Journal of Multivariate Analysis 169 (2019) 21–32

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Projection tests for high-dimensional spiked covariance matrices

Wenwen Guo *, Hengjian Cui

Capital Normal University, Beijing, China

A R T I C L E I N F O

Article history:
Received 7 December 2017
Available online 27 August 2018

Keywords:
Large p small n
Power enhancement technique
Projection
Spiked covariance matrix

AMS subject classification:
62H99

A B S T R A C T

Testing the existence of low-dimensional perturbations or signals is very important, e.g., in factor analysis and signal processing. This paper aims to develop new tests for high-dimensional spiked covariance matrices based on a projection approach. The asymptotic distribution of the proposed tests is obtained under regularity conditions. We further explore a power enhancement technique under covariance matrix sparsity. The finite-sample enhanced power performance of the proposed tests is shown through simulations. A microarray dataset is used for illustration purposes.

1. Introduction

Considerable attention has been paid to testing the identity and sphericity of covariance matrices. In modern applications, the matrix is often large, with the number p of variables comparable to, or much larger than, the sample size n. In order to accommodate high dimensionality, new tests based on entropy or quadratic loss have been proposed. Under entropy loss, Bai et al. [2] proposed a modified likelihood ratio statistic for testing identity when $p/n \to c \in (0, 1)$. Further work about the likelihood ratio test in high-dimensional settings is reported in [13]. Zhang et al. [26] proposed empirical likelihood ratio procedures for testing whether a covariance matrix equals a given one or has a banded structure. Under quadratic loss, Ledoit and Wolf [17] built a test statistic by measuring the Euclidean distance between the sample covariance matrix and the null matrix, and they obtained asymptotic properties in a Gaussian framework when $p/n \to c < \infty$.

To avoid estimating the covariance matrix when $p > n$, Chen et al. [10] investigated testing procedures based on U-statistics for both the identity and the sphericity hypothesis. Li and Chen [18] proposed tests for equality of covariance matrices applicable for "large p, small n" situations. Cai and Ma [8] considered testing the equality of a covariance matrix to a given one; they characterized the boundary separating the testable region from the non-testable region when p/n is bounded. Recently, Peng et al. [24] considered improving the power of testing for identity and sphericity by employing banding estimators for high-dimensional covariance matrices.

Recent studies place an emphasis on the existence of low-dimensional perturbations or signals in the data, which is a special alternative to the identity or sphericity of the covariance matrix. To be exact, the corresponding alternative covariance matrix is proportional to the sum of the identity matrix and a matrix of finite rank K. Johnstone [14] calls it a "spiked covariance". The practical justification for analyzing such matrices comes from various sources, such as signal processing and factor analysis; see [4,5,15,16,21–23] for a few examples.

In a high-dimensional setting, a sparsity assumption was proposed to ensure the asymptotic properties of the statistics. Cai et al. [9] considered both minimax and adaptive estimation of the principal subspace in sparse principal component

* Corresponding author.
E-mail addresses: wwguo@cnu.edu.cn (W. Guo), hjcui@bnu.edu.cn (H. Cui).

https://doi.org/10.1016/j.jmva.2018.08.009
0047-259X/

analysis and established the optimal rates of convergence for estimation of the principal subspace. Berthet and Rigollet [4] investigated the detection of sparse principal component problems, and they proposed a test based on a conservative critical value in order to control the type I error of the test. However, to the best of our knowledge, there has been no attempt to address the problem of testing spiked covariance matrices, which motivates this work.

In high-dimensional contexts, many asymptotic properties of the classical test statistics no longer hold. Therefore, we propose a two-step testing procedure to detect the existence of perturbations or signals in spiked covariance matrices. First, we project the original sample to a lower dimensional space; second, we carry out an efficient test with the projected sample. If there is only one spike, so that the largest eigenvalue of the covariance matrix is unique, we derive the optimal projection space, i.e., the eigen-subspace associated with the largest eigenvalue, and the projection direction can be estimated by a data driven method. If there is more than one spike, accounting only for the largest eigenvalue will result in the loss of some useful information. Motivated by the power enhancement approach in [11], we modify our statistic to improve performance using all the perturbations of the alternative hypothesis.

The methodology of the projection tests for the spiked covariance matrices is described in Section 2, where the theoretical optimal direction and the performance of the proposed test are also explored. In Section 3, a power enhancement technique is introduced. Section 4 reports Monte Carlo simulations and an example on the performance of the proposed tests. Technical proofs are relegated to the Appendix.

2. Projection tests and their performance

In this section, testing for the high-dimensional spiked covariance model introduced in [14] is discussed. Suppose that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent and identically distributed (iid) p-dimensional random vectors with covariance matrix $\boldsymbol{\Sigma} = \mathbf{I}_p + \mathbf{V}\boldsymbol{\Omega}\mathbf{V}^\top$, where $\mathbf{I}_p$ is the p-dimensional identity matrix, $\boldsymbol{\Omega} = \operatorname{diag}(\omega_1, \ldots, \omega_K)$ with $\omega_1 \geq \cdots \geq \omega_K > 0$, and $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_K)$ is a $p \times K$ matrix with orthogonal and unit columns. In this model, K is called the number of spikes.

In [15], it is stated that for each $k \in \{1, \ldots, K\}$, the kth largest eigenvalue of $\boldsymbol{\Sigma}$ is $\omega_k + 1$, and the other eigenvalues are all equal to 1. Moreover, $\mathbf{v}_1, \ldots, \mathbf{v}_K$ are the eigenvectors associated to the largest K eigenvalues of the covariance matrix $\boldsymbol{\Sigma}$. In this work, the object is to detect the existence of the spikes, i.e., to test the hypotheses

$$\mathcal{H}_{10} : \boldsymbol{\Sigma} = \mathbf{I}_p \quad \text{vs.} \quad \mathcal{H}_{11} : \boldsymbol{\Sigma} = \mathbf{I}_p + \mathbf{V}\boldsymbol{\Omega}\mathbf{V}^\top \tag{1}$$

or

$$\mathcal{H}_{20} : \boldsymbol{\Sigma} = \sigma^2\mathbf{I}_p \quad \text{vs.} \quad \mathcal{H}_{21} : \boldsymbol{\Sigma} = \sigma^2(\mathbf{I}_p + \mathbf{V}\boldsymbol{\Omega}\mathbf{V}^\top), \tag{2}$$

where σ^2 is an unknown but finite positive constant. Note that the null hypothesis in (1) covers the hypothesis $\mathcal{H}_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ for a specific known invertible matrix $\boldsymbol{\Sigma}_0$.

First consider the hypothesis test in (1). The objective is to discriminate between the null and the alternative hypothesis in (1) from a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. When p is fixed, a natural idea is to use the difference of the sample covariance and the null matrix, i.e.,

$$T_{n,p} = \lambda_{\max}(\mathbf{S}) - 1,$$

to perform the test, where $\lambda_{\max}(\mathbf{S})$ is the largest eigenvalue of $\mathbf{S}$. However, due to the curse of dimensionality, many desirable properties of classical statistics no longer hold, e.g., the sample covariance matrix is no longer a consistent estimator of $\boldsymbol{\Sigma}$. A straightforward idea is to reduce the dimension to an acceptable level and then perform the test on the observations with relatively lower dimension. This idea is pursued here: we project the original sample to a lower dimensional space first, and then carry out a test with the projected sample.

2.1. Projection procedure

Let $\mathbf{P}$ be a $p \times s$ matrix with $\mathbf{P}^\top\mathbf{P} = \mathbf{I}_s$, $s \leq p$ and $s < n$. For each $i \in \{1, \ldots, n\}$, set $\mathbf{Y}_i = \mathbf{P}^\top\mathbf{X}_i$. Then $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are independent and identically distributed with covariance matrix $\boldsymbol{\Sigma}_\mathbf{P} = \mathbf{P}^\top\boldsymbol{\Sigma}\mathbf{P}$. Obviously, under $\mathcal{H}_0$, $\operatorname{var}(\mathbf{Y}_i) = \mathbf{I}_s$. Define the projection test of $T_{n,p}$ to be

$$T^{\mathbf{P}}_{n,p} = \lambda_{\max}(S_Y) - 1,$$

which is constructed based on the $\mathbf{Y}_i$s. Then $T^{\mathbf{P}}_{n,p}$ can be used to test $\mathcal{H}_{0P} : \boldsymbol{\Sigma}_\mathbf{P} = \mathbf{I}_s$. Here, we want to determine $\mathbf{P}$ such that $T^{\mathbf{P}}_{n,p}$ maximizes the power under $\mathcal{H}_{11}$ in (1). The following proposition provides a solution.

Proposition 1. *If $\lambda_1(\boldsymbol{\Sigma}_\mathbf{P}) \geq \cdots \geq \lambda_s(\boldsymbol{\Sigma}_\mathbf{P})$, then the projection test $T^{\mathbf{P}}_{n,p}$ reaches its maximum power for $\mathcal{H}_{11}$ at $s = 1$ and $\mathbf{P} = \mathbf{v}_1$.*

Proposition 1 sheds some light on how to construct a projection test for a spiked covariance matrix. In addition, it is also seen that $T^{\mathbf{P}}_{n,p} = \mathbf{v}_1^\top\mathbf{S}_\mathbf{X}\mathbf{v}_1 - 1$ when $\mathbf{P} = \mathbf{v}_1$. Since $\mathbf{P}$ depends on an unknown eigenvector of the covariance matrix ($\mathbf{v}_1$), we choose an estimation value of $\mathbf{v}_1$ in applications.

