Statistica Sinica **33** (2023), 1673-1696
doi:https://doi.org/10.5705/ss.202020.0514

ADAPTIVE TESTS FOR BANDEDNESS OF HIGH-DIMENSIONAL COVARIANCE MATRICES

Xiaoyi Wang, Gongjun Xu and Shurong Zheng

Beijing Normal University, University of Michigan and Northeast Normal University

Abstract: Estimations of high-dimensional banded covariance matrices are widely used in multivariate statistical analysis. To ensure the validity of such estimations, we test the hypothesis that the covariance matrix is banded with a certain bandwidth under a high-dimensional framework. Though several testing methods have been proposed in the literature, these tests are only powerful for some alternatives with certain sparsity levels, but not others. Here, we propose two adaptive tests for the bandedness of a high-dimensional covariance matrix that is powerful for alternatives with various sparsity levels. The proposed methods can also be used to test the banded structure of the covariance matrices of the error vectors in high-dimensional factor models. Based on these statistics, we introduce a consistent bandwidth estimator for a banded high-dimensional covariance matrix. We use simulation studies and an application to a prostate cancer data set to evaluate the effectiveness of the proposed adaptive tests and bandwidth estimator.

Key words and phrases: Asymptotic normality, banded covariance matrix, high-dimensional hypothesis testing, sparsity level, U-statistics.

1. Introduction

Statistical tests for covariance matrices play an important role in multivariate and high-dimensional statistical analysis, for example, in principle component analysis, multivariate regression analysis, and factor analysis (see Anderson (2003); Bai and Yin (1993); Johnstone (2001); Cai and Liu (2011); Fan, Liao and Yao (2015)).

Tests of high-dimensional covariance matrices have been studied from various aspects, for example, testing $H_{01} : \mathbf{\Sigma} = \mathbf{\Sigma}_0$, where $\mathbf{\Sigma}$ is a population covariance matrix, and $\mathbf{\Sigma}_0$ is a given positive-definite matrix. For instance, Ledoit and Wolf (2002) proposed a robust statistic for testing H_{01} based on the Frobenius norm under a Gaussian assumption. Without the Gaussian assumption, Bai et al. (2009) developed a corrected likelihood ratio test (LRT) when the dimen-

Corresponding author: Shurong Zheng, School of Mathematics & Statistics and KLAS, Northeast Normal University, Changchun, Jilin 130024, China. E-mail: zhengsr@nenu.edu.cn.

sion p of the data is smaller than the sample size n. Jiang, Jiang and Yang (2012) studied the asymptotic distribution of the corrected LRT for normal random vectors when $p/n \to y \in (0, 1]$. Later, Wang et al. (2013) redefined the above statistics, and introduced two tests that accommodate data with an unknown mean and a nonGaussian distribution. Moreover, Chen, Zhang and Zhong (2010) and Cai and Ma (2013) constructed sum-of-squared-type tests using U-statistics as $n, p \to \infty$. Another aspect is to test $H_{02} : \mathbf{\Sigma} = c\mathbf{\Sigma}_0$, where c is an unknown positive number. For sphericity testing, Ledoit and Wolf (2002) and Chen, Zhang and Zhong (2010) proposed sum-of-squared-type statistics, where the former substituted the sample covariance matrix $\mathbf{S}_n$ for $\mathbf{\Sigma}$; and the latter used unbiased U-statistics. Furthermore, Wang and Yao (2013) developed a corrected LRT ($p < n$) and John's test. Jiang and Yang (2013) extended the corrected LRT to the case of $p/n \to y \in (0, 1]$, and Li and Yao (2016) proposed a quasi-LRT allowing $p/n \to \infty$. In addition, researchers have tested general linear structures of covariance matrices. Zheng et al. (2019) studied the problem of testing $H_{03} : \mathbf{\Sigma} = \theta_1\mathbf{A}_1 + \cdots + \theta_K\mathbf{A}_K$, where $\theta_1, \ldots, \theta_K$ are unknown parameters and $\mathbf{A}_1, \ldots, \mathbf{A}_K$ are known basis matrices. Furthermore, Zhong et al. (2017) introduced an adjusted goodness-of-fit test that examines a broad range of covariance structures to assess the adequacy of specified covariance structures.

Here, we test the banded structure of covariance matrices, which has numerous applications in areas such as biological science, climate, econometrics, and finance (see Andrews (1991); Ligeralde and Brown (1995)). For instance, in high-dimensional data analyses, a popular covariance matrix estimation method is to band or taper the sample covariance matrix (e.g., Bickel and Levina (2008)). Although the large-sample consistency of the corresponding estimators has been established for covariance matrices in the "bandable" class, it remains questionable whether or not the underlying covariance matrix belongs to the "bandable" class. The proposed hypothesis test for the banded structure of a covariance matrix may provide a practical statistical guideline for this issue.

Several methods have been proposed for testing the bandedness of a high-dimensional covariance matrix. In particular, Qiu and Chen (2012) developed a test using a linear combination of U-statistics by collecting the sum-of-squares of all covariance differences between the null and alternative hypotheses. This test is powerful against dense alternatives, because there are many nonzero components in the aforementioned covariance differences. However, it is not powerful when the alternative is sparse. To address this problem, Cai and Jiang (2011) proposed a maximum-type statistic by capturing the maximum componentwise covariance difference for multivariate normal random vectors. Shao and Zhou

(2014) restudied the above statistic, and suggested using a chi-squared distribution instead of the type-I extreme distribution to improve the convergence rate of the maximum-type statistic. Furthermore, Xiao and Wu (2013) relaxed the normality assumption for the normalized maximum componentwise sample covariance difference. The maximum-type test is powerful when the alternative is sparse, but is less powerful for dense alternatives. In practice, however, it is often unclear how dense or sparse the alternative hypothesis is. In addition, neither test is powerful when the alternative hypothesis is denser or less sparse, as shown in Section 3.1.

Motivated by this, we propose two adaptive tests based on a series of unbiased U-statistics for the banded structure of high-dimensional covariance matrices, following the idea of the adaptive test in Xu et al. (2016) and He et al. (2021). Our contributions are as follows:

(i) We derive the joint asymptotic distribution of the series of U-statistics under the null hypothesis. Furthermore, we show that the U-statistics are asymptotically independent and jointly normally distributed under certain regularity conditions.

(ii) We establish the asymptotic distribution of the series of finite-order U-statistics under a local alternative hypothesis. Furthermore, we compare the power performance of these single U-statistic-based tests and show their consistency.

(iii) We propose two adaptive tests that combine the p-values of the U-statistics. The consistency of the adaptive tests is also guaranteed. These adaptive tests select the test with the most significant result, and yield high power under a range of alternative hypothesis scenarios.

(iv) We provide an adaptive estimator for the bandwidth of a high-dimensional banded covariance matrix and establish its consistency.

The rest of the paper is organized as follows. In Section 2, we first introduce a series of U-statistics, and then derive their joint asymptotic distributions under the null and local alternative hypotheses. Furthermore, we propose two adaptive tests for testing the banded structure of high-dimensional covariance matrices. We also present a bandwidth estimator for the banded covariance matrix and show its consistency. In Section 3, we present our simulation studies, and in Section 4, we demonstrate the proposed tests by analyzing a prostate cancer data set. Section 5 concludes this paper. The main technical proofs and additional simulation results are provided in the Supplementary Material.

2. The Proposed Test Methods

Let $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^{\mathrm{T}}$, for $i = 1, \ldots, n$, be independent and identically distributed (i.i.d.) samples from a p-dimensional population $\mathbf{x} = (x_1, \ldots, x_p)^{\mathrm{T}}$ with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^{\mathrm{T}}$ and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{j_1 j_2})_{p\times p}$. The population covariance matrix $\boldsymbol{\Sigma} = (\sigma_{j_1 j_2})_{p\times p}$ is said to be banded if there exists an integer $k \in \{0, \ldots, p-2\}$ such that $\sigma_{j_1 j_2} = 0$, for $|j_1 - j_2| > k$. The smallest k such that $\boldsymbol{\Sigma}$ is banded is called the bandwidth of $\boldsymbol{\Sigma}$. Let $\boldsymbol{B}_k(\boldsymbol{\Sigma}) = (\sigma_{j_1 j_2}\mathbf{1}_{\{|j_1-j_2|\le k\}})_{p\times p}$ be the banded version of $\boldsymbol{\Sigma}$ with bandwidth k, where $\mathbf{1}_{\{\cdot\}}$ is an indicator function. When $k = 0$, $\boldsymbol{B}_0(\boldsymbol{\Sigma})$ is the diagonal version of $\boldsymbol{\Sigma}$. Here, we test

$$H_{k,0} : \boldsymbol{\Sigma} = \boldsymbol{B}_k(\boldsymbol{\Sigma}) \quad \text{vs.} \quad H_{k,1} : \boldsymbol{\Sigma} \neq \boldsymbol{B}_k(\boldsymbol{\Sigma}), \tag{2.1}$$

for a certain positive integer k. We further rewrite the hypothesis test in (2.1) as

$$H_{k,0} : \mathcal{E} = \mathbf{0} \quad \text{vs.} \quad H_{k,1} : \mathcal{E} \neq \mathbf{0},$$

where $\mathcal{E} = \{\sigma_{j_1 j_2} : k < |j_1 - j_2| < p\}$.

