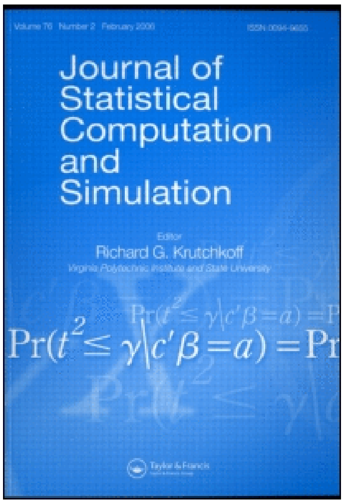

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713650378

A parametric bootstrap solution to the MANOVA under heteroscedasticity

K. Krishnamoorthy[a]; Fei Lu[a]
[a] Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA, USA

First published on: 03 September 2009

To cite this Article Krishnamoorthy, K. and Lu, Fei(2010) 'A parametric bootstrap solution to the MANOVA under heteroscedasticity', Journal of Statistical Computation and Simulation, 80: 8, 873 — 887, First published on: 03 September 2009 (iFirst)
To link to this Article: DOI: 10.1080/00949650902822564
URL: http://dx.doi.org/10.1080/00949650902822564

Journal of Statistical Computation and Simulation
Vol. 80, No. 8, August 2010, 873–887

A parametric bootstrap solution to the MANOVA under heteroscedasticity

K. Krishnamoorthy* and Fei Lu

Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504-1010, USA

(*Received 22 November 2007; final version received 13 February 2009*)

In this article, we consider the problem of comparing several multivariate normal mean vectors when the covariance matrices are unknown and arbitrary positive definite matrices. We propose a parametric bootstrap (PB) approach and develop an approximation to the distribution of the PB pivotal quantity for comparing two mean vectors. This approximate test is shown to be the same as the invariant test given in [Krishnamoorthy and Yu, *Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem*, Stat. Probab. Lett. 66 (2004), pp. 161–169] for the multivariate Behrens–Fisher problem. Furthermore, we compare the PB test with two existing invariant tests via Monte Carlo simulation. Our simulation studies show that the PB test controls Type I error rates very satisfactorily, whereas other tests are liberal especially when the number of means to be compared is moderate and/or sample sizes are small. The tests are illustrated using an example.

Keywords: generalized *p*-value; generalized variable test; Johansen test; moment approximation; modified Nel–Van der Merwe test; Type I error

1. Introduction

The problem of comparing the mean vectors of several multivariate normal populations is referred to as the multivariate analysis of variance (MANOVA). If the population covariance matrices are assumed to be equal, then there are some popular tests available to test the equality of the mean vectors. The tests that are commonly used are Roy's [1] largest root, the Lawley-Hotelling trace [2,3], Wilks' [4] likelihood ratio, and the Pillai–Bartlett trace [5,6]. When there are some departures from the standard assumption, that is, unequal population covariance matrices, Olson [7,8] recommended the Pillai–Bartlett trace because it was most robust to such violations. If only two population means are to be compared assuming that covariance matrices are equal, then a uniformly most powerful invariant test, known as the Hotelling T^2 test, is available. This test, however, may become seriously biased when the assumption of equality of covariance matrices is not satisfied, resulting in spurious decisions about the null hypothesis of equal means. Furthermore, the assumption of variance homogeneity is very unlikely to be satisfied in practice.

*Corresponding author. Email: krishna@louisiana.edu

ISSN 0094-9655 print/ISSN 1563-5163 online

DOI: 10.1080/00949650902822564
http://www.informaworld.com

The problem of making inference about the difference between two normal mean vectors without assuming equality of population covariance matrices is referred to as the multivariate Behrens–Fisher problem, and it has been well addressed in the literature. The usual practice, regarding the choice between the tests for comparing two normal mean vectors, is to first test the equality of covariance matrices, and if the equality is tenable then use the Hotelling T^2 test, otherwise use one of the procedures for the multivariate Behrens–Fisher problem. Recently, Krishnamoorthy and Xia [9] showed that this usual practice may lead to erroneous conclusions. Another criticism of the conventional approach is the appropriateness of the usual variance ratio test; this test is heavily dependent on the normality assumption, and sometimes rejection of the null hypothesis of equality of variances may be attributed to non-normality rather than to inequality of variances. Therefore, test procedures for comparing mean vectors without imposing any assumption on covariance matrices are warranted. Many tests for the multivariate Behrens–Fisher problem have been proposed. We refer to the following tests in the literature: Bennett's [10], Brown and Forsythe [11], James' [12], Yao's [13], Johansen's [14], Nel and Van der Merwe's [15], and Kim's [16] tests. Christensen and Rencher [17] compared seven tests and recommended Kim's test. However, Krishnamoorthy and Yu [18] noted that Kim's test is not non-singular invariant, and recently Park and Sinha [19] pointed out that Kim's test is in general conservative. Recent comparison studies by Krishnamoorthy and Yu [18], Hirokazu and Ke-Hai [20], Park and Sinha [19], and Belloni and Didier [21] indicate that the modified Nel and Van der Merwe's (MNV) test proposed in Krishnamoorthy and Yu [18] is comparable to, or better than, other affine invariant tests. We will see in the sequel that this MNV test is a special case of the parametric bootstrap (PB) test that we propose for the MANOVA problem.

If the covariance matrices are unknown and arbitrary, then the problem of testing equality of the mean vectors is more complex, and only approximate solutions are available. James [12] and Johansen [14] proposed multivariate tests for the situation in which the covariance matrices could be unequal. James' [12] tests, which include a first-order and a second-order approximation to the null distribution, are an extension of his series solution to the univariate Behrens–Fisher problem. Johansen [14] generalized Welch's univariate *approximate degrees of freedom solution* [22,23] to the present problem of comparing several normal mean vectors. All the proposed tests are based on a natural invariant test statistic but used different approaches to approximate its null distribution. Gamage, Mathew, and Weerahandi [24] used the *generalized variable* (GV) approach that was used to solve the multivariate Behrens–Fisher problem. Tang and Algina [25] compared James' first- and second-order tests, Johansen's test, and the Pillai–Bartlett trace and concluded that none of them is satisfactory for all sample size and parameter configurations. Overall, they recommended the James second-order test followed by the Johansen test. Our preliminary study showed that James' second-order test is computationally very involved and offered little improvement over the Johansen test. In particular, the second-order test is difficult to apply when the number of means to be compared is four or more.

In this article, we propose a PB approach for comparing k normal mean vectors when the covariance matrices are unknown and arbitrary positive definite. The PB solution is an extension of our solution to the univariate case [26]. For the univariate case, we found via simulation studies that the PB test was very satisfactory for all sample size and parameter configurations. Indeed, the PB test is the only test that controls Type I error rates when k is moderate or large and the sample sizes are as small as three; other tests, including James' second-order test have inflated Type I error rates for values of k moderate to large and/or the sample sizes are small. In view of our univariate results, it is of interest to develop a PB test for the multivariate case and study its size properties.

This article is organized as follows. In the following section, we provide some preliminaries and distributional results. In Section 3, we outline the Johansen test, the test based on the GV approach [24] and derive a PB pivotal quantity. For the case of $k = 2$, we also provide a moment approximation to the distribution of the PB pivotal quantity. The test based on the moment approximation is the same as the MNV test given in Krishnamoorthy and Yu [18] which is, as mentioned earlier, seems to be the best for the multivariate Behrens–Fisher problem. In Section 4, we compare the Johansen test, the GV test, and the PB test with respect to Type I error rates. Our comparison studies show that the PB test is the only test that performs very satisfactorily for all dimension, sample size and parameter configurations considered. The tests are illustrated using an example in Section 5, and some concluding remarks are given in Section 6.

2. Some preliminaries

Let $\boldsymbol{Y}_{i1}, \ldots, \boldsymbol{Y}_{in_i}$ be a sample from a p-variate normal distribution with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, k$. Assume that all the samples are independent. Let $\bar{\boldsymbol{Y}}_i$ and $\boldsymbol{S}_i$ denote, respectively, the sample mean and sample covariance matrix based on the ith sample. That is,

$$\bar{\boldsymbol{Y}}_i = \frac{1}{n}\sum_{j=1}^{n} Y_{ij} \text{ and } \boldsymbol{S}_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(Y_{ij} - \bar{\boldsymbol{Y}}_i)(Y_{ij} - \bar{\boldsymbol{Y}}_i)', \quad i = 1, \ldots, k. \tag{1}$$

Define $\tilde{\boldsymbol{\Sigma}}_i = 1/n_i \boldsymbol{\Sigma}_i$ and $\tilde{\boldsymbol{S}}_i = 1/n_i \boldsymbol{S}_i$. We note that $\bar{\boldsymbol{Y}}_i$'s and $\tilde{\boldsymbol{S}}_i$'s are mutually independent with

$$\bar{\boldsymbol{Y}}_i \sim N_p\left(\boldsymbol{\mu}_i, \frac{1}{n_i}\boldsymbol{\Sigma}_i\right) \text{ and } \tilde{\boldsymbol{S}}_i \sim W_p\left(n_i - 1, \frac{1}{n_i - 1}\tilde{\boldsymbol{\Sigma}}_i\right), \quad i = 1, \ldots, k, \tag{2}$$

where $W_p(r, \boldsymbol{\Sigma})$ denotes the p-dimensional Wishart distribution with degrees of freedom (df) = r and scale parameter matrix $\boldsymbol{\Sigma}$.

The problem of interest here is to test

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k \quad vs. \quad H_a : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j \text{ for some } i \neq j, \tag{3}$$

based on the sample means and covariance matrices that are minimal sufficient statistics.

