A Computational Approach Test for the Equality of Two Multivariate Normal Mean Vectors under Heterogeneity of Covariance Matrices *

Authors: Esra Gökpinar
– Department of Statistics, Gazi University, Turkey (eyigit@gazi.edu.tr)

Sezen Karanfil
– Department of Statistics, Gazi University, Turkey (sznkaranfil06@gmail.com)

Meral Ebegil
– Department of Statistics, Gazi University, Turkey (mdemirel@gazi.edu.tr)

Yaprak Arzu Ozdemir
– Department of Statistics, Gazi University, Turkey (yaprak@gazi.edu.tr)

Fikri Gökpinar
– Department of Statistics, Gazi University, Turkey (fikri@gazi.edu.tr)

Abstract:

- In this paper, a computational approach test (CAT) was proposed to test the equality of two multivariate normal mean vectors under heterogeneity of covariance matrices. The proposed test was compared with the other popular tests as well as their CAT versions in terms of estimated type I error rate and power. Simulation study shows that the proposed test and CAT versions of tests can be used as a good alternative test to test the equality of two multivariate normal mean vectors under heterogeneity of covariance matrices.

Key-Words:

- *Computational Approach Test; Parametric Bootstrap Approach; Simulation Study.*

AMS Subject Classification:

- 62F03, 65C05, 65C60.

*The opinions expressed in this text are those of the authors and do not necessarily reflect the views of any organization.

1. INTRODUCTION

In statistical analysis, the comparison of the means of two or more groups is a very common problem. However, in many real world problems, there can be more than one variable which are related with each other. In this case, it is not appropriate to use univariate statistical methods. Therefore, it is required to utilize some multivariate statistical methods. One of the most common methods is the Hotelling T^2 statistic to compare the mean vectors of two independent groups under multivariate normality. It is known that Hotelling T^2 statistic requires the assumption of equality of two covariance matrices. However, this assumption is not valid in many statistical application areas. The violation of this assumption is called multivariate Behrens–Fisher problem in statistical analysis. In the case of multivariate Behrens-Fisher problem, the type I error rates of Hotelling T^2 statistic are not close to the nominal level and this also affects the power of the test negatively. Ito and Schull [12] indicated that when variances of two groups are not equal, the type I error rate for Hotelling's T^2 is approximately equal to the nominal level rate only when the sample sizes of groups are large and equal. Therefore, many solutions can be found for this problem in the literature.

Bennett [2] is one of the pioneers who presented the exact solution to the multivariate Behrens-Fisher problem. Since the Bennett's test depends on the order of the observations, it is not useful for larger sample sizes. In addition, James [13] improved the simple chi-square approximation by the Cornish–Fisher expansion until the third order term. Yao [27] suggested the approximate degrees-of-freedom solution and indicated that type I error rate of this test is lower than that of James' test in almost all cases. Subrahmaniam and Subrahmaniam [21, 22] compared the tests of Bennett, James, and Yao according to type I error rates and powers. Johansen [14] also studied the Behrens–Fisher problem in the context of general linear models. Christensen and Rencher [6] compared the seven tests given by Bennett [2], James [13], Yao [27], Johansen [14], Nel and Van der Merwe [18], Hwang and Paulson [11] and Kim [15] for the multivariate Behrens-Fisher problem in terms of type I error rates and powers. Algina, Oshima and Tang [1] also compared the tests given by Yao [27] , James [13] and Johansen [14] under various conditions of heteroscedasticity and non-normality. They showed that the type I error rate of Johansen's test is roughly equivalent to that of Yao's solution. In addition, the type I error rates of Johansen's test improve as number of variables increases. Kim [15] showed that the type I error rate of the test is more conservative than that of Yao's test in almost every situation. However, Kim's test has higher power than Yao's test when the smaller sample size is associated with the large variance [15].

De la Rey and Nel's [7] compared the tests given by Bennett [2], James [13], Yao [27] and Nel and Van der Merwe [18] and showed that Nel and Van der Merwe [18] and Yao [27] gave better solutions. According to the results of the comparative papers mentioned above, apparently there is no definitive solution that shows good performance in all circumstances. Finally, Krishnamoorthy and

Yu [17] modified Nel and Van Der Merwe's [18] procedure by providing an invariant statistic. Recently, bootstrap-based methods for multivariate hypothesis testing were proposed. For example, Smaga [20] developed bootstrap methods of some test statistics based on different weight matrices for testing the mean vector of a multivariate distribution. Konietschke et al. [16] developed parametric and nonparametric bootstrap methods of Wald-type test for multi-factor multivariate data which also includes multivariate Behren Fisher problem. They compared these tests via simulation study under both normality and non-normality models.

The purpose of this paper is to test the equality of two normal mean vectors under heterogeneity of covariance matrices by using computational approach test (CAT). This method which was firstly introduced by Pal et al. [19] is used in situations where traditional approaches do not provide useful solutions. The CAT is a special case of parametric bootstrap and based on restricted maximum likelihood estimation under null hypothesis. One of the most important advantages of this procedure is that it does not require the knowledge of any sampling distribution. Pal et al. [19] showed the application of the CAT to Gamma and Weibull distributions for hypothesis testing and interval estimations. CAT was also applied by Chang and Pal [3] for testing the equality of two normal population means under heteroscedastisity. Chang et al. [4, 5] suggested test procedures based on CAT for hypotheses testing of the Poisson and Gamma models. Gökpınar and Gökpınar [8] applied CAT to test the equality of several normal population means when the variances are unknown and arbitrary and Gökpınar et al. [9] proposed CAT for the equality of several inverse Gaussian means under heterogeneity of scale parameters. Moreover, Gökpınar and Gökpınar [10] proposed CAT for the equality of coefficient of variations in k populations. In these studies, it was shown that the CAT procedure is a good alternative for other testing procedures for various statistical problems.

For this reason, in this study, the CAT method to the equality of two normal mean vectors under heterogeneity of covariance matrices was applied and this method was also compared with Bennett [2], Johansen [14], Nel and Van Der Merwe [18], Krishnamoorthy and Yu [17] tests in terms of their type I error rates and powers under various situations.