2.2. Testing procedure

In practice, estimating the direction $\mathbf{v}_1$ is required. Motivated by the single sample-splitting strategy proposed in [25], we split the sample into two separate parts, viz. $\mathcal{S}_1 = \{X_{11}, \ldots, X_{1n_1}\}$ and $\mathcal{S}_2 = \{X_{21}, \ldots, X_{2n_2}\}$ with $n_1 + n_2 = n$. We use $\mathcal{S}_1$ to estimate the direction $\mathbf{v}_1$ and $\mathcal{S}_2$ to construct the test statistic. Denote the estimation of direction and the statistic by $\widehat{\mathbf{v}}_{1(1)}$ and $T^{\mathbf{P}}_{n_2,p}$, respectively. Then, the statistic

$$T^{\mathbf{P}}_{n_2,p} = \widehat{\mathbf{v}}^{\top}_{1(1)} \mathbf{S}^{(2)}_{\mathbf{X}} \widehat{\mathbf{v}}_{1(1)} - 1 \tag{3}$$

can be used to test hypothesis (1), where $\mathbf{S}^{(2)}_{\mathbf{X}}$ denotes the sample covariance matrix of $\mathcal{S}_2$. Next, a test of hypothesis (2) is considered. In this case, the largest eigenvector $\mathbf{v}_1$ and another parameter σ^2 need to be estimated. Here, the statistic

$$W^{\mathbf{P}}_{n_2,p} = \frac{\widehat{\mathbf{v}}^{\top}_{1(1)} \mathbf{S}^{(2)}_{\mathbf{X}} \widehat{\mathbf{v}}_{1(1)}}{\mathrm{tr}(\mathbf{S}_{\mathbf{X}})/p} - 1 \tag{4}$$

is used to test (2).

At this point, the normality assumption on the variables is dropped, and we explore the asymptotic properties of the test statistics under the assumption that the covariates follow an elliptical distribution. This broad class includes many useful distributions, such as multivariate Gaussian, Kotz-type, and Student t distributions; see, e.g., [12]. Next, Conditions (C1)–(C4) are introduced.

(C1) The random vector follows a p-dimensional elliptically contoured distribution, i.e., $\mathbf{X}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma} R_i \mathbf{U}_i$, where $\boldsymbol{\Gamma}$ is a $p \times p$ matrix, $\mathbf{U}_i$ is a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$, and R_i is a nonnegative random variable independent of $\mathbf{U}_i$ such that $\mathrm{E}(R_i^2) = p$ and $\mathrm{var}(R_i^2) = O(p)$.
(C2) As $n \to \infty$, $p = p(n) \to \infty$, and $n_2/n \to b \in (0, 1)$.

The following theorem gives the asymptotic null distribution of the statistics $T^{\mathbf{P}}_{n_2,p}$ and $W^{\mathbf{P}}_{n_2,p}$ in (3) and (4) under (C1) and (C2). In what follows, $\rightsquigarrow$ stands for convergence in distribution.

Theorem 1. *If Conditions (C1)–(C2) hold, then* $\sqrt{n_2}\, T^{\mathbf{P}}_{n_2,p} \rightsquigarrow \mathcal{N}(0, 2)$ *under the null hypothesis in* (1), *while* $\sqrt{n_2}\, W^{\mathbf{P}}_{n_2,p} \rightsquigarrow \mathcal{N}(0, 2)$ *under the null hypothesis in* (2).

Thus, for sufficiently large n, the proposed tests reject $\mathcal{H}_{10}$ and $\mathcal{H}_{20}$ at the significant level α if

$$\sqrt{n_2} T^{\mathbf{P}}_{n_2,p}/\sqrt{2} \geq z_\alpha \quad \text{and} \quad \sqrt{n_2} W^{\mathbf{P}}_{n_2,p}/\sqrt{2} \geq z_\alpha,$$

respectively, where z_α is the upper α quantile of the standard normal distribution.

2.3. Performance of the tests

In this subsection, we study the power of the proposed tests under local alternatives and the condition that the estimator of the largest eigenvector is consistent, as specified below.

(C3) Local alternative: $\omega_1 = o(1)$.
(C4) There exists $\mathbf{v}_1 \in \mathcal{R}_1$ such that $\|\mathbf{v}_1\| = 1$ and $\widehat{\mathbf{v}}^{\top}_{1(1)} \mathbf{v}_1 \to 1$ in probability.

Here, $\mathcal{R}_1$ denotes the spanning space of the largest eigenvector(s). Finding a consistent estimator of $\mathbf{v}_1$ is a challenge but previous work on matrix estimation and principal component analysis based on high-dimensional setting allows us to obtain a consistent estimator under certain sparse conditions; see [4,6,7,9,14,23] for a few examples. If $\widehat{\mathbf{v}}_{1(1)}$ is inconsistent, we may also use these methods to test the hypotheses (1) and (2), but it might not be so efficient, as shown in the following theorem.

Theorem 2. *Assume Conditions (C1)–(C4) hold and set* $\boldsymbol{\Sigma}_1 = \mathbf{I}_p + \mathbf{V}\boldsymbol{\Omega}\mathbf{V}^{\top}$. *Then*

$$\sqrt{n_2}\,(1+\omega_1)^{-1} \left\{ T^{\mathbf{P}}_{n_2,p} - \left(\widehat{\mathbf{v}}^{\top}_{1(1)} \boldsymbol{\Sigma}_1 \widehat{\mathbf{v}}_{1(1)} - 1 \right) \right\} \rightsquigarrow \mathcal{N}(0, 2) \tag{5}$$

and

$$\sqrt{n_2}\,(1+\omega_1)^{-1} \left\{ W^{\mathbf{P}}_{n_2,p} - \left(\widehat{\mathbf{v}}^{\top}_{1(1)} \boldsymbol{\Sigma}_1 \widehat{\mathbf{v}}_{1(1)} - 1 \right) \right\} \rightsquigarrow \mathcal{N}(0, 2). \tag{6}$$

Theorem 2 provides the asymptotic power of the two tests, viz.

$$\beta_1^{Ptest}(\omega_1) = \beta_2^{Ptest}(\omega_1) = \Phi \left\{ -\frac{z_\alpha}{1+\omega_1} + \sqrt{\frac{n_2}{2}} \left(1 - \frac{1}{1+\omega_1} \right) \right\}, \tag{7}$$

and both converge to 1 as $n_1, n_2 \to \infty$ if $\omega_1 > 0$.

Remark 1. Our proposed test might be affected by the result of random splitting. To solve this issue, Meinshausen et al. [20] introduced a multi-splitting technique. A single random splitting procedure is used to construct a test statistic, and this procedure is then repeated N times, leading to N p-values. These p-values could then be aggregated using Eq. (2.3) in that paper. Under this multi-splitting method, the effect of random splitting will be eliminated, and the false discovery rate will also be controlled. In this paper, the multi-splitting method can also be used to eliminate the effect of random splitting, where the results and their proofs are quite similar to the corresponding parts in [20].

Remark 2. Theorems 1 and 2 are obtained under the elliptical distribution assumption (C1). However, the results could be extended to the pseudo-independence assumption listed in [3].

3. Power enhancement technique

In Section 2, the statistic is constructed based on the largest eigenvalue of the spiked covariance matrix, and we obtain the asymptotic distribution of the statistics under regularity conditions (C1)–(C4), which required that the covariance matrices are sparse enough to obtain consistent estimator(s) of the largest eigenvector(s). If the number K of spikes is greater than 1, it will lead to information loss of the kth largest eigenvalue for $k \in \{2, \ldots, K\}$. Similar to Assumption 3.1 in [11], a power enhancement technique is proposed if the eigenvalues of the estimated covariance matrix $\widehat{\mathbf{\Sigma}}$ satisfy the following condition:

(C5) $\inf_{\Sigma \in \mathcal{U}^*} \Pr\left(\max_{i\in\{1,\ldots,p\}} |\hat{\lambda}_i - \lambda_i| < \delta_{n,p} |\lambda/\sqrt{n}\right) \to 1$

where $\delta_{n,p} = C_1 \ln\{\ln(n)\}\sqrt{\ln(p)}$, $\mathcal{U}^*$ is a collection of spiked covariance matrices $\mathbf{\Sigma}$, λ is the set of all the eigenvalues of $\mathbf{\Sigma}$, and $\lambda_1 \geq \cdots \geq \lambda_p$ and $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p$ stand for the eigenvalues of $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}$, respectively. Condition (C5) requires that the estimations of the eigenvalues be consistent. The power enhancement technique can then be applied.

