Science in China Series A: Mathematics
Jun., 2009, Vol. 52, No. 6, 1285–1304
www.scichina.com math.scichina.com
www.springerlink.com

On the *k*-sample Behrens-Fisher problem for high-dimensional data

Dedicated to Professor Zhidong Bai on the occasion of his 65th birthday

ZHANG JinTing† & XU JinFeng

Department of Statistics and Applied Probability, National University of Singapore, 3 Science Drive 2, 117546 Singapore
(email: stazjt@nus.edu.sg, staxj@nus.edu.sg)

Abstract For several decades, much attention has been paid to the two-sample Behrens-Fisher (BF) problem which tests the equality of the means or mean vectors of two normal populations with unequal variance/covariance structures. Little work, however, has been done for the k-sample BF problem for high dimensional data which tests the equality of the mean vectors of several high-dimensional normal populations with unequal covariance structures. In this paper we study this challenging problem via extending the famous Scheffe's transformation method, which reduces the k-sample BF problem to a one-sample problem. The induced one-sample problem can be easily tested by the classical Hotelling's T^2 test when the size of the resulting sample is very large relative to its dimensionality. For high dimensional data, however, the dimensionality of the resulting sample is often very large, and even much larger than its sample size, which makes the classical Hotelling's T^2 test not powerful or not even well defined. To overcome this difficulty, we propose and study an L^2-norm based test. The asymptotic powers of the proposed L^2-norm based test and Hotelling's T^2 test are derived and theoretically compared. Methods for implementing the L^2-norm based test are described. Simulation studies are conducted to compare the L^2-norm based test and Hotelling's T^2 test when the latter can be well defined, and to compare the proposed implementation methods for the L^2-norm based test otherwise. The methodologies are motivated and illustrated by a real data example.

Keywords: χ^2**-approximation,** χ^2**-type mixtures, high-dimensional data analysis, Hotelling's** T^2 **test,** k**-sample test,** L^2**-norm based test**

MSC(2000): Primary 62H15, Secondary 62E17, 62E20

1 Introduction

The problem of testing the equality of the means or mean vectors of two univariate/multivariate normal populations with unequal variance/covariances is referred to as the Behrens-Fisher (BF) problem. It is very challenging since it can be proven that there is no exact solution satisfying the classical criteria for good tests. This problem was very interesting and has drawn a lot of attention for several decades. For the univariate BF problem, various approaches have been proposed. Behrens[1] was the first to address this problem whose solution was justified by

Received October 29, 2008; accepted April 22, 2009
DOI: 10.1007/s11425-009-0091-x
† Corresponding author
The work was supported by the National University of Singapore Academic Research Grant (Grant No. R-155-000-085-112)

Citation: Zhang J T, Xu J F. On the *k*-sample Behrens-Fisher problem for high-dimensional data. Sci China Ser A, 2009, 52(6): 1285–1304, DOI: 10.1007/s11425-009-0091-x

Fisher[2] using his fiducial theory of inference. Scheffe[3] proposed a transformation method which reduces the original two-sample problem to a one-sample problem. This one-sample problem can be tested via the usual t-test. The transformation method is simple and the distribution of the test statistic is exact. Scheffe[3] also showed that this method gives the shortest confidence interval for the difference between the two means. Welch[4] proposed an approximate degrees of freedom method based on Student's t-distribution. Scheffe[5] and Lee and Gurland[6] provided reviews of many other methods. In the recent decade, several new methods have been proposed for the BF problem. For example, the generalized P-value method was studied by Weerahandi[7, 8], Tang and Tsui[9] and references therein; the Bayes method was studied by Ghosh and Kim[10] and references therein, and the empirical likelihood method was studied by Dong[11] among others.

For the multivariate BF problem, Bennett[12] extended Scheffe's transformation method[3]. Anderson[13, p. 178] pointed out that the advantage of the transformation method is that the sample mean vector difference, which is used in the test statistic, is most relevant to the population mean vector difference; the sacrifice of observations in estimating a covariance matrix is not so important, especially when the sample sizes for the two multivariate samples are about the same. Welch's approximate degrees of freedom method[4] was also extended to the multivariate BF problem by several authors, including James[14], Yao[15], Johansen[16], and Nel and Van der Merwe[17] among others. The type-I errors of Yao's and James' test were compared by Algina and Tang[18]. Recently, Krishnamoorthy and Yu[19] examined the affine invariant properties of these tests and pointed out that the Nel and Van der Merwe's[17] test was not properly defined. They then proposed the modified Nel and van der Merwe's test. Krishnamoorthy and Xia[20] considered selecting tests for the multivariate BF problem.

In this paper, we are interested in the k-sample high-dimensional BF problem, which tests the equality of the mean vectors of several high-dimensional normal populations with unequal covariance structures. This problem was motivated by the dog potassium data considered by Grizzle and Allen[21] and re-analyzed by Wang[22]. The data are coronary sinus potassium concentrations measured on each of 36 dogs. The measurements on each dog were taken every 2 mins from 1 to 13 mins after occlusion. These 36 dogs were divided into 4 treatment groups with $9, 10, 8, 9$ dogs respectively. We are interested in testing if the treatment effects are different from each other. Figure 1 displays the pointwise group means (solid with o) with 95% pointwise confidence intervals (dashed with $\times$) for the dog potassium data. From Figure 1, it can be seen that the pointwise confidence intervals have different lengths for different groups, which suggests that the four group dog measurements may have different covariance structures. Therefore, this is a k-sample high-dimensional BF problem with $k = 4$ since the dimensionality of the data is large compared with the sample sizes of the four groups. The challenges of the k-sample high-dimensional BF problem include: (1) it involves several normal populations and hence is more complicated than the usual two-sample BF problem; (2) the dimensionality of the data may be very large, or even larger than the sample sizes; and (3) when the dimensionality is too large, traditional testing procedures such as the classical Hotelling's T^2-test may fail to work (see [23]). To our best knowledge, there is little work in the literature available for this problem.

Due to the simplicity and good properties of Scheffe's transform method[3] for the usual two-sample BF problem, following Bennett[12] and Anderson[13], we extend the transform method to the k-sample high-dimensional BF problem. The transformation method has the following advantages. The transformation is simple. It reduces the k-sample BF problem to a one-sample problem. The observations in the resulting sample are independently normally distributed with mean vector consisting of the differences between the mean vector of one population and the mean vectors of the other populations. The sample mean vector consists of the differences between the sample mean vector of one sample and the sample mean vectors of the other samples, which is most relevant to test the null hypothesis. In addition, the one-sample problem can be easily tested by the classical Hotelling's T^2 test when the sample size is very large relative to its dimensionality.

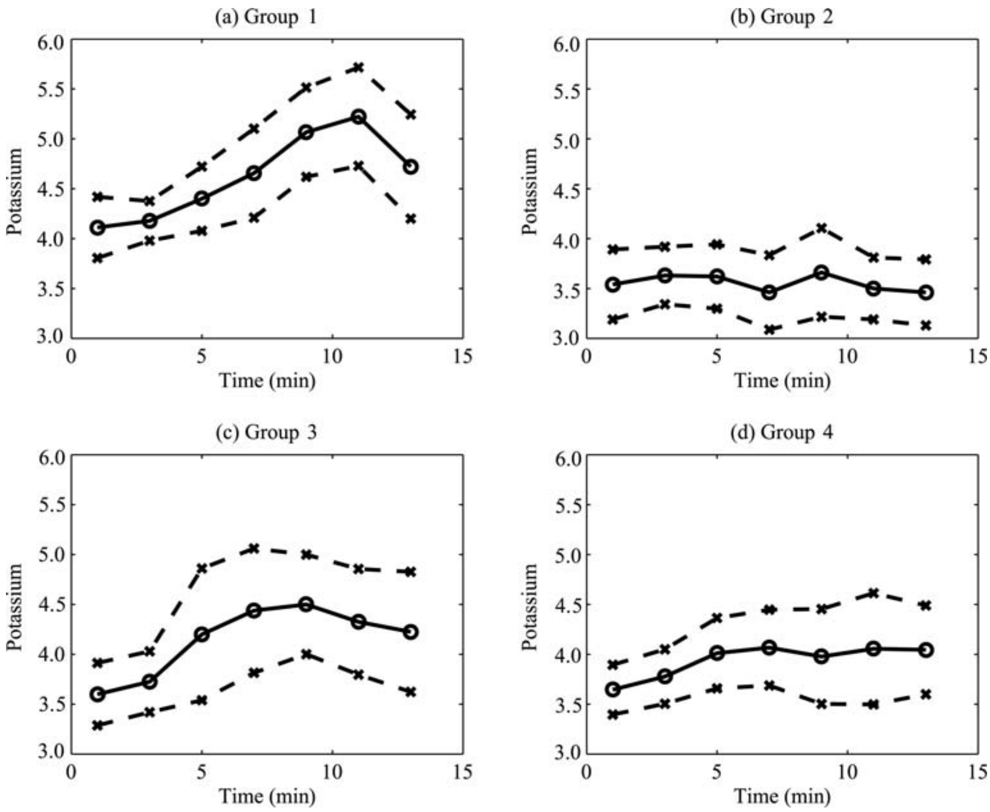

Figure 1 Pointwise group means with 95% pointwise confidence intervals for the dog potassium data.

Notice that the transformation method has several disadvantages. The size of the resulting sample is reduced to the smallest sample size of the k samples. This indicates that some observations in the k samples are "sacrificed" by the transformation method and it does affect the accuracy of the covariance structure estimation to some extent. As pointed out by Anderson[13], this sacrifice is not so important, especially when the sample sizes of the k samples are about the same. Another problem is that the dimensionality of the resulting sample is enlarged to $(k-1)$ times of the dimensionality of the original samples so that the dimensionality may be close to or even larger than the sample size. In these two cases, the classical Hotelling's T^2 test is not powerful or cannot even be well defined. To overcome this difficulty, following Bai and Saranadasa[23], we propose and study an L^2-norm based test for the resulting one-sample problem.