2.1. A series of U-statistics

Motivated by He et al. (2021), we consider a series of measurements of $\mathcal{E}$, defined by $||\mathcal{E}||_a = [\sum_{k<|j_1-j_2|<p}(\sigma_{j_1 j_2})^a]^{1/a}$, and construct test statistics that are powerful against $||\mathcal{E}||_a$, for a finite positive integer a. Because $\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2} - x_{i_1,j_1}x_{i_2,j_2}) = \sigma_{j_1 j_2}$ for $1 \le i_1 \neq i_2 \le n$, we propose the U-statistic

$$\mathcal{U}(a) = \sum_{k<|j_1-j_2|<p} (P^n_{2a})^{-1} \sum_{1\le i_1\neq\cdots\neq i_{2a}\le n} \prod_{l=1}^{a} (x_{i_{2l-1},j_1}x_{i_{2l-1},j_2} - x_{i_{2l-1},j_1}x_{i_{2l},j_2})$$

as an unbiased estimator of $||\mathcal{E}||_a^a$, where $P^n_{2a} = n!/(n-2a)!$ denotes the number of $2a$-permutations of n. A straightforward calculation shows that

$$\begin{aligned}\mathcal{U}(a) = &\sum_{k<|j_1-j_2|<p} \sum_{c=0}^{a} \binom{a}{c} (-1)^c (P^n_{a+c})^{-1} \sum_{1\le i_1\neq\cdots\neq i_{a+c}\le n} \\ &\prod_{l=1}^{a-c}(x_{i_l,j_1}x_{i_l,j_2}) \prod_{s=a-c+1}^{a} x_{i_s,j_1} \prod_{t=a+1}^{a+c} x_{i_t,j_2}.\end{aligned} \tag{2.2}$$

The form of $\mathcal{U}(a)$ in (2.2) plays an essential role in deriving the theoretical properties of our proposed statistics. Specifically, to obtain the expression in (2.2), we define $\varphi_{j_1 j_2} = \mathrm{E}(x_{i,j_1}x_{i,j_2})$ and $\sigma_{j_1 j_2} = \mathrm{E}[(x_{i,j_1} - \mu_{j_1})(x_{i,j_2} - \mu_{j_2})] = \varphi_{j_1 j_2} - \mu_{j_1}\mu_{j_2}$. For any finite positive integer a, we have

$$\sum_{k<|j_1-j_2|<p} \sigma_{j_1j_2}^a = \sum_{k<|j_1-j_2|<p} (\varphi_{j_1j_2} - \mu_{j_1}\mu_{j_2})^a$$
$$= \sum_{k<|j_1-j_2|<p} \sum_{c=0}^{a} \binom{a}{c} (-1)^c \varphi_{j_1j_2}^{a-c} \mu_{j_1}^c \mu_{j_2}^c. \tag{2.3}$$

Because $x_{i,j}$ and $x_{i,j_1}x_{i,j_2}$ are unbiased estimators of μ_j and $\varphi_{j_1j_2}$, respectively, for $1 \leq i_1 \neq \cdots \neq i_{a+c} \leq n$, it follows that $\mathrm{E}\big(\prod_{l=1}^{a-c} x_{i_l,j_1}x_{i_l,j_2} \prod_{s=a-c+1}^{a} x_{i_s,j_1} \prod_{t=a+1}^{a+c} x_{i_t,j_2}\big) = \varphi_{j_1j_2}^{a-c}\mu_{j_1}^c\mu_{j_2}^c$. Thus, we obtain the expression (2.2).

Remark 1. If we consider only $c = 0$ in (2.2), we have

$$\tilde{\mathcal{U}}(a) = (P_a^n)^{-1} \sum_{k<|j_1-j_2|<p} \sum_{1\leq i_1\neq\cdots\neq i_a\leq n} \prod_{l=1}^{a} (x_{i_l,j_1}x_{i_l,j_2}). \tag{2.4}$$

In Section 2.2, we show that this is a leading term of (2.2), under certain regularity conditions, and use it for our theoretical analysis presented in the Supplementary Material.

2.2. Asymptotic properties of the U-statistics under the null hypothesis

Before deriving the theoretical properties of the U-statistics under the null hypothesis, we first introduce the following notation: $u_{n,p} = o(v_{n,p})$ if $\limsup_{n,p\to\infty} |u_{n,p}/v_{n,p}| = 0$, and $u_{n,p} = \Theta(v_{n,p})$ if $0 < \liminf_{n,p\to\infty} |u_{n,p}/v_{n,p}| \leq \limsup_{n,p\to\infty} |u_{n,p}/v_{n,p}| < \infty$ and

$$\Pi_{j_1,\ldots,j_t} = \mathrm{E}[(x_{1,j_1} - \mu_{j_1}) \cdots (x_{1,j_t} - \mu_{j_t})]. \tag{2.5}$$

We assume the following regularity conditions in our analysis:

Condition 1. $\lim_{p\to\infty} \max_{1\leq j\leq p} \mathrm{E}[(x_j - \mu_j)^8] < \infty$ and $\lim_{p\to\infty} \min_{1\leq j\leq p} \mathrm{E}[(x_j - \mu_j)^2] > 0$.

Condition 2. A sequence of random variables $\mathbf{z} = \{z_j, j \geq 1\}$ is said to be α-mixing if $\lim_{s\to\infty} \alpha_{\mathbf{z}}(s) = 0$, where $\alpha_{\mathbf{z}}(s) = \sup_{t\geq 1}\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{F}_1^t, B \in \mathcal{F}_{t+s}^\infty\}$, with $\mathcal{F}_a^b$ being the σ-algebra generated by $\{z_a, z_{a+1}, \ldots, z_b\}$. Under H_0, we assume $\mathbf{x}$ is α-mixing with $\alpha_{\mathbf{x}}(s) \leq M\delta^s$, where $\delta \in (0,1)$ and M is some positive constant.

The regularity conditions are similar to those of Theorem 2.1 in He et al. (2021), who studied the problem of testing the diagonality of a covariance matrix, a special case of (2.1) with $k = 0$. Specifically, Condition 1 requires that the eighth marginal moments of $\mathbf{x}$ are uniformly bounded from above, and that the second marginal moments are uniformly bounded from below. Condition 2 prescribes weak dependence between the components of $\mathbf{x}$ in α-mixing type, which is satisfied when $\mathbf{x}$ is an m-dependent random vector or a Gaussian distributed random vector with a banded covariance matrix. In addition, there are several strong mixing conditions, such as ϕ-mixing, ψ-mixing, ρ-mixing, and β-mixing. In the classical theory, these five strong mixing conditions have emerged as the most prominent, with the α-mixing condition being the weakest of them (e.g., Bradley (2005)). The α-mixing condition is also used by, among others, Xu et al. (2016) and Chen, Li and Zhong (2019). In our work, the α-mixing condition to these mixture moments $\mathrm{E}(\prod_{t=1}^{s} x_{j_t})$, for $2 \le s \le 8$, ensures the asymptotic independence of different finite-order U-statistics. In addition, Bai and Saranadasa (1996) assumed the independent component structure $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{z}$ to describe the weak dependence between the components of $\mathbf{x}$. The random vector $\mathbf{x}$ is α-mixing when $\boldsymbol{\Gamma} = (\gamma_{j_1 j_2})_{p\times p}$ is an upper-triangular matrix with $\gamma_{j_1 j_2} = 0$, for $j_2 - j_1 > k$.

Theorem 1. *Under Conditions* 1, 2, *and* $H_{k,0}$, *for any finite positive integers* $a_1, \ldots, a_m$, *we have*

$$\left(\frac{\mathcal{U}(a_1)}{\sigma(a_1)}, \ldots, \frac{\mathcal{U}(a_m)}{\sigma(a_m)}\right)^{\mathrm{T}} \xrightarrow{D} \mathcal{N}(0, I_m), \quad n, p \to \infty, \tag{2.6}$$

where

$$\sigma^2(a) = \mathrm{Var}[\mathcal{U}(a)] = (P_a^n)^{-1} a! \sum_{\substack{k<|j_1-j_2|<p,\\ k<|j_3-j_4|<p}} (\Pi_{j_1,j_2,j_3,j_4})^a + o(n^{-a}p^2), \tag{2.7}$$

with $\prod_{j_1,j_2,j_3,j_4}$ *defined in* (2.5).

Because $\sigma^2(a)$ is unknown, we obtain the following theorem, where $\sigma^2(a)$ is replaced with the estimator $\hat{\sigma}^2(a)$ provided in (2.9). To ensure the consistency of $\hat{\sigma}^2(a)$, we require Condition 3.

Condition 3. For a finite positive integer a, $\lim_{p\to\infty} \max_{1\le j\le p} \mathrm{E}[(x_j - \mu_j)^{8a}] < \infty$.

Theorem 2. *Under Conditions* 1, 2, *and* $H_{k,0}$, *for any positive integers* $a_1, \ldots, a_m$ *satisfying Condition* 3, *we have*

$$\left(\frac{\mathcal{U}(a_1)}{\hat{\sigma}(a_1)}, \ldots, \frac{\mathcal{U}(a_m)}{\hat{\sigma}(a_m)}\right)^{\mathrm{T}} \xrightarrow{D} \mathcal{N}(0, I_m), \quad n, p \to \infty, \tag{2.8}$$

where $\hat{\sigma}^2(a)/\sigma^2(a) \xrightarrow{P} 1$ *and*

$$\hat{\sigma}^2(a) = 2(P_a^n)^{-2}a! \sum_{\substack{k<|j_1-j_2|<p,\\ k<|j_3-j_4|<p,\\ |j_1-j_3|\leq k,|j_2-j_4|\leq k}} \sum_{1\leq i_1\neq\cdots\neq i_a\leq n} \prod_{l=1}^{a}[(x_{i_l,j_1}-\bar{x}_{j_1})\cdots(x_{i_l,j_4}-\bar{x}_{j_4})]. \tag{2.9}$$

Theorem 2 shows that $\mathcal{U}(a_1)/\hat{\sigma}(a_1), \ldots, \mathcal{U}(a_m)/\hat{\sigma}(a_m)$ are *asymptotically independent and normally distributed.* The theoretical results in Theorems 1 and 2 extend those in He et al. (2021) from testing the diagonality of a covariance matrix to testing a general banded structure, which is often of practical interest in high-dimensional covariance matrix estimation. The general banded structure makes the analysis more technically involved; the details are presented in the Supplementary Material.