Remark 3. In this paper, Conditions (C4) and (C5) are listed separately. Often these two conditions can be eliminated if we assume that the covariance matrices are sparse. For example, Cai and Liu [7] defined a special class of sparse matrices, viz.

$$\mathcal{U}_q^* = \left\{ \mathbf{\Sigma} : \mathbf{\Sigma} \succ 0, \max_{i\in\{1,\ldots,p\}} \sum_{j=1}^{p} (\sigma_{ii}\sigma_{jj})^{(1-q)/2} |\sigma_{ij}|^q \leq C \right\}$$

for $q = 1/2$ and C an unknown constant. Under this condition, they proposed an adaptive thresholding technique on the sample covariance matrix. In more detail, denote the sample covariance matrix $\mathbf{S} = (\hat{\sigma}_{ij})$ and the estimator proposed by Cai and Liu [7], viz.

$$\widehat{\mathbf{\Sigma}} = (\hat{\sigma}_{ij}^*),$$

with $\hat{\sigma}_{ij}^* = \hat{\sigma}_{ij}\mathbf{1}(|\widehat{\sigma}_{ij}| \geq \lambda_{ij})$, where $\mathbf{1}$ is an indicator function, $\lambda_{ij} = \delta\sqrt{\hat{\xi}_{ij}\ln(p)/n}$,

$$\hat{\xi}_{ij} = \frac{1}{n_1}\sum_{k=1}^{n_1}\left\{\left(X_{1ki} - \overline{X}^{1i}\right)\left(X_{1kj} - \overline{X}^{1j}\right) - \hat{\sigma}_{ij}\right\}^2, \quad \overline{X}^{1i} = \frac{1}{n_1}\sum_{k=1}^{n_1} X_{1ki}$$

and δ is a tuning parameter that could be taken as fixed ($\delta = 2$). In their paper, Conditions (C4) and (C5) are satisfied.

If (C5) is satisfied, a power enhancement technique motivated by the study in [11] can be proposed. Let J_1 be a test statistic that has a correct asymptotic size, e.g., the projected tests introduced in Section 2, and then augment the test by adding a power enhancement component $J_0 \geq 0$. Define

$$S(\lambda) = \left\{i \in \{1, \ldots, p\} : \lambda_i - 1 > 2\delta_{n,p}/\sqrt{n}\right\}, \tag{8}$$

$$\widehat{S} = \left\{i \in \{1, \ldots, p\} : \widehat{\lambda}_i - 1 > \delta_{n,p}/\sqrt{n}\right\}. \tag{9}$$

Then, choose

$$J_0 = n\sum_{i=1}^{p}(\widehat{\lambda}_i - 1)\mathbf{1}(\widehat{\lambda}_i - 1 > \delta_{n,p}/\sqrt{n}) = n\sum_{i\in\widehat{S}}(\widehat{\lambda}_i - 1). \tag{10}$$

The following theorem states the asymptotic behavior of J_0.

Theorem 3. *Suppose the covariance matrix satisfies Condition (C5). Then as $n \to \infty$, $\Pr(\widehat{S} = \emptyset|\mathcal{H}_0) \to 1$ under $\mathcal{H}_0$. Hence,*

$$\Pr(J_0 = 0|\mathcal{H}_0) \to 1 \quad \text{and} \quad \inf_{\{\lambda:S(\lambda)\neq\emptyset\}} \Pr(J_0 > \sqrt{n}|\lambda) \to 1.$$

Theorem 3 gives the asymptotic behavior of J_0, and at the same time the sure screening property of $\widehat{S}$ which indicates that the proposed procedure selects all the significant components whose indices are in $S(\lambda)$ rather than only the largest one. Therefore, adding the term J_0 will take more information from the alternative, and then the power of the test will increase accordingly. We know from the analysis in the last section that $\sqrt{n_2}\,T^{\mathbf{P}}_{n_2,p}/\sqrt{2}$ follows an asymptotic Gaussian distribution under the null hypothesis. Similar to the discussion in [11], a general form of the test statistics is proposed. Assume there is a test statistic J_1 following a known asymptotic null distribution denoted by F, and which rejects $\mathcal{H}_0$ if $J_1 > F_\alpha$. Then, the following result can be deduced from Theorem 3.

Theorem 4. *Let Condition (C5) hold. Define $\Lambda_s = \{\lambda : \boldsymbol{\Sigma} \in \mathcal{U}^*, S(\lambda) \neq \emptyset\}$. Then J_1 and $J_1 + J_0$ share the same asymptotic null distribution F. In addition, as $n \to \infty$,*

$$\inf_{\lambda\in\Lambda_s} \Pr(J_1 + J_0 \geq F_\alpha|\lambda) \to 1.$$

Theorem 4 shows that the test can attain high power for $\lambda \in \Lambda_s$. Similarly, we can define $S(\lambda)$ and $\hat{S}$ to improve the power for testing hypothesis (2). It is worth mentioning that although the proposed projection test statistics in Section 2 and the power enhancement technique in Section 3 are closely connected, the latter can also be used to enhance the power through combination with tests in [10,17] and many others.

4. Simulations

In this section, Monte Carlo simulations are used to assess the finite-sample performance of the proposed projection tests (PTest) compared with the tests proposed by Jiang and Yang [13], Ledoit and Wolf [17], and Chen et al. [10]. For convenience, these three tests are denoted as CLRT, LW test, and CZZ test, respectively. Note that these procedures were designed to test identity and sphericity, so it is not quite fair for them to be applied to detect the existence of spikes. Here, they are just used for comparison.

4.1. Example 1

As the projection matrix $\mathbf{P} = \mathbf{v}_1$ is unknown, an estimation value can be used. In this paper, a fraction of the samples is taken out to estimate the direction $\mathbf{v}_1$, but the percentage samples should be investigated. Denote $\kappa \in (0, 1)$ as the splitting parameter with $n_1 = \lfloor n\kappa \rfloor$ and $n_2 = n - n_1$, where $\lfloor n\kappa \rfloor$ denotes the integer part of $n\kappa$. It is expected that a larger value of κ will produce a more accurate estimation of the direction, yet it will reduce the number of samples to construct the test statistic. Therefore, we explore the effect of κ on the power of the projection test by taking a grid of values of κ over $(0, 1)$. Fig. 1 depicts the power function of κ under two different alternative covariance structures with sample size $n \in \{100, 200\}$ and $p \in \{50, 100, 200, 400\}$. They are

$$\mathcal{H}^1_{11} : \boldsymbol{\Sigma} = \mathrm{diag}(2, 1.5, 1.5, 1, 1, \ldots, 1), \quad \mathcal{H}^2_{11} : \boldsymbol{\Sigma} = \mathbf{I}_p + 2\boldsymbol{\xi}_1\boldsymbol{\xi}_1^\top + 1.5\mathbf{e}_2\mathbf{e}_2^\top + 1.5\mathbf{e}_3\mathbf{e}_3^\top,$$

where $\xi_1 = (3, 2, 1, 0, 0, \ldots, 0)^\top/\sqrt{14}$ and for $i \in \{2, 3\}$, $\mathbf{e}_i$ is a p-dimensional vector with the ith element equal to 1 and the other elements are all 0. It is assumed that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent and identically distributed following two p-dimensional elliptically contoured distributions: the multivariate Gaussian distribution $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the multivariate Student t distribution $\sqrt{1 - 2/q}\,\mathcal{T}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q)$ with $q = p/2$. Each element of the mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ is generated once and for all using a $\mathcal{U}(2, 3)$. The tests are repeated 500 times to simulate the power.

Figs. 1–2 depict the effect of κ on the power under the two alternatives. When κ is relatively small, say less than 0.5, the power increases with κ because the estimation of the projection direction is more accurate. However, an overly large κ, say more than 0.7, adversely affects the power, because the power of the proposed projection test largely depends on the sample size used to construct the statistic, especially for relatively small samples. The empirical "optimal" splitting percentage varies from case to case. Figs. 1–2 illustrate that $\kappa = 0.6$ is a reasonable choice.

Table 1 shows the empirical sizes of the proposed test (PTest) and the related tests (LW, CLRT, and CZZ). Tables 2–3 report the empirical power of the PTest and some other tests for the two alternative settings respectively. It can be deduced from Table 1 that if the random vectors follow the multivariate Gaussian distribution, see Scenario (I), all the tests maintain the nominal 5% level. However, under the multivariate Student t assumption, Scenario (II) of Table 1, the CLRT and LW test suffered serious size distortion while the CZZ test and the proposed test still had reasonable size. This could be because the CLRT and LW test were constructed based on the normality assumption. Thus, we do not report the power of CLRT and LW test in Table 3 under the multivariate Student t distribution setting. We note that there were slight size distortions for the CZZ and the proposed tests when p and n are relatively small, e.g., $n = 100$, $p \in \{20, 50\}$. After all, the test is asymptotic; as p and n tend to infinity, the sizes of the two tests were quite close to the 5% nominal level.