This paper is organized as follows. In Section 2, we describe the transformation method which reduces the k-sample high-dimensional BF problem to a high-dimensional one-sample problem. Section 3 derives and compares the asymptotic powers of Hotelling's T^2 test and the L^2-norm based test for the high-dimensional one-sample problem under the cases when the dimensionality of the data is fixed or when it tends to ∞ with the sample size. Methods for implementing the L^2-norm based test are also described. Two simulation studies are presented in Section 4. In Section 5, the methodologies proposed in this paper are illustrated using the dog potassium data.

2 The transformation method

In this section, following Bennett[12] and Anderson[13], we shall extend Scheffe's transformation method[3] to the k-sample high-dimensional BF problem. Assume that we have the following k independent normal samples $\mathbf{x}_{lj}, j = 1, 2, \ldots, n_l \sim N_p(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l),\ l = 1, 2, \ldots, k$, where and throughout the paper, $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a p-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. For high-dimensional data, we mean p is very large compared with the sample sizes $n_l, l = 1, 2, \ldots, k$ or even larger than them. The BF problem refers to the problem of testing whether the k mean vectors are equal:

$$H_0:\ \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k, \quad \text{vs} \quad H_1:\ \ H_0 \text{ is not true}, \tag{2.1}$$

without assuming that the covariance matrices $\boldsymbol{\Sigma}_l, l = 1, 2, \ldots, k$ are equal.

Without loss of generality, we assume that $n_1 \leqslant n_2 \leqslant \cdots \leqslant n_k$. The transformation method is described as follows. Denote $\bar{\mathbf{x}}_l(m) = m^{-1}\sum_{j=1}^m \mathbf{x}_{lj}$ as the partial sample mean of the first m observations of the l-th sample. Obviously, $\bar{\mathbf{x}}_l(m)$ is an unbiased estimator of $\boldsymbol{\mu}_l$. For $l = 2, \ldots, k$, define

$$\mathbf{y}_{lj} = [\mathbf{x}_{1j} - \bar{\mathbf{x}}_l(n_l)] + \sqrt{\frac{n_1}{n_l}}[\mathbf{x}_{lj} - \bar{\mathbf{x}}_l(n_1)], \quad j = 1, 2, \ldots, n_1.$$

This is Scheffe's transformation method[3] which transforms two normal samples into one. Notice that the first partial sample mean $\bar{\mathbf{x}}_l(n_l)$ uses all the observations in the l-th sample while the second partial sample mean $\bar{\mathbf{x}}_l(n_1)$ uses only the first n_1 observations. Combining the resulting $(k-1)$ samples into one, we have

$$\mathbf{z}_j = [\mathbf{y}_{2j}^T, \mathbf{y}_{3j}^T, \ldots, \mathbf{y}_{kj}^T]^T, \quad j = 1, 2, \ldots, n_1.$$

It is easy to show that

$$\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{n_1} \text{ i.i.d. } \sim N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2.2}$$

where $q = (k-1)p$, and

$$\begin{aligned} \boldsymbol{\mu} &= [\boldsymbol{\mu}_1^T - \boldsymbol{\mu}_2^T, \boldsymbol{\mu}_1^T - \boldsymbol{\mu}_3^T, \ldots, \boldsymbol{\mu}_1^T - \boldsymbol{\mu}_k^T]^T, \\ \boldsymbol{\Sigma} &= \mathbf{J}_{k-1} \otimes \boldsymbol{\Sigma}_1 + n_1 \mathrm{diag}\left(\frac{\boldsymbol{\Sigma}_2}{n_2}, \frac{\boldsymbol{\Sigma}_3}{n_3}, \ldots, \frac{\boldsymbol{\Sigma}_k}{n_k}\right). \end{aligned} \tag{2.3}$$

In the above expressions, $\mathbf{A} \otimes \mathbf{B}$ denotes the Kroneck product of two matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{J}_{k-1}$ the $(k-1) \times (k-1)$ matrix with all entries being 1, and $\mathrm{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_l)$ the block diagonal matrix with entries $\mathbf{A}_1, \ldots, \mathbf{A}_l$. In this way, the original k-sample BF problem (2.1) is reduced to the one-sample problem (3.5) described in next section with $\boldsymbol{\mu}_0 = \mathbf{0}$ based on the i.i.d. high-dimensional normal sample (2.2).

The transformation method has several advantages. First of all, the induced one-sample problem is simpler than the original k-sample BF problem (2.1). In particular, when q is much smaller than n_1, the one-sample problem can be tested by the classical Hotelling's T^2 test which has a known and exact distribution. Secondly, the observations in the induced sample (2.2) are independently normally distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ as described in (2.3). Notice that $\boldsymbol{\mu} = [\boldsymbol{\mu}_1^T - \boldsymbol{\mu}_2^T, \boldsymbol{\mu}_1^T - \boldsymbol{\mu}_3^T, \ldots, \boldsymbol{\mu}_1^T - \boldsymbol{\mu}_k^T]^T$, consisting of the differences between $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_l, l = 2, 3, \ldots, k$. This indicates that $\boldsymbol{\mu} = 0$ if and only if $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k$. Therefore, testing the null hypothesis of the original k-sample BF problem (2.1) is equivalent to testing the null hypothesis of the one-sample problem (3.5). Thirdly, $\boldsymbol{\mu}$ can be estimated by the associated sample mean vector $\bar{\mathbf{z}} = [\bar{\mathbf{x}}_1^T - \bar{\mathbf{x}}_2^T, \ldots, \bar{\mathbf{x}}_1^T - \bar{\mathbf{x}}_k^T]^T$, which consists of the differences between $\bar{\mathbf{x}}_1$and $\bar{\mathbf{x}}_l, l = 2, 3, \ldots, k$. Obviously, $\bar{\mathbf{z}}$ is invariant to the orders of the k-samples and is most relevant to test the original null hypothesis and the null hypothesis of the induced one-sample problem. In addition, we shall know in next section that there is no need to estimate $\boldsymbol{\Sigma}_l, l = 1, 2, \ldots, k$ to test the induced one-sample problem.

Notice that the transformation method has some disadvantages too. First of all, the sample size of the induced sample (2.2) is n_1, the smallest sample size among the k samples. This indicates that some observations are sacrificed by the transformation method. This does not affect the estimation accuracy of the mean vector $\boldsymbol{\mu}$ but it does affect the estimation accuracy of the covariance matrix $\boldsymbol{\Sigma}$ to some degree. However, it is known that the sample covariance matrix $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + O_P(n_1^{-1/2})$, indicating that this effect will be small when the sizes of the k samples are large and are about the same. Moreover, during the data transformation, some information of the k samples may have already been taken into account, as shown in (2.3) where the factors $\boldsymbol{\Sigma}_l/n_l, l = 2, 3, \ldots, k$ are involved in $\boldsymbol{\Sigma}$. Therefore, as pointed out by Anderson[13], the sacrifice of the observations for estimating the covariance matrix is not so important especially when the sizes of the k samples are large and are about the same. Another disadvantage is that the dimensionality $q = (k-1)p$ of the induced sample may be very large even when k and p are moderately large. When this is the case, as pointed out by Bai and Saranadasa[23] via studying a high-dimensional two-sample problem, the classical Hotelling's T^2 test may be not powerful or cannot even be properly defined. In this case, we shall propose an L^2-norm based test. Details for testing a general high-dimensional one-sample problem will be given in the next section.

3 Testing a general high-dimensional one-sample problem

The induced one-sample problem is useful not only for testing the k-sample BF problem (2.1) for high-dimensional data, but also has its own merits. A lot of other testing problems can be reduced to it. For example, a pairwise two-sample problem for high dimensional data can be easily reduced to such a one-sample problem. Therefore, in this section, we shall treat a general one-sample problem for high-dimensional data, which is described as follows. Given a sample

$$\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n \text{ i.i.d. } \sim N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3.4}$$

where q is large, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. We want to test

$$H_0: \ \boldsymbol{\mu} = \boldsymbol{\mu}_0, \quad \text{vs} \quad H_1: \quad \boldsymbol{\mu} \neq \boldsymbol{\mu}_0, \tag{3.5}$$

where $\boldsymbol{\mu}_0$ is a pre-specified q-dimensional vector. In this section, we shall study two testing procedures: Hotelling's T^2-test and an L^2-norm based test, under two cases when q is fixed and when q varies with n. The second case aims to study the asymptotic behaviors of the two tests for high-dimensional data.

3.1 Hotelling's T^2 test

When $q < n$ and $\boldsymbol{\Sigma}$ is invertible, the famous Hotelling's T^2 test[24] for the one-sample problem (3.5) is well defined. Hotelling's T^2 test is well understood and its test statistic is defined as

$$T_n^2 = n(\bar{\mathbf{z}} - \boldsymbol{\mu}_0)^T \hat{\boldsymbol{\Sigma}}^{-1} (\bar{\mathbf{z}} - \boldsymbol{\mu}_0), \tag{3.6}$$

where

$$\bar{\mathbf{z}} = n^{-1} \sum_{i=1}^n \mathbf{z}_i, \ \ \hat{\boldsymbol{\Sigma}} = (n-1)^{-1} \sum_{i=1}^n (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T, \tag{3.7}$$

are respectively the unbiased estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. It is well known that

$$\frac{n-q}{(n-1)q} T_n^2 \sim F_{q,n-q}(\delta^2), \ \ \delta^2 = n(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0),$$

where $F_{q,n-q}(\delta^2)$ denotes an noncentral F-distribution with q and $n-q$ degrees of freedom and the noncentrality parameter δ^2. Under H_0, the noncentrality parameter $\delta^2 = 0$. Therefore it is easy to conduct Hotelling's test. For a given significance level α, the critical value for T_n^2 is

$$T_{\alpha,n}^2 = \frac{(n-1)q}{n-q} F_{\alpha,q,n-q}.$$

Notice that we generally use W_α to denote the upper (100α) percentile of the random variable W. For example, $T_{\alpha,n}^2$ and $F_{\alpha,q,n-q}$ in the above expression denote the upper (100α) percentiles of T_n^2 and the central F-distribution with q and $n-q$ degrees of freedom respectively; and in Theorem 1, z_α denotes the upper (100α) percentile of the standard normal distribution.