The rest of this study was organized as follows. In Section 2, the method was described to obtain the maximum likelihood estimates (MLEs) over unrestricted parameter space and over a restricted parameter space. Simple fixed point iteration was proposed to compute the MLEs under restricted parameters space. In section 3, the tests given by Bennett [2], Johansen [14], Nel and Van Der Merwe [18], Krishnamoorthy and Yu [17] and Konietschke et al. [16] were presented. In section 4, the concept of CAT procedure and its application to the equality of two normal mean vectors under heterogeneity of covariance matrices were given. In section 5, simulation studies were presented to assess the performance of the proposed test in terms of the type I error rates and powers under multivariate normal distribution with different parameter combinations. Furthermore, to see robustness of all tests under non-normal distribution, the estimated type I error rates and powers of all tests were calculated. Finally, concluding remarks were summarized in Section 6.

2. The Maximum Likelihood Estimates

Let $\boldsymbol{Y_{i1}}, \boldsymbol{Y_{i2}}, ..., \boldsymbol{Y_{in_i}}$, have p-variate normal distribution with mean vector $\boldsymbol{\mu_i} = (\mu_{i1}, \mu_{i2}, ..., \mu_{ip})^\top$ and covariance matrix $\boldsymbol{\Sigma_i}$ $i = 1, 2$. Assume that sample units are independent from each other. n_i $\quad (i = 1, 2)$ is the sample size of i-th group.

$\boldsymbol{\bar{Y}_i}$ is the sample mean vector of i-th group and $\boldsymbol{S_i}$ is the maximum likelihood estimation of covariance matrix of i-th group. $\boldsymbol{S_{(i)}}$ is the unbiased estimation of covariance matrix of i-th group. Thus,

$$\bar{\mathbf{Y}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{Y}_{ij} \quad i = 1, 2 \tag{2.1}$$

$$\mathbf{S}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}(\boldsymbol{Y}_{ij} - \bar{\mathbf{Y}}_i)(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)^\top \quad i = 1, 2 \tag{2.2}$$

The log-likelihood function under the unrestricted parameter space is given by

(2.3)

$$lnL = \frac{-p(n_1+n_2)}{2}ln(2\pi) - \frac{n_1}{2}ln(|\boldsymbol{\Sigma_1}|) - \frac{n_2}{2}ln(|\boldsymbol{\Sigma_2}|) - \frac{1}{2}\sum_{j=1}^{n_1}[\mathbf{Y_{1j}} - \mu_\mathbf{1}]^\top\boldsymbol{\Sigma}_1^{-1}[\mathbf{Y_{1j}} - \mu_\mathbf{1}]$$
$$- \frac{1}{2}\sum_{j=1}^{n_2}[\mathbf{Y_{2j}} - \mu_\mathbf{2}]^\top\boldsymbol{\Sigma}_2^{-1}[\mathbf{Y_{2j}} - \mu_\mathbf{2}]$$

To find the unrestricted MLEs, the partial derivatives of Eq.(2.3) with respect to $\boldsymbol{\Sigma_1}$, $\boldsymbol{\Sigma_2}$, $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ yield the following equations:

$$\hat{\mu}_\mathbf{1} = \bar{\mathbf{Y}}_1, \quad \hat{\mu}_\mathbf{2} = \bar{\mathbf{Y}}_2, \quad \hat{\boldsymbol{\Sigma}}_\mathbf{1} = \frac{\sum_{j=1}^{n_1}[\mathbf{Y_{1j}} - \bar{\mathbf{Y}}_\mathbf{1}]^\top[\mathbf{Y_{1j}} - \bar{\mathbf{Y}}_\mathbf{1}]}{n_1} = \mathbf{S_1}$$
$$\hat{\boldsymbol{\Sigma}}_\mathbf{2} = \frac{\sum_{j=1}^{n_2}[\mathbf{Y_{2j}} - \bar{\mathbf{Y}}_\mathbf{2}]^\top[\mathbf{Y_{2j}} - \bar{\mathbf{Y}}_\mathbf{2}]}{n_2} = \mathbf{S_2}$$

To find restricted MLE (RMLE), under $H_0 : \boldsymbol{\mu_1} = \boldsymbol{\mu_2} = \boldsymbol{\mu}$, the log-likelihood function can be expressed as

(2.4)

$$lnL = \frac{-p(n_1+n_2)}{2}ln(2\pi) - \frac{n_1}{2}ln(|\mathbf{\Sigma_1}|) - \frac{n_2}{2}ln(|\mathbf{\Sigma_2}|) - \frac{1}{2}tr\left(\mathbf{\Sigma}_1^{-1}\sum_{j=1}^{n_1}[\mathbf{Y_{1j}}-\mu]^\top[\mathbf{Y_{1j}}-\mu]\right)$$
$$-\frac{1}{2}tr\left(\mathbf{\Sigma}_2^{-1}\sum_{j=1}^{n_2}[\mathbf{Y_{2j}}-\mu]^\top[\mathbf{Y_{2j}}-\mu]\right)$$

where $\boldsymbol{\mu}$ denotes the unknown common mean under H_0. To find the restricted MLEs (RMLEs), by using the following equations as $\frac{\partial tr(\boldsymbol{X^{-1}}A)}{\partial \boldsymbol{X}} = -\boldsymbol{X^{-1}}A\boldsymbol{X^{-1}}$ or $\frac{\partial tr(\boldsymbol{X^{-1}}A)}{\partial \boldsymbol{X}} = -(\boldsymbol{X^{-1}})^\top A^\top(\boldsymbol{X^{-1}})^\top$ the partial derivatives of Eq.(2.4) with respect to $\mathbf{\Sigma_1}$, $\mathbf{\Sigma_2}$ and $\boldsymbol{\mu}$ yield the following equations:

$$\frac{\partial lnL}{\partial \mathbf{\Sigma}_1} = -\frac{n_1}{2}\mathbf{\Sigma}_1^{-1} + \frac{1}{2}\mathbf{\Sigma}_1^{-1}\sum_{j=1}^{n_1}[\mathbf{Y_{1j}}-\mu]^\top[\mathbf{Y_{1j}}-\mu]\mathbf{\Sigma}_1^{-1}$$

$$\frac{\partial lnL}{\partial \mathbf{\Sigma}_2} = -\frac{n_2}{2}\mathbf{\Sigma}_2^{-1} + \frac{1}{2}\mathbf{\Sigma}_2^{-1}\sum_{j=1}^{n_2}[\mathbf{Y_{2j}}-\mu]^\top[\mathbf{Y_{2j}}-\mu]\mathbf{\Sigma}_2^{-1}$$

$$\frac{\partial lnL}{\partial \mu} = n_1\mathbf{\Sigma}_1^{-1}\bar{\mathbf{Y}}_1 - \frac{1}{2}\left(2n_1\mathbf{\Sigma}_1^{-1}\mu^\top\right) + n_2\mathbf{\Sigma}_2^{-1}\bar{\mathbf{Y}}_2 - \frac{1}{2}\left(2n_2\mathbf{\Sigma}_2^{-1}\boldsymbol{\mu}^\top\right)$$