Let $\bar{\mathbf{Y}} = (\bar{\boldsymbol{Y}}_1', \ldots, \bar{\boldsymbol{Y}}_k')'$, $\tilde{\mathbf{S}} = \text{diag}(\tilde{\boldsymbol{S}}_1, \ldots, \tilde{\boldsymbol{S}}_k)$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_k')'$, and $\boldsymbol{\Delta} = \text{diag}(\tilde{\boldsymbol{\Sigma}}_1, \ldots, \tilde{\boldsymbol{\Sigma}}_k)$. Under H_0 given in Equation (3), let $\boldsymbol{\mu}_0$ denote the common value of the $\boldsymbol{\mu}_i$'s. If $\boldsymbol{\Sigma}_i$'s are known, then

$$\widehat{\boldsymbol{\mu}}_0 = \left(\sum_{i}^{k} \tilde{\boldsymbol{\Sigma}}_i^{-1}\right)^{-1} \sum_{i=1}^{k} \tilde{\boldsymbol{\Sigma}}_i^{-1}\bar{\mathbf{Y}}_i \tag{4}$$

is the best linear unbiased estimator of $\boldsymbol{\mu}_0$, and a natural test statistic is given by

$$\begin{aligned} T(\bar{\mathbf{Y}}_i; \tilde{\boldsymbol{\Sigma}}_i) &= \sum_{i=1}^{k}(\bar{\mathbf{Y}}_i - \widehat{\boldsymbol{\mu}}_0)'\tilde{\boldsymbol{\Sigma}}_i^{-1}(\bar{\mathbf{Y}}_i - \widehat{\boldsymbol{\mu}}_0) \\ &= \sum_{i=1}^{k} \bar{\mathbf{Y}}_i'\tilde{\boldsymbol{\Sigma}}_i^{-1}\bar{\mathbf{Y}}_i - \left(\sum_{i=1}^{k} \tilde{\boldsymbol{\Sigma}}_i^{-1}\bar{\mathbf{Y}}_i\right)' \left(\sum_{i=1}^{k} \tilde{\boldsymbol{\Sigma}}_i^{-1}\right)^{-1} \left(\sum_{i=1}^{k} \tilde{\boldsymbol{\Sigma}}_i^{-1}\bar{\mathbf{Y}}_i\right) \\ &= \bar{\mathbf{Y}}'\boldsymbol{\Delta}^{-1/2}\left[\mathbf{I}_{kp} - \boldsymbol{\Delta}^{-1/2}\mathbf{J}(\mathbf{J}'\boldsymbol{\Delta}^{-1}\mathbf{J})^{-1}\mathbf{J}'\boldsymbol{\Delta}^{-1/2}\right]\boldsymbol{\Delta}^{-1/2}\bar{\mathbf{Y}}, \end{aligned} \tag{5}$$

where $\mathbf{J} = (\boldsymbol{I}_p, \ldots, \boldsymbol{I}_p)'_{kp\times p}$. Let $\boldsymbol{B} = [\mathbf{I}_{kp} - \boldsymbol{\Delta}^{-1/2}\mathbf{J}(\mathbf{J}'\boldsymbol{\Delta}^{-1}\mathbf{J})^{-1}\mathbf{J}'\boldsymbol{\Delta}^{-1/2}]$. Notice that $\boldsymbol{\Delta}^{-1/2}\bar{\mathbf{Y}} \sim N_{kp}(\boldsymbol{\Delta}^{-1/2}\boldsymbol{\mu}, \boldsymbol{I}_{kp})$ and $\boldsymbol{B}$ is an idempotent matrix with rank $kp - p = p(k-1)$, and so

$$T(\bar{\mathbf{Y}}_i; \tilde{\boldsymbol{\Sigma}}_i) \sim \chi^2_{p(k-1)}(\boldsymbol{\mu}'\boldsymbol{\Delta}^{-1}\boldsymbol{B}\boldsymbol{\Delta}^{-1}\boldsymbol{\mu}),$$

where $\chi^2_m(\delta)$ denotes the non-central chi-square random variable with $df = m$ and the non-centrality parameter δ. We also observe that the noncentrality parameter $\boldsymbol{\mu}'\boldsymbol{\Delta}^{-1}\boldsymbol{B}\boldsymbol{\Delta}^{-1}\boldsymbol{\mu} = \sum_{i=1}^{k}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)'\tilde{\boldsymbol{\Sigma}}_i(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)$, and is equal to zero only when $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$.

If $\tilde{\boldsymbol{\Sigma}}_i$'s are unknown, then replacing them in Equation (5) by $\tilde{\boldsymbol{S}}_i$'s, we can get the test statistic $T(\bar{\boldsymbol{Y}}_i; \tilde{\boldsymbol{S}}_i)$. Letting $\boldsymbol{W}_i = \tilde{\boldsymbol{S}}_i^{-1}$, $i = 1, \ldots, k$ and $\boldsymbol{W} = \sum_{i=1}^{k}\boldsymbol{W}_i$, and defining

$$\widehat{\boldsymbol{\mu}}_0^* = \boldsymbol{W}^{-1}\sum_{i=1}^{k}\boldsymbol{W}_i\bar{\boldsymbol{Y}}_i,$$

we can express

$$\begin{aligned} T(\bar{\boldsymbol{Y}}_i; \tilde{\boldsymbol{S}}_i) &= \sum_{i=1}^{k}(\bar{\boldsymbol{Y}}_i - \widehat{\boldsymbol{\mu}}_0^*)'\boldsymbol{W}_i(\bar{\boldsymbol{Y}}_i - \widehat{\boldsymbol{\mu}}_0^*) \\ &= \sum_{i=1}^{k}\bar{\boldsymbol{Y}}_i'\boldsymbol{W}_i\bar{\boldsymbol{Y}}_i - \left(\sum_{i=1}^{k}\boldsymbol{W}_i\bar{\boldsymbol{Y}}_i\right)'\boldsymbol{W}^{-1}\left(\sum_{i=1}^{k}\boldsymbol{W}_i\bar{\boldsymbol{Y}}_i\right). \end{aligned} \quad (6)$$

3. The tests

We shall now describe three tests that use $T(\bar{\boldsymbol{Y}}_i; \tilde{\boldsymbol{S}}_i) = T(\bar{\boldsymbol{Y}}_i, \ldots, \bar{\boldsymbol{Y}}_k; \tilde{\boldsymbol{S}}_1, \ldots, \tilde{\boldsymbol{S}}_k)$ in Equation (6) as a test statistic.

3.1. *Johansen's test*

Johansen's [14] test is based on the test statistic

$$J_{\text{OH}} = \frac{T(\bar{\boldsymbol{Y}}_1, \ldots, \bar{\boldsymbol{Y}}_k; \tilde{\boldsymbol{S}}_1, \ldots, \tilde{\boldsymbol{S}}_k)}{c}, \quad (7)$$

where

$$c = p(k-1) + 2A - \frac{6A}{p(k-1)+2} \quad (8)$$

and

$$A = \sum_{i=1}^{k}\frac{\text{tr}(\boldsymbol{I} - \boldsymbol{W}^{-1}\boldsymbol{W}_i)^2 + \left[\text{tr}(\boldsymbol{I} - \boldsymbol{W}^{-1}\boldsymbol{W}_i)\right]^2}{2(n_i - 1)}. \quad (9)$$

Johansen showed that, under H_0, J_{OH} is approximately distributed as F_{f_1,f_2} random variable, where the dfs $f_1 = p(k-1)$ and $f_2 = p(k-1)[p(k-1)+2]/(3A)$.

Thus, the Johansen test rejects the null hypothesis in Equation (3) whenever $J_{\text{OH}} > F_{f_1,f_2,1-\alpha}$, where $F_{m,n;q}$ denotes the qth quantile of an F distribution with dfs m and n.

3.2. *The generalized variable test*

Gamage *et al.* [24] proposed a test referred as the GV test, which is based on the concept of generalized p-value introduced by Tsui and Weerahandi [27]. To describe the test, let $(\bar{\boldsymbol{y}}_i, \tilde{\boldsymbol{s}}_i)$ be an observed value of $(\bar{\boldsymbol{Y}}_i, \tilde{\boldsymbol{S}}_i)$, and let

$$\boldsymbol{R}_i^* = \left[\tilde{\boldsymbol{s}}_i^{-1/2}\tilde{\boldsymbol{\Sigma}}_i\tilde{\boldsymbol{s}}_i^{-1/2}\right]^{-1/2}\left[\tilde{\boldsymbol{s}}_i^{-1/2}\tilde{\boldsymbol{S}}_i\tilde{\boldsymbol{s}}_i^{-1/2}\right]\left[\tilde{\boldsymbol{s}}_i^{-1/2}\tilde{\boldsymbol{\Sigma}}_i\tilde{\boldsymbol{s}}_i^{-1/2}\right]^{-1/2}, \quad i = 1, \ldots, k. \quad (10)$$

For given $\tilde{\boldsymbol{s}}_i$'s, $\boldsymbol{R}_i^*$'s are independent, and as $\tilde{\boldsymbol{s}}_i^{-1/2}\tilde{\boldsymbol{S}}_i\tilde{\boldsymbol{s}}_i^{-1/2} \sim W_p(n_i - 1, \tilde{\boldsymbol{s}}_i^{-1/2}\tilde{\boldsymbol{\Sigma}}_i\tilde{\boldsymbol{s}}_i^{-1/2})$, $\boldsymbol{R}_i^* \sim W_p(n_i - 1, 1/(n_i-1)\boldsymbol{I}_p)$, $i = 1, \ldots, k$.