We now study the asymptotic powers of T_n^2 under two cases when q is fixed and when q varies with n. For this end, we specify a sequence of alternatives with detecting difficulty increasing with n as follows:

$$H_{1n}: \boldsymbol{\mu} = \boldsymbol{\mu}_0 + n^{-\omega/2}\mathbf{u}, \tag{3.8}$$

where ω is some constant satisfying $0 < \omega < 1$ and $\mathbf{u}$ is any fixed real vector with $\|\mathbf{u}\| \in (0, \infty)$. When q varies with n, we impose the following assumptions:

Assumption A. (A1) $q/n \to \gamma \in (0,1)$ as $n, q \to \infty$;

(A2) $\mathbf{u}^T \boldsymbol{\Sigma}^{-1} \mathbf{u} \to c_1 \in (0, \infty)$ as $q \to \infty$.

Theorem 1. *Assume* $0 < \omega < 1$. *When q is fixed, the asymptotic power of T_n^2 is*

$$P(T_n^2 \geqslant T_{\alpha,n}^2 | H_{1n}) = \Phi(n^{(1-\omega)/2} \sqrt{\mathbf{u}^T \boldsymbol{\Sigma}^{-1} \mathbf{u}/4}) + o(1),$$

which tends to 1 *as* $n \to \infty$*. However, when* q *tends to* ∞ *with* n*, under Assumption A, the asymptotic power of* T_n^2 *is*

$$P(T_n^2 \geqslant T_{\alpha,n}^2 | H_{1n}) = \Phi\bigg(- z_\alpha + n^{1/2-\omega}\sqrt{\frac{1-\gamma}{2\gamma}}c_1\bigg) + o(1).$$

As $n \to \infty$*, the above power tends to* 1 *only for* $0 < \omega < 1/2$ *and tends to* α *for* $1/2 < \omega < 1$*.*

Notice that in the above theorem and the rest of this section, $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. By the above theorem, it is seen that when q is fixed, Hotelling's T^2 test is root-n consistent. However, when q tends to ∞ with n, Hotelling's T^2 test is no longer root-n consistent due to the "curse of dimensionality". This effect of large q is not only reflected by $1 - \gamma$ but also by the consistency rate. It is seen that when $1/2 < \omega < 1$, the asymptotic power of T_n^2 is always α, the nominal significance level.

3.2 L^2-norm based test

When $q > n$ or when $\mathbf{\Sigma}$ is degenerate, Hotelling's T^2 test is not well defined since in this case $\hat{\mathbf{\Sigma}}$ is degenerate. For the two-sample testing problem, Bai and Saranadasa[23] proposed an L^2-norm based test to improve Hotelling's T^2 test. For the one-sample problem (3.5), the L^2-norm based test statistic can be specified as:

$$R_n = n\|\bar{\mathbf{z}} - \boldsymbol{\mu}_0\|^2, \tag{3.9}$$

where $\bar{\mathbf{z}}$ is the sample mean vector as defined in (3.7). The above test statistic is proper since under H_0, it is expected that R_n will be small, and otherwise large. To conduct the test (3.5), we need to derive the asymptotic null distribution of R_n. For this purpose, throughout this paper, let $\mathbf{v}_1, \dots, \mathbf{v}_q$ and $\lambda_1, \dots, \lambda_q$ be the eigenvectors and eigenvalues of $\mathbf{\Sigma}$. Let m denote the number of all the positive eigenvalues. When all the eigenvalues are positive, we have $m = q$; otherwise $\lambda_r > 0$ for $r \leqslant m$ and $\lambda_r = 0$ for all $m < r \leqslant q$. In addition, let $\stackrel{d}{=}$ denote that the left and right hand-side random variables have the same distribution, and let $\chi_d^2(\delta^2)$ denote a chi-squared distribution with d degrees of freedom and the noncentrality parameter δ^2. First of all, we can show the following result:

Theorem 2.

$$R_n \stackrel{d}{=} \sum_{r=1}^m \lambda_r A_r + n\bigg(\|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|^2 - \sum_{r=1}^m \pi_r^2\bigg),$$

where $A_r \sim \chi_1^2(n\lambda_r^{-1}\pi_r^2), r = 1, 2, \dots, m$ *are independent, and* $\pi_r = (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T\mathbf{v}_r, r = 1, 2, \dots, q$*.*

We now study the asymptotic power of the L^2-norm based test. Like Hotelling's T^2 test, the behavior of R_n is different for fixed q and varying q. We first deal with the case when q is fixed. By Theorem 2 and under H_{1n}, we have

$$R_n \stackrel{d}{=} \sum_{r=1}^m \lambda_r A_r + n^{1-\omega}\bigg(\|\mathbf{u}\|^2 - \sum_{r=1}^m \delta_r^2\bigg), \quad A_r \sim \chi_1^2(n^{1-\omega}\lambda_r^{-1}\delta_r^2), \tag{3.10}$$

where $\delta_r = \mathbf{u}^T\mathbf{v}_r, r = 1, 2, \dots, q$. According to the values of m and $\delta_r^2, r = 1, 2, \dots, q$, we need to consider only three possible cases: (1) $m < q$ and $\delta_r = 0$ for all $r \in \{1, 2, \dots, m\}$, (2) $m < q$ and $\delta_r \neq 0$ for at least one $r \in \{1, 2, \dots, m\}$, and (3) $m = q$. We shall show that the asymptotic power of R_n for H_{1n} tends to 1 as $n \to \infty$ under any of the three cases. That is, the proposed L^2-norm based test is root n-consistent when q is fixed.

We first consider the asymptotic power of R_n under Case (1), in which (3.10) can be simplified as

$$R_n \stackrel{d}{=} R^* + n^{1-\omega}\|\mathbf{u}\|^2, \tag{3.11}$$

where

$$R^* \stackrel{d}{=} \sum_{r=1}^{m} \lambda_r A_r, \quad A_r \stackrel{\text{i.i.d}}{\sim} \chi_1^2, \tag{3.12}$$

which is the null random expression of R_n as derived from Theorem 2. Let R_α^* denote the upper 100α percentile of R^*, which is a fixed number when q is fixed. We have the following result:

Theorem 3. *Assume* $0 < \omega < 1$. *When* q *is fixed, the asymptotic power of* R_n *under Case* (1) *is*

$$P(R_n \geqslant R_\alpha^* | H_{1n}) = P(R^* \geqslant R_\alpha^* - n^{1-\omega}\|\mathbf{u}\|^2),$$

which tends to 1 *as* $n \to \infty$.

We now study the asymptotic power of R_n under Cases (2) and (3). In these two cases, we first show that R_n is asymptotically normally distributed and then give the asymptotic power of R_n.

Theorem 4. *Assume* $0 < \omega < 1$. *When* q *is fixed, under Cases* (2) *and* (3), *as* $n \to \infty$, *we have*

$$\frac{R_n - \left[\text{tr}(\mathbf{\Sigma}) + n^{1-\omega}\|\mathbf{u}\|^2\right]}{\sqrt{2\left[\text{tr}(\mathbf{\Sigma}^2) + 2n^{1-\omega}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u}\right]}} \xrightarrow{L} N(0,1). \tag{3.13}$$

In addition, the asymptotic power of R_n *is*

$$P(R_n \geqslant R_\alpha^* | H_{1n}) = \Phi\left[\frac{n^{(1-\omega)/2}\|\mathbf{u}\|^2}{2\sqrt{\mathbf{u}^T\mathbf{\Sigma}\mathbf{u}}}\right] + o(1), \tag{3.14}$$

which tends to 1 *as* $n \to \infty$.

We now study the case when q tends to ∞ with n. In this case, the quantities $\text{tr}(\mathbf{\Sigma}), \text{tr}(\mathbf{\Sigma}^2)$, $\|u\|^2$ and $\lambda_{\max} = \max_{1\leqslant r \leqslant m} \lambda_r$ will vary with q. To derive the asymptotic null distribution of R_n and its asymptotic power, we impose the following regular assumptions:

Assumption B. (B1) $\text{tr}(\mathbf{\Sigma})/q \to c_2 \in (0,\infty)$ and $\text{tr}(\mathbf{\Sigma}^2)/q \to c_3 \in (0,\infty)$ as $q \to \infty$;

(B2) $\|\mathbf{u}\|^2 \to c_4 \in (0,\infty)$ and $\mathbf{u}^T\mathbf{\Sigma}\mathbf{u} \to c_5 \in (0,\infty)$ as $q \to \infty$;

(B3) $\lambda_{\max}/\sqrt{q} \to 0$ as $q \to \infty$.

These assumptions are similar to those imposed by Bai and Saranadasa[23] for the study of their two-sample testing procedures.

Theorem 5. *Assume* $0 < \omega < 1$ *and Assumptions* (A1) *and* B *are satisfied. Then as* $n \to \infty$, *we have the expression* (3.13). *Moreover, the upper* 100α *percentile of* R_n *under* H_0 *can be expressed as*

$$R_\alpha^* = \text{tr}(\mathbf{\Sigma}) + \sqrt{2\text{tr}(\mathbf{\Sigma}^2)}z_\alpha + o[\text{tr}^{1/2}(\mathbf{\Sigma}^2)]. \tag{3.15}$$

In addition, the asymptotic power of R_n *is*

$$P\left(R_n \geqslant R_\alpha^* | H_{1n}\right) = \Phi\left[-z_\alpha + \frac{n^{1/2-\omega}c_4}{\sqrt{2\gamma c_3}}\right] + o(1).$$

As $n \to \infty$, the above power tends to 1 only for $0 < \omega < 1/2$ and tends to α for $1/2 < \omega < 1$.