Remark 2. For an extreme case, as an even number $a \to \infty$, we have

$$||\mathcal{E}||_a = \left[\sum_{k<|j_1-j_2|<p} \sigma_{j_1j_2}^a\right]^{1/a} \longrightarrow ||\mathcal{E}||_\infty = \max_{k<|j_1-j_2|<p} |\sigma_{j_1j_2}|.$$

Thus, the statistic $\mathcal{U}(a)$ and the maximum-type statistic perform similarly when the even order a is large enough, as shown in Xu et al. (2016) and He et al. (2021). He et al. (2021) provide the asymptotic independence between the finite-order U-statistics and the infinite-order U-statistic (maximum-type statistic) when the components of the random vector $\mathbf{x}$ are uncorrelated. We expect a similar result under certain regular conditions in our setting. However, it is challenging to establish the asymptotic joint distribution of the maximum-type statistic and the finite-order U-statistics, because the banded covariance structure is much more complicated than it is in the i.i.d. case, owing to the dependence, as pointed out in Cai and Jiang (2011). We leave this for future work.

2.3. Power analysis

In this section, we investigate the asymptotic distributions of the series of U-statistics under the local alternative hypothesis $H_{k,A} : \mathbf{\Sigma} = \mathbf{\Sigma}_A$, which is described in Condition 4. For a given bandwidth k, we denote the set of locations of the signals by $J_A = \{(j_1, j_2) : \sigma_{j_1j_2} \neq 0, k < |j_1 - j_2| < p, j_1, j_2 = 1, \ldots, p\}$, and the cardinality of J_A by $|J_A|$, which represents the sparsity level of $\mathbf{\Sigma}_A$. The sparsity level of the alternative hypothesis decreases as $|J_A|$ increases. We introduce two conditions for the asymptotic distribution under the local alternative hypothesis.

Condition 4. Assume $|J_A| = o(p^2)$ and, for any $(j_1, j_2) \in J_A$, $|\sigma_{j_1 j_2}| = \Theta(\rho)$, where $\rho = \sum_{(j_1,j_2)\in J_A} |\sigma_{j_1 j_2}|/|J_A|$.

Condition 5. For $t \le 8$, we assume there exists a constant κ such that $\Pi_{j_1,\dots,j_t} = \kappa \mathrm{E}(\prod_{k=1}^{t} z_{j_k})$, where $1 \le j_1, \dots, j_t \le p$ and $(z_1, \dots, z_p)^{\mathrm{T}} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_A)$.

Theorem 3. *Under Conditions* 1*,* 4*, and* 5*, for any positive integers* $a_1, \dots, a_m$*, if* $\rho = O(|J_A|^{-1/a_t} p^{1/a_t} n^{-1/2})$*, for* $t = 1, \dots, m$*, we have*

$$\left(\frac{\mathcal{U}(a_1) - \mathrm{E}_A[\mathcal{U}(a_1)]}{\sigma_A(a_1)}, \dots, \frac{\mathcal{U}(a_m) - \mathrm{E}_A[\mathcal{U}(a_m)]}{\sigma_A(a_m)}\right)^{\mathrm{T}} \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{I}_m), \quad n, p \to \infty,$$

where $\mathrm{E}_A[\mathcal{U}(a)] = \sum_{(j_1,j_2)\in J_A} \sigma_{j_1 j_2}^a$ *and* $\sigma_A^2(a) \simeq 2(P_a^n)^{-1} a! \kappa^a \sum_{k<|j_1-j_2|<p, k<|j_3-j_4|<p, |j_1-j_3|\le k, |j_2-j_4|\le k} \sigma_{j_1 j_3}^a \sigma_{j_2 j_4}^a$*, with order* $\Theta(n^{-a} p^2)$*.*

Following Theorem 3, the power function of a single U-statistic $\mathcal{U}(a)$ is

$$\beta(a) = P\left(\frac{\mathcal{U}(a)}{\sigma(a)} > z_{1-\alpha} \middle| H_{k,A}\right) \to \Phi\left(-z_{1-\alpha} + \frac{\mathrm{E}_A[\mathcal{U}(a)]}{\sigma_A(a)}\right), \tag{2.10}$$

where $z_{1-\alpha}$ and $\Phi(\cdot)$ are the $(1-\alpha)$th quantile and the cumulative distribution function of the standard normal distribution, respectively. The signal-to-noise ratio $\mathrm{SNR}_a = \mathrm{E}_A[\mathcal{U}(a)]/\sigma_A(a)$ plays an important role in the power performance of the U-statistic $\mathcal{U}(a)$. For any finite order a, we define the corresponding average standardized signal as $\bar{\rho}_a = \sum_{(j_1,j_2)\in J_A} n^{a/2} \sigma_{j_1 j_2}^a / |J_A|$. The asymptotic power function $\beta(a) \to 1$ if $p^{-1}|J_A|\bar{\rho}_a \to \infty$, because $\mathrm{E}_A[\mathcal{U}(a)] = \sum_{(j_1,j_2)\in J_A} \sigma_{j_1 j_2}^a$ and $\sigma_A(a)$ is of order $n^{-a/2}p$. In other words, if $\bar{\rho}_a$ is of order higher than $p|J_A|^{-1}$, $\beta(a) \to 1$ as $n \to \infty$.

Another attractive work is to investigate the relationship between the order of the U-statistic with the highest asymptotic power and the sparsity level $|J_A|$. We give a criterion for comparing the power performance of two finite-order U-statistics, $\mathcal{U}(a_1)$ and $\mathcal{U}(a_2)$. We say $\mathcal{U}(a_1)$ is better than $\mathcal{U}(a_2)$ when they attain the same asymptotic power and satisfy $\rho_{a_1} < \rho_{a_2}$. In particular, we consider a special case in which the signal strength is fixed at the same level, $\sigma_{j_1 j_2} = \rho > 0$ for $(j_1, j_2) \in J_A$, and $\sigma_{j_1 j_3} = \sigma_{j_2 j_4} = \nu > 0$ for $|j_1 - j_3| \le k$ and $|j_2 - j_4| \le k$. In this case,

$$\mathrm{SNR}_a \simeq \frac{\sum_{(j_1,j_2)\in J_A} \sigma_{j_1 j_2}^a}{\left\{2(P_a^n)^{-1} a! \kappa^a \sum_{k<|j_1-j_2|<p, k<|j_3-j_4|<p, |j_1-j_3|\le k, |j_2-j_4|\le k} (\sigma_{j_1 j_3}\sigma_{j_2 j_4})^a\right\}^{1/2}}.$$

Hence, the power function $\beta(a) \to \Phi(-z_{1-\alpha} + |J_A|\rho^a/(\sqrt{2a!\kappa^a}\nu^a n^{-a/2} p'))$, where

$p' = (2k+1)\sqrt{(p-k-1)(p-k)}$. Thus,

$$\rho_a = \left(\frac{Mp'}{|J_A|}\right)^{1/a} (a!)^{1/2a} \kappa^{1/2} \nu n^{-1/2} \tag{2.11}$$

achieves the asymptotic power $\Phi(-z_{1-\alpha} + M/\sqrt{2})$ of $\mathcal{U}(a)$, where M is some constant. Proposition 1 establishes the relationship between the sparsity level and the order of the U-statistic.

Proposition 1. *For a given bandwidth k, under the special case described above, given $n, p, |J_A|$, and M, by considering* (2.11) *as a function of integer order a, we have*

(i) when $|J_A| \geq Mp'$, the minimum of ρ_a is achieved at $a = 1$;

(ii) when $|J_A| < Mp'$, the minimum of ρ_a is achieved at some a, which increases as $Mp'/|J_A|$ increases.

When $|J_A| \geq Mp'$, the alternative is very dense, and $\mathcal{U}(1)$ is the most powerful test. When $|J_A| < Mp'$, as $Mp'/|J_A|$ increases, the sparsity level of the alternative hypothesis increases, and the U-statistic with the larger order performs better. This result is consistent with the analysis in He et al. (2021), and we extend their result to the banded covariance matrix setting.

2.4. Two adaptive testing procedures

For the proposed family of U-statistics, $\mathcal{U}(a)$ is powerful against the alternative with large $||\mathcal{E}||_a^a = \sum_{k<|j_1-j_2|<p} \sigma_{j_1 j_2}^a$. The power performance of $\mathcal{U}(a)$ is determined by the sparsity and the strength of the signals. For a denser alternative, we prefer a test with a smaller order a. For example, $\mathcal{U}(1)$ is the most powerful test when the alternative is very dense, as shown in Section 3.1. In practice, it is often unclear which test to choose because the true alternative is usually unknown. Therefore, motivated by the work of Xu et al. (2016) and He et al. (2021), we develop adaptive tests by combining the information from U-statistics with different orders, which yield higher power against various alternatives.

