorname{tr}(\Upsilon\Gamma) = 2\theta^2(p-2)(p-3) + 4(\theta^2+\theta)(p-2) + 2(\theta^2+1)$ and $\operatorname{tr}(\Gamma - \Upsilon\Gamma)^2 \asymp 4(\theta-\theta^2)^2 p(p-1)(p-2)$. 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 $\operatorname{tr}(\Upsilon\Gamma) \asymp 1$, $\operatorname{tr}(\Upsilon\Gamma)^2 \asymp 1$ and $\operatorname{tr}(\Gamma - \Upsilon\Gamma)^2 \asymp 1$. 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, $\operatorname{tr}(\Upsilon\Gamma) \asymp p^{2-2\theta_1}$, $\operatorname{tr}(\Upsilon\Gamma)^2 \asymp p^{4-4\theta_1}$, $\operatorname{tr}(\Gamma^2) \asymp p^{4-4\theta_1}$ and $\operatorname{tr}(\Gamma - \Upsilon\Gamma)^2 \asymp p^{4-4\theta_1}$. Then $\kappa_0 \asymp 1$. On the other hand, if $\theta_1 > 1/2$, then $\operatorname{tr}(\Upsilon\Gamma) \asymp p^{3-4\theta_1}\log^2(p) + 1$, $\operatorname{tr}(\Upsilon\Gamma)^2 \asymp p^{6-8\theta_1}\log^4(p) + 1$, $\operatorname{tr}(\Gamma^2) \asymp p^2$ and $\operatorname{tr}(\Gamma - \Upsilon\Gamma)^2 \asymp p^2$. 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),

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

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 $\Phi_{a,\alpha,\hat{\theta}}$ 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

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

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

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

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

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

The following theorem states the asymptotic properties of $\hat{G}_n$. Let $|\mathcal{S}|$ be the cardinality of $\mathcal{S}$; let $\lambda_1 \geq \cdots \geq \lambda_{p^2} \geq 0$ 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 $\operatorname{tr}(\Lambda\Gamma) = f_1$ and $|\Lambda\Gamma|_F = f_2$.

THEOREM 4.1. *Consider the linear process model $X_{ij} = \sum_{l=1}^{N} b_{j,l}\xi_{il}$, $1 \leq j \leq p$, where ξ_{il} are i.i.d. and satisfy Assumption 4.1. Suppose that Assumption 4.2 holds and $\tau_\varrho < \infty$ with $0 < \delta \leq 1$, $\varrho \geq 0$. (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},*

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

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, that is, *for any $\varepsilon > 0$,*

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

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 $\mathsf{E}|\boldsymbol{W}_1^T \Lambda \boldsymbol{W}_1|^2 \asymp p^2|\mathcal{S}|^2$, $f_1 \asymp p|\mathcal{S}|$ 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 $B_1, \ldots, B_L$ be i.i.d. uniformly sampled from the class $\mathcal{B} := \{B : B \subset \{1, \ldots, n\}, |B| = m\}$, 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 $F_G(t) := \mathsf{P}(n\hat{G}_n \leq t)$ by

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

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 $\{1, \ldots, n\}$. 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 Supplementary 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:

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

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

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

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

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 nonnormality and heavy tails.

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 $\rho_\Sigma \nrightarrow 0$, 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.

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

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)

		Proposed Test $\Phi_{a,\alpha}$			Qiu-Chen			CCK		
p	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

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 10,000 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:

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

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.

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 $Z_i = (Z_{i1}, \dots, Z_{ip})'$ are i.i.d. We consider the following covariance structure $\Sigma_0(\theta)$,

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

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

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

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*

		Normal				Gamma		Student t	
		$\Phi_{a,\alpha,\theta}$		ZLST		$\Phi_{a,\alpha,\theta}$		$\Phi_{a,\alpha,\theta}$	
p	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

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) standardized 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

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*

		Normal				Gamma		Student t	
		$\Phi_{a,\alpha,\theta}$		ZLST		$\Phi_{a,\alpha,\theta}$		$\Phi_{a,\alpha,\theta}$	
p	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

		Normal				Gamma		Student t	
		$\Phi_{a,\alpha,\theta}$		ZLST		$\Phi_{a,\alpha,\theta}$		$\Phi_{a,\alpha,\theta}$	
p	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

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 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 preassigned. 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}, $j = 1, \dots, p$, are chosen from Unif$(0.5, 2.5)$.

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

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

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	n:	Proposed Test $\Phi_{b,\alpha}$ 20	50	100	CCK 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

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 reports the empirical sizes of the proposed test $\Phi_{b,\alpha}$ (cf. 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 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.

Acknowledgment. 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.

SUPPLEMENTARY MATERIAL

Supplement to "Test for high dimensional covariance matrices" (DOI: 10.1214/20-AOS1943SUPP; .pdf). Supplementary information.

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	n:	Proposed Test $\Phi_{b,\alpha}$ 20	50	100	CCK 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

REFERENCES

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662

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. MR2572444 https://doi.org/10.1214/09-AOS694

BENDAT, J. S. (1990). *Nonlinear System Analysis and Identification from Random Data. A Wiley-Interscience Publication*. Wiley, New York. MR1045198

BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008 https://doi.org/10.1214/08-AOS600

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. MR2850210 https://doi.org/10.1214/11-AOS879

CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 https://doi.org/10.3150/12-BEJ455

CAI, T., MA, Z. and WU, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815. MR3334281 https://doi.org/10.1007/s00440-014-0562-z

CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885 https://doi.org/10.1214/09-AOS752

CASTI, J. L. (1985). *Nonlinear System Theory*. *Mathematics in Science and Engineering* **175**. Academic Press, Orlando, FL. MR0777143

CHATTERJEE, S. (2008). A new method of normal approximation. *Ann. Probab.* **36** 1584–1610. MR2435859 https://doi.org/10.1214/07-AOP370

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. MR2604697 https://doi.org/10.1214/09-AOS716

CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863 https://doi.org/10.1198/jasa.2010.tm09560

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. MR3468501 https://doi.org/10.1515/amcs-2015-0062

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. MR3161448 https://doi.org/10.1214/13-AOS1161

DE LA FUENTE, A. (2010). From 'differential expression' to 'differential networking'—identification of dysfunctional regulatory networks in diseases. *Trends Genet.* **26** 326–333.