Comparing the asymptotic powers of T_n and R_n, it can be seen that their asymptotical consistency rates are the same regardless of whether q is fixed or tends to ∞ with n. However, when q tends to ∞ with n, the asymptotic power of Hotelling's T^2 test may be lower due to the fact that the term $1-\gamma$ appears in the expression of the asymptotic power of T_n but it does not appear in the expression of the asymptotic power of R_n. When q and n are about the same, this term is about 0, making the classical Hotelling's T^2 test less powerful.

We now study how to approximate the null distribution of the L^2-norm based test. For this purpose, we need first to address a few issues. First of all, the asymptotic null distribution of R_n, i.e., the distribution of R^*, depends on m unknown positive eigenvalues of the underlying covariance matrix $\mathbf{\Sigma}$. Secondly, the number m is unknown. Finally, when m is very large, say, $m > 50$, it is not easy to compute the distribution of R^* even when the m positive eigenvalues are known. In what follows, we show how to address them.

First of all, to address the last issue, we approximate the distribution of R^* by the 2-cumulant matched χ^2-approximation. The key idea of the method is to approximate the distribution of R^* by that of a random variable of form $S = \beta\chi^2_d$. The parameters β and d are determined, via matching the first two cumulants of R^* and S where d is usually referred to as the approximate degrees of freedom of R^*. Simple calculation leads to

$$\beta = \frac{\sum_{r=1}^m \lambda_r^2}{\sum_{r=1}^m \lambda_r} = \frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma})}, \quad d = \frac{(\sum_{r=1}^m \lambda_r)^2}{\sum_{r=1}^m \lambda_r^2} = \frac{\mathrm{tr}^2(\mathbf{\Sigma})}{\mathrm{tr}(\mathbf{\Sigma}^2)}. \tag{3.16}$$

This indicates that for estimating β and d, we only need to estimate $\mathrm{tr}(\mathbf{\Sigma}), \mathrm{tr}^2(\mathbf{\Sigma})$ and $\mathrm{tr}(\mathbf{\Sigma}^2)$ and we do not actually need to estimate the eigenvalues of $\mathbf{\Sigma}$ and the number of positive eigenvalues, m. Therefore, the first two issues are already addressed.

Let $\chi^2_{\alpha,d}$ denote the upper 100α percentile of χ^2_d. Then the α-level critical value of R^* can be approximately specified by

$$\beta\chi^2_{\alpha,d}, \quad \text{or} \quad \beta d + \sqrt{2d} z_\alpha. \tag{3.17}$$

The first formula corresponds to the 2-cumulant matched χ^2-approximation method and it can be used even when d is relatively small or moderately large, say $d \geqslant 2$. The second formula corresponds to the normal approximation method of Bai and Saranadasa[23] and it can be used only when d is relatively large, say $d \geqslant 30$. Notice that d may not always be an integer. This will not be a problem for users who compute the P-values using some statistical software since popular statistical software such as Matlab does allow non-integer degrees of freedom for chi-squared distributions. However, this may cause a problem for those users who conduct the proposed L^2-norm based test by looking at the chi-squared table for proper critical values. To avoid this inconvenience, we may truncate the d to its nearest integer, i.e., to approximate the distribution of R^* by that of $\beta\chi^2_{[d]}$ where $[d]$ denotes the closest integer to d. That is, the α-level critical values of R^* can be obtained from (3.17) by replacing d by $[d]$.

Buckley and Eagleson[25] and Zhang[26] showed that one can also approximate the distributions of R^* via matching three cumulants, which is known as the three-cumulant matched χ^2-approximation method. Asymptotically, we can show that the three-cumulant matched χ^2-approximation is more accurate than the 2-cumulant matched χ^2-approximation. However, in this paper, we shall recommend to use the 2-cumulant matched χ^2-approximation due to: (1) the latter is simpler since it involves only two parameters while the former involves three, and (2) the parameters β and d can be better estimated by a bias-reduced method described below.

A natural way for estimating the parameters β and d is obtained via replacing $\mathbf{\Sigma}$ in (3.16) with $\hat{\mathbf{\Sigma}}$, the unbiased estimator of $\mathbf{\Sigma}$ given in (3.7). This method is called as the naive method. It is biased since $\mathrm{tr}^2(\hat{\mathbf{\Sigma}})$ and $\mathrm{tr}(\hat{\mathbf{\Sigma}}^2)$ are biased upward for $\mathrm{tr}^2(\mathbf{\Sigma})$ and $\mathrm{tr}(\mathbf{\Sigma}^2)$ respectively. In fact, we can show that

$$\mathrm{Etr}^2(\hat{\mathbf{\Sigma}}) = \mathrm{tr}^2(\mathbf{\Sigma}) + \mathrm{Var}(\mathrm{tr}(\hat{\mathbf{\Sigma}})),\ \mathrm{Etr}(\hat{\mathbf{\Sigma}}^2) = \mathrm{tr}(\mathbf{\Sigma}^2) + \sum_{i=1}^{q}\sum_{j=1}^{q}\mathrm{Var}(\hat{\sigma}_{ij}),$$

where $\hat{\sigma}_{ij}$ is the (i,j)-th entry of $\hat{\mathbf{\Sigma}}$. To address this problem, we propose to replace $\mathrm{tr}^2(\mathbf{\Sigma})$ and $\mathrm{tr}(\mathbf{\Sigma}^2)$ in (3.16) by their unbiased estimators respectively. We call this latter method as the bias-reduced method.

The unbiased estimators of $\mathrm{tr}^2(\mathbf{\Sigma})$ and $\mathrm{tr}(\mathbf{\Sigma}^2)$ may be found in [23] and are

$$\begin{aligned} \widehat{\mathrm{tr}^2(\mathbf{\Sigma})} &= \frac{(n-1)n}{(n-2)(n+1)}\left[\mathrm{tr}^2(\hat{\mathbf{\Sigma}}) - \frac{2}{n}\mathrm{tr}(\hat{\mathbf{\Sigma}}^2)\right], \\ \widehat{\mathrm{tr}(\mathbf{\Sigma}^2)} &= \frac{(n-1)^2}{(n-2)(n+1)}\left[\mathrm{tr}(\hat{\mathbf{\Sigma}}^2) - \frac{1}{n-1}\mathrm{tr}^2(\hat{\mathbf{\Sigma}})\right]. \end{aligned} \tag{3.18}$$

Plugging these into (3.16), we obtain the bias-reduced estimators for β and d. A simulation study conducted in Section 4 shows that the 2-cumulant matched bias-reduced method indeed outperforms the 2-cumulant (and 3-cumulant) matched naive method. However, by (3.18) this advantage may disappear when n is large or when $\mathrm{tr}^2(\hat{\mathbf{\Sigma}})$ and $\mathrm{tr}(\hat{\mathbf{\Sigma}}^2)$ are large.

4 Simulation studies

In this section, we shall present two simulation studies. Simulation 1 aims to compare the approximate powers of Hotelling's T^2 test, the L^2-norm based test with the 2-cumulant matched bias-reduced χ^2-approximation, and that with the normal approximation (see [23]). Simulation 2 aims to compare the approximate powers of the L^2-norm based test with the χ^2-approximation using the 2-cumulant matched naive, the 2-cumulant matched bias-reduced and the 3-cumulant matched naive methods respectively.

In these two simulation studies, we shall generate simulated samples according to (3.4). That is, for given mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, simulated samples will be generated from $N_q(\boldsymbol{\mu}, \mathbf{\Sigma})$. For the given hypothetical mean vector $\boldsymbol{\mu}_0$, the mean vector $\boldsymbol{\mu}$ is specified as $\boldsymbol{\mu} = \boldsymbol{\mu}_0 + \delta\mathbf{u}$ where δ is a tuning parameter taking values over $[0, \delta_0]$ and $\mathbf{u}$ is some fixed nonzero q-dimensional vector. Notice that when $\delta = 0$, the simulated samples follow the null distribution $N_q(\boldsymbol{\mu}_0, \mathbf{\Sigma})$ so that we can study if the test under consideration has the nominal significant level α, and when $\delta \neq 0$, the simulated samples follow an alternative distribution $N_q(\boldsymbol{\mu}_0 + \delta\mathbf{u}, \mathbf{\Sigma})$ so that we can study what the power of the test at δ is. The power of the test will generally increase up to 1 with δ increasing. The constant δ_0 was chosen so that the associated power is about 1 when $\delta = \delta_0$. For different covariance matrix $\mathbf{\Sigma}$, the performance of a test may be different. We specified $\mathbf{\Sigma}$ as $\mathbf{\Sigma} = (\sigma_{ij})$ with $\sigma_{ij} = aw^{|i-j|}$ where a is a constant and w is a tuning parameter. The constant a specifies the diagonal entries of $\mathbf{\Sigma}$, the variances of the components of $\mathbf{z} \sim N_q(\boldsymbol{\mu}_0+\delta\mathbf{u}, \boldsymbol{\Sigma})$ and we used $a = 40$ for simplicity. The tuning parameter w determines the correlation size of $\mathbf{z}$. When $w = 0$, we defaulted $\boldsymbol{\Sigma}$ as $a\mathbf{I}_q$, indicating that the components of $\mathbf{z}$ are independently normally distributed. When w increases, the correlations of the components of $\mathbf{z}$ also increase.