The RMLEs are given by

$$\hat{\mathbf{\Sigma}}_{1(RML)} = \frac{\sum_{j=1}^{n_1}[\mathbf{Y_{1j}}-\hat{\mu}_{(RML)}]^\top[\mathbf{Y_{1j}}-\hat{\mu}_{(RML)}]}{n_1} \quad \hat{\mathbf{\Sigma}}_{2(RML)} = \frac{\sum_{j=1}^{n_2}[\mathbf{Y_{2j}}-\hat{\mu}_{(RML)}]^\top[\mathbf{Y_{2j}}-\hat{\mu}_{(RML)}]}{n_2}$$

(2.5)

$$\hat{\mu}_{(RML)}^\top = \left(n_1\hat{\mathbf{\Sigma}}_{1(RML)}^{-1} + n_2\hat{\mathbf{\Sigma}}_{2(RML)}^{-1}\right)^{-1}\left(n_1\hat{\mathbf{\Sigma}}_{1(RML)}^{-1}\bar{\mathbf{Y}}_1 + n_2\hat{\mathbf{\Sigma}}_{2(RML)}^{-1}\bar{\mathbf{Y}}_2\right)$$

Since there are no close forms of these equations, these estimators can be obtained iteratively as follows: updating the estimates from l-step estimates ($\mathbf{\Sigma}_1^{(l)}, \mathbf{\Sigma}_2^{(l)}$ and $\boldsymbol{\mu}^{(l)}$) by

$$\mathbf{\Sigma}_1^{(l+1)} = \frac{\sum_{j=1}^{n_1}[\mathbf{Y_{1j}}-\mu^{(l+1)}]^\top[\mathbf{Y_{1j}}-\mu^{(l+1)}]}{n_1}$$

$$\mathbf{\Sigma}_2^{(l+1)} = \frac{\sum_{j=1}^{n_2} [\mathbf{Y_{2j}} - \mu^{(l+1)}]^\top [\mathbf{Y_{2j}} - \mu^{(l+1)}]}{n_2}$$

$$\mu^{(l+1)} = \left(n_1 \left(\mathbf{\Sigma}_1^{(l+1)}\right)^{-1} + n_2 \left(\mathbf{\Sigma}_2^{(l+1)}\right)^{-1}\right)^{-1} \left(n_1 \left(\mathbf{\Sigma}_1^{(l+1)}\right)^{-1} \bar{\mathbf{Y}}_1 + n_2 \left(\mathbf{\Sigma}_2^{(l+1)}\right)^{-1} \bar{\mathbf{Y}}_2\right)$$

where initial value $\boldsymbol{\mu}^{(0)}$ could set as $\mu^{(0)} = \left(n_1\mathbf{S}_1^{-1} + n_2\mathbf{S}_2^{-1}\right)^{-1} \left(n_1\mathbf{S}_1^{-1}\bar{\mathbf{Y}}_1 + n_2\mathbf{S}_2^{-1}\bar{\mathbf{Y}}_2\right)$. $\mathbf{\Sigma}_1^{(l)}$, $\mathbf{\Sigma}_2^{(l)}$ and $\mu^{(l)}$ converge to the RMLEs under H_0 denoted as $\hat{\mathbf{\Sigma}}_{i(RML)}$ and $\hat{\mu}_{(RML)}$. For example, let,

$p = 3$, $n_1 = n_2 = 5$, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = [1 \quad 1 \quad 1]$,

$$\mathbf{\Sigma}_1 = \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 1 \end{bmatrix}$$

and

$$\mathbf{\Sigma}_2 = \begin{bmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{bmatrix}$$

The Monte Carlo estimates of the expected value of $\hat{\mathbf{\Sigma}}_{i(RML)}$ and $\hat{\boldsymbol{\mu}}_{(RML)}$ are

$$\tilde{E}(\hat{\mu}_{RML}) = [1.005 \quad 1.006 \quad 1.003],$$

$$\tilde{E}(\hat{\mathbf{\Sigma}}_{1(RML)}) = \begin{bmatrix} 0.969 & 0.190 & 0.199 \\ 0.190 & 0.981 & 0.197 \\ 0.199 & 0.197 & 0.969 \end{bmatrix}$$

and

$$\tilde{E}(\hat{\mathbf{\Sigma}}_{2(RML)}) = \begin{bmatrix} 0.999 & 0.795 & 0.798 \\ 0.795 & 0.994 & 0.794 \\ 0.798 & 0.794 & 0.992 \end{bmatrix}$$

As seen from above, the obtained result is well.

3. Test Statistics

Let $\boldsymbol{Y_{i1}}, \boldsymbol{Y_{i2}}, ..., \boldsymbol{Y_{in_i}}$, have p-variate normal distribution with mean vector $\boldsymbol{\mu_i}$ and covariance matrix $\boldsymbol{\Sigma_i}$, $i = 1, 2$. Let us denote

$$\mathbf{S}_{(i)} = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)^\top \quad i = 1, 2$$

$$\tilde{\boldsymbol{\Sigma}}_i = \frac{1}{n_i}\boldsymbol{\Sigma}_i \quad \text{and} \quad \tilde{\mathbf{S}}_i = \frac{1}{n_i}\mathbf{S}_i \quad i = 1, 2$$

Since the sample mean vector of i-th group, $\bar{\boldsymbol{Y}}_i$ and the maximum likelihood estimation of covariance matrix of i-th group, $\boldsymbol{S}_i$'s are independent from each other, the following equations can be written as follows:

$$\bar{\mathbf{Y}}_i \sim N_p(\mu_i, \frac{1}{n_i}\boldsymbol{\Sigma}_i) \quad \text{and} \quad \tilde{\mathbf{S}}_i \sim W_p(n_i - 1, \frac{1}{n_i - 1}\tilde{\boldsymbol{\Sigma}}_i) \quad i = 1, 2$$

Here $W_p(r, \boldsymbol{\Sigma})$ is p-variate Wishart distribution with degrees of freedom r. This distribution is also known as generalized chi-square distribution which is obtained by Wishart [25].