We propose two adaptive tests: one based on the minimum combination method, and the other based on Fisher's method. Suppose we have a candidate set $\mathcal{I} = \{a_1, \ldots, a_m\}$, $|\mathcal{I}| = m$, where $|\mathcal{I}|$ denotes the cardinality of $\mathcal{I}$. Let p_a be the p-value of test $\mathcal{U}(a)$ as $p_a = 2(1 - \Phi(|\mathcal{U}(a)/\hat{\sigma}(a)|))$.

Minimum combination method: Reject H_0 if $p_{\text{adpUmin}} < \alpha$, where

$$p_{\text{adpUmin}} = 1 - (1 - T_{\text{adpUmin}})^{|\mathcal{I}|}, \quad T_{\text{adpUmin}} = \min_{a \in \mathcal{I}} p_a, \tag{2.12}$$

with nominal significance level α. The type-I error of the minimum combination method can be controlled using $P(p_{\text{adpUmin}} < \alpha) = P(T_{\text{adpUmin}} < p^*_\alpha) \to \alpha$, where $p^*_\alpha = 1 - (1-\alpha)^{1/|\mathcal{I}|}$ and we use the asymptotic independence of $\mathcal{U}(a_1)/\hat{\sigma}(a_1)$, $\ldots, \mathcal{U}(a_m)/\hat{\sigma}(a_m)$.

Fisher's method: We have $T_{\text{adpUf}} = -2\sum_{a\in\mathcal{I}} \log\, p_a \xrightarrow{D} \chi^2_{2|\mathcal{I}|}$, where $\chi^2_{2|\mathcal{I}|}$ follows a chi-squared distribution with degrees of freedom $2|\mathcal{I}|$. We reject H_0 if $p_{\text{adpUf}} < \alpha$, with

$$p_{\text{adpUf}} = 1 - \Psi(T_{\text{adpUf}}), \tag{2.13}$$

where $\Psi(\cdot)$ is the cumulative distribution function of $\chi^2_{2|\mathcal{I}|}$ with degrees of freedom $2|\mathcal{I}|$.

Remark 3. For the two adaptive statistics, we have

(i) $P(T_{\text{adpUmin}} = \min_{a\in\mathcal{I}} p_a < p^*_\alpha) \geq P(p_a < p^*_\alpha) \to \Phi\big(-z_{1-p^*_\alpha} + \mathrm{E}_A[\mathcal{U}(a)]/\sigma_A(a)\big)$,

(ii) $P(T_{\text{adpUf}} = -2\sum_{a\in\mathcal{I}} \log p_a > c_{1-\alpha}) \geq P(-2\log p_a > c_{1-\alpha}) \to \Phi\big(-z_{1-c^*_\alpha} + \mathrm{E}_A[\mathcal{U}(a)]/\sigma_A(a)\big)$, where $c^*_\alpha = e^{-(1/2)c_{1-\alpha}}$ and $c_{1-\alpha}$ is the $(1-\alpha)$th quantile of $\chi^2_{2|\mathcal{I}|}$.

The asymptotic power of the proposed adaptive tests converges to one as $n \to \infty$ if there is a U-statistic $\mathcal{U}(a)$ that satisfies the average standardized signal $\bar{\rho}_a$ of order higher than $p|J_A|^{-1}$, with $\bar{\rho}_a = \sum_{(j_1,j_2)\in J_A} n^{a/2}\sigma^a_{j_1j_2}/|J_A|$.

Remark 4. Our proposed adaptive tests are versatile in the sense that they can adapt to the unknown sign and sparsity level of the signal set $\mathcal{E}$ under the alternative hypotheses. Their performance depends on the selection of the order set $\mathcal{I}$. U-statistics with an odd order may lose power quickly, because the different signs of the elements in $\mathcal{E}$ lead to a cancellation of positive and negative $\sigma^a_{j_1j_2}$. In this case, we suggest using U-statistics with an even order to construct the adaptive tests. However, U-statistics with an odd order are still more suitable when the elements in $\mathcal{E}$ are all in the same direction. For example, $\mathcal{U}(1)$ is a representative of the burden tests based on genotype pooling or collapsing, as discussed in Morgenthaler and Thilly (2007), Li and Leal (2008), and Pan et al. (2014). In the absence of information about the directions of the signals, we suggest using both odd- and even-order U-statistics. Furthermore, the theoretical arguments on the power analysis and the results of our simulation studies indicate that the order of the best U-statistic increases as the sparsity level decreases. To address sparse alternative hypotheses, we select the biggest order to be six because the performance of $\mathcal{U}(6)$ is good enough compared with that of the maximum-type statistic, as shown in the first figure in Figure 1. The discussions in Xu et al. (2016) and He et al. (2021) support our suggestion.

Remark 5. It is of interest to study whether the proposed U-statistics can achieve the optimal detection/testing boundary at different sparsity levels. However, this problem for U-statistics, differs from those in previous (e.g., Donoho and Jin (2004)), because of differences between the studied testing problems. Thus, we require a new theoretical development to handle the dependence structure of a banded covariance matrix. When testing $\mathbf{\Sigma} = \mathbf{I}$, Cai and Ma (2013) show that $\mathcal{U}(2)$ is rate optimal in terms of the Frobenius norm for both the testable and the non-testable regions. It would be interesting to extend this result to U-statistics with different orders when testing banded covariance matrices. This is left to future work.

2.5. Simplifying computation

The costs of directly calculating $\mathcal{U}(a)$ in (2.2) and $\hat{\sigma}^2(a)$ in (2.9) are as expensive as $O(n^{2a}p^2)$ and $O(n^a p^2)$, respectively. To reduce the computational cost, we use Algorithm 1 in He et al. (2021) to change the input $s_{i,l}$, thus reducing the computation cost across i from $O(n^a)$ to $O(n)$.

When $\mathrm{E}(\mathbf{x})$ is known, we assume $\mathrm{E}(\mathbf{x}) = 0$, without loss of generality. In this case, $\mathcal{U}(a)$ degenerates into $\tilde{\mathcal{U}}(a)$ in (2.4). **(1).** To compute $\tilde{\mathcal{U}}(a)$, we specify $s_{i,l} = x_{i,j_1}x_{i,j_2}$ in Algorithm 1, where $i = 1,\ldots,n$ and $l \in \mathcal{L} = \{(j_1,j_2) : k < |j_1 - j_2| < p\}$. **(2).** Similarly, we compute $\hat{\sigma}^2(a)$ with $s_{i,l} = \prod_{s=1}^4 (x_{i,j_s} - \bar{x}_{j_s})$, where $i = 1,\ldots,n$ and $l \in \mathcal{L} = \{(j_1,j_2,j_3,j_4) : k < |j_1 - j_2| < p, k < |j_3 - j_4| < p, |j_1 - j_3| \le k, |j_2 - j_4| \le k\}$. When $\mathrm{E}(\mathbf{x})$ is unknown, Proposition 2 explains how to compute $\mathcal{U}(a)$.

Proposition 2. *The forms of* $\mathcal{U}(a)$ *with different order* a *are as follows:*

(i) *When* $a = 1$, $\mathcal{U}(1) = \sum_{k<|j_1-j_2|<p}\{n^{-1}\sum_{i=1}^n x_{i,j_1}x_{i,j_2} - (P_2^n)^{-1}[(\sum_{i=1}^n x_{i,j_1})(\sum_{i=1}^n x_{i,j_2}) - \sum_{i=1}^n x_{i,j_1}x_{i,j_2}]\}$.

(ii) *When* $a = 2$, $\mathcal{U}(2) = \sum_{k<|j_1-j_2|<p}\{(P_2^n)^{-1}U_0(2) - 2(P_3^n)^{-1}U_1(2) + (P_4^n)^{-1}U_2(2)\}$, *with* $U_0(2) = (\sum_{i=1}^n x_{i,j_1}x_{i,j_2})^2 - \sum_{i=1}^n (x_{i,j_1}x_{i,j_2})^2$, $U_1(2) = (\sum_{i=1}^n x_{i,j_1}x_{i,j_2})U_{11}(2) - U_{12}(2) - U_{13}(2)$, *and* $U_2(2) = \prod_{s=1}^2\{(\sum_{i=1}^n x_{i,j_s})^2 - (\sum_{i=1}^n x_{i,j_s}^2)\} - 2U_0(2) - 4U_1(2)$, *with* $U_{11}(2) = (\sum_{i=1}^n x_{i,j_1})(\sum_{i=1}^n x_{i,j_2}) - \sum_{i=1}^n x_{i,j_1}x_{i,j_2}$, $U_{12}(2) = (\sum_{i=1}^n x_{i,j_1}^2 x_{i,j_2})(\sum_{i=1}^n x_{i,j_2}) - \sum_{i=1}^n x_{i,j_1}^2 x_{i,j_2}^2$, *and* $U_{13}(2) = (\sum_{i=1}^n x_{i,j_1}x_{i,j_2}^2)(\sum_{i=1}^n x_{i,j_2}) - \sum_{i=1}^n x_{i,j_1}^2 x_{i,j_2}^2$.

(iii) *When* $a \ge 3$, *let* $\mathcal{U}_c(a) = (P_a^n)^{-1}\sum_{k<|j_1-j_2|<p}\sum_{1\le i_1\ne\cdots\ne i_a\le n}\prod_{l=1}^a (x_{i_l,j_1} - \bar{x}_{j_1})(x_{i_l,j_2} - \bar{x}_{j_2})$. *Under Conditions* 1, 2, 3, *and* H_0, *if* a *is odd*, $p =$

$o(n^{1+a/2})$; if a is even, $p = o(n^{a/2})$. Then, $\{\mathcal{U}(a) - \mathcal{U}_c(a)\}/\sigma(a) \xrightarrow{P} 0$.