Specifically, we first randomly generated $\boldsymbol{\mu}_0$ and $\mathbf{u}$. We then took $w = 0, 0.3, 0.6$ or 0.9. For each fixed w, we let δ take values uniformly in $[0, \delta_0]$, e.g., $\delta = 0, 1/3, 2/3, \ldots, \delta_0$. For each pair (w, δ), $N = 10000$ samples of size n were generated from $N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. For each sample, the test statistics of the testing procedures were computed and the associated P-values were calculated using some method under consideration. When the test statistics are larger than the critical values or when the P-values are smaller than the nominal significance level α, the null hypothesis is rejected. The simulated power of a testing procedure is the proportion of the number of rejections based on the simulated critical values, computed based on the $N = 10000$ samples when $\delta = 0$ and with the same a and w. The approximate power of a testing procedure is the proportion of the number of rejections based on the calculated P-values.

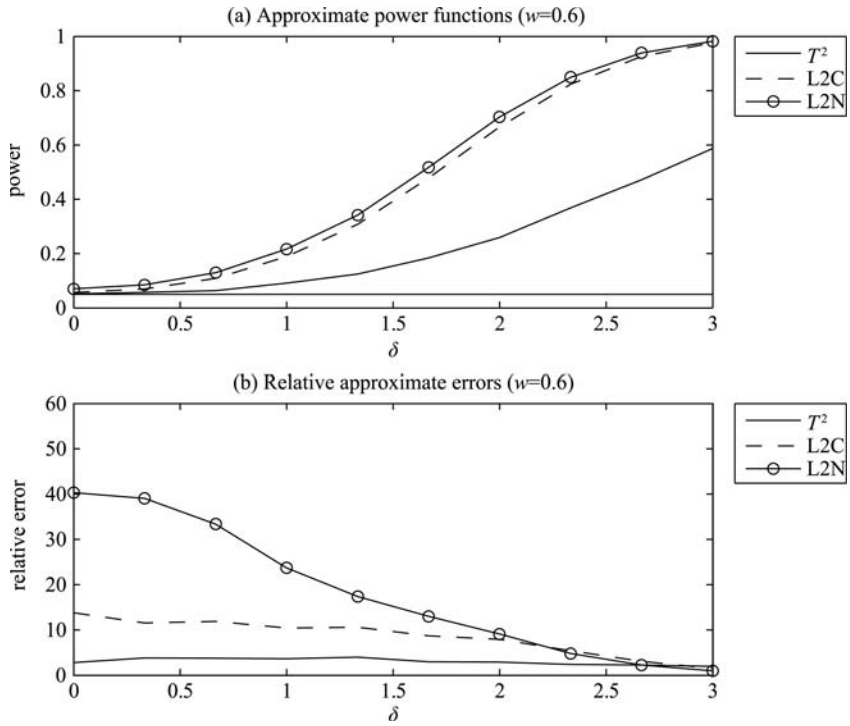

Figure 2 Approximate power functions (upper panel) and relative approximate errors (lower panel) of the "T^2", "L2C" and "L2N" tests for Simulation 1 when $w = 0.6$.

In Simulation 1, we took $n = 40$ and $q = 30$ so that Hotelling's T^2 test can be well defined. We aimed to compare the approximate powers of Hotelling's T^2 test, the L^2-norm based test with the 2-cumulant matched bias-reduced χ^2-approximation, and that with the normal approximation, namely the "T^2", "L2C" and "L2N" tests respectively. Figure 2 displays the simulation results when $w = 0.6$. The upper panel presents the approximate power functions of the three tests. It is seen that the approximate powers of Hotelling's T^2 test are much lower than the approximate powers of the other two tests. This observation is also valid for the cases when $w = 0$ and $w = 0.3$ (not shown here). It indicates that when the correlation is moderate or small, for high-dimensional data, Hotelling's T^2 test has much lower powers than the L^2-norm based tests. This is consistent with Theorems 1 and 5, and the remarks following Theorem 5. For two-sample tests for high-dimensional data, similar phenomenon was observed by Bai and Saranadasa[23].

From the upper panel of Figure 2, it seems that the powers of the "L2N" test are slightly higher than those of the "L2C" test. This phenomenon, unfortunately, was caused by the normal approximation used for the L^2-norm based test. This can be seen from the lower panel of Figure 2 where the relative approximate errors (in percentage) of the three tests are displayed. The relative approximate errors are defined as the errors made by the approximation method, relative to the simulated power (the assumed true power of a test), i.e., the percentage of the differences between the approximate power and the simulated power over the simulated power. It is seen that the relative approximate errors of Hotelling's T^2 test are much smaller than those of the "L2C" test, while the latter's relative approximate errors are much smaller than those of the "L2N" test, especially for smaller δ. Therefore, the normal approximation is less attractive for the L^2-norm based test especially when the null hypothesis is nearly valid.

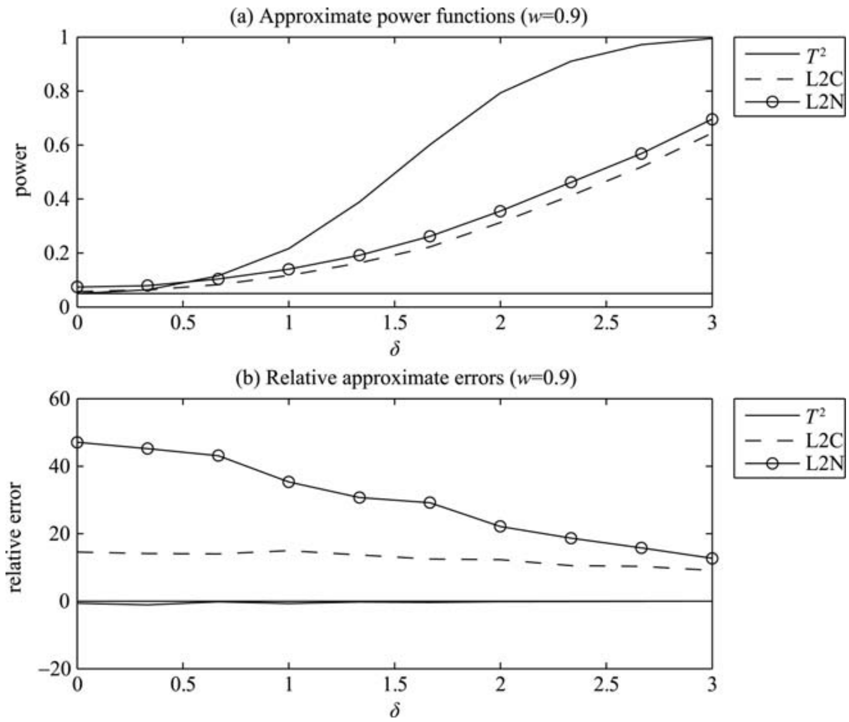

Figure 3 Same caption as that of Figure 2 but now for Simulation 1 with $w = 0.9$.

When the correlation is very large, it is expected that Hotelling's T^2 test performs better than the L^2-norm based test when Hotelling's T^2 test can be well defined. Figure 3 presents such a case when $w = 0.9$, indicating large correlation for the simulated data. From the upper panel, it is seen that the approximate powers of Hotelling's T^2-test are higher than those of the other two tests, especially when δ is large. Moreover, the relative approximation errors of Hotelling's T^2-test are smaller than those of the other two tests. This indicates that although the dimensionality is very large relative to the sample size, when the correlation is very large, it is still a good choice to employ Hotelling's test which takes the correlation into account.

In Simulation 2, we took $n = 40$ and $q = 50$ in which Hotelling's T^2-test can not be well defined and we aimed to compare the approximate powers of the L^2-norm based test with the χ^2-approximation using the 2-cumulant matched naive, the 2-cumulant matched bias-reduced and the 3-cumulant matched naive methods, namely, the "L2C, 2-c, naive", "L2C, 2-c, bias-reduced" and "L2C, 3-c, naive" tests respectively. Figure 4 shows the simulation results for the case when $w = 0.6$. The upper panel displays the approximate power functions of the three tests. It is seen that the "L2C, 2-c bias-reduced" test has higher powers than the other two tests. The lower panel displays the relative approximate errors of the three tests. It is seen that the "L2C, 2-c, bias-reduced" test has much smaller relative approximate errors than the other two tests, especially when δ is small. This phenomenon was also observed for the cases when $w = 0$ and $w = 0.3$ (not shown here). Therefore, for the moderate and small correlation, the bias-reduced method is preferred.

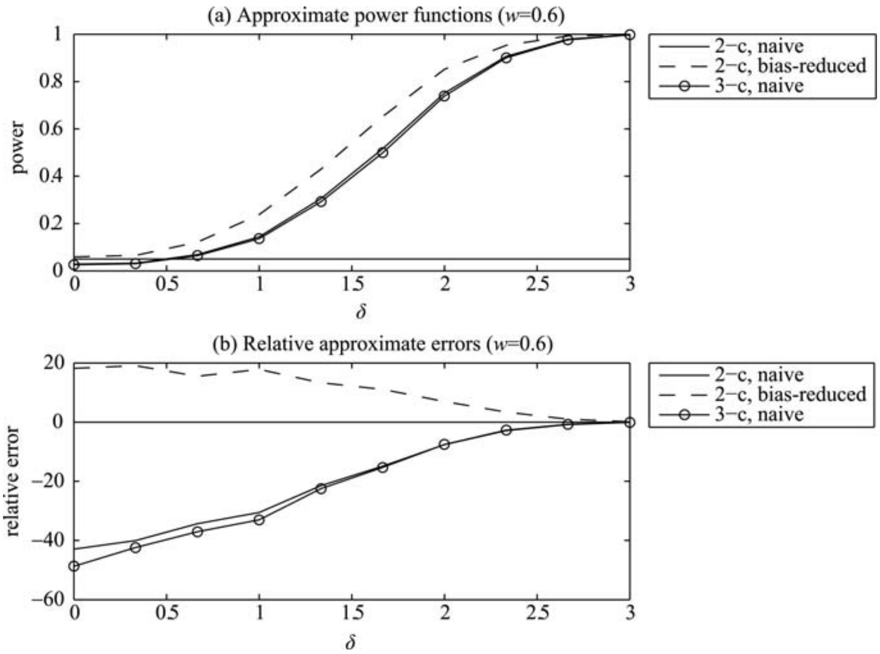

Figure 4 Approximate power functions (upper panel) and relative approximate errors (lower panel) of the "L2C, 2-c, naive", "L2C, 2-c, bias-reduced" and " L2C, 3-c, naive" tests for Simulation 2 when $w = 0.6$.