The generalized test variable is defined as

$$G = \frac{T(\bar{\boldsymbol{Y}}_1, \ldots, \bar{\boldsymbol{Y}}_k; \tilde{\boldsymbol{\Sigma}}_1, \ldots, \tilde{\boldsymbol{\Sigma}}_k)}{T\left(\bar{\boldsymbol{y}}_1, \ldots, \bar{\boldsymbol{y}}_k; \tilde{\boldsymbol{s}}_1^{1/2}\boldsymbol{R}_1^{*-1}\tilde{\boldsymbol{s}}_1^{1/2}, \ldots, \tilde{\boldsymbol{s}}_k^{1/2}\boldsymbol{R}_k^{*-1}\tilde{\boldsymbol{s}}_k^{1/2}\right)}. \tag{11}$$

It follows from Equation (5) that, under H_0, $T(\bar{\boldsymbol{Y}}_1, \ldots, \bar{\boldsymbol{Y}}_k; \tilde{\boldsymbol{\Sigma}}_1, \ldots, \tilde{\boldsymbol{\Sigma}}_k)$ has a chi-square distribution with df $= p(k-1)$. Furthermore, for a given $(\bar{\boldsymbol{y}}_1, \ldots, \bar{\boldsymbol{y}}_k; \tilde{\boldsymbol{s}}_1, \ldots, \tilde{\boldsymbol{s}}_k)$, the *generalized p-value* is given by

$$P_{\chi^2_{p(k-1)}, \boldsymbol{R}_1^*, \ldots, \boldsymbol{R}_k^*}\left(\frac{\chi^2_{p(k-1)}}{T_N(\bar{\boldsymbol{y}}_1, \ldots, \bar{\boldsymbol{y}}_k; \tilde{\boldsymbol{s}}_1^{1/2}\boldsymbol{R}_1^{*-1}\tilde{\boldsymbol{s}}_1^{1/2}, \ldots, \tilde{\boldsymbol{s}}_k^{1/2}\boldsymbol{R}_k^{*-1}\tilde{\boldsymbol{s}}_k^{1/2})} > 1\right). \tag{12}$$

The GV test rejects the null hypothesis in Equation (3) whenever the generalized p-value in Equation (12) is less than a given nominal level α. Notice that, for a given $(\bar{\boldsymbol{y}}_1, \ldots, \bar{\boldsymbol{y}}_k; \tilde{\boldsymbol{s}}_1, \ldots, \tilde{\boldsymbol{s}}_k)$, the probability distribution in Equation (12) does not depend on any unknown parameters, so the generalized p-value can be estimated using Monte Carlo simulation. An unappealing feature of this test is that it is not non-singular invariant.

3.3. *The parametric bootstrap (PB) test*

The PB test involves sampling from the estimated models. That is, samples or sample statistics are generated from parametric models with the parameters replaced by their estimates, and the generated samples are used to approximate the null distribution of a test statistic. Recall that under $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$ all $\bar{\boldsymbol{Y}}_i$'s have the same mean. As the test statistic T in Equation (6) is location invariant, without loss of generality, we can take this common mean to be the vector of zeroes to find the null distribution of T. Using these facts, the PB pivotal quantity can be obtained as follows. Let $\bar{\boldsymbol{Y}}_{Bi} \sim N_p(\boldsymbol{0}, \tilde{\boldsymbol{s}}_i)$ and $\tilde{\boldsymbol{S}}_{B_i} \sim W_p(n_i - 1, (1/(n_i-1))\tilde{\boldsymbol{s}}_i)$, where $(\tilde{\boldsymbol{s}}_1, \ldots, \tilde{\boldsymbol{s}}_k)$ is an observed value of $(\tilde{\boldsymbol{S}}_1, \ldots, \tilde{\boldsymbol{S}}_k)$. In terms of these random quantities, we define the PB pivotal quantity as

$$T_B(\bar{\boldsymbol{Y}}_{Bi}, \tilde{\boldsymbol{S}}_{Bi}) = \sum_{i=1}^{k}(\bar{\boldsymbol{Y}}_{Bi} - \widehat{\boldsymbol{\mu}}_B^*)'\tilde{\boldsymbol{S}}_{Bi}^{-1}(\bar{\boldsymbol{Y}}_{Bi} - \widehat{\boldsymbol{\mu}}_B^*), \tag{13}$$

where

$$\widehat{\boldsymbol{\mu}}_B^* = \left(\sum_{i}^{k}\tilde{\boldsymbol{S}}_{B_i}^{-1}\right)^{-1}\sum_{i=1}^{k}\tilde{\boldsymbol{S}}_{B_i}^{-1}\bar{\boldsymbol{Y}}_{B_i}, \tag{14}$$

or equivalently,

$$T_B(\bar{\boldsymbol{Y}}_{Bi}, \tilde{\boldsymbol{S}}_{Bi}) = \sum_{i=1}^{k}\bar{\boldsymbol{Y}}_{Bi}'\tilde{\boldsymbol{S}}_{Bi}^{-1}\bar{\boldsymbol{Y}}_{Bi} - \left(\sum_{i=1}^{k}\bar{\boldsymbol{Y}}_{Bi}'\tilde{\boldsymbol{S}}_{Bi}^{-1}\right)\left(\sum_{i=1}^{k}\tilde{\boldsymbol{S}}_{Bi}^{-1}\right)^{-1}\left(\sum_{i=1}^{k}\tilde{\boldsymbol{S}}_{Bi}^{-1}\bar{\boldsymbol{Y}}_{Bi}\right). \tag{15}$$

For an observed value T_0 of T in Equation (6), the PB p-value is defined as

$$P(T_B(\bar{\boldsymbol{Y}}_{Bi}, \tilde{\boldsymbol{S}}_{Bi}) > T_0), \tag{16}$$

and the null hypothesis in Equation (3) is rejected whenever the above p-value is less than the nominal level α. Notice that, for a given $(\tilde{\boldsymbol{s}}_1, \ldots, \tilde{\boldsymbol{s}}_k)$, the probability in Equation (16) does not depend on any unknown parameters, and so it can be estimated using the Monte Carlo simulation as described below.

Let $\boldsymbol{t}_i$ be the Cholesky factor of $\tilde{\boldsymbol{s}}_i$, so that $\tilde{\boldsymbol{s}}_i = \boldsymbol{t}_i\boldsymbol{t}_i'$, $i = 1, \ldots, k$. Then $\bar{\boldsymbol{Y}}_{Bi} \sim \boldsymbol{t}_i\boldsymbol{Z}_i$ and $\tilde{\boldsymbol{S}}_{Bi} \sim \boldsymbol{t}_i\boldsymbol{V}_i\boldsymbol{t}_i'/(n_i - 1)$, where $\boldsymbol{Z}_i$ and $\boldsymbol{V}_i$ are independent with $\boldsymbol{Z}_i \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ and $\boldsymbol{V}_i \sim W_p(n_i - 1, \boldsymbol{I}_p)$. In terms of these variables, we see that the PB pivotal quantity in Equation (15) is distributed as

$$T_B(\boldsymbol{Z}_i, \boldsymbol{V}_i) = \sum_{i=1}^{k} f_i\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i - \left(\sum_{i=1}^{k} f_i\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\boldsymbol{t}_i^{-1}\right)\left(\sum_{i=1}^{k} f_i(\boldsymbol{t}_i\boldsymbol{V}_i\boldsymbol{t}_i')^{-1}\right)^{-1}\left(\sum_{i=1}^{k} f_i\boldsymbol{t}_i'^{-1}\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i\right), \tag{17}$$

where $f_i = n_i - 1$, $i = 1, \ldots, k$.

For a given dimension p, values of k, $(n_1, n_2, \ldots, n_k)$, $(\bar{\boldsymbol{y}}_1, \ldots, \bar{\boldsymbol{y}}_k)$, and $(\tilde{\boldsymbol{s}}_1, \ldots, \tilde{\boldsymbol{s}}_k)$, the PB p-value can be estimated using the following steps.

(1) Compute the observed value T_0 using Equation (6).
(2) Compute the Cholesky factor $\boldsymbol{t}_i$, so that $\boldsymbol{t}\boldsymbol{t}_i' = \tilde{\boldsymbol{s}}_i$, $i = 1, \ldots, k$.
(3) Generate $\boldsymbol{Z}_i \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ and $\boldsymbol{V}_i \sim W_p(n_i - 1, \boldsymbol{I}_p)$, $i = 1, \ldots, k$.
(4) Set $\bar{\boldsymbol{Y}}_{Bi} = \boldsymbol{t}_i\boldsymbol{Z}_i$ and $\tilde{\boldsymbol{S}}_{Bi} = (\boldsymbol{t}_i\boldsymbol{V}_i\boldsymbol{t}_i')/(n_i - 1)$, $i = 1, \ldots, k$.
(5) Compute $T_B(\boldsymbol{Z}_i, \boldsymbol{V}_i)$ using Equation (17).
(6) Repeat the steps 3, 4, and 5 for a large number (say, $M = 10{,}000$) of times.

The proportion of times T_B exceeds the observed value T_0 is an estimate of the PB p-value defined in Equation (16).