The empirical power of the tests is given in Tables 2–3. When p is small, the CLRT, LW, and CZZ tests perform well. When $n = 200$ and $p = 20$, their power is nearly 1. As p increases, the proposed PTest performs much better than other tests. For example, when $\mathbf{X}_i \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $(n, p) = (100, 400)$, the empirical power of PTest could be as high as 73.8% and 67.8% for the two alternatives, respectively. In contrast, the LW and CZZ tests only have power about 9% for both alternatives. When it comes to the case $(n, p) = (200, 400)$, the merit of the projection test is more obvious, as expected. One reason is that we

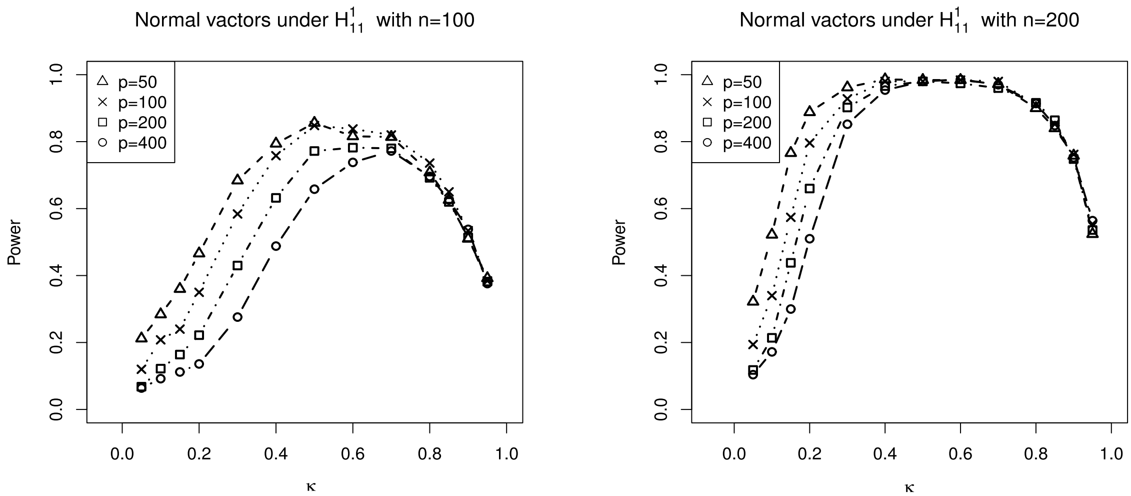

Fig. 1. Power with different sample splitting percentages under $\mathcal{H}^1_{11}$.

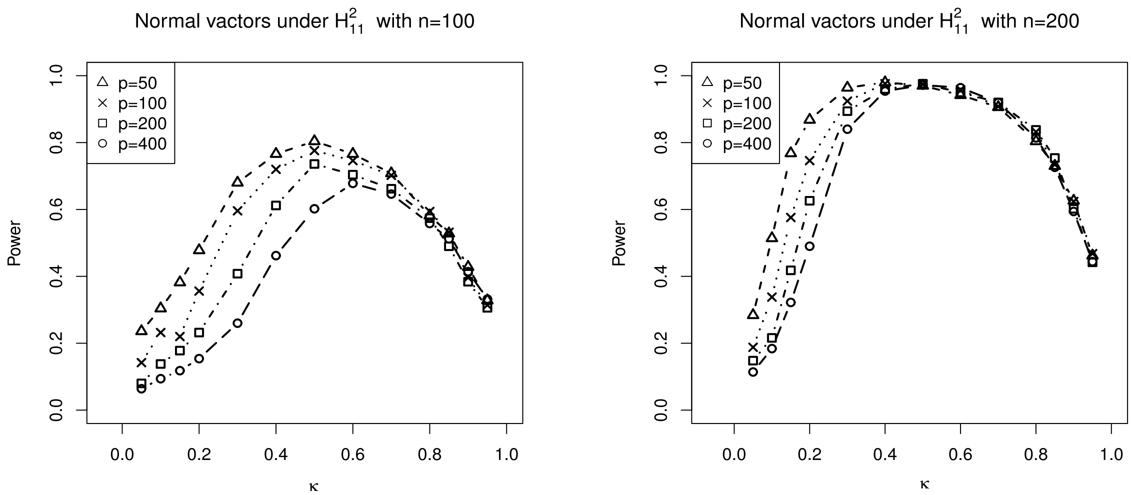

Fig. 2. Power with different sample splitting percentages under $\mathcal{H}^2_{11}$.

use the additional information that the covariance matrix is sparse. Another reason is that the CLRT, LW, and CZZ procedures are designed to test for all the alternatives, hence the inadequacy of these three tests to detect the existence of spikes. It is interesting that the power of the CLRT, LW, and CZZ tests decreases as p increases, whereas the power of the PTest remains stable. This could be explained from Eq. (7): the power functions of PTest are unrelated with the value of p when n and p are large enough.

For the last column of Tables 1–3, we construct a relatively conservative bound in Eqs. (8)–(9) in order to ensure the right size of the tests. We use $J_{CZZ} + J_0$ and $J_{PTest} + J_0$ for comparison – see the definition of J_0 in (10) – and observe that the statistic $J_{CZZ} + J_0$ brings significant power increase compared with the CZZ test, which indicates the efficiency of the power enhancement technique. The statistic $J_{PTest} + J_0$ also leads to better performance than the PTest, but the power increase is relatively small compared with the enhancement from CZZ to $J_{CZZ} + J_0$. This is partly because the PTest performs quite well, so that the potential power enhancement is relatively small. Another reason might be that the PTest holds information from the largest eigenvalue, thus the information between the two terms J_{PTest} and J_0 overlaps to some degree.

4.2. Data example

In this section, we consider a dataset of an experiment for 24 six-month-old Yorkshire gilts. In the experiment, the gilts are divided into four groups equally. The gilts are genetyped by the melanocortin-4 receptor gene, half of them with D298 and the other half with N298. For each genetyped gilt, two diet treatments were assigned randomly; six of them were feeding without

Table 1
Empirical size of the tests for multivariate Gaussian and Student t vectors.

n	p	CLRT	LW	CZZ	PTest	$J_{CZZ}+J_0$	$J_{PTest}+J_0$	$\Pr(\widehat{S}=\emptyset)$
(a): Normal random vectors								
100	20	0.060	0.046	0.048	0.074	0.048	0.074	100%
	50	0.056	0.056	0.068	0.062	0.068	0.062	100%
	80	0.056	0.044	0.050	0.064	0.050	0.064	100%
	100	–	0.066	0.058	0.060	0.058	0.060	100%
	200	–	0.036	0.042	0.066	0.042	0.066	100%
	400	–	0.046	0.056	0.064	0.056	0.064	100%
200	20	0.060	0.052	0.060	0.044	0.060	0.044	100%
	50	0.062	0.064	0.066	0.052	0.066	0.052	100%
	80	0.062	0.078	0.074	0.062	0.074	0.062	100%
	100	0.046	0.062	0.058	0.066	0.058	0.066	100%
	200	–	0.048	0.050	0.064	0.050	0.064	100%
	400	–	0.072	0.074	0.064	0.074	0.064	100%
(b): Multivariate Student t random vectors								
100	20	0.806	0.804	0.096	0.082	0.098	0.084	99.6%
	50	0.568	0.708	0.090	0.086	0.090	0.086	100%
	80	0.422	0.694	0.090	0.060	0.090	0.060	100%
	100	–	0.652	0.054	0.072	0.054	0.072	100%
	200	–	0.684	0.068	0.064	0.068	0.064	100%
	400	–	0.602	0.060	0.070	0.060	0.070	100%
200	20	0.858	0.862	0.094	0.082	0.094	0.082	100%
	50	0.680	0.756	0.092	0.076	0.092	0.076	100%
	80	0.580	0.694	0.070	0.054	0.070	0.054	100%
	100	0.494	0.660	0.068	0.046	0.068	0.046	100%
	200	–	0.650	0.052	0.052	0.052	0.052	100%
	400	–	0.624	0.052	0.060	0.052	0.060	100%

Table 2
Empirical power of the tests for multivariate Gaussian vectors.

	n	p	CLRT	LW	CZZ	PTest	$J_{CZZ}+J_0$	$J_{PTest}+J_0$	$\Pr(\widehat{S}=\emptyset)$
$\mathcal{H}_{11}^1$	100	20	0.644	0.836	0.846	0.868	0.916	0.922	18.2%
		50	0.194	0.446	0.444	0.816	0.720	0.858	35.6%
		80	0.118	0.282	0.284	0.828	0.626	0.860	44.6%
		100	–	0.224	0.216	0.838	0.630	0.856	45.6%
		200	–	0.086	0.094	0.782	0.468	0.818	57.2%
		400	–	0.086	0.088	0.738	0.382	0.778	67.0%
	200	20	0.986	0.996	0.998	0.976	1.000	0.998	0.8%
		50	0.538	0.804	0.806	0.984	0.966	0.994	5.4%
		80	0.286	0.578	0.588	0.976	0.954	0.988	7.2%
		100	0.204	0.446	0.450	0.984	0.922	0.994	9.6%
		200	–	0.194	0.186	0.974	0.870	0.984	15.0%
		400	–	0.136	0.126	0.986	0.776	0.990	24.4%
$\mathcal{H}_{11}^2$	100	20	0.732	0.924	0.918	0.790	0.938	0.882	29.6%
		50	0.222	0.528	0.530	0.766	0.664	0.810	56.6%
		80	0.128	0.322	0.330	0.766	0.526	0.802	62.8%
		100	–	0.246	0.250	0.746	0.446	0.786	67.0%
		200	–	0.106	0.120	0.704	0.290	0.728	79.0%
		400	–	0.092	0.086	0.678	0.210	0.690	86.2%
	200	20	0.996	0.998	0.998	0.964	0.998	0.994	2.4%
		50	0.606	0.900	0.894	0.942	0.956	0.980	13.6%
		80	0.344	0.694	0.704	0.964	0.888	0.982	22.6%
		100	0.230	0.560	0.560	0.958	0.834	0.978	25.4%
		200	–	0.250	0.250	0.950	0.634	0.968	45.0%
		400	–	0.150	0.148	0.964	0.486	0.968	60.6%

restrictions, and the others were fasting. This dataset was previously analyzed in [10,19]. More details about the experiment can be found in [19]. In this dataset, there are 24,123 genes of the liver tissues, and 6176 gene sets with dimensions ranging from 1 to 5158. Many of the gene sets share common genes. The goal of this study was to identify treatment effects on the gene expression structure levels. Here, our interest is testing the correlation structure among the genes for each gene set.