FAN, J., FAN, Y. and LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics* **147** 186–197. MR2472991 https://doi.org/10.1016/j.jeconom.2008.09.017

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. With 33 discussions by 57 authors and a reply by Fan, Liao and Mincheva. MR3091653 https://doi.org/10.1111/rssb.12016

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. MR2301751 https://doi.org/10.1016/j.jmva.2006.08.003

GNEITING, T., KLEIBER, W. and SCHLATHER, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.* **105** 1167–1177. MR2752612 https://doi.org/10.1198/jasa.2010.tm09420

HAN, Y. and WU, W. B. (2020). Supplement to "Test for high dimensional covariance matrices." https://doi.org/10.1214/20-AOS1943SUPP.

HE, J. and CHEN, S. X. (2016). Testing super-diagonal structure in high dimensional covariance matrices. *J. Econometrics* **194** 283–297. MR3536981 https://doi.org/10.1016/j.jeconom.2016.05.007

HE, J. and CHEN, S. X. (2018). High-dimensional two-sample covariance matrix testing via super-diagonals. *Statist. Sinica* **28** 2671–2696. MR3839879

HEDAYAT, A. and WALLIS, W. D. (1978). Hadamard matrices and their applications. *Ann. Statist.* **6** 1184–1238. MR0523759

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. MR3229522 https://doi.org/10.1016/j.jmva.2014.04.020

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 Comput. Biol.* **5** e1000382. https://doi.org/10.1371/journal.pcbi.1000382

IDEKER, T. and KROGAN, N. J. (2012). Differential network biology. *Mol. Syst. Biol.* **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 Syst. Biol.* **5 Suppl 3** S12. https://doi.org/10.1186/1752-0509-5-S3-S12

JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906 https://doi.org/10.1214/105051604000000143

JOHN, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127. MR0275568 https://doi.org/10.1093/biomet/58.1.123

JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59** 169–173. MR0312619 https://doi.org/10.1093/biomet/59.1.169

KALOUPTSIDIS, N. and KOUKOULAS, P. (2005). Blind identification of Volterra–Hammerstein systems. *IEEE Trans. Signal Process.* **53** 2777–2787. MR2170655 https://doi.org/10.1109/TSP.2005.850357

KELLEY, R. and IDEKER, T. (2005). Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.* **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. MR1926169 https://doi.org/10.1214/aos/1031689018

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. MR2985938 https://doi.org/10.1214/12-AOS993

MASSART, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. MR1062069

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling. Springer Series in Statistics*. Springer, New York. MR1707286 https://doi.org/10.1007/978-1-4612-1554-7

POURAHMADI, M. (2013). *High-Dimensional Covariance Estimation. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3235948 https://doi.org/10.1002/9781118573617

PRIESTLEY, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*. Academic Press [Harcourt Brace Jovanovich, Publishers], London. MR0991969

QIU, Y. and CHEN, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *Ann*. *Statist*. **40** 1285–1314. MR3015026 https://doi.org/10.1214/12-AOS1002

RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

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. MR3346695 https://doi.org/10.1214/14-AOS1286

RUGH, W. J. (1981). *Nonlinear System Theory*: *The Volterra–Wiener Approach*. *Johns Hopkins Series in Information Sciences and Systems*. Johns Hopkins Univ. Press, Baltimore, MD. MR0713994

SCHETZEN, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley, New York–Chichester–Brisbane. A Wiley-Interscience Publication. MR0562917

STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. *Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

WIESEL, A., BIBI, O. and GLOBERSON, A. (2013). Time varying autoregressive moving average models for covariance estimation. *IEEE Trans*. *Signal Process*. **61** 2791–2801. MR3064093 https://doi.org/10.1109/TSP.2013.2256900

WU, C.-F. J. (1990). On the asymptotic properties of the jackknife histogram. *Ann*. *Statist*. **18** 1438–1452. MR1062718 https://doi.org/10.1214/aos/1176347759

WU, W. B., LOU, Z. and HAN, Y. (2018). Hypothesis testing for high-dimensional data. In *Handbook of Big Data Analytics*. *Springer Handb*. *Comput*. *Stat*. 203–224. Springer, Cham. MR3932004

WU, W. B. and POURAHMADI, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist*. *Sinica* **19** 1755–1768. MR2589209

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. MR3371002 https://doi.org/10.1093/biomet/asu074

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. MR3803468 https://doi.org/10.1080/01621459.2016.1251930

XIAO, H. and WU, W. B. (2013). Asymptotic theory for maximum deviations of sample covariance matrix estimates. *Stochastic Process*. *Appl*. **123** 2899–2920. MR3054550 https://doi.org/10.1016/j.spa.2013.03.012

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. MR1492991 https://doi.org/10.1007/978-1-4615-6313-6

ZHANG, R., PENG, L. and WANG, R. (2013). Tests for covariance matrix with fixed or divergent dimension. *Ann*. *Statist*. **41** 2075–2096. MR3127858 https://doi.org/10.1214/13-AOS1136

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. MR3316190 https://doi.org/10.1214/14-AOS1292

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. MR3662452 https://doi.org/10.1214/16-AOS1481