An approximation to the PB test: For $k = 2$, we can find an approximation to the distribution of the PB pivotal quantity as follows. For $k = 2$, the T_B in Equation (13) can be expressed as

$$T_B = (\bar{\boldsymbol{Y}}_{B1} - \bar{\boldsymbol{Y}}_{B2})'\left(\tilde{\boldsymbol{S}}_{B1} + \tilde{\boldsymbol{S}}_{B2}\right)^{-1}(\bar{\boldsymbol{Y}}_{B1} - \bar{\boldsymbol{Y}}_{B2}).$$

Recall that $\bar{\boldsymbol{Y}}_{B1} - \bar{\boldsymbol{Y}}_{B2} \sim N_p(\boldsymbol{0}, \tilde{\boldsymbol{s}}_1 + \tilde{\boldsymbol{s}}_2)$ and $\tilde{\boldsymbol{S}}_{B_i} \sim W_p(n_i - 1, (1/(n_i - 1))\tilde{\boldsymbol{s}}_i)$, and so

$$\begin{aligned} T_B &\sim \boldsymbol{Z}'(\tilde{\boldsymbol{s}}_1 + \tilde{\boldsymbol{s}}_2)^{1/2}(\tilde{\boldsymbol{S}}_{B1} + \tilde{\boldsymbol{S}}_{B2})^{-1}(\tilde{\boldsymbol{s}}_1 + \tilde{\boldsymbol{s}}_2)^{1/2}\boldsymbol{Z} \\ &= \boldsymbol{Z}'(\boldsymbol{Q}_1 + \boldsymbol{Q}_2)^{-1}\boldsymbol{Z}, \end{aligned} \tag{18}$$

where $\boldsymbol{Q}_i = (\tilde{\boldsymbol{s}}_1 + \tilde{\boldsymbol{s}}_2)^{-1/2}\tilde{\boldsymbol{S}}_{Bi}(\tilde{\boldsymbol{s}}_1 + \tilde{\boldsymbol{s}}_2)^{-1/2} \sim W_p(n_i - 1, (\tilde{\boldsymbol{s}}_1 + \tilde{\boldsymbol{s}}_2)^{-1/2}\tilde{\boldsymbol{s}}_i(\tilde{\boldsymbol{s}}_1 + \tilde{\boldsymbol{s}}_2)^{-1/2}/(n_i - 1))$, $i = 1, 2$. As $E(\boldsymbol{Q}_1 + \boldsymbol{Q}_2) = \boldsymbol{I}_p$, a reasonable approximation to the distribution of $\boldsymbol{Q}_1 + \boldsymbol{Q}_2$ could be $W_p(\nu, (1/\nu)\boldsymbol{I}_p)$, where ν is to be determined by matching the expectation of $\mathrm{tr}(\boldsymbol{Q}_1 + \boldsymbol{Q}_2)^2$ with that of $\mathrm{tr}(\boldsymbol{Q}^2)$, where $\boldsymbol{Q} \sim W_p(\nu, (1/\nu)\boldsymbol{I}_p)$. By matching these expectations (see the Appendix), we found

$$\nu = \frac{p^2 + p}{1/(n_1 - 1)\left\{\mathrm{tr}\left[\left(\tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}^{-1}\right)^2\right] + \left[\mathrm{tr}\left(\tilde{\boldsymbol{s}}_1\tilde{s}^{-1}\right)\right]^2\right\} + 1/(n_2 - 1)\left\{\mathrm{tr}\left[\left(\tilde{\boldsymbol{s}}_2\tilde{\boldsymbol{s}}^{-1}\right)^2\right] + \left[\mathrm{tr}\left(\tilde{\boldsymbol{s}}_2\tilde{\boldsymbol{s}}^{-1}\right)\right]^2\right\}}, \tag{19}$$

where $\tilde{\boldsymbol{s}} = \tilde{\boldsymbol{s}}_1 + \tilde{\boldsymbol{s}}_2$. Thus, we conclude that

$$\boldsymbol{Q}_1 + \boldsymbol{Q}_2 \sim W_p\left(\nu, \frac{1}{\nu}\boldsymbol{I}_p\right) \text{ approximately}$$

and independently of $\boldsymbol{Z}$ in Equation (18). Note that

$$\boldsymbol{Z}'(\boldsymbol{Q}_1 + \boldsymbol{Q}_2)^{-1}\boldsymbol{Z} = \frac{\boldsymbol{Z}'\boldsymbol{Z}}{\boldsymbol{Z}'\boldsymbol{Z}/\boldsymbol{Z}'(\boldsymbol{Q}_1 + \boldsymbol{Q}_2)^{-1}\boldsymbol{Z}}, \tag{20}$$

and $\boldsymbol{Z}'\boldsymbol{Z} \sim \chi^2_p$ independently of $\boldsymbol{Z}'\boldsymbol{Z}/\boldsymbol{Z}'(\boldsymbol{Q}_1 + \boldsymbol{Q}_2)^{-1}\boldsymbol{Z}$, which is distributed as $\chi^2_{\nu-p+1}/\nu$ approximately (see [28, p. 98], for the distributional results of the Hotelling T^2). Thus, it follows from Equation (20) that

$$\boldsymbol{Z}'(\boldsymbol{Q}_1 + \boldsymbol{Q}_2)^{-1}\boldsymbol{Z} \sim \frac{\nu p}{\nu - p + 1} F_{p,\nu-p+1} \text{ approximately.}$$

We reject the null hypothesis in Equation (3) whenever an observed value of the test statistic in Equation (6) is greater than or equal to $\nu p F_{p,\nu-p+1;1-\alpha}/(\nu - p + 1)$. This approximate test is the same as the invariant test given in Krishnamoorthy and Yu [18] who obtained it by modifying the Nel–Van der Merwe [15] test. Furthermore, Krishnamoorthy and Yu showed that $\min\{n_1 - 1, n_2 - 1\} \le \nu \le n_1 + n_2 - 2$.

4. Monte Carlo studies

As all the tests are location invariant, we can take $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k)'$ to be vector of zeroes for evaluating Type I error rates. For the case of comparing two group means using invariant tests, we can assume $\boldsymbol{\Sigma}_1$ to be identity matrix, and $\boldsymbol{\Sigma}_2$ to be $\mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, where λ_i's are the eigenvalues of $\Sigma_1^{-1}\Sigma_2$. This is because there exists a non-singular matrix $\mathbf{M}$ such that $\boldsymbol{\Sigma}_1 = \mathbf{M}\mathbf{M}'$ and $\boldsymbol{\Sigma}_2 = \mathbf{M}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_p)\mathbf{M}'$, and the test procedures are affine invariant. For the case of comparing more than two population mean vectors, the parameter space cannot be simplified much, except that we can take $\boldsymbol{\Sigma}_1 = \boldsymbol{I}_p$, $\boldsymbol{\Sigma}_2 = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, and other matrices are arbitrary positive definite. Even though the GV test is not non-singular invariant, for simplicity and convenience we shall estimate the sizes of the tests for the parameter space described above. To compute the sizes of the various tests via Monte Carlo simulation, we used the IMSL subroutine RNMVN to generate $p \times 1$ multivariate normal random vectors and the *Applied Statistics* Algorithm (AS 53) due to Smith and Hocking [29] to generate Wishart random matrices. To estimate the sizes of the Johansen test, we used simulation consisting of 10,000 runs. Notice that two nested 'do loops' are required to estimate the sizes of the GV and PB tests; we used 2500 runs for outer 'do loop' (for generating the observed statistics $(\bar{\boldsymbol{y}}_1, \ldots, \bar{\boldsymbol{y}}_k)$ and $(\tilde{\boldsymbol{s}}_1, \ldots, \tilde{\boldsymbol{s}}_k)$) and 5000 runs for 'inner loop' (for generating standard normal random vectors and Wishart matrices).

Several articles compared different solutions for the multivariate Behrens–Fisher problem, that is, for the case of $k = 2$. As noted in the introduction, the modified Nel–Van der Merwe test proposed in Krishnamoorthy and Yu [18] seems to be one of the best invariant tests. As the approximation to the PB test in the preceding section is the same as the modified Nel–Van der Merwe test, comparison study for the case of $k = 2$ is not necessary, and we shall compare the tests when the number of groups is three or more.

For the bivariate case, Type I error rates are estimated for $k = 3$ and 5, and are presented in Table 1. We observe from this table that for smaller sample sizes, the Johansen test and the GV test have inflated Type I error rates. The GV test appears to be liberal even when the sample sizes are moderately large and not much different, and Johansen's test appears to be satisfactory in these cases. It is also clear that the PB test controls Type I error rates (close to the nominal level 0.05) and behaves like an exact test for all sample size and parameter configurations considered. To compare the tests for $k = 10$, for convenience and simplicity, we take the covariance matrices are of the form $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $-1 < \rho^2 < 1$. In this case, Johansen's test appears to be satisfactory only when the sample sizes are moderate ($\ge$20) and not much different. The GV test appears to be liberal even for moderate samples. In particular, Type I error rates of the GV test exceed 0.4 for small samples (see the case of $n_1 = \ldots = n_{10} = 5$). The behaviours of the GV test are similar to those given in Krishnamoorthy *et al.* [26] for the univariate case.

Table 1. Monte Carlo estimates of Type I error rates for comparing bivariate normal mean vectors.