In this paper, we assume that the dimension of the covariates tends to infinity, so in this example we only choose the 324 gene sets with dimension more than 50. Let $\mathcal{S}_1, \ldots, \mathcal{S}_{324}$ be the sets of qualified genes, whose sizes are denoted $p_1, \ldots, p_{324}$, respectively. In order to make the marginal variances of the gene sets homogeneous, we first standardized the data for each gene. Similar to the analysis in [10], let y_{ijkt}^g be a p_g-dimensional vector for the ith treatment, jth block with kth genetype of

Table 3
Empirical power of the tests for multivariate Student t vectors.

	n	p	CZZ	PTest	$\mathrm{J}_{CZZ}+\mathrm{J}_0$	$\mathrm{J}_{PTest}+\mathrm{J}_0$	$\Pr(\widehat{S}=\emptyset)$
H_{11}^1	100	20	0.772	0.792	0.880	0.898	23.8%
		50	0.460	0.854	0.730	0.900	36.8%
		80	0.262	0.796	0.632	0.854	43.8%
		100	0.224	0.796	0.616	0.844	45.4%
		200	0.134	0.738	0.478	0.784	58.6%
		400	0.092	0.716	0.390	0.756	67.0%
	200	20	0.980	0.978	0.992	1.000	2.2%
		50	0.816	0.972	0.968	0.984	6.0%
		80	0.552	0.980	0.942	0.988	10%
		100	0.426	0.982	0.920	0.988	9.6%
		200	0.204	0.986	0.868	0.992	15.4%
		400	0.110	0.980	0.764	0.988	26.8%
H_{11}^2	100	20	0.832	0.710	0.868	0.822	37.2%
		50	0.576	0.754	0.696	0.798	52.2%
		80	0.326	0.740	0.514	0.778	66.0%
		100	0.250	0.698	0.464	0.742	67.6%
		200	0.120	0.714	0.318	0.724	77.0%
		400	0.086	0.656	0.226	0.670	85.6%
	200	20	0.996	0.908	0.996	0.978	6.6%
		50	0.890	0.930	0.954	0.972	16.0%
		80	0.686	0.942	0.870	0.970	23.4%
		100	0.532	0.960	0.832	0.974	27.2%
		200	0.252	0.948	0.650	0.962	44.0%
		400	0.132	0.946	0.486	0.950	58.8%

the tth pig, and the gth gene set. Then we postulate the following factor model:

$$y_{ijkt}^g=\tau^g+\mu_i^g+\beta_j\mathbf{1}_{p_g}+\alpha_k\mathbf{1}_{p_g}+\eta_G\mathbf{1}_{p_g}+\varepsilon_{ijkt}^g. \tag{11}$$

Our aim is to test for the covariance structure of $\eta_G\mathbf{1}_{p_g}+\varepsilon_{ijkt}^g$. It is assumed that $\mathrm{var}(\varepsilon_{ijkt}^g)=\sigma_\varepsilon^2\mathbf{I}_{p_g}$. Thus, the covariance of $\eta_G\mathbf{1}_{p_g}+\varepsilon_{ijkt}^g$ is constructed as

$$\Sigma_g=\sigma_\varepsilon^2\mathbf{I}_{p_g}+\sigma_{\eta_G}^2\mathbf{1}_{p_g}\mathbf{1}_{p_g}^\top,$$

where $\sigma_{\eta_G}^2=\mathrm{var}(\eta_G)$ and Σ_g is the covariance matrix of the gene set $\mathcal{S}_g$. We want to test the identity hypothesis $\mathcal{H}_{0g}:\Sigma_g=\sigma_\varepsilon^2\mathbf{I}_{p_g}$ and $\mathcal{H}_{1g}:U_g^\top\Sigma_gU_g=\sigma_\varepsilon^2\mathbf{I}_{p_g}+\sigma_{\eta_G}^2\mathbf{e}_1\mathbf{e}_1^\top$, where $g\in\{1,\ldots,324\}$ and U_g is an orthogonal matrix defined as $U_g=[\mathbf{u}_1\ \cdots\ \mathbf{u}_{p_g}]$, where $\mathbf{u}_1=\mathbf{1}_{p_g}/\sqrt{p_g}$, $\mathbf{u}_2=(1,-1,0,\ldots,0)^\top/\sqrt{2}$, and similarly $\mathbf{u}_{p_g}=(1/(p_g-1),\ldots,1/(p_g-1),-1)^\top/\sqrt{1+1/(p_g-1)}$. To eliminate the effects of treatment, block and the genotype, the parameters μ_i^g and τ^g should be estimated as in [10]. The parameters β and α are estimated next. Here the "residuals" represent $\eta_G\mathbf{1}_{p_g}+\varepsilon_{ijkt}^g$ via the least square regression.

The histogram of the p-values of the sphericity tests is displayed in Fig. 3. If the FDR is controlled at 0.05, there are about 174 gene sets that are significant for the sphericity test, which indicates that the factor model listed in (11) is not sufficient to describe the gene-wise dependence, and further studies should be investigated.

Acknowledgments

The research was supported by the Project funded by China Postdoctoral Science Foundation (2017M620827, 2018T110118), National Natural Science Foundation of China (NNSFC) grants (11471223, 11231010, 11071022), and Capacity Building for Sci-Tech Innovation-Fundamental Scientific Research Funds, China (No. 025185305000/204) and Youth Innovative Research Team of Capital Normal University, China. The authors are grateful to the Editor-in-Chief, the Associate Editor and referees for comments, suggestions, and detailed advice that led to the present, vastly improved paper. As usual, all remaining errors are our own.

Appendix. Proofs

This appendix includes the proofs of the paper's results. In what follows, $c_1,c_2,\ldots$ stand for some finite constants.

Proof of Proposition 1. From Theorem 13.5.1 in [1],

$$\sqrt{n}\left\{\widehat{\lambda}_{\max}(S_Y)/\lambda_{\max}(\mathbf{\Sigma_P})-1\right\}\rightsquigarrow\mathcal{N}(0,2).$$

Therefore, the power of $T_{n,p}^{\mathbf{P}}$ increases with $\lambda_{\max}(\mathbf{\Sigma_P})$. Note that $\lambda_{\max}(\mathbf{\Sigma_P})\le\lambda_{\max}(\mathbf{\Sigma})$, and equality follows if $\mathbf{P}=\mathbf{v}_1$. This implies that $s=1$ with $\mathbf{P}=\mathbf{v}_1$ is the best choice to achieve optimal power. □

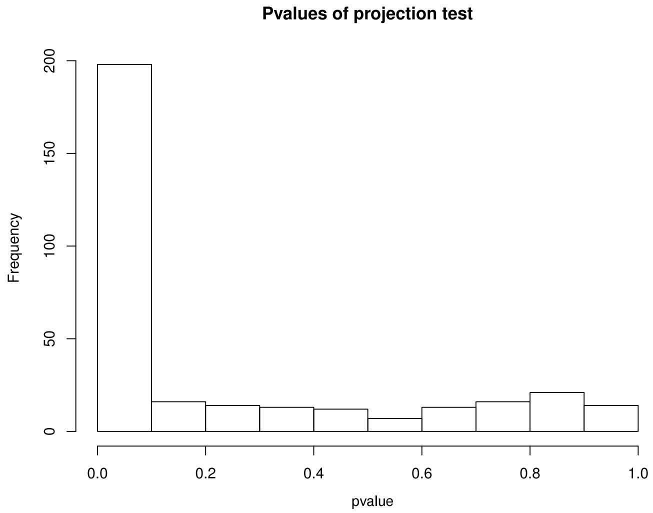

Fig. 3. Histograms of p-values for the projection tests.