When $a = 1, 2$, we compute $\mathcal{U}(a)$ directly using Proposition 2.(i)–(ii). When $a \geq 3$, $\mathcal{U}(a)$ can be replaced by $\mathcal{U}_c(a)$ from Proposition 2.(iii). We compute $\mathcal{U}_c(a)$ using Algorithm 1 by setting $s_{i,l} = (x_{i,j_1} - \bar{x}_{j_1})(x_{i,j_2} - \bar{x}_{j_2})$, for $i = 1, \ldots, n$ and $l \in \mathcal{L} = \{(j_1, j_2) : k < |j_1 - j_2| < p\}$.

2.6. Adaptive bandwidth estimation

Based on previously studied U-statistics, we propose a method for estimating the bandwidth parameter of a high-dimensional banded covariance matrix $\mathbf{\Sigma}$. Our method is motivated by Qiu and Chen (2012). To facilitate the illustration, we first define some notation. For a given bandwidth parameter k, we denote the corresponding statistic $\mathcal{U}(a)$ in (2.2) as $\mathcal{U}_{a,k}$, and its asymptotic standard deviation $\sigma(a)$ and asymptotic standard deviation estimator $\hat{\sigma}(a)$ as $\sigma_{a,k}$ and $\hat{\sigma}_{a,k}$, respectively. Following Qiu and Chen (2012), we consider a banded covariance matrix with true bandwidth k_0. We define $\mathcal{T}_{a,k} = n^{-1}\mathcal{U}_{a,k}/\hat{\sigma}_{a,k}$, and rewrite it as $\mathcal{T}_{a,k} = \mathcal{T}_{a,k,1} + \mathcal{T}_{a,k,2}$, where

$$\mathcal{T}_{a,k,1} = n^{-1}\frac{\mathcal{U}_{a,k} - \mu_{a,k}}{\sigma_{a,k}}\frac{\sigma_{a,k}}{\hat{\sigma}_{a,k}} \quad \text{and} \quad \mathcal{T}_{a,k,2} = n^{-1}\frac{\mu_{a,k}}{\sigma_{a,k_0}}\frac{\sigma_{a,k_0}}{\sigma_{a,k}}\frac{\sigma_{a,k}}{\hat{\sigma}_{a,k}}.$$

Because $\{\mathcal{U}_{a,k} - \mu_{a,k}\}/\sigma_{a,k}$ is stochastically bounded and $\sigma_{a,k}/\hat{\sigma}_{a,k} \xrightarrow{P} 1$, $\mathcal{T}_{a,k,1} = O_p(n^{-1})$. In addition, because σ_{a,k_0} and $\sigma_{a,k}$ are both $\Theta(n^{-a/2}p)$, $\mathcal{T}_{a,k,2}$ is determined using

$$n^{-1}\frac{\mu_{a,k}}{\sigma_{a,k_0}} = \frac{n^{-1}\sum_{k<|j_1-j_2|<p}\sigma_{j_1 j_2}^a}{\left[(P_a^n)^{-1}a!\sum_{k_0<|j_1-j_2|<p,k_0<|j_3-j_4|<p}(\Pi_{j_1,j_2,j_3,j_4})^a\right]^{1/2}}.$$

In particular, if the signs of the covariances σ_{j_1,j_2} are all positive, with $|j_1 - j_2| \leq k$, it can be checked that $n^{-1}\mu_{a,k}/\sigma_{a,k_0} > 0$ for $k < k_0$, and $n^{-1}\mu_{a,k}/\sigma_{a,k_0} = 0$ for $k \geq k_0$. Therefore, we consider an estimator based on the difference between successive statistics $d_{a,k} = \mathcal{T}_{a,k} - \mathcal{T}_{a,k+1}$, for a given finite order $a \in \mathcal{I}$. We multiply n^δ by $\mathcal{T}_{a,k}$, with a small positive $\delta \in (0, 1)$, to increase the magnitude of $\mathcal{T}_{a,k,2}$, and to ensure that $\mathcal{T}_{a,k,1}$ converges to zero in probability at a quick rate. For any $a \in \mathcal{I}$, we define $d_{a,k}^\delta = n^\delta(\mathcal{T}_{a,k} - \mathcal{T}_{a,k+1})$ yielding the bandwidth estimator

$$\hat{k}_{a,\delta,\theta} = \min\{k : |d_{a,k}^\delta| < \theta\}. \tag{2.14}$$

By combining the effects of different orders, we propose an adaptive bandwidth estimator $\hat{k}_{\delta,\theta} = \max_{a\in\mathcal{I}}\hat{k}_{a,\delta,\theta}$. In the Supplementary Material, we present a simulation study in Section S10 to illustrate the motivation of $\hat{k}_{\delta,\theta}$, as well as the consistency of the bandwidth estimator $\hat{k}_{\delta,\theta}$.

Proposition 3. *Under Conditions* 1, 2, 3, *and* $\liminf_n \{\inf_{k<k_0}(\mu_{a,k} - \mu_{a,k+1})\} > 0$, *for any banded covariance matrix with bandwidth* k_0, $\hat{k}_{\delta,\theta} - k_0 \xrightarrow{P} 0$, *for any* $\theta > 0$ *and* $\delta \in (0, 1)$.

In Proposition 3, $\liminf_n \{\inf_{k<k_0}(\mu_{a,k} - \mu_{a,k+1})\} > 0$ excludes the case of a zero sub-diagonal followed by nonzero sub-diagonals as one moves away from the main diagonal. The performance of the adaptive estimator $\hat{k}_{\delta,\theta}$ may be affected by the tuning parameters θ and δ. As pointed out in Qiu and Chen (2012), the multiplier n^δ leads to θ being "free ranged" as long as $\theta > 0$. We suggest $\delta = 0.5$ as a balance between the convergence rate of $\mathcal{T}_{a,k,1}$ and the performance of $\mathcal{T}_{a,k,2}$. The performance of our adaptive bandwidth estimator $\hat{k}_{\delta,\theta}$ in Monte Carlo simulation studies is presented in Section 3.2 .

3. Simulation Study

In this section, we use simulation studies to evaluate the performance of our adaptive tests and estimator. We generate n random vectors $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^{\mathrm{T}}$ from two populations: (i). a multivariate normal distribution: $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$; and (ii). a multivariate t-distribution with seven degrees of freedom: $t_7(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{T}}$. We choose the index set $\mathcal{I} = \{1, \ldots, 6\}$.

3.1. Adaptive testing methods

For $a \in \mathcal{I}$, let "$\mathcal{U}(a)$" denote the testing procedure with the rejection region $\{\mathbf{x}_1, \ldots, \mathbf{x}_n : |\mathcal{U}(a)|/\hat{\sigma}(a) > q_{1-\alpha/2}\}$ and $q_{1-\alpha/2}$ the $(1-\alpha/2)100\%$ quantile of $\mathcal{N}(0, 1)$. Denote "adpUmin" and "adpUf" as our proposed testing procedures in (2.12) and (2.13), respectively. We also compare "adpUmin" and "adpUf" with "$\mathcal{U}(1)$," "$\mathcal{U}(2)$," "$\mathcal{U}(3)$," "$\mathcal{U}(4)$," "$\mathcal{U}(5)$," "$\mathcal{U}(6)$," and "QC" in Qiu and Chen (2012) and "XW" in Xiao and Wu (2013). We take $n = 100$, $p = 50, 100, 200, 400, 600, 800, 1000$ to represent the empirical size, and $n = 100$, $p = 600, 1000$ to investigate the empirical power. The population covariance matrix $\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{T}}$ varies under three settings. Before introducing the settings, we use $J_{|J_A|,k}$ to present a set of $|J_A|$ random positions (j_1, j_2) that satisfy $j_2 - j_1 > k$.

Setting 1. Let $\mathbf{\Gamma} = (\gamma_{j_1 j_2})_{p\times p}$; when $j_2 - j_1 = 1$, $\gamma_{j_1 j_2} = 1$; when $(j_1, j_2) \in J_{|J_A|,1}$, $\gamma_{j_1 j_2} = \rho$; otherwise, $\gamma_{j_1 j_2} = 0$. We investigate the empirical size with $|J_A| = 0$, and the empirical power by varying the signal magnitude $\rho \in (0, 1)$ and the sparsity level $|J_A| = 2, 400, 1200, 2400$. In this setting, the bandwidth

Table 1. Empirical sizes under Setting 1 for $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ and $n = 100$ (in percentage).

p	50	100	200	400	600	800	1,000
adpUmin	4.70	6.40	6.70	7.20	5.60	5.10	4.30
adpUf	5.60	6.80	6.30	6.90	5.80	5.70	4.80
$\mathcal{U}(1)$	4.60	5.70	4.90	5.60	6.10	5.00	5.00
$\mathcal{U}(2)$	5.40	4.40	4.60	5.20	4.80	5.50	5.50
$\mathcal{U}(3)$	5.10	5.10	4.40	5.50	5.60	4.80	5.40
$\mathcal{U}(4)$	5.10	6.20	7.40	6.40	6.10	7.00	4.10
$\mathcal{U}(5)$	4.80	6.10	5.10	5.70	4.70	5.80	4.60
$\mathcal{U}(6)$	3.20	3.90	4.80	6.20	6.10	6.70	5.40
QC	4.50	3.90	4.90	5.00	4.90	5.90	6.20
XW	4.40	4.00	5.50	6.00	3.70	4.90	5.50

Table 2. Empirical sizes under Setting 1 for $t_7(\mathbf{0}, \mathbf{\Sigma})$ and $n = 100$ (in percentage).

p	50	100	200	400	600	800	1,000
adpUmin	6.50	6.10	6.10	6.00	8.30	5.80	7.20
adpUf	8.00	7.00	5.70	5.60	7.70	5.80	6.50
$\mathcal{U}(1)$	3.80	4.60	4.30	5.40	5.70	5.00	4.40
$\mathcal{U}(2)$	6.00	6.30	5.50	4.40	5.90	5.90	6.60
$\mathcal{U}(3)$	6.00	5.00	4.70	6.10	5.30	4.30	5.00
$\mathcal{U}(4)$	5.20	6.00	5.10	5.30	5.90	5.50	5.40
$\mathcal{U}(5)$	5.40	5.90	5.70	6.00	5.80	5.10	6.40
$\mathcal{U}(6)$	4.30	5.20	4.90	4.50	5.30	5.90	5.60
QC	17.4	18.5	20.3	19.6	19.2	21.5	19.6
XW	1.20	1.60	0.80	0.80	0.90	0.80	0.70

$k = 1$ when $|J_A| = 0$.