$k = 3,\ p = 2,\ \mathbf{\Sigma}_1 = \boldsymbol{I}_2,\ \mathbf{\Sigma}_2 = \text{diag}(\lambda_1, \lambda_2),\ \mathbf{\Sigma}_3 = \begin{pmatrix} 1 & \rho_3 \\ \rho_3 & 1 \end{pmatrix}$				
(n_1, n_2, n_3)	$(\lambda_1, \lambda_2, \rho_3)$	Johansen	GV	PB
(7, 7, 7)	(1,1,0)	0.057	0.054	0.052
	(1,0.9,0.1)	0.054	0.052	0.041
	(1,0.5,0.2)	0.058	0.047	0.049
	(1,0.1,0.3)	0.068	0.061	0.056
	(0.2,0.6,0.5)	0.067	0.073	0.048
	(0.9,0.9,0.6)	0.055	0.056	0.044
	(0.7,0.8,−0.2)	0.056	0.058	0.045
(7, 10, 20)	(1,1,0)	0.060	0.095	0.052
	(1,0.9,0.1)	0.059	0.072	0.056
	(1,0.5,0.2)	0.062	0.089	0.054
	(1,0.1,0.3)	0.070	0.086	0.056
	(0.2,0.6,0.5)	0.067	0.078	0.056
	(0.9,0.9,0.6)	0.061	0.096	0.055
	(0.7,0.8,−0.2)	0.064	0.079	0.054
(10, 10, 10)	(1,1,0)	0.052	0.055	0.043
	(1,0.9,0.1)	0.052	0.050	0.040
	(1,0.5,0.2)	0.052	0.054	0.039
	(1,0.1,0.3)	0.058	0.053	0.054
	(0.2,0.6,0.5)	0.056	0.058	0.049
	(0.9,0.9,0.6)	0.050	0.046	0.054
	(0.7,0.8,−0.2)	0.052	0.045	0.052
(10, 10, 40)	(1,1,0)	0.055	0.100	0.058
	(1,0.9,0.1)	0.055	0.090	0.049
	(1,0.5,0.2)	0.054	0.093	0.043
	(1,0.1,0.3)	0.055	0.096	0.052
	(0.2,0.6,0.5)	0.054	0.110	0.052
	(0.9,0.9,0.6)	0.055	0.111	0.057
	(0.7,0.8,−0.2)	0.055	0.099	0.045
(25, 20, 20)	(1,1,0)	0.049	0.043	0.041
	(1,0.9,0.1)	0.050	0.059	0.049
	(1,0.5,0.2)	0.049	0.048	0.051
	(1,0.1,0.3)	0.052	0.054	0.052
	(0.2,0.6,0.5)	0.053	0.054	0.051
	(0.9,0.9,0.6)	0.050	0.059	0.052
	(0.7,0.8,−0.2)	0.050	0.058	0.054
$k = 5,\ p = 2,\ \mathbf{\Sigma}_1 = \boldsymbol{I}_2,\ \mathbf{\Sigma}_2 = \text{diag}(\lambda_1, \lambda_2),\ \mathbf{\Sigma}_i = \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix},\ i = 3, 4, 5$				
$(n_1, \ldots, n_5)$	$(\lambda_1, \lambda_2, \rho_3, \rho_4, \rho_5)$	Johansen	GV	PB
(7,7,7,7,7)	(1, 1, 0.5, 0.5, 0.5)	0.071	0.104	0.050
	(0.1, 0.1, 0.3, 0.3, 0.3)	0.072	0.114	0.050
	(0.1, 0.7, 0, 0, 0)	0.072	0.113	0.048
	(0.1, 0.9, 0.1, 0.4, 0.9)	0.074	0.120	0.048
	(0.1, 0.3, −0.1, 0.1, 0.9)	0.076	0.126	0.051
	(0.4, 0.4, 0.5, −0.3, 0.4, 0.3)	0.072	0.119	0.053
	(0.9, 0.9, −0.4, 0.6, 0.9)	0.077	0.138	0.052
(12,12,12,12,12)	(1, 1, 0.5, 0.5, 0.5)	0.055	0.075	0.050
	(0.1, 0.1, 0.3, 0.3, 0.3)	0.056	0.078	0.053
	(0.1, 0.7, 0, 0, 0)	0.056	0.084	0.052
	(0.1, 0.9, 0.1, 0.4, 0.9)	0.056	0.084	0.051
	(0.1, 0.3, −0.1, 0.1, 0.9)	0.057	0.083	0.050
	(0.4, 0.4, 0.5, −0.3, 0.4, 0.3)	0.056	0.087	0.050
	(0.9, 0.9, −0.4, 0.6, 0.9)	0.057	0.082	0.047

(*Continued*)

Table 1. Continued.

(n_1, n_2, n_3)	$(\lambda_1, \lambda_2, \rho_3)$	Johansen	GV	PB
(20,20,20,20,20)	(1, 1, 0.5, 0.5, 0.5)	0.053	0.054	0.054
	(0.1, 0.1, 0.3, 0.3, 0.3)	0.051	0.057	0.051
	(0.1, 0.7, 0, 0, 0)	0.052	0.065	0.052
	(0.1, 0.9, 0.1, 0.4, 0.9)	0.052	0.061	0.047
	(0.1, 0.3, −0.1, 0.1, 0.9)	0.052	0.062	0.051
	(0.4, 0.4, 0.5, −0.3, 0.4, 0.3)	0.052	0.055	0.053
	(0.9, 0.9, −0.4, 0.6, 0.9)	0.053	0.068	0.048
(15,20,10,32,7)	(1, 1, 0.5, 0.5, 0.5)	0.068	0.114	0.047
	(0.1, 0.1, 0.3, 0.3, 0.3)	0.068	0.139	0.049
	(0.1, 0.7, 0, 0, 0)	0.068	0.099	0.054
	(0.1, 0.9, 0.1, 0.4, 0.9)	0.063	0.117	0.052
	(0.1, 0.3, −0.1, 0.1, 0.9)	0.062	0.111	0.051
	(0.4, 0.4, 0.5, −0.3, 0.4, 0.3)	0.067	0.099	0.049
	(0.9, 0.9, −0.4, 0.6, 0.9)	0.064	0.111	0.051

$$k = 10,\ p = 2,\ \boldsymbol{\Sigma}_i = \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix},\ i = 1, \ldots, 10$$

$(n_1, \ldots, n_{10})$	$(\rho_1, \ldots, \rho_{10})$	Johansen	GV	PB
(5,5,5,5,5,	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)	0.205	0.429	0.030
5,5,5,5,5)	(0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9)	0.216	0.415	0.044
	(0.1, 0.2, 0.1, 0.2, 0.9, 0.9, 0.9, −0.9, −0.8, 0.5)	0.217	0.453	0.047
	(0.1, 0.1, 0.1, −0.2, −0.2, −0.2, 0, 0, 0, 0)	0.205	0.427	0.045
	(0.1, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9, 0.9, 0.9)	0.219	0.426	0.047
	(−0.1, 0.1, −0.4, 0.4, −0.5, 0.5, −0.7, 0.7, −0.9, 0.9)	0.221	0.415	0.046
	(0.5, 0.5, 0.5, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9)	0.210	0.431	0.045
(10,10,10,10,10,	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)	0.079	0.156	0.041
10,10,10,10,10)	(0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9)	0.082	0.168	0.050
	(0.1, 0.2, 0.1, 0.2, 0.9, 0.9, 0.9, −0.9, −0.8, 0.5)	0.078	0.170	0.051
	(0.1, 0.1, 0.1, −0.2, −0.2, −0.2, 0, 0, 0, 0)	0.077	0.172	0.052
	(0.1, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9, 0.9, 0.9)	0.079	0.166	0.047
	(−0.1, 0.1, −0.4, 0.4, −0.5, 0.5, −0.7, 0.7, −0.9, 0.9)	0.076	0.167	0.048
	(0.5, 0.5, 0.5, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9)	0.079	0.171	0.051
(10,7,12,7,11,	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)	0.096	0.218	0.049
10,8,12,7,15)	(0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9)	0.098	0.209	0.051
	(0.1, 0.2, 0.1, 0.2, 0.9, 0.9, 0.9, −0.9, −0.8, 0.5)	0.096	0.215	0.048
	(0.1, 0.1, 0.1, −0.2, −0.2, −0.2, 0, 0, 0, 0)	0.093	0.198	0.051
	(0.1, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9, 0.9, 0.9)	0.094	0.212	0.050
	(−0.1, 0.1, −0.4, 0.4, −0.5, 0.5, −0.7, 0.7, −0.9, 0.9)	0.092	0.208	0.047
	(0.5, 0.5, 0.5, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9)	0.095	0.218	0.051
(10,10,10,5,5,	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)	0.151	0.288	0.055
5,20,20,20,20)	(0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9)	0.155	0.261	0.056
	(0.1, 0.2, 0.1, 0.2, 0.9, 0.9, 0.9, −0.9, −0.8, 0.5)	0.154	0.287	0.061
	(0.1, 0.1, 0.1, −0.2, −0.2, −0.2, 0, 0, 0, 0)	0.147	0.277	0.060
	(0.1, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9, 0.9, 0.9)	0.150	0.291	0.057
	(−0.1, 0.1, −0.4, 0.4, −0.5, 0.5, −0.7, 0.7, −0.9, 0.9)	0.170	0.256	0.058
	(0.5, 0.5, 0.5, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9)	0.161	0.281	0.054
(25,23,20,27,21,	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)	0.052	0.082	0.051
25,26,22,20,25)	(0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9)	0.055	0.078	0.048
	(0.1, 0.2, 0.1, 0.2, 0.9, 0.9, 0.9, −0.9, −0.8, 0.5)	0.049	0.102	0.051
	(0.1, 0.1, 0.1, −0.2, −0.2, −0.2, 0, 0, 0, 0)	0.057	0.107	0.050
	(0.1, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9, 0.9, 0.9)	0.056	0.081	0.047
	(−0.1, 0.1, −0.4, 0.4, −0.5, 0.5, −0.7, 0.7, −0.9, 0.9)	0.056	0.087	0.048
	(0.5, 0.5, 0.5, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9)	0.051	0.101	0.052

Table 2. Monte Carlo estimates of Type I error rates for comparing trivariate normal mean vectors.