Lemma 1. *Under Conditions (C1)–(C4), there exists* $\mathbf{v}_1 \in \mathcal{R}_1$ *such that*

$$\sqrt{n_2}\,(\widehat{\mathbf{v}}_{1(1)}^\top \mathbf{S}_{\mathbf{X}}^{(2)} \widehat{\mathbf{v}}_{1(1)} - \widehat{\mathbf{v}}_{1(1)}^\top \boldsymbol{\Sigma} \widehat{\mathbf{v}}_{1(1)}) = \sqrt{n_2}\,(\mathbf{v}_1^\top \mathbf{S}_{\mathbf{X}}^{(2)} \mathbf{v}_1 - \mathbf{v}_1^\top \boldsymbol{\Sigma} \mathbf{v}_1) + o_P(1). \tag{A.1}$$

Proof. Assume that $\mathcal{R}_1$ is the column space of $\boldsymbol{Z}_1 = (\mathbf{v}_1, \ldots, \mathbf{v}_s)$, and denote $\mathbf{v}_1 = \boldsymbol{Z}_1 \boldsymbol{Z}_1^\top \widehat{\mathbf{v}}_{1(1)}$. Decompose the term on the left-hand side of Eq. (A.1) as

$$\sqrt{n_2}(\widehat{\mathbf{v}}_{1(1)}^\top \mathbf{S}_{\mathbf{X}}^{(2)} \widehat{\mathbf{v}}_{1(1)} - \widehat{\mathbf{v}}_{1(1)}^\top \boldsymbol{\Sigma} \widehat{\mathbf{v}}_{1(1)}) = \sqrt{n_2}(\mathbf{v}_1^\top \mathbf{S}_{\mathbf{X}}^{(2)} \mathbf{v}_1 - \mathbf{v}_1^\top \boldsymbol{\Sigma} \mathbf{v}_1) + 2W_1 + W_2,$$

where

$$\begin{aligned} W_1 &= \sqrt{n_2}\,\{\mathbf{v}_1^\top \mathbf{S}_{\mathbf{X}}^{(2)} (\widehat{\mathbf{v}}_{1(1)} - \mathbf{v}_1) - \mathbf{v}_1^\top \boldsymbol{\Sigma} (\widehat{\mathbf{v}}_{1(1)} - \mathbf{v}_1)\}, \\ W_2 &= \sqrt{n_2}\,\{(\widehat{\mathbf{v}}_{1(1)} - \mathbf{v}_1)^\top \mathbf{S}_{\mathbf{X}}^{(2)} (\widehat{\mathbf{v}}_{1(1)} - \mathbf{v}_1) - (\widehat{\mathbf{v}}_{1(1)} - \mathbf{v}_1)^\top \boldsymbol{\Sigma} (\widehat{\mathbf{v}}_{1(1)} - \mathbf{v}_1)\}. \end{aligned}$$

Then it is sufficient to prove $W_1 = o_P(1)$ and $W_2 = o_P(1)$. For any $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in \mathbb{R}^p$,

$$\begin{aligned} \mathrm{var}(\boldsymbol{\eta}_1^\top \mathbf{S}_{\mathbf{X}}^{(2)} \boldsymbol{\eta}_2) &= \frac{1}{(n_2-1)^2} \left[n_2 \mathrm{E}\{\boldsymbol{\eta}_1^\top (\mathbf{X}_{21} - \boldsymbol{\mu})(\mathbf{X}_{21} - \boldsymbol{\mu})^\top \boldsymbol{\eta}_2\}^2 + n_2(n_2-1)(\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma}^\top \boldsymbol{\eta}_2)^2 \right] \\ &\quad - \frac{2n_2^2}{(n_2-1)^2} \mathrm{E}\{\boldsymbol{\eta}_1^\top (\mathbf{X}_{21} - \boldsymbol{\mu})(\mathbf{X}_{21} - \boldsymbol{\mu})^\top \boldsymbol{\eta}_2 \boldsymbol{\eta}_1^\top (\overline{\mathbf{X}}_2 - \boldsymbol{\mu})(\overline{\mathbf{X}}_2 - \boldsymbol{\mu})^\top \boldsymbol{\eta}_2\} \\ &\quad + \frac{n_2^2}{(n_2-1)^2} \mathrm{E}\{\boldsymbol{\eta}_1^\top (\overline{\mathbf{X}}_2 - \boldsymbol{\mu})(\overline{\mathbf{X}}_2 - \boldsymbol{\mu})^\top \boldsymbol{\eta}_2\}^2 - (\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma} \boldsymbol{\eta}_2)^2 \\ &= \frac{\mathrm{E}(R_1^4)}{p(p+2)} \left\{ \frac{2}{n_2} (\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma} \boldsymbol{\eta}_2)^2 + \frac{1}{n_2} (\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma} \boldsymbol{\eta}_1)(\boldsymbol{\eta}_2^\top \boldsymbol{\Sigma} \boldsymbol{\eta}_2) \right\} - \frac{(n_2-2)}{n_2(n_2-1)} (\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma} \boldsymbol{\eta}_2)^2 \\ &\quad + \frac{1}{n_2(n_2-1)} (\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma} \boldsymbol{\eta}_1)(\boldsymbol{\eta}_2^\top \boldsymbol{\Sigma} \boldsymbol{\eta}_2) \\ &\le c_1 n_2^{-1} \left\{ (\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma} \boldsymbol{\eta}_2)^2 + (\boldsymbol{\eta}_1^\top \boldsymbol{\Sigma} \boldsymbol{\eta}_1)(\boldsymbol{\eta}_2^\top \boldsymbol{\Sigma} \boldsymbol{\eta}_2) \right\}, \end{aligned}$$

where the last equation holds by the Cauchy–Schwarz inequality and Condition (C1). Combining the fact that $\mathbf{v}_1^\top \boldsymbol{\Sigma} (\widehat{\mathbf{v}}_{1(1)} - \mathbf{v}_1) = -\lambda_{\max}(\boldsymbol{\Sigma}) \|\widehat{\mathbf{v}}_{1(1)} - \mathbf{v}_1\|^2/2$, we find that $\mathrm{var}(W_1 | \widehat{\mathbf{v}}_{1(1)}) \le c_2 \lambda_{\max}(\boldsymbol{\Sigma}) \|\widehat{\mathbf{v}}_{1(1)} - \mathbf{v}_1\|$ and $\mathrm{var}(W_2 | \widehat{\mathbf{v}}_{1(1)}) \le c_3 \lambda_{\max}(\boldsymbol{\Sigma}) \|\widehat{\mathbf{v}}_{1(1)} - \mathbf{v}_1\|^2$. Therefore, using Markov's inequality, we conclude that $W_1 = o_P(1)$ and $W_2 = o_P(1)$ are true under Condition (C4). □

Lemma 2. *Under Conditions (C1)–(C3),*

$$\sqrt{n_2}\,(\mathbf{v}_1^\top \mathbf{S}_{\mathbf{X}}^{(2)} \mathbf{v}_1 - \mathbf{v}_1^\top \boldsymbol{\Sigma} \mathbf{v}_1) = \frac{1}{\sqrt{n_2}} \sum_{i=1}^{n_2} \left[\{\mathbf{v}_1^\top (\mathbf{X}_{2i} - \boldsymbol{\mu})\}^2 - \mathbf{v}_1^\top \boldsymbol{\Sigma} \mathbf{v}_1 \right] + o_P(1). \tag{A.2}$$

Proof. Write $W^{(2)} = \sqrt{n_2}(\mathbf{v}_1^\top \mathbf{S}_\mathbf{X}^{(2)}\mathbf{v}_1 - \mathbf{v}_1^\top \boldsymbol{\Sigma}\mathbf{v}_1)$, and write

$$W^{(2)} - \frac{1}{\sqrt{n_2}}\sum_{i=1}^{n_2}[\{\mathbf{v}_1^\top(\mathbf{X}_{2i}-\boldsymbol{\mu})\}^2 - \mathbf{v}_1^\top\boldsymbol{\Sigma}\mathbf{v}_1] = W_3 + W_4,$$

where

$$W_3 = \frac{1}{\sqrt{n_2}(n_2-1)}\sum_{i=1}^{n_2}[\{\mathbf{v}_1^\top(\mathbf{X}_{2i}-\boldsymbol{\mu})\}^2 - \mathbf{v}_1^\top\boldsymbol{\Sigma}\mathbf{v}_1], \quad W_4 = -\frac{n_2^{3/2}}{(n_2-1)}[\{\mathbf{v}_1^\top(\bar{\mathbf{X}}_2-\boldsymbol{\mu})\}^2 - \mathbf{v}_1^\top\boldsymbol{\Sigma}\mathbf{v}_1/n].$$

Eq. (A.2) holds because $\mathrm{E}(|W_3|^2) = O\{\lambda_{\max}^2(\boldsymbol{\Sigma})n_2^{-2}\}$ and $\mathrm{E}(|W_4|^2) = O\{\lambda_{\max}^2(\boldsymbol{\Sigma})n_2^{-1}\}$, both of which go to zero under Condition (C3). □

Lemma 3. *Under Conditions (C1) and (C3), one has, as $n, p \to \infty$,*

$$\sqrt{n}\left\{\frac{1}{p}\mathrm{tr}(\mathbf{S}_\mathbf{X}) - \frac{1}{p}\mathrm{tr}(\boldsymbol{\Sigma})\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{p}\left\{(\mathbf{X}_i-\boldsymbol{\mu})^\top(\mathbf{X}_i-\boldsymbol{\mu}) - \mathrm{tr}(\boldsymbol{\Sigma})\right\} + o_P(1). \tag{A.3}$$

Proof. Write

$$\sqrt{n}\left\{\frac{1}{p}\mathrm{tr}(\mathbf{S}_\mathbf{X}) - \frac{1}{p}\mathrm{tr}(\boldsymbol{\Sigma})\right\} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{p}\left\{(\mathbf{X}_i-\boldsymbol{\mu})^\top(\mathbf{X}_i-\boldsymbol{\mu}) - \mathrm{tr}(\boldsymbol{\Sigma})\right\} = W_5 + W_6,$$

where

$$W_5 = \frac{\sqrt{n}}{n(n-1)p}\sum_{i=1}^{n}\left\{(\mathbf{X}_i-\boldsymbol{\mu})^\top(\mathbf{X}_i-\boldsymbol{\mu}) - \mathrm{tr}(\boldsymbol{\Sigma})\right\}, \quad W_6 = -\frac{n\sqrt{n}}{(n-1)p}\left\{(\bar{\mathbf{X}}-\boldsymbol{\mu})^\top(\bar{\mathbf{X}}-\boldsymbol{\mu}) - \frac{1}{n}\mathrm{tr}(\boldsymbol{\Sigma})\right\}.$$