The null and alternative hypotheses for testing the equality of two multivariate normal mean vectors are as follows:

$$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2 \tag{3.1}$$

For this problem, a natural statistic, the multivariate version of the statistic considered by Welch [24], is given as follows:

$$T = (\bar{\mathbf{Y}}_{\mathbf{1}} - \bar{\mathbf{Y}}_{\mathbf{2}})^\top(\mathbf{S}_e)^{-1}(\bar{\mathbf{Y}}_{\mathbf{1}} - \bar{Y}_2) \tag{3.2}$$

where $\boldsymbol{S}_e = \frac{\boldsymbol{S}_1}{n_1} + \frac{\boldsymbol{S}_2}{n_2}$.

T statistic is asymptotically distributed as chi-square with degrees of freedom p when n_1 and n_2 approach to infinite. This approach is not valid for the small values of n_1 and n_2. Under H_0 and the assumption of the homogeneity of covariance matrices ($\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$), $(n - p - 1)T/(p(n - 2))$ is distributed as F with degrees of freedom p and $n - p - 1$, where $n = n_1 + n_2$.

In the rest of this section, the tests given by Bennett [2], Johansen [14], Nel and Van Der Merwe [18], Krishnamoorthy and Yu [17] were introduced briefly.

3.1. Bennett Test

Bennett [2] proposed a test for the equality of two mean vectors for $n_2 \geq n_1$. This test statistic can be given as follows:

$$T_B = n_1 \bar{\mathbf{z}}^\top \mathbf{S}_{\mathbf{z}}^{-1} \bar{\mathbf{z}} \tag{3.3}$$

where $\boldsymbol{z_j} = \boldsymbol{Y_{1j}} - \sqrt{\frac{n_1}{n_2}} \boldsymbol{Y_{2j}} + \frac{1}{\sqrt{n_1 n_2}} \sum_{k=1}^{n_1} \boldsymbol{Y_{1k}} - \frac{1}{n_2} \sum_{k=1}^{n_2} \boldsymbol{Y_{2k}} \quad j = 1, ..., n_1$

and also $\bar{\boldsymbol{z}}$ and $\boldsymbol{S_z}$ are the mean and variance-covariance matrix of $\boldsymbol{z_j}$, $j = 1, .., n_1$, respectively. By using the following transformation, the distribution of test statistic can be obtained as follows:

$$F = \frac{n_1 - p}{p(n_1 - 1)} T_B \sim F_{p, n_1 - p}.$$

3.2. Johansen Test

Johansen [14] obtained a test given below:

$$T_J = \frac{T}{C} \tag{3.4}$$

Here, T is given in Eq. (2.6),

$$C = p - 2D - 6D/[p(p-1) + 2]$$

$$D = \sum_{i=1}^{2} \frac{1}{2(n_i - 1)} \left\{ tr(\mathbf{I} - \mathbf{V}^{-1}\mathbf{V}_i)^2 + [tr(\mathbf{I} - \mathbf{V}^{-1}\mathbf{V}_i)]^2 \right\}$$

where $\boldsymbol{V}_i = (\boldsymbol{S_i}/n_i)^{-1}, \quad i = 1, 2$ and $\boldsymbol{V} = \boldsymbol{V}_1 + \boldsymbol{V}_2$.

This test statistic is distributed as F with degrees of freedom p and f=$p(p+2)/(3\mathrm{D})$.

3.3. Nel and Van der Merwe Test

Nel and Van der Merwe [18] modified the test statistic given in Eq. (3.2) as follows:

$$T_{NV} = \frac{v - p + 1}{pv} T \sim F_{p, v-p+1}, \tag{3.5}$$

where $v = \dfrac{tr(\boldsymbol{S_e})^2 + [tr(\boldsymbol{S_e})]^2}{\frac{1}{n_1 - 1}\left\{ tr\left(\frac{\boldsymbol{S_1}}{n_1}\right)^2 + \left[tr\left(\frac{\boldsymbol{S_1}}{n_1}\right)\right]^2 \right\} + \frac{1}{n_2 - 1}\left\{ tr\left(\frac{\boldsymbol{S_2}}{n_2}\right)^2 + \left[tr\left(\frac{\boldsymbol{S_2}}{n_2}\right)\right]^2 \right\}}$

3.4. Krishnamoorthy and Yu Test

Krishnamoorthy and Yu [17] obtained a test statistic by modifying the test statistic given by Nel and Van der Merwe [18]. The test statistic is as follows:

$$T_M = \frac{(\hat{\upsilon}_M - p + 1)T}{p\hat{\upsilon}_M} \tag{3.6}$$

where

$$\hat{\upsilon}_M = \frac{p(p+1)(n-2)}{\hat{\varphi}_1 + \hat{\varphi}_2}$$

$$\hat{\varphi}_1 = \frac{n_2^2(n-2)}{n^2(n_1-1)}\left\{tr(\mathbf{S}_1\bar{\mathbf{S}}^{-\mathbf{1}})\right\}^2 + \frac{n_1^2(n-2)}{n^2(n_2-1)}\left\{tr(\mathbf{S}_2\bar{\mathbf{S}}^{-\mathbf{1}})\right\}^2$$

$$\hat{\varphi}_2 = \frac{n_2^2(n-2)}{n^2(n_1-1)}tr(\mathbf{S}_1\bar{\mathbf{S}}^{-\mathbf{1}}\mathbf{S}_1\bar{\mathbf{S}}^{-\mathbf{1}}) + \frac{n_1^2(n-2)}{n^2(n_2-1)}tr(\mathbf{S}_2\bar{\mathbf{S}}^{-\mathbf{1}}\mathbf{S}_2\bar{\mathbf{S}}^{-\mathbf{1}})$$

and

$$\bar{\mathbf{S}} = \frac{n_2}{n}\mathbf{S}_1 + \frac{n_1}{n}\mathbf{S}_2.$$

T_M is distributed as F with degrees of freedom p and $\hat{\upsilon}_M - p + 1$ [26].