$$k = 3,\ p = 3\ \boldsymbol{\Sigma}_1 = \boldsymbol{I}_3,\ \boldsymbol{\Sigma}_2 = \text{diag}(\lambda_1, \lambda_2, \lambda_3),\ \boldsymbol{\Sigma}_3 = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

(n_1, n_2, n_3)	$(\lambda_1, \lambda_2, \lambda_3, \rho)$	Johansen	GV	PB
(7,7,7)	(1,1,1,0)	0.079	0.077	0.039
	(1,1,0.1,0.1)	0.090	0.081	0.042
	(1,0.1,0.1,0.5)	0.105	0.095	0.049
	(0.2,0.6,0.9,−0.3)	0.088	0.085	0.047
	(0.6,0.6,0.6,0)	0.084	0.074	0.045
	(0.3,0.9,0.1,−0.1)	0.100	0.082	0.045
	(0.8,0.5,0.5,0.1)	0.085	0.080	0.047
(7,10,20)	(1,1,1,0)	0.080	0.111	0.054
	(1,1,0.1,0.1)	0.091	0.112	0.057
	(1,0.1,0.1,0.5)	0.110	0.118	0.062
	(0.2,0.6,0.9,−0.3)	0.093	0.113	0.060
	(0.6,0.6,0.6,0)	0.086	0.107	0.054
	(0.3,0.9,0.1,−0.1)	0.099	0.111	0.052
	(0.8,0.5,0.5,0.1)	0.088	0.116	0.057
(10,10,10)	(1,1,1,0)	0.057	0.063	0.045
	(1,1,0.1,0.1)	0.063	0.069	0.058
	(1,0.1,0.1,0.5)	0.069	0.069	0.052
	(0.2,0.6,0.9,−0.3)	0.061	0.070	0.055
	(0.6,0.6,0.6,0)	0.060	0.066	0.052
	(0.3,0.9,0.1,−0.1)	0.067	0.078	0.057
	(0.8,0.5,0.5,0.1)	0.062	0.071	0.050
(10,10,40)	(1,1,1,0)	0.063	0.131	0.060
	(1,1,0.1,0.1)	0.063	0.146	0.056
	(1,0.1,0.1,0.5)	0.065	0.132	0.051
	(0.2,0.6,0.9,−0.3)	0.064	0.128	0.059
	(0.6,0.6,0.6,0)	0.062	0.137	0.051
	(0.3,0.9,0.1,−0.1)	0.065	0.138	0.056
	(0.8,0.5,0.5,0.1)	0.063	0.146	0.058
(25,20,20)	(1,1,1,0)	0.050	0.057	0.056
	(1,1,0.1,0.1)	0.053	0.054	0.052
	(1,0.1,0.1,0.5)	0.055	0.063	0.047
	(0.2,0.6,0.9,−0.3)	0.052	0.057	0.054
	(0.6,0.6,0.6,0)	0.051	0.045	0.052
	(0.3,0.9,0.1,−0.1)	0.054	0.053	0.059
	(0.8,0.5,0.5,0.1)	0.051	0.042	0.048

$$k = 5,\ p = 3,\ \boldsymbol{\Sigma}_1 = \boldsymbol{I}_3,\ \boldsymbol{\Sigma}_2 = \text{diag}(\lambda_1, \lambda_2, \lambda_3),\ \boldsymbol{\Sigma}_i = \begin{pmatrix} 1 & \rho_i & \rho_i \\ \rho_i & 1 & \rho_i \\ \rho_i & \rho_i & 1 \end{pmatrix},\ i = 3, 4, 5$$

$(n_1, \ldots, n_5)$	$(\lambda_1, \lambda_2, \lambda_3, \rho_3, \rho_4, \rho_5)$	Johansen	GV	PB
(7,7,7,7,7)	(1, 1, 1, 0.5, 0.5, 0.5)	0.112	0.208	0.047
	(0.1, 0.1, 0.1, 0.3, 0.3, 0.3)	0.178	0.253	0.051
	(0.1, 0.4, 0.7, 0, 0, 0)	0.148	0.227	0.049
	(0.1, 0.3, 0.9, 0.1, 0.4, 0.9)	0.143	0.232	0.051
	(0.1, 0.2, 0.3, −0.1, 0.1, 0.9)	0.160	0.235	0.048
	(0.4, 0.4, 0.5, −0.3, 0.4, 0.3)	0.131	0.206	0.051
	(0.9, 0.9, 0.9, −0.4, 0.6, 0.9)	0.130	0.248	0.050
(12,12,12,12,12)	(1, 1, 1, 0.5, 0.5, 0.5)	0.073	0.133	0.047
	(0.1, 0.1, 0.1, 0.3, 0.3, 0.3)	0.083	0.142	0.051
	(0.1, 0.4, 0.7, 0, 0, 0)	0.074	0.128	0.049
	(0.1, 0.3, 0.9, 0.1, 0.4, 0.9)	0.080	0.128	0.051
	(0.1, 0.2, 0.3, −0.1, 0.1, 0.9)	0.082	0.146	0.048
	(0.4, 0.4, 0.5, −0.3, 0.4, 0.3)	0.069	0.122	0.051
	(0.9, 0.9, 0.9, −0.4, 0.6, 0.9)	0.073	0.139	0.050

(*Continued*)

Table 2. Continued.

(n_1, n_2, n_3)	$(\lambda_1, \lambda_2, \lambda_3, \rho)$	Johansen	GV	PB
(20,20,20,20,20)	(1, 1, 1, 0.5, 0.5, 0.5)	0.061	0.086	0.047
	(0.1, 0.1, 0.1, 0.3, 0.3, 0.3)	0.065	0.105	0.051
	(0.1, 0.4, 0.7, 0, 0, 0)	0.065	0.078	0.049
	(0.1, 0.3, 0.9, 0.1, 0.4, 0.9)	0.060	0.087	0.051
	(0.1, 0.2, 0.3, −0.1, 0.1, 0.9)	0.058	0.084	0.048
	(0.4, 0.4, 0.5, −0.3, 0.4, 0.3)	0.062	0.086	0.051
	(0.9, 0.9, 0.9, −0.4, 0.6, 0.9)	0.066	0.093	0.050
(15,20,10,32,7)	(1, 1, 1, 0.5, 0.5, 0.5)	0.092	0.158	0.047
	(0.1, 0.1, 0.1, 0.3, 0.3, 0.3)	0.126	0.225	0.051
	(0.1, 0.4, 0.7, 0, 0, 0)	0.116	0.178	0.049
	(0.1, 0.3, 0.9, 0.1, 0.4, 0.9)	0.086	0.183	0.051
	(0.1, 0.2, 0.3, −0.1, 0.1, 0.9)	0.101	0.183	0.048
	(0.4, 0.4, 0.5, −0.3, 0.4, 0.3)	0.111	0.197	0.051
	(0.9, 0.9, 0.9, −0.4, 0.6, 0.9)	0.076	0.189	0.050

We also observe from the first two sets of rows in Table 1 (the case of $k = 10$) that even for smaller samples of equal size, the PB test controls Type I error rates within the nominal level, whereas the other tests have inflated Type I error rates. When the sample sizes are very different (see the fourth set of rows in Table 1, $k = 10$), the Johansen test and the GV test are still liberal while the PB test appears to be slightly liberal. In general for a moderate k, the PB test is the only test appears to be satisfactory.

For the case of $p = 3$, we evaluated the sizes of the tests for the number of groups $k = 3$ and 5. Type I error rates are reported in Table 2. The tests exhibit similar performance as in the case of $p = 2$. Specifically, the PB test is the only test that controls Type I error rates very close to the nominal level. The Johansen test performs satisfactorily when the sample sizes are moderate and close to each other. The GV test seems to be the worst among these three tests. To judge the behaviours of the tests for higher dimension, we estimated the sizes for $p = 10$ and $k = 3$, and reported them in Table 3. The tests exhibit similar behaviours as they did for the case of $p = 3$ and $k = 3$. It appears that Type-I error rates of the tests are affected by the number of means to be compared, not by the dimension.

Overall, we see that the PB test is the only test that appears to be satisfactory for all the sample size and parameter configurations, and the number of groups to be compared.

Table 3. Monte Carlo estimates of Type I error rates when $p = 10$.

$(k = 3;\ \mathbf{\Sigma}_1 = \boldsymbol{I}_{10}, \mathbf{\Sigma}_2 = \text{diag}(\lambda_1, \ldots, \lambda_{10}), \mathbf{\Sigma}_3 = \text{diag}(\eta_1, \ldots, \eta_{10}))$					
(n_1, n_2, n_3)	$(\lambda_1, \ldots, \lambda_{10})$	$(\eta_1, \ldots, \eta_{10})$	Johansen	GV	PB
(15,15,15)	(1,. . .,1)	(1,. . .,1)	0.084	0.090	0.052
	(1,. . .,1)	(0.1,0.1,0.1,0.2,0.2,0.2,0.3,0.3,0.3,0.1)	0.078	0.085	0.044
	(1,2,2,8,8,8,10,10,10,10)	(10,10,10,10,2,3,6,6,10,10)	0.081	0.088	0.051
	(1,1,1,3,3,3,9,9,9,20)	(5,5,5,15,15,15,45,45,45,100)	0.077	0.081	0.055
	(12,12,12,1,1,1,24,24,24,1)	(1,1,1,0.1,0.1,0.1,2,2,24,21)	0.085	0.095	0.046
(25,35,50)	(1,. . .,1)	(1,. . .,1)	0.061	0.077	0.051
	(1,. . .,1)	(0.1,0.1,0.1,0.2,0.2,0.2,0.3,0.3,0.3,0.1)	0.066	0.081	0.049
	(1,2,2,8,8,8,10,10,10,10)	(10,10,10,10,2,3,6,6,10,10)	0.055	0.072	0.051
	(1,1,1,3,3,3,9,9,9,20)	(5,5,5,15,15,15,45,45,45,100)	0.067	0.071	0.046
	(12,12,12,1,1,1,24,24,24,1)	(1,1,1,0.1,0.1,0.1,2,2,24,21)	0.066	0.069	0.047