Setting 2. Let $\mathbf{\Gamma} = (\gamma_{j_1 j_2})_{p\times p}$; when $j_2 - j_1 = 1$, $\gamma_{j_1 j_2} = 0.8$; when $j_2 - j_1 = 2$, $\gamma_{j_1 j_2} = 0.6$; when $(j_1, j_2) \in J_{|J_A|,2}$, $\gamma_{j_1 j_2} = \rho_A$; otherwise, $\gamma_{j_1 j_2} = 0$. We investigate the empirical size with $|J_A| = 0$, and the empirical power by varying the signal magnitudes ρ_A, which are generated from Unif$(0, 2\rho)$ with $\rho \in (0, 1)$, and the sparsity level $|J_A| = 2, 400, 1200, 2400$. In this setting, the bandwidth $k = 2$ when $|J_A| = 0$.

Setting 3. Let $\mathbf{\Gamma} = (\gamma_{j_1 j_2})_{p\times p}$; when $j_2 - j_1 = 1, \ldots, 5$, $\gamma_{j_1 j_2} = 0.6$; when $j_2 - j_1 = 5 + a$, $\gamma_{j_1 j_2} = \rho$; otherwise, $\gamma_{j_1 j_2} = 0$. We investigate the empirical size with $a = 0$, and the empirical power by varying the signal magnitude ρ and the sparsity level $a = 1, 3, 6, 10, 15, 25$. In this setting, the bandwidth $k = 5$ when $a = 0$.

The simulation are replicated 1,000 times and the nominal test level $\alpha = 5\%$. Table 1 presents the empirical sizes with multivariate normal populations for

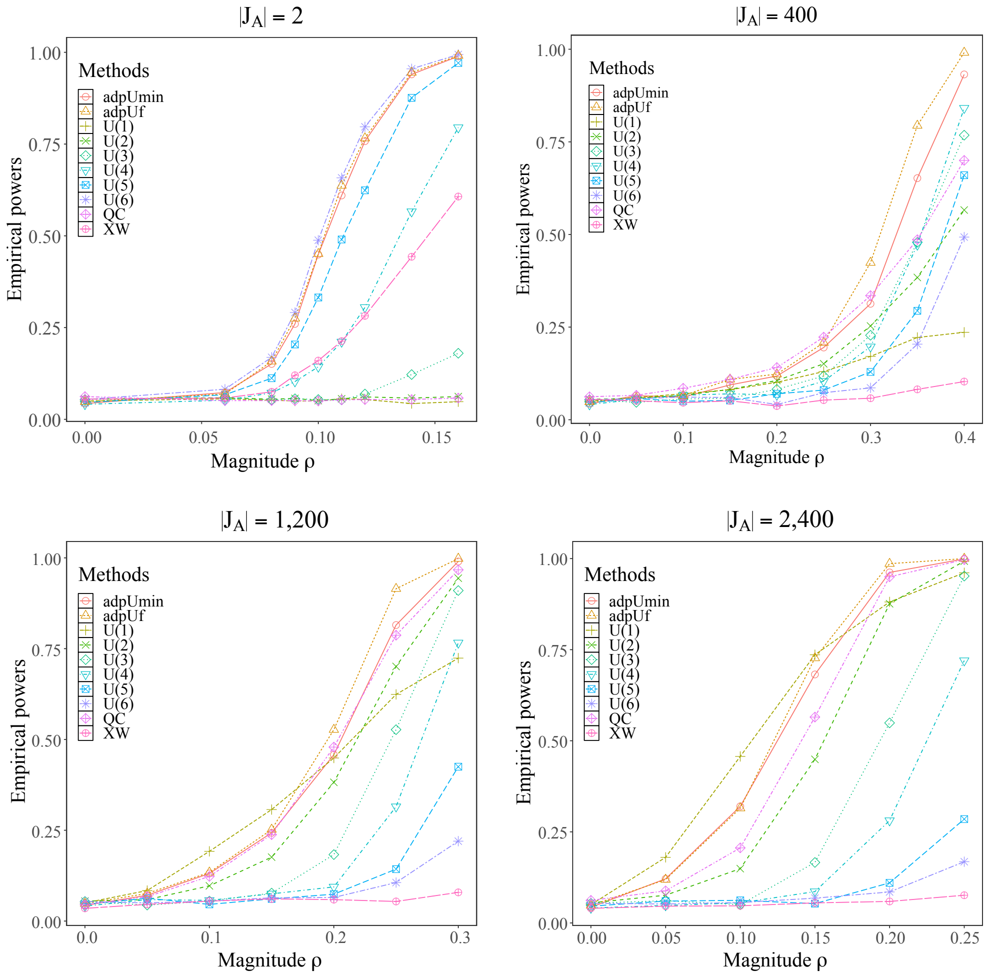

Figure 1. Empirical power comparison under Setting 1 for multivariate normal distribution: $n = 100$, $p = 1000$.

different combinations of n and p under Setting 1. The simulation results show that the empirical sizes of the compared tests are all close to 5%. Table 2 shows the empirical sizes with multivariate t random samples with seven degrees of freedom. The empirical sizes of these single U-statistic tests and our proposed adaptive tests are still close to 5%. However, the empirical sizes of "QC" and "XW" are far from the nominal level. Figure 1 summarizes the empirical power under Setting 1 for multivariate normal random vectors. The empirical power profiles in Figure 1 show that

- for extremely sparse alternatives with $|J_A| = 2$, "$\mathcal{U}(6)$" performs well;
- for moderately sparse alternatives with $|J_A| = 400$, "$\mathcal{U}(4)$" performs well;
- for dense alternatives with $|J_A| = 1200$ and 2400, "$\mathcal{U}(1)$" and "$\mathcal{U}(2)$" perform well;
- when $|J_A|$ increases, the empirical power of "QC" also increases, but the empirical power of "XW" decreases. Nonetheless, our proposed two testing procedures "adpUmin" and "adpUf" always maintain high empirical power regardless the magnitude of $|J_A|$. It also appears that "adpUf" performs better than "adpUmin," in general.

In summary, the adaptive tests either achieved the highest power or were close to the test with the highest power in all settings, indicating their good performance across a wide range of situations. Other simulation results are presented in Section S9.1 of the Supplementary Material. The conclusions are similar to those drawn from Tables 1–2 and Figure 1.

3.2. Adaptive bandwidth estimator

We compare our proposed adaptive bandwidth estimator (Adaptive) with the estimator (BLa) discussed in Bickel and Levina (2008) and the fixed estimator (QC) in Qiu and Chen (2012). For the bandwidth estimation, we set the parameters $\delta = 0.5$ and $\theta = 0.06$ in our proposed estimator $\hat{k}_{\delta,\theta}$ and the QC estimator. The parameters for the BLa estimator are set as in the original study. We set $n = 100, 200$ and $p = 50, 200, 400, 600, 1000$ in Models 1–4, with true bandwidths $k = 2, 5, 10, 15$, respectively.

Model 1. Let $\mathbf{\Gamma} = (\gamma_{j_1 j_2})_{p\times p}$; when $j_2 - j_1 = 1$, $\gamma_{j_1 j_2} = 0.8$; when $j_2 - j_1 = 2$, $\gamma_{j_1 j_2} = 0.6$; otherwise, $\gamma_{j_1 j_2} = 0$.

Model 2. Let $\mathbf{\Gamma} = (\gamma_{j_1 j_2})_{p\times p}$; when $j_2 - j_1 = 1, \ldots, 5$, $\gamma_{j_1 j_2} = 0.6$; otherwise, $\gamma_{j_1 j_2} = 0$.

Model 3. Let $\mathbf{\Gamma} = (\gamma_{j_1 j_2})_{p\times p}$; when $j_2 - j_1 = 1, \ldots, 5$, $\gamma_{j_1 j_2} = 0.2$; when $j_2 - j_1 = 6, \ldots, 10$, $\gamma_{j_1 j_2} = 0.4$; otherwise $\gamma_{j_1 j_2} = 0$.

Model 4. Let $\mathbf{\Gamma} = (\gamma_{j_1 j_2})_{p\times p}$; when $j_2 - j_1 = 1, \ldots, 10$, $\gamma_{j_1 j_2} = 0.2$; when $j_2 - j_1 = 11, \ldots, 15$, $\gamma_{j_1 j_2} = 0.4$; otherwise, $\gamma_{j_1 j_2} = 0$.