3.5. Yao Test

Yao [27] proposed a test which is an extension of the Welch test provided by Tukey [23]. This test statistic T_Y based on T in Eq. (3.2) can be given as:

$$T_Y = \frac{m-p+1}{pm}T^2 \sim F_{p,m-p+1} \tag{3.7}$$

where $\frac{1}{m} = \frac{1}{(T)^2}\sum_{i=1}^{2}\frac{1}{n_i-1}\left[(\bar{\boldsymbol{Y}}_1 - \bar{\boldsymbol{Y}}_2)^\top(\boldsymbol{S_e})^{-1}\frac{\boldsymbol{S}_i}{n_i}(\boldsymbol{S_e})^{-1}(\bar{\boldsymbol{Y}}_1 - \bar{\boldsymbol{Y}}_2)\right]^2$

3.6. Wald Test and its Bootstrap Approach

Konietschke et al. [16] developed parametric bootstrap methods for the multivariate Behrens–Fisher problem. According to this, for two multivariate normal mean vectors , H_0 stated in Eq. (3.1) is equivalent to testing $H_0^\top$: $\boldsymbol{H}\boldsymbol{\mu}^* = \mathbf{0}$, where $\boldsymbol{\mu}^* = (\boldsymbol{\mu}_\mathbf{1}^\top, \boldsymbol{\mu}_\mathbf{2}^\top)^\top$ and contrast matrix is given by $\boldsymbol{H} = \boldsymbol{P} \otimes \boldsymbol{I}_p$.

Here

$$\mathbf{P} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

and $\boldsymbol{I}_p$ is the p-dimensional unit matrix.

The Wald test statistic for testing $H_0^\top$ is

$$Q_n(H) = n\bar{\mathbf{Y}}^\top \mathbf{H}(\mathbf{H}\hat{\mathbf{V}}_\mathbf{n}\mathbf{H})^+ \boldsymbol{H}\bar{\boldsymbol{Y}} \tag{3.8}$$

where $(.)^+$ denotes the Moore-Penrose inverse and $\hat{\boldsymbol{V}}_{\boldsymbol{n}} = diag\left(\frac{n}{n_i}\boldsymbol{S}_{(\boldsymbol{i})}; 1 \leq i \leq 2\right)$. $Q_n(H)$ is asymptotically distributed as χ^2 with degrees of freedom rank($\boldsymbol{H}$).

Konietschke et al. [16] applied the nonparametric and parametric bootstrap of the Wald test. The algorithm of these tests, respectively, are given below:

The nonparametric bootstrap of the Wald test:

1. For given data, calculate value of test statistic given in Eq.(3.8).
2. Generate nonparametric bootstrap sample, $\boldsymbol{Y}^\bullet_{i1}, \boldsymbol{Y}^\bullet_{i2}, ..., \boldsymbol{Y}^\bullet_{in_i}$, which are drawn with replacement from the pooled observation vectors, $\boldsymbol{Y}_{11}, ..., \boldsymbol{Y}_{2n_2}$.
3. Compute value of test statistic given in Eq. (3.8) from the nonparametric bootstrap sample and denote it by $Q_n^{\bullet(l)}(H)$.
4. Repeat the steps 2 and 3 for a large number of times (say, L times).
5. Compute the Monte Carlo estimates of the p-values as $\hat{p} = \sum_{l=1}^{L} I(Q_n^{\bullet(l)}(H) > Q_n(H))/L$, where I is the indicator function.
6. If $\hat{p} < \alpha$, H_0 is rejected.

We refer to the nonparametric bootstrap of the Wald test as WB in rest of study.

The parametric bootstrap of the Wald test:

1. For given data, calculate value of test statistic given in Eq.(3.8).
2. Generate parametric bootstrap variables as,

$$\boldsymbol{Y}^*_{i1}, \boldsymbol{Y}^*_{i2}, ..., \boldsymbol{Y}^*_{in_i} \sim N(0, \boldsymbol{S}_{(i)}), \quad i = 1, 2.$$

3. Compute value of test statistic given in Eq. (3.8) from the parametric bootstrap vectors, and denote it by $Q_n^{*(l)}(H)$.
4. Repeat the steps 2 and 3 for a large number of times (say, L times).

5. Compute the Monte Carlo estimates of the p-values as $\hat{p} = \sum_{l=1}^{L} I(Q_n^{*(l)}(H) > Q_n(H))/L$, where I is the indicator function.

6. If $\hat{p} < \alpha$, H_0 is rejected.

We refer to the parametric bootstrap of the Wald test as WPB in rest of study.

4. Computational Approach Test

Initially, before applying CAT for testing the null hypothesis given in Eq. (3.1), the general technique in CAT was first given.

Let $X_1, X_2, ..., X_n$ be a random sample having a probability density function as $f(x|\theta)$, where the functional form of f is assumed to be known. Let $\theta = (\theta^{(1)}, \theta^{(2)})$ be the parameter vector and our primary interest lies in $\theta^{(1)}$, i.e., $\theta^{(2)}$ is the nuisance parameter. Our goal is to test $H_0^\top : \theta^{(1)} = \theta_0^{(1)}$ versus a suitable alternative. To test $H_0^\top : \theta^{(1)} = \theta_0^{(1)}$ against $H_1^\top$, $H_0^\top$ was first expressed as $H_0^{\top *} : \eta(\theta^{(1)}, \theta_0^{(1)}) = 0$ against $H_1^{\top *}$, where η is a scalar valued function. The general methodology of the CAT for testing $H_0^{\top *} : \eta(\theta^{(1)}, \theta_0^{(1)}) = 0$ against a suitable alternative at a desired level α was given through the following steps [19].

1. Obtain the MLEs of the parameters, $\theta^{(1)}$ and $\theta^{(2)}$. Obtain a suitable $\eta(\theta^{(1)}, \theta_0^{(1)})$ and the MLE of η, $\hat{\eta} = \hat{\eta}(\hat{\theta}^{(1)}, \theta_0^{(1)})$ can be used as a test s-tatistic.

2. Under H_0, find the RMLEs of $\theta^{(2)}$ parameter, which is denoted by $\tilde{\theta}^{(2)}$. Generate artificial sample $Y_1, Y_2, ..., Y_n$ from $f(y|\theta_0^{(1)}, \tilde{\theta}^{(2)})$ large number of times, say L times.