5. Illustrative examples

We shall illustrate the three tests using the data sets originally discussed by Thomson and Randall-Maciver [30], so that we can compare our results and understand the behaviour of these three tests described in Section 3 for comparing several groups. There are five samples of 30 skulls from each of the early predynastic period (*circa* 4000 BC), the late predynastic period (*circa* 3300 BC), the 12th and 13th dynasties (*circa* 1850 BC), the Ptolemaic period (*circa* 200 BC), and the Roman period (*circa* AD 150). Four measurements are available on each skull, namely, $X_1 =$ maximum breadth, $X_2 =$ borborygmatic height, $X_3 =$ dentoalveolar length, and $X_4 =$ nasal height (all in mm). The measurements made on male Egyptian skulls from the area of Thebes are available at *Statlib* (http://lib.stat.cmu.edu/DASL/Stories/EgyptianSkullDevelopment.html), and they do provide evidence of multivariate normality. In order to see the performance of these four tests in the case of small sample size and for ease of presenting the numerical results, we only take the first 15 skull observations, and consider only the first three samples (we discard the sample from the Roman period). The purpose of this study is to find whether the differences in the sample means for the variables reflect gradual changes with time. In our present set-up, we have $n_1 = \cdots = n_4 = 15$, the number of groups is $k = 4$, and the number of variables is $p = 4$.

The null hypothesis of interest is whether the mean vectors for the four variables are the same across the four periods. The hypothesis may be written as

$$H_0 : \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{14} \end{pmatrix} = \begin{pmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \\ \mu_{24} \end{pmatrix} = \begin{pmatrix} \mu_{31} \\ \mu_{32} \\ \mu_{33} \\ \mu_{34} \end{pmatrix} = \begin{pmatrix} \mu_{41} \\ \mu_{42} \\ \mu_{43} \\ \mu_{44} \end{pmatrix} \quad vs. \quad H_0 \text{ is not true.}$$

Based on the sample data, we computed the summary statistics for the four groups as

$$\left(\bar{\boldsymbol{Y}}_1, \ldots, \bar{\boldsymbol{Y}}_4\right) = \begin{pmatrix} 131.40 & 133.07 & 134.27 & 136.33 \\ 134.07 & 134.00 & 135.47 & 132.47 \\ 97.73 & 99.13 & 96.60 & 94.87 \\ 50.27 & 49.93 & 49.67 & 51.87 \end{pmatrix}.$$

The matrices

$$\boldsymbol{W}_1 = \begin{pmatrix} 0.862 & -0.173 & -0.210 & -1.174 \\ - & 0.604 & 0.138 & 0.076 \\ - & - & 0.493 & 0.308 \\ - & - & - & 3.508 \end{pmatrix},$$

$$\boldsymbol{W}_2 = \begin{pmatrix} 0.573 & 0.205 & -0.327 & -0.110 \\ - & 0.953 & -0.146 & -0.922 \\ - & - & 2.223 & -0.868 \\ - & - & - & 2.717 \end{pmatrix},$$

$$\boldsymbol{W}_3 = \begin{pmatrix} 0.925 & 0.091 & 0.070 & 0.015 \\ - & 0.610 & 0.025 & -0.193 \\ - & - & 0.625 & -0.227 \\ - & - & - & 1.587 \end{pmatrix},$$

$$\boldsymbol{W}_4 = \begin{pmatrix} 1.409 & 0.085 & 0.121 & -0.430 \\ - & 0.964 & -0.095 & -0.666 \\ - & - & 0.640 & -0.362 \\ - & - & - & 2.174 \end{pmatrix},$$

and

$$\boldsymbol{W}^{-1} = \begin{pmatrix} 0.294 & 0.013 & 0.042 & 0.057 \\ - & 0.355 & 0.027 & 0.066 \\ - & - & 0.268 & 0.043 \\ - & - & - & 0.126 \end{pmatrix}.$$

Using these matrices, we computed

$$\widehat{\boldsymbol{\mu}}_0^* = \boldsymbol{W}^{-1} \sum_{i=1}^{k} \boldsymbol{W}_i \bar{\boldsymbol{Y}}_i = \begin{pmatrix} 134.09 \\ 134.10 \\ 98.349 \\ 50.832 \end{pmatrix}.$$

Finally, the value of the test statistic is computed as

$$T_0 = \sum_{i=1}^{4} (\bar{\boldsymbol{Y}}_i - \widehat{\boldsymbol{\mu}}_0^*)' \boldsymbol{W}_i (\bar{\boldsymbol{Y}}_i - \widehat{\boldsymbol{\mu}}_0^*) = 32.90.$$

5.1. *Johansen's test*

Using Equations (9) and (8), respectively, we can get $A = 1.6227$ and $c = 14.5500$. The observed value of Johansen's test statistic is

$$J_{\mathrm{OH}} = \frac{T(\bar{\boldsymbol{Y}}_1, \ldots, \bar{\boldsymbol{Y}}_k; \tilde{\boldsymbol{S}}_1, \ldots, \tilde{\boldsymbol{S}}_k)}{c} = \frac{32.90}{14.55} = 2.2612.$$

Taking α to be 0.05, we have

$$F_{f_1, f_2; 0.05} = 2.0454$$

with $f_1 = p(k-1) = 12$ and $f_2 = p(k-1)[p(k-1)+2]/(3A) = 34.51$ degrees of freedom. Furthermore, the p-value is computed as $P(F_{12,34.51} > 2.2612) = 0.0304$. Thus, on the basis of the F critical value (or the p-value), the null hypothesis of equal mean vector is rejected at the nominal level 0.05.

5.2. *Generalized variable test*

To apply this test, we first computed $\tilde{\boldsymbol{s}}_i$ and mean vectors $\bar{\boldsymbol{y}}_i$, $i = 1, \ldots, 4$, and then generated 100,000 values of G variable in Equation (11). We estimated the p value by the proportion of these 100,000 generated values that are greater than or equal to 1, and is given by 0.0009. Obviously, we reject H_0. That is, the Egyptian skulls experienced a significant change over those four periods.

5.3. *Parametric bootstrap test*

To compute PB p-value, we computed the Cholesky factors $\boldsymbol{t}_i$'s (so that $\boldsymbol{t}_i \boldsymbol{t}_i' = \tilde{\boldsymbol{s}}_i$, $i = 1, \ldots, 4$) as

$$\boldsymbol{t}_1 = \begin{pmatrix} 1.542 & 0 & 0 & 0 \\ 0.320 & 1.330 & 0 & 0 \\ 0.264 & -0.374 & 1.465 & 0 \\ 0.486 & 0.004 & -0.129 & 0.534 \end{pmatrix}, \quad \boldsymbol{t}_2 = \begin{pmatrix} 1.433 & 0 & 0 & 0 \\ -0.226 & 1.361 & 0 & 0 \\ 0.216 & 0.309 & 0.717 & 0 \\ 0.050 & 0.561 & 0.229 & 0.607 \end{pmatrix},$$

$$t_3 = \begin{pmatrix} 1.053 & 0 & 0 & 0 \\ -0.167 & 1.305 & 0 & 0 \\ -0.128 & 0.005 & 1.299 & 0 \\ -0.048 & 0.159 & 0.185 & 0.794 \end{pmatrix}, \quad t_4 = \begin{pmatrix} 0.871 & 0 & 0 & 0 \\ 0.036 & 1.206 & 0 & 0 \\ -0.062 & 0.428 & 1.314 & 0 \\ 0.173 & 0.441 & 0.219 & 0.678 \end{pmatrix}.$$

Using the t_i's and the steps of Section 3.3, the PB p-value was obtained (using a simulation consisting of 10,000 runs) as 0.0410. Therefore, we reject H_0, the same conclusion as previous tests.

Finally, we also note that the above results of the tests are in agreement with our size studies in Section 4. More specifically, we observed in Section 4 that the GV test is more liberal than the Johansen test followed by the PB test, and this is reflected by the magnitude of the p-values of the tests.

6. Concluding remarks

We extended the univariate results of Krishnamoorthy *et al.* [26] to the MANOVA, and showed that the available approximate methods are not satisfactory, and the GV test, that was developed recently, performed poorly with respect to Type I error rates. The proposed PB test is the only test that performs satisfactorily for all the situations considered. Furthermore, the PB test is as simple as other tests to use in applications. For the special case of comparing two mean vectors, we developed an approximate test that is the same as the existing satisfactory test in Krishnamoorthy and Yu [18]. It is plausible that the PB approach can be used to obtain analytical approximate test for a general case of comparing several normal mean vectors. We used the moment matching method to find an approximate test for the multivariate Behrens–Fisher problem. At present, we are unable to extend this moment matching method to get an approximate test for the general case, and plan to investigate this problem in the future.