Table 3 reports the average empirical bias and standard deviations with the innovations from a normal distribution, based on 100 replications. From Table 3, we observe that our proposed estimator performs well compared with BLa, owing to the smaller bias and standard deviation. The bias and standard deviation of the BLa estimator increase with the dimension p, owing to the inappropriate estimation of the covariance matrix in the high-dimensional setting. Similar results for these estimators with $t_7(\mathbf{0}, \mathbf{\Sigma})$ are presented in Table 5 in Section S9.2 of the Supplementary Material.

4. Data Analysis

In this section, we apply our proposed procedures to a prostate cancer data set from a protein mass spectroscopy study (Adam et al. (2002)). The study analyzed the constituents of the proteins in the blood for two groups of people, namely, the healthy group and the cancer group. The data set has also been studied in Levina, Rothman and Zhu (2008) and Qiu and Chen (2012). For each blood serum sample i, the data consist of the intensity X_{ij} for a large number of time-of-flight values t_j, which is related to the mass-over-charge ratio of the constituent proteins. We analyzed the standardized data set, which consists of 157 healthy and 167 cancer patients, with a 218-dimensional intensity vector for each individual.

We focus on testing a string of null hypotheses $H_{k,0} : \mathbf{\Sigma} = \boldsymbol{B}_k(\mathbf{\Sigma})$, for $k = 0, \ldots, 216$ and estimating the bandwidths of the covariance matrices of the healthy and cancer groups. In particular, we choose $\delta = 0.5$ and $\theta = 0.005$ when analyzing the real data using our adaptive estimator, which is consistent with the choice of Qiu and Chen (2012). We show some representative p-values in Table 4 and bandwidth estimates in Table 5.

All p-values of the "adpUmin" and "adpUf" tests are very close to zero, borrowing strength from $\mathcal{U}(5)$ and $\mathcal{U}(6)$, and the estimated values of our proposed adaptive estimator are 203 and 216 for the healthy group and the cancer group, respectively. In practice, a covariance matrix with a large bandwidth may not be valuable because it will not significantly reduce the number of parameters. Therefore, these small p-values and bandwidth estimates suggest that the covariances of the healthy group and the cancer group may not be banded. Furthermore, the heatmaps in Figure 2 show that most of the sample correlations in the matrices of the two groups are nonnegligible, leading to nonbanded structures, thus supporting our conclusion.

A similar conclusion is obtained from the XW test, which also provides very small p-values for all hypotheses. However, the QC test yields different conclusions, where the smallest k such that $H_{k,0}$ is not rejected is 116 for the

Table 3. Average empirical bias and standard deviation (in parentheses) of three bandwidth estimators with normal innovations: our proposed adaptive bandwidth estimator with $\delta = 0.5$ and $\theta = 0.06$, and the estimators proposed in Bickel and Levina (2008) (BLa) and Qiu and Chen (2012) (QC).

			Bandwidth			
n	p	Methods	2	5	10	15
100	50	Adaptive	0.04(0.243)	0(0)	0(0)	−0.03(0.171)
		BLa	0.15(0.411)	−0.37(0.691)	−0.89(1.144)	−0.96(1.809)
		QC	0(0)	0(0)	0(0)	−0.02(0.141)
	200	Adaptive	0(0)	0(0)	0(0)	0(0)
		BLa	0.38(0.663)	0.27(1.062)	0.14(1.128)	−0.44(1.641)
		QC	0(0)	0(0)	0(0)	0(0)
	400	Adaptive	0(0)	0(0)	0(0)	0.01(0.100)
		BLa	0.56(1.258)	0.84(1.631)	0.50(1.554)	0.12(1.653)
		QC	0(0)	0(0)	0(0)	0(0)
	600	Adaptive	0(0)	0(0)	0(0)	0(0)
		BLa	0.91(1.518)	0.74(1.574)	0.41(1.615)	0.59(2.216)
		QC	0(0)	0(0)	0(0)	0(0)
	1,000	Adaptive	0(0)	0(0)	0(0)	0(0)
		BLa	1.61(2.260)	1.44(2.328)	1.22(2.389)	0.69(2.862)
		QC	0(0)	0(0)	0(0)	0(0)
200	50	Adaptive	0(0)	0.01(0.100)	0.04(0.243)	0.04(0.315)
		BLa	0.11(0.345)	0.09(0.637)	0.19(0.929)	0.22(1.630)
		QC	0.01(0.100)	0.01(0.100)	0(0)	0(0)
	200	Adaptive	0(0)	0(0)	0(0)	0(0)
		BLa	0.29(0.537)	0.34(0.879)	0.14(0.899)	0.05(1.290)
		QC	0(0)	0(0)	0(0)	0(0)
	400	Adaptive	0(0)	0(0)	0(0)	0(0)
		BLa	0.71(1.028)	0.70(1.087)	0.44(1.122)	0.43(1.513)
		QC	0(0)	0(0)	0(0)	0(0)
	600	Adaptive	0(0)	0(0)	0(0)	0(0)
		BLa	0.88(1.17)	1.02(1.463)	0.75(1.344)	1.14(1.809)
		QC	0(0)	0(0)	0(0)	0(0)
	1,000	Adaptive	0(0)	0(0)	0(0)	0(0)
		BLa	1.23(1.399)	1.50(1.957)	1.15(1.822)	1.41(2.396)
		QC	0(0)	0(0)	0(0)	0(0)

healthy group, and 191 for the cancer group. Note that the statistic values of the $\mathcal{U}(2)$ test are the same as those of the QC test. We found that the performance of the QC test differs from that of the $\mathcal{U}(2)$ test with large k. One possible reason for this is that the variance estimation of the statistic in the

Table 4. The p-values (%) of various tests applied to the prostate cancer data set.

		Bandwidith								
	Test	5	86	106	109	116	125	150	200	216
Health Group	adpUmin	0	0	0	0	0	0	0	0	0
	adpUf	0	0	0	0	0	0	0	0	0
	$\mathcal{U}(1)$	0	6.03	99.54	75.87	32.32	8.39	0.19	1.21e-03	1.08e-03
	$\mathcal{U}(2)$	0	<1.0e-13	<1.0e-13	<1.0e-13	1.4e-07	4.47e-6	1.7e-12	<1.0e-13	<1.0e-13
	$\mathcal{U}(3)$	0	<1.0e-13	16.71	16.87	2.4e-09	<1.0e-13	<1.0e-13	<1.0e-13	<1.0e-13
	$\mathcal{U}(4)$	0	0	0	<1.0e-13	<1.0e-13	<1.0e-13	<1.0e-13	0	0
	$\mathcal{U}(5)$	0	0	<1.0e-13	2.4e-03	<1.0e-13	<1.0e-13	0	0	0
	$\mathcal{U}(6)$	0	0	0	0	0	0	0	0	0
	QC	0	2.06e-7	2.06e-7	0	5.64	10.36	12.92	42.69	49.95
	XW	0	0	3.4e-07	4.9e-06	1.6e-03	2.2e-2	2.2e-2	3.1e-2	8.5e-2
		Bandwidith								
	Test	5	52	61	67	91	150	191	200	216
Cancer Group	adpUmin	0	0	0	0	0	0	0	0	0
	adpUf	0	0	0	0	0	0	0	0	0
	$\mathcal{U}(1)$	0	6.19	45.20	96.22	3.3e-01	2.8e-10	4.1e-10	4.1e-09	2.4e-07
	$\mathcal{U}(2)$	0	<1.0e-13	<1.0e-13	<1.0e-13	<1.0e-13	<1.0e-13	<1.0e-13	<1.0e-13	<1.0e-13
	$\mathcal{U}(3)$	0	<1.0e-13	19.85	<1.0e-13	0	0	0	0	0
	$\mathcal{U}(4)$	0	0	0	0	0	0	0	0	0
	$\mathcal{U}(5)$	0	0	0	0	0	0	0	0	0
	$\mathcal{U}(6)$	0	0	0	0	0	0	0	0	0
	QC	0	0	0	0	0	1.3e-8	5.66	22.52	49.79
	XW	0	0	6.0e-12	2.7e-11	2.7e-11	8.0e-11	8.0e-11	8.0e-11	7.8e-07

$\mathcal{U}(a)$, the proposed tests based on U-statistics with different values of a; adpUmin, the adaptive test based on the minimax method; adpUf, the adaptive test based on the Fisher combination; QC, test of Qiu and Chen (2012); XW, test of Xiao and Wu (2013).

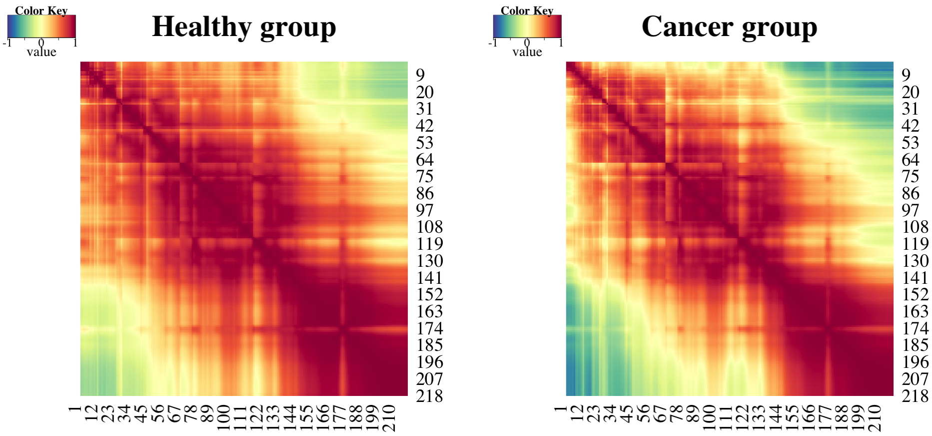

Figure 2. Heatmaps for the sample covariance matrices of the healthy group and the cancer group. Blue represents a negative correlation, red represents a positive correlation, and the color deepens as the correlation increases.