3. For each of these replicated samples, recalculate the MLE of η, $\hat{\eta}^{*(l)}$ $l = 1, ..., L$.

4. Estimate the p-value as, $\hat{p} = \sum_{l=1}^{L} (\hat{\eta}^{*(l)} > \hat{\eta})/L$. In the case of $\hat{p} < \alpha$, H_0 is rejected.

CAT is based on restricted maximum likelihood estimations (RMLEs) under null hypothesis. There is no need to obtain theoretical distribution of test statistic and the p value can be calculated directly; therefore, this method is quite easy to apply. Then, the testing procedure based on CAT for the equality of two multivariate normal mean vectors under heterogeneity of covariance matrix can be given as follows:

The observed value of the test statistic based on random sample is calculated as follows:

Step 1 The observed value of the test statistic based on random sample is calculated as follows:

$$\hat{\eta}_{ML} = (\bar{\mathbf{Y}}_{\mathbf{1}} - \bar{\mathbf{Y}}_{\mathbf{2}})^\top \left(\frac{\hat{\Sigma}_1}{n_1} + \frac{\hat{\Sigma}_2}{n_2} \right)^{-1} (\bar{\mathbf{Y}}_{\mathbf{1}} - \bar{\mathbf{Y}}_{\mathbf{2}}) \tag{4.1}$$

Step 2 Under H_0, the RMLEs of $(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$ are obtained as $\hat{\boldsymbol{\mu}}_{(RML)}$ and $\hat{\boldsymbol{\Sigma}}_{i(RML)}$ in Eq. (2.5), iteratively.

Step 3 A large number is generated, say L, of artificial sample from $N_p(\hat{\boldsymbol{\mu}}_{(RML)}, \hat{\boldsymbol{\Sigma}}_{i(RML)})$ i=1,2. For every artificial sample $\hat{\eta}_{ML}^{(l)}$, $l = 1, ..., L$ are calculated.

Step 4 p value is calculated as $\hat{p} = \sum_{l=1}^{L} \frac{(\hat{\eta}_{ML}^{(l)} > \hat{\eta}_{ML})}{L}$. H_0 is rejected when $\hat{p} < \alpha$.

By using these steps, a simulation study was carried out.

First, we carry out the CAT for testing the H_0 using the test statistic $\hat{\eta}_{ML}$ in Eq. (4.1) and refer to this test statistics as CAT. Beside, we carry out the CAT using the test statistics T_B in Eq. (3.3), T_M in Eq. (3.6), T_J in Eq. (3.4), T_Y in Eq. (3.7) and T_{NV} in Eq. (3.5), we refer to these test statistics as B-CAT, M-CAT, J-CAT, Y-CAT, NV-CAT, respectively.

5. Simulation Study

In this section, all tests were compared with respect to their estimated type-I error rates and powers for multivariate normality and non-normality. For this purpose, the cases of p= 2, 3, 4 with different combinations of equal and unequal sample sizes were considered. To estimate type-I error rates and the powers of all tests under multivariate normality assumption, 2000 random numbers with a sample size n_i (i=1,2) from the multivariate normal distribution were generated. Mean vectors were used as $\boldsymbol{\mu}_1 = (0, 0, ..., 0)_{1\text{x}p}$ and $\boldsymbol{\mu}_2 = (\Delta, \Delta, ..., \Delta)_{1\text{x}p}$.

Following Konietschke et al.[16], we considered six covariance structure as:

Setting 1: $\boldsymbol{\Sigma}_1 = I_p + 0.5(J_p - I_p) = \boldsymbol{\Sigma}_2$,
Setting 2: $\boldsymbol{\Sigma}_1 = [\sigma_{rs}] = (0.6)^{|r-s|} = \boldsymbol{\Sigma}_2$,
Setting 3: $\boldsymbol{\Sigma}_1 = I_p + 0.5(J_p - I_p)$ and $\boldsymbol{\Sigma}_2 = 3I_p + 0.5(J_p - I_p)$,
Setting 4: $\boldsymbol{\Sigma}_1 = [\sigma_{rs}] = (0.6)^{|r-s|}$ and $\boldsymbol{\Sigma}_2 = (0.6)^{|r-s|} + 2I_p$,
Setting 5: $\boldsymbol{\Sigma}_1 = I_p + 0.5(J_p - I_p)$ and $\boldsymbol{\Sigma}_2 = 9I_p + 0.5(J_p - I_p)$,
Setting 6: $\boldsymbol{\Sigma}_1 = [\sigma_{rs}] = (0.6)^{|r-s|}$ and $\boldsymbol{\Sigma}_2 = (0.6)^{|r-s|} + 8I_p$,

where I_p is an identity vector with dimension p and J_p is the pxp matrix of 1's. While setting 1 represents a scenario with homoscedastic compound symmetric, settings 3 and 5 represent the scenarios with moderate and severe heteroscedastic versions of this structure, respectively. While setting 2 represents a scenario with homoscedastic autoregressive covariance structure, setting 4 and 6 represent the scenarios with moderate and severe heteroscedastic versions of this structure, respectively.

To calculate the $\hat{p}$-values of the CAT and CAT versions of the tests, m was taken as 2000. The simulation study was conducted in MATLAB. Initially, the estimated type-I error rates of all tests under the null hypothesis were calculated. The simulation results were provided in Tables 1-3 under the multivariate normal model at the nominal level 0.05.

Tables 1-3 are here.

It can be seen in Table 1 that for $p = 2$, in cases of homoscedastic and moderate heteroscedastic structures (setting 1, 2, 3 and 4), all tests except Wald test have the estimated type I error rates close to the nominal level 0.05. As heterogeneity is getting severe (setting 5 and 6), in cases of small sample sizes, the estimated type I error rates of the WB and T_Y tests as well as the Wald test exceed the nominal level 0.05, that is, these tests tend to be liberal.

It can be seen in Table 2 that in case of homoscedastic structure, the results of the tests for $p = 3$ show similar pattern as those for the cases of $p = 2$. However, in case of heterogeneity even when moderate heterogeneity, the T_Y, T_J, Wald, WB tests tend to be liberal for small sample sizes. Furthermore, as heterogeneity is getting severe, these tests tend to be highly liberal for small sample size. A remarkable consequence is that when sample sizes are different, the estimated type I error rates of the WB test are considerably lower than the nominal level 0.05.