Under Condition (C1), we have $\mathrm{E}|W_5|^2 = O\{\mathrm{tr}(\boldsymbol{\Sigma}^2)/(n^2p^2)\}$ and $\mathrm{E}(|W_6|^2) = O\{\mathrm{tr}(\boldsymbol{\Sigma}^2)/(np^2)\}$, and they go to zero under Condition (C3). Thus, (A.3) is true and the proof is complete. □

Proof of Theorem 1. Since $\hat{\mathbf{v}}_1$ is independent of $\mathcal{S}_2$, then under the null hypothesis we have

$$\sqrt{n_2}(\widehat{\mathbf{v}}_{1(1)}^\top\mathbf{S}_\mathbf{X}^{(2)}\widehat{\mathbf{v}}_{1(1)} - \widehat{\mathbf{v}}_{1(1)}^\top\boldsymbol{\Sigma}\widehat{\mathbf{v}}_{1(1)}) = \frac{1}{\sqrt{n_2}}\sum_{i=1}^{n_2}\left\{(\widehat{\mathbf{v}}_{1(1)}^\top(\mathbf{X}_{2i}-\boldsymbol{\mu}))^2 - \widehat{\mathbf{v}}_{1(1)}^\top\boldsymbol{\Sigma}\widehat{\mathbf{v}}_{1(1)}\right\} + o_P(1).$$

The derivation of the last equation is quite similar to the proof of Lemma 2, so we omit it. Then the asymptotic normality of $T_{n_2,p}^{\mathbf{P}}$ follows because the Lindeberg condition is immediately verified under Condition (C1). It follows from Lemma 3 that

$$\sqrt{n}\left\{\frac{1}{p}\mathrm{tr}(\mathbf{S}_\mathbf{X}) - \sigma^2\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{p}\left\{(\mathbf{X}_i-\boldsymbol{\mu})^\top(\mathbf{X}_i-\boldsymbol{\mu}) - \mathrm{tr}(\boldsymbol{\Sigma})\right\} + o_P(1),$$

under the null hypothesis. Invoking Slutsky's Lemma, we get

$$\sqrt{n_2}\,W_{n_2,p}^{\mathbf{P}} = \frac{1}{\sqrt{n_2}\sigma^2}\sum_{i=1}^{n_2}\left[\{\widehat{\mathbf{v}}_{1(1)}^\top(\mathbf{X}_{2i}-\boldsymbol{\mu})\}^2 - \widehat{\mathbf{v}}_{1(1)}^\top\boldsymbol{\Sigma}\widehat{\mathbf{v}}_{1(1)}\right] - \frac{1}{\sqrt{n}\sigma^2}\sum_{i=1}^{n}\frac{1}{p}\left\{(\mathbf{X}_i-\boldsymbol{\mu})^\top(\mathbf{X}_i-\boldsymbol{\mu}) - \mathrm{tr}(\boldsymbol{\Sigma})\right\} + o_P(1). \tag{A.4}$$

It is simple to deduce that

$$\mathrm{var}\left\{\frac{1}{p}(\mathbf{X}_1-\boldsymbol{\mu})^\top(\mathbf{X}_1-\boldsymbol{\mu})\right\} = \frac{\mathrm{E}(R_1^4)}{p(p+2)}\left\{\frac{2}{p^2}\mathrm{tr}(\boldsymbol{\Sigma}^2)\right\} + \left\{\frac{\mathrm{E}(R_1^4)}{p(p+2)} - 1\right\}\left\{\frac{1}{p^2}\mathrm{tr}^2(\boldsymbol{\Sigma})\right\} \le c_2\frac{\mathrm{tr}(\boldsymbol{\Sigma}^2)}{p^2} \to 0. \tag{A.5}$$

Combining (A.4)–(A.5), the proof is completed by invoking the Central Limit Theorem. □

Proof of Theorem 2. Using Lemmas 1–2, we have

$$\sqrt{\frac{n_2}{2}}\left\{\frac{T_{n_2,p}^{\mathbf{P}} - (\widehat{\mathbf{v}}_{1(1)}^\top\boldsymbol{\Sigma}\widehat{\mathbf{v}}_{1(1)} - 1)}{\lambda_{\max}(\boldsymbol{\Sigma})}\right\} = \frac{1}{\sqrt{2n_2}}\sum_{i=1}^{n_2}\frac{\{\mathbf{v}_1^\top(\mathbf{X}_{2i}-\boldsymbol{\mu})\}^2 - \mathbf{v}_1^\top\boldsymbol{\Sigma}\mathbf{v}_1}{\lambda_{\max}(\boldsymbol{\Sigma})} + o_P(1),$$

then the proof is completed using the fact that

$$\mathrm{var}[\{\mathbf{v}_1^\top(\mathbf{X}_1-\boldsymbol{\mu})\}^2 - \mathbf{v}_1^\top\boldsymbol{\Sigma}\mathbf{v}_1] = \frac{3\mathrm{E}(R_1^4)}{p(p+2)}\lambda_{\max}^2(\boldsymbol{\Sigma}) - \lambda_{\max}^2(\boldsymbol{\Sigma}) = \{2 + O(p^{-1})\}\lambda_{\max}^2(\boldsymbol{\Sigma}),$$

where the last equation follows by Condition (C1). Eq. (5) follows by the Central Limit Theorem. In the following, we will prove the asymptotic distribution of the $W_{n,p}^{\mathbf{P}}$. Using Lemmas 1–3 and Slutsky's Lemma, we find

$$\sqrt{n_2}\left[W^{\mathbf{P}}_{n_2,p}-\left\{\frac{\widehat{\mathbf{v}}_{1(1)}^\top\boldsymbol{\Sigma}\widehat{\mathbf{v}}_{1(1)}}{\mathrm{tr}(\boldsymbol{\Sigma})/p}-1\right\}\right]=\sqrt{n_2}\left\{\frac{\widehat{\mathbf{v}}_{1(1)}^\top\mathbf{S}_{\mathbf{X}}^{(2)}\widehat{\mathbf{v}}_{1(1)}}{\mathrm{tr}(\mathbf{S}_{\mathbf{X}})/p}-\frac{\widehat{\mathbf{v}}_{1(1)}^\top\boldsymbol{\Sigma}\widehat{\mathbf{v}}_{1(1)}}{\mathrm{tr}(\boldsymbol{\Sigma})/p}\right\}$$

and the right-hand side can be rewritten as follows:

$$\begin{aligned}
&\frac{1}{\sqrt{n_2}\,\mathrm{tr}(\boldsymbol{\Sigma})/p}\sum_{i=1}^{n_2}[\{\mathbf{v}_1^\top(\mathbf{X}_{2i}-\boldsymbol{\mu})\}^2-\mathbf{v}_1^\top\boldsymbol{\Sigma}\mathbf{v}_1]-\frac{1}{\sqrt{n}\,\mathrm{tr}(\boldsymbol{\Sigma})/p}\left\{\frac{\widehat{\mathbf{v}}_{1(1)}^\top\boldsymbol{\Sigma}\widehat{\mathbf{v}}_{1(1)}}{\mathrm{tr}(\boldsymbol{\Sigma})/p}\right\}\\
&\quad\times\sum_{i=1}^{n}\frac{1}{p}\{(\mathbf{X}_i-\boldsymbol{\mu})^\top(\mathbf{X}_i-\boldsymbol{\mu})-\mathrm{tr}(\boldsymbol{\Sigma})\}+o_P(1)\\
&=\frac{1}{\sqrt{n_2}\,\sigma^2}\sum_{i=1}^{n_2}[\{\mathbf{v}_1^\top(\mathbf{X}_{2i}-\boldsymbol{\mu})\}^2-\mathbf{v}_1^\top\boldsymbol{\Sigma}\mathbf{v}_1]-\frac{1}{\sqrt{n}\sigma^2}\left\{\frac{\widehat{\mathbf{v}}_{1(1)}^\top\boldsymbol{\Sigma}\widehat{\mathbf{v}}_{1(1)}}{\sigma^2}\right\}\sum_{i=1}^{n}\frac{1}{p}\{(\mathbf{X}_i-\boldsymbol{\mu})^\top(\mathbf{X}_i-\boldsymbol{\mu})-\mathrm{tr}(\boldsymbol{\Sigma})\}+o_P(1)\\
&=\frac{1}{\sqrt{n_2}\,\sigma^2}\sum_{i=1}^{n_2}[\{\mathbf{v}_1^\top(\mathbf{X}_{2i}-\boldsymbol{\mu})\}^2-\mathbf{v}_1^\top\boldsymbol{\Sigma}\mathbf{v}_1]+o_P(1).
\end{aligned}$$

The last equation holds under Condition (C3) because $\mathrm{var}\{(\mathbf{X}_i-\boldsymbol{\mu})^\top(\mathbf{X}_i-\boldsymbol{\mu})\}=O\{\mathrm{tr}(\boldsymbol{\Sigma}^2)\}$. Note that $\mathrm{var}[\{\mathbf{v}_1^\top(\mathbf{X}_1-\boldsymbol{\mu})\}^2]=2\lambda^2_{\max}(\boldsymbol{\Sigma})$. The asymptotic property in (6) is obtained using the Central Limit Theorem. In view of the definitions of $\boldsymbol{\Sigma}_1$ and ω_1, the proof is complete. □