It is interesting to notice that when the correlation is very large, the naive method for the χ^2-approximation can perform as well as the bias-reduced method. Figure 5 presents such a case where $w = 0.9$, indicating large correlation for the simulated data. From the upper panel, it is seen that the approximate powers of the three tests are similar, with the approximate powers of the "L2C, 2-c, bias-reduced" test slightly higher than those of the other two tests. However, the relative approximation errors of the "L2C, 2-c, bias-reduced" test are larger than those of the other two tests. This indicates that when the correlation is large, it is sufficient to employ the L^2-norm based test with the χ^2-approximation using the 2 or 3-cumulant matched naive methods. From the lower panels of Figures 2–5, we can also see that the bias-reduced method performs quite stably whenever $w = 0.6$ or $w = 0.9$.

5 The dog potassium data

We now apply the proposed methodologies to the dog potassium data briefly described in Section 1. This is a 4-sample BF problem. We first transformed this 4-sample BF problem into a one-sample problem using the transformation method described in Section 2. The resulting sample has 8 observations, with dimension $(4-1)\times 7 = 21$. Thus, Hotelling's T^2 test is not applicable. We then applied the L^2-norm based test using the 2-cumulant matched bias-reduced χ^2-approximation, resulting in a P-value 0.0022, indicating that there is very strong evidence to reject the null hypothesis. That is, the four groups of dog potassium measurements unlikely have the same treatment effects. This conclusion is expected if one observes Figure 1 carefully.

We also applied the L^2-norm based test using the 2-cumulant matched naive χ^2-approximation and the normal approximation, resulting in P-values 0.0175 and 0 respectively. They show that there is indeed strong evidence to reject the null hypothesis although the normal approximation P-value is less trustful. In fact, the approximate degrees of freedom for the 2-cumulant matched naive and bias-reduced χ^2-approximations are 3.053 and 5.682 respectively, which are too small to make the normal approximation adequate.

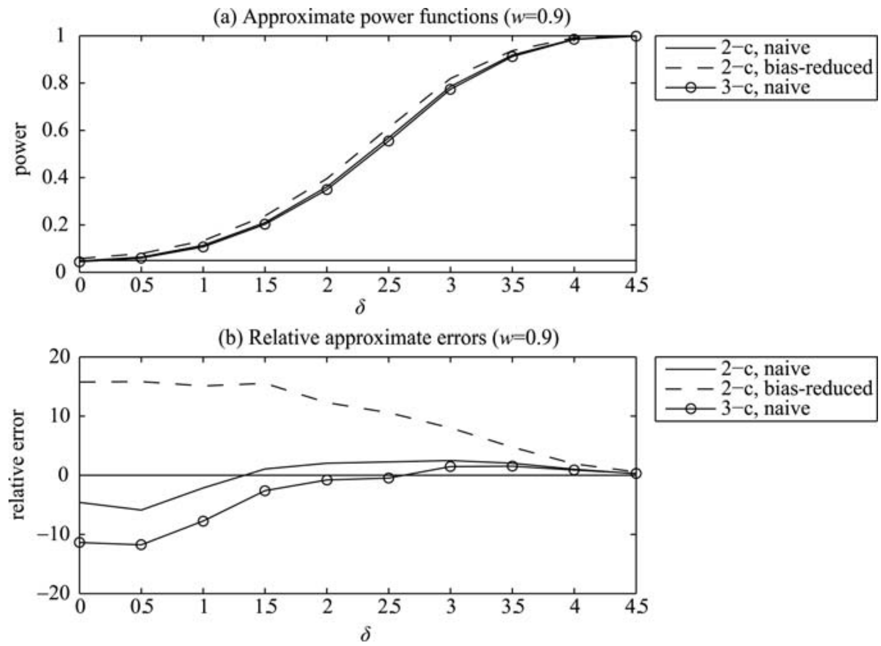

Figure 5 Same caption as that of Figure 4 but now for Simulation 2 with $w = 0.9$.

We also applied the proposed methodologies to test if the mean differences of any two groups of the dog potassium data are zero. These mean differences with pointwise 95% confidence intervals are displayed in Figure 6. Panel (a) shows that the mean differences between Groups 1 and 2 should be statistically highly significant since the pointwise confidence intervals do not contain 0. Panel (f) shows that the mean differences between Groups 3 and 4 should not be statistically significant since the pointwise confidence intervals always contain 0. Panels (b)–(e) indicate that the mean differences of the other pairwise groups should be somewhat significant since some of the pointwise confidence intervals contain 0 and some do not. Table 1 shows the P-values for the pairwise tests. Column 1 shows the pairwise group mean difference tests. Columns 2–5 respectively list the P-values of Hotelling's T^2 test, the L^2-norm based tests with P-values approximated by the three methods: the 2-cumulant matched naive and bias-reduced methods and the normal approximation method. The P-values of Hotelling's T^2 tests indicate that all these pairwise tests are not statistically significant. This indicates that Hotelling's T^2-test fails to detect the pairwise group mean differences and it yields misleading results. This is consistent with Theorem 1 since the sample sizes of the induced samples are either 8 or 9, which are very close to the associated dimensionality 7. However, in these cases, the L^2-norm based tests, especially the L^2-norm based test using the 2-cumulant matched bias-reduced χ^2-approximation can powerfully detect the pairwise group mean differences indicated by Figure 6. These examples show that the L^2-norm based tests are indeed useful and can be more powerful than Hotelling's T^2 test when the dimensionality is very close to the sample sizes.

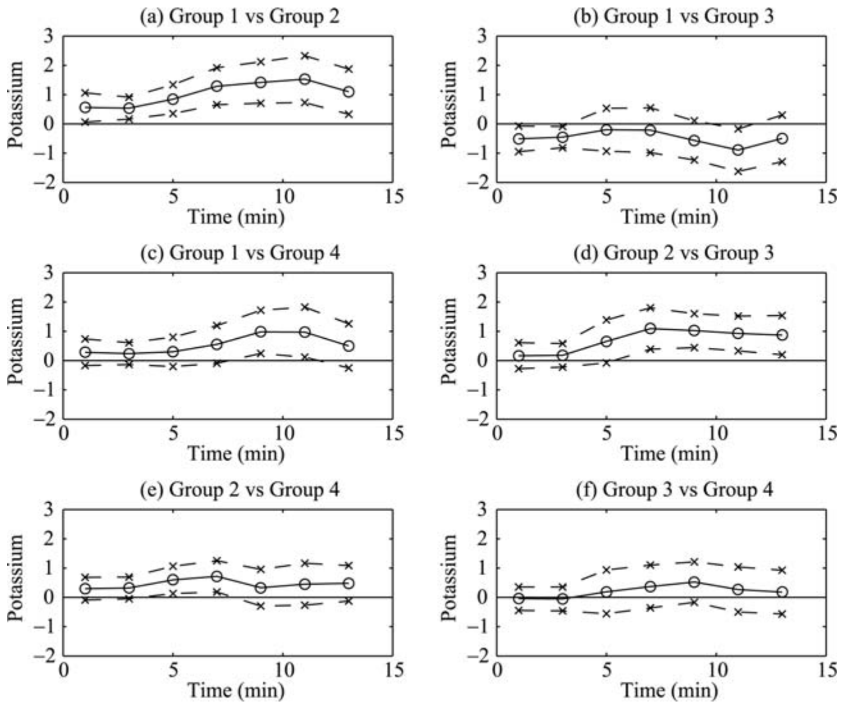

Figure 6 Pairwise group mean differences with 95% confidence intervals for the dog potassium data.

Table 1 P-values for pairwise group mean tests for the dog potassium data.

Test for groups	T^2	L2C, 2-c naive	L2C, 2-c bias-reduced	L2N
1 vs 2	0.1178	0.0000	0.0000	0.0000
1 vs 3	0.3399	0.0532	0.0241	0.0082
1 vs 4	0.5999	0.0003	0.0000	0.0000
2 vs 3	0.2383	0.0010	0.0001	0.0000
2 vs 4	0.3368	0.0343	0.0155	0.0022
3 vs 4	0.2126	0.3962	0.3802	0.5345

Acknowledgements The authors thank Prof-Hu Feifang and a referee for their comments which help improve the paper substantially.