It can be seen in Table 3 that for $p = 4$, in case of homoscedastic structure and different sample sizes, the T_Y and T_J, tests tend to be liberal. Furthermore, when small and equal sample sizes, these tests tend to be highly liberal in case of heterogeneity. As sample sizes increase, these tests have the estimated type I error rates close to the nominal level 0.05. Besides, in case of heteroscedastic structure, the estimated type I error rates of the WB test exhibit similar pattern as those for the cases of $p = 2$ and $p = 3$. However, as p increases, this test tend to be highly liberal. It can be seen from all tables that as p increase, the T_Y, T_J, and WB tests tend to be highly liberal. Furthermore, as heterogeneity is getting severe, these tests also tend to be highly liberal. The CAT versions of the $T_Y$and T_J, tests, J-CAT and Y-CAT, greatly improved these tests' behavior. In this cases, these tests have the estimated type I error rates close to the nominal level 0.05 in most cases. As for the WPB test, in homoscedastic structure, the estimated type I error rates of this test is slightly higher than the nominal level 0.05 when sample size is small and unequal. Furthermore, as heterogeneity is getting severe, the estimated type I error rates of this test is getting quite higher than the nominal level 0.05.

In summary, according to the results obtained from all tables, the CAT, K-CAT, J-CAT, Y-CAT, NV-CAT, T_M and WPB tests have the estimated type I error rates close to the nominal level 0.05 in most cases. Furthermore, as p increase and heterogeneity is getting severe, the estimated type I error rates of all tests are affected negatively from this case, that is, these tests behaviors depend on the p and degree of heterogeneity.

The simulated powers of all tests were provided in Tables 4-6 under the normal model. The tests attaining nominal level closely can be compared meaningfully in terms of power. Since the estimated type I error rates of the Wald test exceed 6% in all considered cases, the Wald test was ignored and excluded from tables. While the powers of the tests were interpreted, the tests which had greater the estimated type I error rates than 6% given in Tables 1-3 were disregarded. Thus, the estimated powers of these tests were denoted by $*$.

Tables 4-6 are here.

When all tests were compared in terms of their powers in Table 4-6, the T_B test and the CAT version of this test, B-CAT, has a smaller power than the other tests in most cases. In cases where the estimated type I error rates of the T_J, T_Y and WB tests are close the nominal level, the powers of these tests are close to the CAT, K-CAT, J-CAT, Y-CAT, NV-CAT, T_M and WPB tests, even sometimes the powers of these tests are slightly higher than those of these tests. However, as p and the degree of heterogeneity increase, and in case of small sample sizes, the T_J, T_Y and WB tests tend to be highly liberal, which is a disadvantage for them.

When the CAT, K-CAT, J-CAT, Y-CAT, NV-CAT, T_M and WPB tests were compared in terms of their powers, these tests have powers close to each other. Besides, while the CAT has a bit higher power than the other tests in some cases, the NV-CAT has a bit higher power than the other tests in some cases. Since the CAT has simple form than the other CAT versions of the tests, it can be preferred instead of the others where they have similar powers.

To get idea about robustness of the above tests against multivariate non-normality we conduct simulation study under multivariate non-normal models. Following Konietschke et al. [16], we generated data as

$$\mathbf{Y}_{ij} = \mu_i + \mathbf{\Sigma}_i^{1/2}\boldsymbol{\varepsilon}_{ij},\ i = 1, ..., n_i;\ j = 1, 2$$

using the Cholesky decomposition $\mathbf{\Sigma}_i^{1/2}$ of a given covariance matrix $\mathbf{\Sigma}_i$. The independent and identically distributed random error vectors $\varepsilon_{ij} = (\varepsilon_{ij}^{(1)}, ..., \varepsilon_{ij}^{(p)})^\top$

were generated from different standardized symmetric or skewed distributions by

$$\varepsilon_{ij}^{(s)} = \frac{W_{ij}^{(s)} - E(W_{ij}^{(s)})}{\sqrt{Var(W_{ij}^{(s)})}}$$

Here $W_{ij}^{(s)}$ are double exponential distribution (DE), t-distribution with degrees of freedom 7 (t_7), χ^2 distribution with degrees of freedom 15 (χ^2_{15}) and χ^2 distribution with degrees of freedom 20 (χ^2_{20}). We refer to these distributions as distribution 1, 2, 3 and 4 in tables, respectively. We estimated the type-I error rates of all tests under these distributions, respectively. The simulated results were provided in Tables 7-18.

Tables 7-18 are here.

The results under the double exponential model are almost similar to those under the normal model. However, unlike normal distribution, it can be seen that the estimated type I error rates of all tests are smaller than the nominal level 0.05 especially in small sample size. Besides, while the estimated type I error rates of the WB test are close to the nominal level in cases of homogeneity structure, those of this test are higher than the nominal level as the degree of heterogeneity.

The results under t_7-model are quite similar to those under the normal model. While the results under the χ^2_{15}-model are somewhat similar to those under the normal model, it can be seen that the estimated type I error rates of tests increase significantly. Because of skewed distribution, the T_J, T_Y and WB tests tend to be highly liberal when heterogeneity is severe and sample size is small. A remarkable consequence is that as p increase, especially in cases of small sample size, the estimated type I error rates of all tests are higher than the nominal level. However, the B-CAT test has the estimated type I error rate close to the nominal level. Also, note that as the degree of freedom increase, since χ^2_{20} distribution is more close to a symmetric distribution than χ^2_{15} distribution, the results under this model are more similar to those under the normal model.

In cases of normal distributed and symmetric distributed models, it can be seen that the CAT, the CAT versions of tests, T_M and WPB tests have the estimated type I error rates close to the nominal level 0.05. The T_J, T_Y and WB tests tend to be highly liberal in cases of small sample size and heterogeneity. In cases of model with skewed distribution, that is, the χ^2_{15}-model, the estimated type I error rates of many tests significantly exceed the nominal level 0.05. However, the B-CAT performs well than other tests in terms of type I error rate under this model.

As Konietschke et al. [16] noted and our simulation study can be seen, the WB test's behavior depends on the p, degree of heterogeneity and the amount of skewness. Furthermore, T_J and T_Y tests' behavior also depend on the p, degree of heterogeneity and the amount of skewness. Thus, as seen from simulation study, CAT method can be uses as a good alternative for the equality of two multivariate normal mean vectors under heterogeneity of covariance.