Proof of Theorem 3. Define event

$$\mathcal{A}=\left\{\max_{i\in\{1,\ldots,p\}}|\widehat{\lambda}_i-\lambda_i|<\delta_{n,p}/\sqrt{n}\right\}.$$

For any $i\in S(\lambda)$, one has $\lambda_i-1>2\delta_{n,p}/\sqrt{n}$ from the definition of $S(\lambda)$. Then under event $\mathcal{A}$,

$$\widehat{\lambda}_i-1\geq|\lambda_i-1|-|\widehat{\lambda}_i-\lambda_i|>2\delta_{n,p}/\sqrt{n}-\delta_{n,p}/\sqrt{n}=\delta_{n,p}/\sqrt{n}.$$

Then $i\in\widehat{S}$, and hence $S(\lambda)\subseteq\widehat{S}$. In fact, we have proved this statement on the event $\mathcal{A}$ uniformly for $\boldsymbol{\Sigma}\in\mathcal{U}^*$:

$$\inf_{\Sigma\in\mathcal{U}^*}\Pr(S(\lambda)\subseteq\widehat{S}\mid\lambda)\to 1.$$

Under $\mathcal{H}_0$,

$$\begin{aligned}
\Pr(J_0=0\mid\mathcal{H}_0)=\Pr(\widehat{S}=\emptyset\mid\mathcal{H}_0)&=\Pr\left(\max_{i\in\{1,\ldots,p\}}\widehat{\lambda}_i-1<\delta_{n,p}/\sqrt{n}\,\middle|\,\mathcal{H}_0\right)\\
&\geq\Pr\left(\max_{i\in\{1,\ldots,p\}}|\widehat{\lambda}_i-1|<\delta_{n,p}/\sqrt{n}\,\middle|\,\mathcal{H}_0\right)\longrightarrow 1.
\end{aligned}$$

In addition, by $\inf_{\Sigma\in\mathcal{U}^*}\Pr(S(\lambda)\subseteq\widehat{S}\mid\lambda)\to 1$,

$$\begin{aligned}
\sup_{\Sigma\in\mathcal{U}^*}\Pr\{J_0\leq\sqrt{n}\mid S(\lambda)\neq\emptyset\}&\leq\sup_{\Sigma\in\mathcal{U}^*}\Pr\{J_0\leq\sqrt{n},\widehat{S}\neq\emptyset\mid S(\lambda)\neq\emptyset\}+\sup_{\lambda\in\mathcal{U}^*}\Pr\{\widehat{S}=\emptyset\mid S(\lambda)\neq\emptyset\}\\
&\leq\sup_{\Sigma\in\mathcal{U}^*}\Pr\left\{\sqrt{n}\sum_{i\in\widehat{S}}\delta_{n,p}\leq\sqrt{n},\widehat{S}\neq\emptyset\,\middle|\,S(\lambda)\neq\emptyset\right\}+o(1)\\
&\leq\sup_{\Sigma\in\mathcal{U}^*}\Pr\{\sqrt{n}\delta_{n,p}\leq\sqrt{n}\mid S(\lambda)\neq\emptyset\}+o(1)\to 0.
\end{aligned}$$

This completes the proof of Theorem 3. □

Proof of Theorem 4. Because $\Pr(J_0=0\mid\mathcal{H}_0)\to 1$ in Theorem 3, $J=J_1+J_0\rightsquigarrow F$ as $n\to\infty$. Thus one just needs to show that $\inf_{\lambda\in\Lambda_s}\Pr(J>F_\alpha\mid\lambda)\to 1$. From the definition of $\widehat{S}$ and J_0, one has $\{J_0<\sqrt{n}\delta_{n,p}\}=\{\widehat{S}=\emptyset\}$. Because $\inf_{\Sigma\in\mathcal{U}^*}\Pr\{S(\lambda)\subseteq\widehat{S}\mid\lambda\}\to 1$ and $\Lambda_s=\{\lambda:S(\lambda)\neq\emptyset\}$, one deduces that

$$\sup_{\lambda\in\Lambda_s}\Pr(J_0<\sqrt{n}\delta_{n,p}\mid\lambda)=\sup_{\lambda\in\Lambda_s}\Pr(\widehat{S}=\emptyset\mid\lambda)\leq\sup_{\{\lambda:S(\lambda)\neq\emptyset\}}\Pr\{\widehat{S}=\emptyset,S(\lambda)\subseteq\widehat{S}\mid\lambda\}+o(1).$$

Note the fact that the first term on the right-hand side of the last inequality goes to zero, which implies $\inf_{\lambda\in\Lambda_s}\Pr(J_0\geq\sqrt{n}\,\delta_{n,p}\mid\lambda)\to 1$. Therefore, as $\delta_{n,p}\to\infty$,

$$\inf_{\lambda\in\Lambda_s}\Pr(J>F_\alpha\mid\lambda)\geq\inf_{\lambda\in\Lambda_s}\Pr(\sqrt{n}\delta_{n,p}+J_1>F_\alpha\mid\lambda)\geq\inf_{\lambda\in\Lambda_s}\Pr(c\sqrt{n}+J_1>F_\alpha\mid\lambda)\to 1.$$

As $\delta_{n,p}\to\infty$, there exists a constant c satisfying the last inequality. This completes the proof. □

References

[1] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, third ed., Wiley, New York, 2003.
[2] Z. Bai, D. Jiang, J.-F. Yao, S. Zheng, Corrections to LRT on large-dimensional covariance matrix by RMT, Ann. Statist. 37 (2009) 3822–3840.
[3] Z. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem, Statist. Sinica 6 (1996) 311–329.
[4] Q. Berthet, P. Rigollet, Optimal detection of sparse principal components in high dimension, Ann. Statist. 41 (2013) 1780–1815.
[5] P. Bianchi, M. Debbah, M. Maida, J. Najim, Performance of statistical tests for single source detection using random matrix theory, IEEE Trans. Inform. Theory 57 (2011) 2400–2419.
[6] P.J. Bickel, E. Levina, Covariance regularization by thresholding, Ann. Statist. 36 (2008) 2577–2604.
[7] T. Cai, W. Liu, Adaptive thresholding for sparse covariance matrix estimation, J. Amer. Statist. Assoc. 106 (2011) 672–684.
[8] T. Cai, Z. Ma, Optimal hypothesis testing for high dimensional covariance matrices, Bernoulli 19 (2013) 2359–2388.
[9] T. Cai, Z. Ma, Y. Wu, Sparse PCA: Optimal rates and adaptive estimation, Ann. Statist. 41 (2013) 3074–3110.
[10] S.X. Chen, L.X. Zhang, P.S. Zhong, Tests for high-dimensional covariance matrices, J. Amer. Statist. Assoc. 105 (2010) 810–819.
[11] J.Q. Fan, Y. Liao, J.W. Yao, Power enhancement in high-dimensional cross-sectional tests, Econometrica 83 (2015) 1497–1541.
[12] K.T. Fang, S. Kotz, K.W. Ng, Symmetric Multivariate and Related Distributions, Chapman & Hall, London, 1990.
[13] T. Jiang, F. Yang, Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions, Ann. Statist. 41 (2013) 2029–2074.
[14] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2001) 295–327.
[15] I.M. Johnstone, B. Nadler, Roy's largest root test under rank-one alternatives, Biometrika 104 (2017) 181–193.
[16] S. Kritchman, B. Nadler, Determining the number of components in a factor model from limited noisy data, Chemometr. Intell. Lab. Syst. 94 (2008) 19–32.
[17] O. Ledoit, M. Wolf, Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size, Ann. Statist. 30 (2002) 1081–1102.
[18] J. Li, S.X. Chen, Two sample tests for high-dimensional covariance matrices, Ann. Statist. 40 (2012) 908–940.
[19] S. Lkhagvadorj, L. Qu, W. Cai, O.P. Couture, C.R. Barb, G.J. Hausman, D. Nettleton, L.L. Anderson, J.C. Dekkers, C.K. Tuggle, Microarray gene expression profiles of fasting induced changes in liver and adipose tissues of pigs expressing the melanocortin-4 receptor D298N variant, Physiol. Genomics 38 (2009) 98–111.
[20] N. Meinshausen, L. Meier, P. Bühlmann, *p*-values for high-dimensional regression, J. Amer. Statist. Assoc. 104 (2009) 1671–1681.
[21] A. Onatski, Testing hypotheses about the number of factors in large factor models, Econometrica 77 (2009) 1447–1479.
[22] A. Onatski, Determining the number of factors from empirical distribution of eigenvalues, Rev. Econom. Statist. 92 (2010) 1004–1016.
[23] A. Onatski, M.J. Moreira, M. Hallin, Asymptotic power of sphericity tests for high-dimensional data, Ann. Statist. 41 (2013) 1204–1231.
[24] L.H. Peng, S.X. Chen, W. Zhou, More powerful tests for sparse high-dimensional covariances matrices, J. Multivariate Anal. 149 (2016) 124–143.
[25] L. Wasserman, K. Roeder, High dimensional variable selection, Ann. Statist. 37 (2009) 2178–2201.
[26] R. Zhang, L. Peng, R. Wang, Tests for covariance matrix with fixed or divergent dimension, Ann. Statist. 41 (2013) 2075–2096.