Table 5. The estimated bandwidths of various procedures applied to the prostate cancer data set.

Method	$\mathcal{U}(1)$	$\mathcal{U}(2)$	$\mathcal{U}(3)$	$\mathcal{U}(4)$	$\mathcal{U}(5)$	$\mathcal{U}(6)$	Adaptive	QC
Healthy Group	132	120	193	123	203	128	203	121
Cancer Group	120	74	171	173	175	216	216	212

QC test is based on the assumption that $\mathrm{tr}(\mathbf{\Sigma}^4)/\mathrm{tr}^2(\mathbf{\Sigma}^2) = O(p^{-1})$, which may not be satisfied by real data, whereas our method does not rely on such an assumption when estimating the variance of $\mathcal{U}(2)$. To explain the difference between the two tests, consider the special case of testing $H_{k,0} : \mathbf{\Sigma} = \boldsymbol{B}_k(\mathbf{\Sigma})$, where $\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{T}}$, $\mathbf{\Gamma} = (\gamma_{j_1 j_2})_{p\times p}$ with $\gamma_{j_1 j_1} = 1$, and $\gamma_{j_1 j_2} \sim$ Unif(0,5) for $0 < j_2 - j_1 \leq k$, with $k = 5$ and 200. We generated 1,000 data sets with sample size $n = 157$ and dimension $p = 218$ from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ under $H_{k,0}$. Table 6 shows that the ASD and AEASD of the QC test are too far away from the MCSD. Therefore, the type-I error of the QC test is very small compared with the nominal level 5% when $k = 200$. In this scenario, $\mathrm{tr}(\mathbf{\Sigma}^4)/\mathrm{tr}^2(\mathbf{\Sigma}^2) = 0.944$ and $1000^{-1}\sum_{b=1}^{1000}\mathrm{tr}(\mathbf{S}_{n,b}^4)/\mathrm{tr}^2(\mathbf{S}_{n,b}^2) = 0.936$, where $\mathbf{S}_{n,b} = (n-1)^{-1}\sum_{i=1}^{n}(\mathbf{x}_i^{(b)} - \bar{\mathbf{x}}^{(b)})(\mathbf{x}_i^{(b)} - \bar{\mathbf{x}}^{(b)})^{\mathrm{T}}$ with $\bar{\mathbf{x}}^{(b)} = n^{-1}\sum_{i=1}^{n}\mathbf{x}_i^{(b)}$, and $\{\mathbf{x}_1^{(b)}, \ldots, \mathbf{x}_n^{(b)}\}$ is the bth sampling from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. This shows that the QC test may overestimate the variance of the statistic when $\mathrm{tr}(\mathbf{\Sigma}^4)/\mathrm{tr}^2(\mathbf{\Sigma}^2)$ is large, thus violating their assumption. The QC test still works well when $\mathrm{tr}(\mathbf{\Sigma}^4)/\mathrm{tr}^2(\mathbf{\Sigma}^2)$ is small, for example, $\mathrm{tr}(\mathbf{\Sigma}^4)/\mathrm{tr}^2(\mathbf{\Sigma}^2) = 0.024$ and $1000^{-1}\sum_{b=1}^{n}\mathrm{tr}(\mathbf{S}_{n,b}^4)\ /\mathrm{tr}^2(\mathbf{S}_{n,b}^2) = 0.041$, with $k = 5$.

Table 6. Results based on 1,000 multivariate normal samples under $H_{k,0} : \mathbf{\Sigma} = \boldsymbol{B}_k(\mathbf{\Sigma})$.

	Case 1 : $k = 5$			
	ASD	AEASD	MCSD	Type I Error
$\mathcal{U}(2)$	20,498.76	20,242.96	21,317.02	0.069
QC	21,059.26	21,079.46	21,317.02	0.054
	Case 2 : $k = 200$			
	ASD	AEASD	MCSD	Type I Error
$\mathcal{U}(2)$	96,270.76	92,689.35	94,453.38	0.046
QC	152,093,323	151,994,013	94,453.38	0

ASD, asymptotic standard deviation of statistic; AEASD, average of estimations of the asymptotic standard deviation of the statistics based on 1,000 replications; MCSD, sample standard deviation of the statistics based on 1,000 replications.

In the real-data analysis, for the health and cancer groups, $\mathrm{tr}(\mathbf{S}_n^4)/\mathrm{tr}^2(\mathbf{S}_n^2) = 0.907$ and 0.820, respectively, indicating a possible overestimation of the variance of the QC test statistic.

5. Conclusion

We have proposed adaptive tests based on a series of U-statistics for testing the bandedness of a high-dimensional covariance matrix. We have investigated the asymptotic joint distribution of the U-statistics under the null hypothesis and specific local alternative hypotheses. Furthermore, we take advantage of the asymptotic independence of multiple U-statistics to construct two proposed adaptive tests by combining the p-values of the U-statistics. Our simulation studies show that the proposed tests are powerful across a wide range of alternatives, whereas the existing tests are powerful only for dense alternatives or sparse alternatives. We also propose a new consistent bandwidth estimator motivated by the by-product of the U-statistics.

The bandwidth k is usually unknown in practice. Instead of testing a general bandedness structure of covariance matrix with a given k, it is of interest to regard the bandwidth k as a tuning parameter and examine the asymptotic properties of a series of U-statistics based on $\sum_{\hat{k}<|j_1-j_2|<p} \sigma_{j_1 j_2}^a$, where $\hat{k}$ is the estimation of the true bandwidth k. As shown in Zhong et al. (2017), the plug-in estimator $\hat{k}$ may incur some leading-order effects. We will examine this topic in the future.

Supplementary Material

The online Supplementary Material contains detailed proofs of the theoretical results and additional simulation results.

Acknowledgements

We would like to thank the editor, associate editor, and two anonymous reviewers for their constructive comments. We are also grateful to Dr. Yumou Qiu for sharing the prostate cancer data set used in Qiu and Chen (2012). Shurong Zheng and Xiaoyi Wang were supported by NSFC grant 12071066, and Gongjun Xu was supported by NSF grant SES 1846747.

References

Adam, B. L., Qu, Y. S., Davis, J. W., Ward, M. D., Clements, M. A., Cazares, L. H. et al. (2002). Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Research* **62**, 3609–3614.

Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. 3nd Edition. Wiley, Hoboken.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.

Bai, Z. D., Jiang, D. D., Yao, J. F. and Zheng, S. R. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *The Annals of Statistics* **37**, 3822–3840.

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

Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *The Annals of Probability* **21**, 1276–1294.

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics* **36**, 199–227.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* **2**, 107–144.

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

Cai, T. T. and Liu, W. D. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* **106**, 672–684.

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

Chen, S. X., Li, J. and Zhong, P. S. (2019). Two-sample and ANOVA tests for high dimensional means. *The Annals of Statistics* **47**, 1443–1474.

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

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics* **32**, 962–994.

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

He, Y. Q., Xu, G. J., Wu, C. and Pan, W. (2021). Asymptotically independent U-statistics in high-dimensional testing. *The Annals of Statistics* **49**, 154–181.

Jiang, D. D., Jiang, T. F. and Yang, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *Journal of Statistical Planning and Inference* **142**, 2241–2256.

Jiang, T. F. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *The Annals of Statistics* **41**, 2029–2074.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics* **29**, 295–327.

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

Levina, E., Rothman, A. and Zhu, J. (2008). Sparse estimation of large covariance matrices via a nested Lasso penalty. *The Annals of Applied Statistics* **2**, 245–263.

Li, B. S. and Leal, S. M. (2008). Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *The American Journal of Human Genetics* **83**, 311–321.

Li, Z. and Yao, J. F. (2016). Testing the sphericity of a covariance matrix when the dimension is much larger than the sample size. *Electronic Journal of Statistics* **10**, 2973–3010.

Ligeralde, A. V. and Brown, B. W. (1995). Band covariance matrix estimation using restricted residuals: A Monte Carlo analysis. *International Economic Review* **36**, 751–767.

Morgenthaler, S. and Thilly, W. G. (2007). A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: A cohort allelic sums test (CAST). *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **615**, 28–56.

Pan, W., Kim, J., Zhang, Y. W., Shen, X. T. and Wei, P. (2014). A powerful and adaptive association test for rare variants. *Genetics* **197**, 1081–1095.

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

Shao, Q. M. and Zhou, W. X. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *The Annals of Probability* **42**, 623–648.

Wang, C., Yang, J., Miao, B. Q. and Cao, L. B. (2013). Identity tests for high dimensional data using RMT. *Journal of Multivariate Analysis* **118**, 128–137.

Wang, Q. W. and Yao, J. F. (2013). On the sphericity test with large-dimensional observations. *Electronic Journal of Statistics* **7**, 2164–2192.

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

Xu, G. J., Lin, L. F., Wei, P. and Pan, W. (2016). An adaptive two-sample test for high-dimensional means. *Biometrika* **103**, 609–624.

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

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

Xiaoyi Wang

Center for Statistics and Data Science, Beijing Normal University, Zhuhai 519087, China.

E-mail: wangxy059@nenu.edu.cn

Gongjun Xu

Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA.

E-mail: gongjun@umich.edu

Shurong Zheng

School of Mathematics & Statistics and KLAS, Northeast Normal University, Changchun, Jilin 130024, China.

E-mail: zhengsr@nenu.edu.cn

(Received December 2020; accepted April 2022)