References

[1] S.N. Roy, *The individual sampling distribution of the maximum, the minimum and any intermediate of the p-statistics on the null-hypothesis*, Sankhya 2, (1945) pp. 133–58.
[2] D.N. Lawley, *A generalization of Fisher's z-test*, Biometrika 30 (1938), pp. 180–187.
[3] H. Hotelling, *A generalized T test and measure of multivariate dispersion*, in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman, (ed.), University of California Press, Berkeley, 1951, pp. 23–41.
[4] S.S. Wilks, *Certain generalizations in the analysis of variance*, Biometrika 24 (1932), pp. 471–494.
[5] M.S. Bartlett, *A note on tests of significance in multivariate analysis*, Math. Proc. Cambridge Philos. Soc., 35 (1939), pp. 180–185.
[6] K.C.S. Pillai, *Some new test criteria in multivariate analysis*, Ann. Math. Stat. 26 (1955), pp. 117–121.
[7] C.L. Olson, *Comparative robustness of six tests in multivariate analysis of variance*, J. Amer. Statist. Assoc. 69 (1974), pp. 894–908.
[8] C.L. Olson, *Practical considerations in choosing a MANOVA test statistic: a rejoinder to Stevens*, Psychol. Bull. 86 (1979), pp. 1350–1352.
[9] K. Krishnamoorthy and Y. Xia, *On selecting tests for equality of two normal mean vectors*, Multivariate Behav. Res. 41 (2009), pp. 533–548.
[10] B.M. Bennett, *Note on a solution of the generalized Behrens–Fisher problem*, Ann. Inst. Statist. Math. 2 (1951), pp. 87–90.
[11] M.B. Brown and A.B. Forsythe, *The small sample behaviour of some statistics which test the equality of several means*, Technometrics 16 (1974), pp. 385–389.
[12] G.S. James, *Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown*, Biometrika 41 (1954), pp. 19–43.
[13] Y. Yao, *An approximate degrees of freedom solution to the multivariate Behrens–Fisher problem*. Biometrika 52 (1980), pp. 139–147.
[14] S. Johansen, *The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression*, Biometrika 67 (1980), pp. 85–92.

[15] D.G. Nel and Van der C.A. Merwe, *A solution to the multivariate Behrens–Fisher problem*, Commun. Stat. Theory Methods 15 (1986), 3719–3735.
[16] S. Kim, *A practical solution to the multivariate Behrens–Fisher problem*, Biometrika, 79 (1992), pp. 171–176.
[17] W.F. Christensen and A.C. Rencher, *A comparison of type I error rates and power levels for seven solutions to the multivariate Behrens–Fisher problem*, Commun. Stat. Theory Methods 26 (1997), pp. 1251–1273.
[18] K. Krishnamoorthy and J. Yu, *Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem*, Stat. Probab. Lett. 66 (2004), pp. 161–169.
[19] J. Park and B.K. Sinha, *Some aspects of multivariate Behrens-Fisher problem*, Technical Report, Deptartment of Mathematics and Statistics, University of Maryland-Baltimore County.
[20] Y. Hirokazu and Y. Ke-Hai, *Three approximate solutions to the multivariate Behrens–Fisher problem*, Commun. Stat. Simulation Comput. 34 (2005), pp. 975–988.
[21] A. Belloni and G. Didier, *On the Behrens–Fisher problem: a globally convergent algorithm and a finite-sample study of the Wald, LR and LM Tests*, Ann. Stat. 36 (2008), pp. 2377–2408.
[22] B.L. Welch, *The significance of the difference between two means when the population variances are unequal*, Biometrika 29 (1938), pp. 350–362.
[23] B.L. Welch, *On the comparison of several mean values: an alternative approach*, Biometrika 38 (1951), pp. 330–336.
[24] J. Gamage, T. Mathew, and S. Weerahandi, *Generalized p-values and generalized confidence regions for the multivariate Behrens–Fisher problem and MANOVA*, J. Multivariate Anal. 88 (2004), pp. 177–189.
[25] K.L. Tang and J. Algina, *Performance of four multivariate tests under variance-covariance heteroscedasticity*, Multivariate Behav. Res. 28 (1993), pp. 391–405.
[26] K. Krishnamoorthy, F. Lu, and T. Mathew, *A parametric bootstrap approach for ANOVA with unequal variances: fixed and random models*, Comput. Stat. Data Anal. 51 (2007), pp. 5731–5742.
[27] K. Tsui and S. Weerahandi, *Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters*, J. Amer. Statist. Assoc. 84 (1989), pp. 602–607.
[28] R.J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, New York, 1982.
[29] W.B. Smith and R.R. Hocking, *Wishart variates generator (Algorithm AS 53)*, Appl. Stat. 21 (1972), pp. 341–345.
[30] A. Thomson and R. Randall-Maciver, *Ancient Races of the Thebaid*, Oxford University Press, 1905.
[31] L.R. Haff, *An identity concerning Wishart distribution with applications*, J. Multivariate Anal. 9 (1979), pp. 531–544.

Appendix

To evaluate $E\text{tr}(\boldsymbol{Q}_1+\boldsymbol{Q}_2)^2$ and $E\text{tr}(\boldsymbol{Q})^2$, we will use the result by Haff [31] that, for $\boldsymbol{V} \sim W_p(m, \boldsymbol{\Delta})$,

$$E(\boldsymbol{V}^2) = m(m+1)\boldsymbol{\Delta}^2 + m\,\text{tr}(\boldsymbol{\Delta})\boldsymbol{\Delta}. \tag{A.1}$$

Recall that $\boldsymbol{Q} \sim W_p(\nu, (1/\nu)\boldsymbol{I}_p)$, and so

$$E(\boldsymbol{Q}^2) = \frac{\nu(\nu+1)\boldsymbol{I}_p}{\nu^2} + \frac{\nu\text{tr}(\boldsymbol{I}_p)\boldsymbol{I}_p}{\nu^2}.$$

Thus

$$\text{tr}E(\boldsymbol{Q}^2) = p + \frac{p+p^2}{\nu}. \tag{A.2}$$

Let $\boldsymbol{C}_1 = (\tilde{\boldsymbol{s}}_1+\tilde{\boldsymbol{s}}_2)^{-1/2}\tilde{\boldsymbol{s}}_1(\tilde{\boldsymbol{s}}_1+\tilde{\boldsymbol{s}}_2)^{-1/2}$ and $\boldsymbol{C}_2 = (\tilde{\boldsymbol{s}}_1+\tilde{\boldsymbol{s}}_2)^{-1/2}\tilde{\boldsymbol{s}}_2(\tilde{\boldsymbol{s}}_1+\tilde{\boldsymbol{s}}_2)^{-1/2}$ so that $\boldsymbol{C}_1+\boldsymbol{C}_2 = \boldsymbol{I}_p$.

As $\boldsymbol{Q}_1 \sim W_p(n_1-1, \boldsymbol{C}_1/(n_1-1))$ independently of $\boldsymbol{Q}_2 \sim W_p(n_2-1, \boldsymbol{C}_2/(n_2-1))$, we have

$$\text{tr}\,E(\boldsymbol{Q}_1\boldsymbol{Q}_2) = \text{tr}[E(\boldsymbol{Q}_1)E(\boldsymbol{Q}_2)] = \text{tr}(\boldsymbol{C}_1\boldsymbol{C}_2)$$

and using Equation (A.1), we get

$$\text{tr}E(\boldsymbol{Q}_i^2) = \text{tr}(\boldsymbol{C}_i^2) + \frac{\text{tr}(\boldsymbol{C}_i^2) + [\text{tr}(\boldsymbol{C}_i)]^2}{n_i-1}, \quad i=1,2.$$

Using the result that for real symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\text{tr}(\boldsymbol{AB}) = \text{tr}(\boldsymbol{BA})$, and the fact that $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ are independent, we get

$$\begin{aligned}\text{tr}E(\boldsymbol{Q}_1+\boldsymbol{Q}_2)^2 &= \text{tr}E(\boldsymbol{Q}_1)^2 + \text{tr}E(\boldsymbol{Q}_2)^2 + 2\,\text{tr}[E(\boldsymbol{Q}_1)E(\boldsymbol{Q}_2)] \\ &= \frac{\text{tr}(\boldsymbol{C}_1^2) + [\text{tr}(\boldsymbol{C}_1)]^2}{n_1-1} + \frac{\text{tr}(\boldsymbol{C}_2^2) + [\text{tr}(\boldsymbol{C}_2)]^2}{n_2-1} \\ &\quad + \text{tr}(\boldsymbol{C}_1^2) + \text{tr}(\boldsymbol{C}_2^2) + 2\,\text{tr}(\boldsymbol{C}_1\boldsymbol{C}_2).\end{aligned} \tag{A.3}$$

Finally, noticing that $\text{tr}(\boldsymbol{C}_1^2) + \text{tr}(\boldsymbol{C}_2^2) + 2\,\text{tr}(\boldsymbol{C}_1\boldsymbol{C}_2) = \text{tr}(\boldsymbol{C}_1+\boldsymbol{C}_2)^2 = \text{tr}(\boldsymbol{I}_p) = p$, we see that Equations (A.2) equals to (A.3) when

$$\nu = \frac{p^2+p}{1/(n_1-1)\left\{\text{tr}(\boldsymbol{C}_1^2) + [\text{tr}(\boldsymbol{C}_1)]^2\right\} + 1/(n_2-1)\left\{\text{tr}(\boldsymbol{C}_2^2) + [\text{tr}(\boldsymbol{C}_2)]^2\right\}}.$$

Replacing $\boldsymbol{C}_i$ by $(\tilde{\boldsymbol{s}}_1+\tilde{\boldsymbol{s}}_2)^{-1/2}\tilde{\boldsymbol{s}}_i(\tilde{\boldsymbol{s}}_1+\tilde{\boldsymbol{s}}_2)^{-1/2}$, $i=1,2$, and using the relation $\text{tr}(\boldsymbol{AB}) = \text{tr}(\boldsymbol{BA})$, we get the expression for ν in Equation (19).