6. Conclusions

In this study, the CAT were proposed and compared it against the other popular tests (T_B, T_M, T_{NV}, T_J, T_Y, WB, WPB) as well as their CAT versions (B-CAT, M-CAT, J-CAT, Y-CAT, NV-CAT) to test the equality of two multivariate normal mean vectors under heterogeneity of covariance matrix. The results of Monte Carlo simulations that were conducted to compare the estimated type I error rates and powers of these tests were presented. The simulation study shows that the CAT, M-CAT, J-CAT, Y-CAT, NV-CAT, T_M and WPB tests performed better than the others in terms of both the estimated type I error rates and power, even the CAT and NV-CAT had a bit higher power than the other tests in some cases. This method can be adapted to the heterogeneity MANOVA models.

ACKNOWLEDGMENTS

The author is thankful to the referee and the editor for their constructive comments that led to a significant improvement of the paper.

REFERENCES

[1] Algina, J., Oshima, T. C. and Tang, K. L. (1991). Robustness of Yao's, James', and Johansen's tests under variance-covariance heteroscedasticity and non-normality, *Journal of Educational and Behavioral Statistics*, **16**, 2, 125–139.

[2] Bennett, B. M. (1951). Note on a Solution of the Generalized Behrens-Fisher Problem, *Annals of the Institute of Statistical Mathematics*, 2, 87–90.

[3] Chang, C. H. and Pal, N. (2008). A revisit to the Behrens–Fisher problem: comparison of five test methods, *Communications in Statistics—Simulation and Computation*, **37**, 6, 1064–1085.

[4] Chang, C. H., Pal, N., Lim, W. K. and Lin, J. J. (2010). Comparing several population means: a parametric bootstrap method, and its comparison with usual ANOVA F test as well as ANOM, *Computational Statistics* , **25**, 1, 71–95.

[5] Chang, C. H., Lin, J. J. and Pal, N. (2011). Testing the equality of several gamma means: a parametric bootstrap method with applications, *Computational Statistics* , **26**, 1, 55–76.

[6] Christensen, W. F. and Rencher, A. C. (1997). A comparison of Type I error rates and power levels for seven solutions to the multivariate Behrens-Fisher problem, *Communications in Statistics-Simulation and Computation*, **26**, 4, 1251–1273.

[7] De la Rey, N. and Nel, D. G. (1993). A comparison of the significance levels and power functions of several solutions to the multivariate Behrens-Fisher problem, *South African Statistical Journal*, **27**, 2, 129–148.

[8] Gökpinar, E.Y., and Gökpinar, F. (2012). A Test Based on the Computational Approach For Equality of Means Under the Unequal Variance Assumption, *Hacettepe Journal of Mathematics and Statistics*, **41**, 4, 1–9.

[9] Gökpinar, E.Y., Polat, E., Gökpinar, F. and Günay, S. (2013). A new computational approach for testing equality of inverse Gaussian means under heterogeneity, *Hacettepe Journal of Mathematics and Statistics*, **42**, 5, 581–590.

[10] Gökpinar, F. and Gökpinar, E (2015). A computational approach for testing equality of coefficients of variation in k normal populations, *Hacettepe Journal of Mathematics and Statistics*, **44**, 5, 1197–1213.

[11] Hwang, H. L. and Paulson, A. S. (1986). Some methods for the multivariate two-sample location problem, *Unpublished manuscript, Oak Ridge National Laboratory, Oak Ridge, TN.*

[12] Ito, K. and Schull, W. J. (1964). On the robustness of the To2 test in multivariate analysis of variance when variance-covariance matrices are not equal, *Biometrika*, **51**, 1-2, 71–82.

[13] James, G. S. (1954). Tests of Linear Hypotheses in Univariate and Multivariate Analysis When the Ratios of the Population Variances are Unknown, *Biometrika*, **41**, 19–43.

[14] Johansen, S. (1980). The Welch-James Approximation to the Distribution of the Residual Sum of Squares in a Weighted Linear Regression, *Biometrika*, **67**, 85–92.

[15] Kim, S. J. (1992) A practical solution to the multivariate Behrens-Fisher problem, *Biometrika*, **79**, 1, 171–176.

[16] Konietschke, F., Bathke, A.C., Harrar, S.W. and Pauly, M. (2015). Parametric and nonparametric bootstrap methods for general MANOVA, *Journal of Multivariate Analysis*, **140**, 291–301.

[17] Krishnamoorthy, K. and Yu, J. (2004). Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem, *Statistics & Probability Letters*, **66**, 2, 161–169.

[18] Nel, D. G. and Van der Merwe, C. A. (1986). A solution to the multivariate Behrens-Fisher problem, *Communications in Statistics-Theory and Methods*, **15**, 12, 3719–3735.

[19] Pal, N., Lim, W. K. and Ling, C. H. (2007). A computational approach to statistical inferences, *Journal of Applied Probability and Statistics*, **2**, 1, 13–35.

[20] Smaga, Ł. (2017). Bootstrap methods for multivariate hypothesis testing, *Communications in Statistics - Simulation and Computation*, **46**, 7654–7667.

[21] Subrahmaniam, K. and Subrahmaniam, K. (1973). On the multivariate Behrens-Fisher problem, *Biometrika*, **60**, 1, 107–11.

[22] Subrahmaniam, K. and Subrahmaniam, K. (1975). On the confidence region comparison of some solutions for the multivariate Behrens-Fisher problem, *Communications in Statistics-Theory and Methods*, **4**, 1, 57–67.

[23] Tukey, J. W. (1959). A quick compact two sample test to Duckworth's specifications, *Technometrics*, **1**, 1, 31–48.

[24] Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal, *Biometrika*, **29**, 3/4 350–362.

[25] Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population, *Biometrika*, **20A**, 1/2 32–52.

[26] Yanagihara, H. andYuan, K. H. (2005). Three approximate solutions to the multivariate Behrens–Fisher problem, *SCommunications in Statistics—Simulation and Computation*, **34**, 4 975–988.

[27] Yao, Y. (1965). An approximate degrees of freedom solution to the multivariate Behrens Fisher problem, *Biometrika*, **52**, 1/2 139–147.