References

1 Behrens B V. Ein Beitrag zur Fehlerberechnung bei wenige Beobachtungen. *Landwirtschaftliches Jahresbuch*, **68**: 807–837 (1929)
2 Fisher R A. The fiducial argument in statistical inference. *Annals of Eugenics*, **11**: 141–172 (1935)
3 Scheffe H. On solutions of the Behrens-Fisher problem, based on the t-distribution. *Ann Math Statist*, **14**: 35–44 (1943)
4 Welch B L. The generalization of Student's problem when several different population variances are involved. *Biometrika*, **34**: 28–35 (1947)
5 Scheffe H. Practical solutions of the Behrens-Fisher problem. *J Amer Statist Assoc*, **65**: 1501–1508 (1970)
6 Lee A, Gurland J. Size and power of tests for equality of means of two normal populations with unequal variances. *J Amer Statist Assoc*, **70**: 933–947 (1975)
7 Weerahandi S. Generalized confidence intervals. *J Amer Statist Assoc*, **88**: 899–905 (1993)
8 Weerahandi S. Exact Statistical Methods in Data Analysis. New York: Springer-Verlag, 1995
9 Tang S, Tsui K W. Distributional properties for the generalized P-value for the Behrens-Fisher problem. *Statist Prob Lett*, **77**: 1–8 (2005)
10 Ghosh M, Kim Y H. The Behrens-Fisher problem revisited: A Bayes-frequentist synthesis. *Canad J Statist*, **29**: 1–14 (2001)
11 Dong L. The Behrens-Fisher problem: an empirical likelihood approach. Technical Report. Department of Economics, University of Victoria, 2004
12 Bennett B M. Note on a solution of the generalized Behrens-Fisher problem. *Ann Inst Statist Math*, **2**: 87–90 (1951)
13 Anderson T W. An Introduction to Multivariate Statistical Analysis. New York: Wiley, 1984
14 James G S. Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika*, **41**: 19–43 (1954)
15 Yao Y. An approximate degrees of freedom solution to the multivariate Behrens-Fisher problem. *Biometrika*, **52**: 139–147 (1965)
16 Johansen S. The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika*, **67**: 85–95 (1980)
17 Nel D G, Van der Merwe C A. A solution to the multivariate Behrens-Fisher problem. *Comm Statist Theory Methods*, **15**: 3719–3735 (1986)
18 Algina J, Tang K. Type-I error rates for Yao's and James' tests of equality of mean vectors under variance-covariance heteroscedasticity. *J Educ Statist*, **13**: 281–290 (1988)
19 Krishnamoorthy K, Yu J. Modified Nel and Van der Merwe test for the multivariate Behrens-Fisher problem. *Statist Prob Lett*, **66**: 161–169 (2004)
20 Krishnamoorthy K, Xia Y. On selecting tests for equality of two normal mean vectors. *Multi Behavioral Res*, **41**: 533–548 (2006)
21 Grizzle J E, Allen D M. Analysis of growth and dose response curves. *Biometrics*, **25**: 357–381 (1969)
22 Wang Y. Mixed effects smoothing spline analysis of variance. *J Royal Statist Soc Ser B*, **60**: 159–174 (1998)
23 Bai Z, Saranadasa H. Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, **6**: 311–329 (1996)
24 Hotelling H. The generalization of student's ratio. *Ann Math Statist*, **2**: 360–378 (1931)
25 Buckley M J, Eagleson G K. An approximation to the distribution of quadratic forms in normal random variables. *Austral J Statist*, **30A**: 150–159 (1988)
26 Zhang J T. Approximate and asymptotic distribution of χ^2-type mixtures with application. *J Amer Statist Assoc*, **100**: 273–285 (2005)

Appendix: technical proofs

In this appendix, we first give a lemma and then give the proofs of some theorems.

Lemma 1. *Let $F \sim F_{r,s}(\delta^2)$. As $r, s \to \infty$ or $\delta^2, s \to \infty$, we have*

$$F \stackrel{d}{=} a + \sqrt{b}Z + o_P(\sqrt{b}), \tag{A.1}$$

where $Z \sim N(0,1), a = 1 + \delta^2/r$ and $b = 2[\frac{1+2\delta^2/r}{r} + \frac{(1+\delta^2/r)^2}{s}]$.

Proof of Lemma 1. Write $F = f(X,Y) = X/Y$, where $rX \sim \chi_r^2(\delta^2)$ and $sY \sim \chi_s^2$ are independent. We can show that, as $r, s \to \infty$ or as $\delta^2, s \to \infty$, both X and Y are asymptotically normally distributed. It follows that F is also asymptotically normally distributed. Notice that $\mathrm{E}(X) = 1 + \delta^2/r, \mathrm{Var}(X) = 2(1+2\delta^2/r)/r,\ \mathrm{E}(Y) = 1, \mathrm{Var}(Y) = 2/s$, and $f_x'(x,y) = 1/y, f_y'(x,y) = -x/y^2$. By Taylor expansion, as $r, s \to \infty$, F can be expressed as (A.1) with

$$\begin{aligned} a &= f(\mathrm{E}(X), \mathrm{E}(Y)) = \mathrm{E}(X)/E(Y) = 1 + \delta^2/r, \\ b &= f_x'^2[\mathrm{E}(X), \mathrm{E}(Y)]\mathrm{Var}(X) + f_y'^2[\mathrm{E}(X), \mathrm{E}(Y)]\mathrm{Var}(Y) \\ &= 2[(1+2\delta^2/r)/r + (1+\delta^2/r)^2/s], \end{aligned}$$

as desired.

Proof of Theorem 1. Notice that when q is fixed and as $n \to \infty$, $F_{\alpha,q,n-q} = \chi_q^2(\alpha)/q[1+o(1)]$ is finite. However, under H_{1n}, by Lemma 1, T_n^2 will be asymptotically normally distributed due to $(n-q) \to \infty$ and the noncentrality parameter

$$\delta^2 = n^{1-\omega}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u} \to \infty, \text{ as } n \to \infty.$$

It follows from Lemma 1 that

$$\frac{n-q}{(n-1)q}T_n^2 \stackrel{d}{=} a + \sqrt{b}Z + o_p(\sqrt{b}), \tag{A.2}$$

where $Z \sim N(0,1)$, and

$$\begin{aligned} a &= 1 + \frac{\delta^2}{q} = n^{1-\omega}\frac{\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}}{q}[1+o(1)], \\ b &= 2\left[\frac{1+2\delta^2/q}{q} + \frac{(1+\delta^2/q)^2}{n-q}\right] = 4n^{1-\omega}\frac{\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}}{q^2}[1+o(1)]. \end{aligned} \tag{A.3}$$

It follows that

$$\begin{aligned} P(T_n^2 \geqslant T_{\alpha,n}^2|H_{1n}) &= P\left(\frac{n-q}{(n-1)q}T_n^2 \geqslant F_{\alpha,q,n-q}|H_{1n}\right) \\ &= P\left(Z \geqslant \frac{F_{\alpha,q,n-q} - a}{\sqrt{b}}\right) + o(1) \\ &= \Phi(n^{(1-\omega)/2}\sqrt{\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}/4}) + o(1), \end{aligned}$$

which tends to 1 as $n \to \infty$.

When q tends to ∞ with n, $\frac{n-q}{(n-1)q}T_n^2$ will be asymptotically normally distributed under both H_0 and H_{1n}. Assume Assumption A is satisfied. Letting $\delta^2 = 0$ in (A.3), we will obtain

$$F_{\alpha,q,n-q} = a + \sqrt{b}z_\alpha + o(\sqrt{b}) = 1 + \sqrt{\frac{2}{n\gamma(1-\gamma)}}z_\alpha + o(n^{-1/2}).$$

On the other hand, under H_{1n}, letting $\delta^2 = n^{1-\omega}\mathbf{u}^T\boldsymbol{\Sigma}^{-1}\mathbf{u} = n^{1-\omega}c_1[1+o(1)]$ in (A.3), we will obtain (A.2) with $a = (1 + n^{1-\omega}c_1/\gamma)[1 + o(1)]$ and $b = \frac{2}{n\gamma(1-\gamma)}[1 + o(1)]$. Therefore,

$$\begin{aligned}P(T_n^2 \geqslant T_{\alpha,n}^2|H_{1n}) &= P\bigg(\frac{n-q}{(n-1)q}T_n^2 \geqslant F_{\alpha,q,n-q}|H_{1n}\bigg)\\ &= P\Bigg(Z \geqslant \frac{1 + \sqrt{\frac{2}{n\gamma(1-\gamma)}}z_\alpha - (1 + n^{1-\omega}c_1/\gamma)}{\sqrt{\frac{2}{n\gamma(1-\gamma)}}}\Bigg) + o(1)\\ &= \Phi\bigg(-z_\alpha + n^{1/2-\omega}\sqrt{\frac{1-\gamma}{2\gamma}}c_1\bigg) + o(1).\end{aligned}$$

The remaining part is obvious. The theorem is proved.

Proof of Theorem 2. Notice that

$$\sqrt{n}[\bar{\mathbf{z}} - \boldsymbol{\mu}] \sim N_q(\mathbf{0}, \boldsymbol{\Sigma}). \tag{A.4}$$

Set $\mathbf{R} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma})$. Recall that $\mathbf{v}_1, \ldots, \mathbf{v}_q$ and $\lambda_1, \ldots, \lambda_q$ are the eigenvectors and eigenvalues of $\boldsymbol{\Sigma}$, and m is the number of all the positive eigenvalues. It follows that

$$\mathbf{R} = \sum_{r=1}^{m} \xi_r \mathbf{v}_r,$$

where $\xi_r = \mathbf{R}^T\mathbf{v}_r, r = 1, 2, \ldots, m$, which are independent and $\mathrm{E}\xi_r = 0, \mathrm{var}(\xi_r) = \lambda_r > 0$ for $r = 1, 2, \ldots, m$. Set $\pi_r = (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T\mathbf{v}_r, r = 1, 2, \ldots, q$. We have

$$\begin{aligned}\|\mathbf{R} + \sqrt{n}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\|^2 &= \bigg\|\sum_{r=1}^{m}[\xi_r + n^{1/2}\pi_r]\mathbf{v}_r + n^{1/2}\sum_{r=m+1}^{q}\pi_r\mathbf{v}_r\bigg\|^2\\ &= \sum_{r=1}^{m}[\xi_r + n^{1/2}\pi_r]^2 + n\sum_{r=m+1}^{q}\pi_r^2,\end{aligned}$$

due to the orthonormality of the eigenvectors $\mathbf{v}_r, r = 1, 2, \ldots, q$. First notice that $\sum_{r=m+1}^{q}\pi_r^2 = \|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|^2 - \sum_{r=1}^{m}\pi_r^2$. On the other hand, since $\mathbf{R}$ is a normal random vector, we have that $\xi_r/\sqrt{\lambda_r}$ i.i.d. $\sim N(0, 1)$ for $r = 1, 2, \ldots, m$. Hence $(\xi_r + n^{1/2}\pi_r)^2 \stackrel{d}{=} \lambda_r A_r, A_r \sim \chi_1^2(n\lambda_r^{-1}\pi_r^2)$. It follows that

$$\begin{aligned}R_n = \|\mathbf{R} + \sqrt{n}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\|^2 &= \sum_{r=1}^{m}(\xi_r + n^{1/2}\pi_r)^2 + n\bigg(\|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|^2 - \sum_{r=1}^{m}\pi_r^2\bigg)\\ &\stackrel{d}{=} \sum_{r=1}^{m}\lambda_r A_r + n\bigg(\|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|^2 - \sum_{r=1}^{m}\pi_r^2\bigg),\end{aligned}$$

as desired.

Proof of Theorem 3. The proof is obvious when one notes that when q is fixed, R_α^* and $\|u\|^2$ are fixed.

Proof of Theorem 4. Under Cases (2) and (3), by (3.10), we have

$$\begin{aligned}\mathrm{E}(R_n) &= \sum_{r=1}^{m}\lambda_r(1 + n^{1-\omega}\lambda_r^{-1}\delta_r^2) + n^{1-\omega}\bigg(\|\mathbf{u}\|^2 - \sum_{r=1}^{m}\delta_r^2\bigg)\\ &= \mathrm{tr}(\boldsymbol{\Sigma}) + n^{1-\omega}\|\mathbf{u}\|^2,\\ \mathrm{Var}(R_n) &= 2\sum_{r=1}^{m}\lambda_r^2(1 + 2n^{1-\omega}\lambda_r^{-1}\delta_r^2) = 2(\mathrm{tr}(\boldsymbol{\Sigma}^2) + 2n^{1-\omega}\mathbf{u}^T\boldsymbol{\Sigma}\mathbf{u}).\end{aligned}$$

Since $\lambda_r A_r \overset{d}{=} \lambda_r(z_r + n^{(1-\omega)/2}\lambda_r^{-1/2}\delta_r)^2,\ z_r \overset{\text{i.i.d}}{\sim} N(0,1)$, we have

$$\sum_{r=1}^{m} \lambda_r A_r \overset{d}{=} \sum_{r=1}^{m} \lambda_r z_r^2 + 2n^{(1-\omega)/2} \sum_{r=1}^{m} \lambda_r^{1/2}\delta_r z_r + n^{1-\omega} \sum_{r=1}^{m} \delta_r^2.$$

It follows that

$$\begin{aligned}
&\frac{R_n - (\text{tr}(\mathbf{\Sigma}) + n\|\mathbf{u}\|^2)}{\sqrt{2(\text{tr}(\mathbf{\Sigma}^2) + 2n^{1-\omega}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u})}} \\
&\quad \overset{d}{=} \frac{\sum_{r=1}^{m} \lambda_r(z_r^2 - 1) + 2n^{(1-\omega)/2}\sum_{r=1}^{m} \lambda_r^{1/2}\delta_r z_r}{\sqrt{2(\text{tr}(\mathbf{\Sigma}) + 2n^{1-\omega}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u})}} \\
&\quad = \frac{\sum_{r=1}^{q} \lambda_r(z_r^2 - 1) + 2n^{(1-\omega)/2}\sum_{r=1}^{q} \lambda_r^{1/2}\delta_r z_r}{\sqrt{2(\text{tr}(\mathbf{\Sigma}^2) + 2n^{1-\omega}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u})}} \\
&\quad = \frac{\sum_{r=1}^{q} \lambda_r(z_r^2 - 1)}{\sqrt{2(\text{tr}(\mathbf{\Sigma}^2) + 2n^{1-\omega}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u})}} + \frac{2n^{(1-\omega)/2}\sum_{r=1}^{q} \lambda_r^{1/2}\delta_r z_r}{\sqrt{2(\text{tr}(\mathbf{\Sigma}^2) + 2n^{1-\omega}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u})}} \\
&\quad \xrightarrow{L} N(0,1).
\end{aligned}$$

In the above equation, we use the facts that the first term is $o_p(1)$ and the second term is asymptotically normally distributed as $N(0,1)$. These two facts can be easily checked by noting that when q is a fixed number, $\text{tr}(\mathbf{\Sigma}), \text{tr}(\mathbf{\Sigma}^2), \|\mathbf{u}\|^2$ and $\mathbf{u}^T\mathbf{\Sigma}\mathbf{u}$ are all fixed. The first part of the proof is finished.

We now move to the proof of the second part. First we have

$$P(R_n \geqslant R_\alpha^*|H_{1n}) = 1 - \Phi\left(\frac{(R_\alpha^* - \text{tr}(\mathbf{\Sigma})) - n^{1-\omega}\|\mathbf{u}\|^2}{\sqrt{2(\text{tr}(\mathbf{\Sigma}^2) + 2n^{1-\omega}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u})}}\right) + o(1).$$

When q is fixed, $\text{tr}(\mathbf{\Sigma}), \text{tr}(\mathbf{\Sigma}^2), \|\mathbf{u}\|^2$ and $\mathbf{u}^T\mathbf{\Sigma}\mathbf{u}$ are all fixed. Moreover, $\mathbf{u}^T\mathbf{\Sigma}\mathbf{u} > 0$. Therefore, the above expression can be further written as

$$P(R_n \geqslant R_\alpha^*|H_{1n}) = \Phi\left[\frac{n^{(1-\omega)/2}\|\mathbf{u}\|^2}{2\sqrt{\mathbf{u}^T\mathbf{\Sigma}\mathbf{u}}}\right] + o(1),$$

which obviously tends to 1 as $n \to \infty$ for $0 < \omega < 1$. The theorem is proved.

Proof of Theorem 5. From the proof of Theorem 4 and under the given conditions, we have

$$\text{E}(R_n) = \text{tr}(\mathbf{\Sigma}) + n^{1-\omega}\|\mathbf{u}\|^2, \quad \text{Var}(R_n) = 2(\text{tr}(\mathbf{\Sigma}^2) + 2n^{1-\omega}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u}), \tag{A.5}$$

and

$$\frac{R_n - \text{E}(R_n)}{\sqrt{\text{Var}(R_n)}} = \frac{\sum_{r=1}^{q} \lambda_r(z_r^2 - 1)}{\sqrt{2(\text{tr}(\mathbf{\Sigma}^2) + 2n^{1-\omega}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u})}} + \frac{2n^{(1-\omega)/2}\sum_{r=1}^{q} \lambda_r^{1/2}\delta_r z_r}{\sqrt{2(\text{tr}(\mathbf{\Sigma}^2) + 2n^{1-\omega}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u})}}, \tag{A.6}$$

where $z_r \overset{\text{i.i.d}}{\sim} N(0,1), r = 1, 2, \ldots, m$. First notice that the second term is $o_p(1)$ as $n \to \infty$ due to the facts that its mean is 0 and its variance tends to 0 under the given conditions. Now we show that the first term in the right-hand side of (A.6) is asymptotically distributed as $N(0,1)$.

Let $S=\sum_{r=1}^q \lambda_r(z_r^2-1)$. Then $\mathrm{E}(S)=0$ and $\mathrm{Var}(S)=2\mathrm{tr}(\boldsymbol{\Sigma}^2)$. Moreover, as $n\to\infty$ and under Assumption (B3), we have

$$\frac{\mathrm{E}\sum_{r=1}^m|\lambda_r(z_r^2-1)|^3}{(2\mathrm{tr}(\boldsymbol{\Sigma}^2))^{3/2}}=\frac{\sum_{r=1}^m\lambda_r^3\mathrm{E}|z_1^2-1|^3}{(2\mathrm{tr}(\boldsymbol{\Sigma}^2))^{3/2}}\leqslant\frac{(\lambda_{\max}/\sqrt{q})^3\mathrm{E}|z_1^2-1|^3}{(2\mathrm{tr}(\boldsymbol{\Sigma}^2)/q)^{3/2}}\to 0.$$

By Liapounov's theorem, the first term in the right-hand side of (A.3) tends to $N(0,1)$ in distribution as $n\to\infty$. The expression (3.13) then follows from (A.5) and (A.6). So does the expression (3.15). Therefore,

$$\begin{aligned}P(R_n\geqslant R_\alpha^*|H_{1n})&=P\left(\frac{R_n-\mathrm{E}(R_n)}{\sqrt{\mathrm{Var}(R_n)}}\geqslant\frac{-n^{1-\omega}\|\mathbf{u}\|^2+\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}z_\alpha}{\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)+2n^{1-\omega}\mathbf{u}^T\boldsymbol{\Sigma}\mathbf{u}}}\right)\\&=\Phi\left(\frac{-\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}z_\alpha+n^{1-\omega}\|\mathbf{u}\|^2}{\sqrt{2(\mathrm{tr}(\boldsymbol{\Sigma}^2)+2n^{1-\omega}\mathbf{u}^T\boldsymbol{\Sigma}\mathbf{u})}}\right)+o(1)\\&=\Phi\left(-z_\alpha+\frac{n^{1-\omega}\|\mathbf{u}\|^2}{\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}}\right)+o(1)\\&=\Phi\left(-z_\alpha+\frac{n^{1/2-\omega}c_4}{\sqrt{2\gamma c_3}}\right)+o(1),\end{aligned}$$

which tends to 1 as $n\to\infty$ only for $0<\omega<1/2$ and tends to α for $1/2<\omega<1$. In the above proof, we use the fact that $n^{1-\omega}\mathbf{u}^T\boldsymbol{\Sigma}\mathbf{u}/\mathrm{tr}(\boldsymbol{\Sigma}^2)=n^{1-\omega}c_5/(n\gamma c_3)[1+o(1)]\to 0$. The theorem is proved.