TEST
DOI 10.1007/s11749-017-0549-z

ORIGINAL PAPER

Testing of multivariate repeated measures data with block exchangeable covariance structure

Ivan Žežula[1] · **Daniel Klein**[1] · **Anuradha Roy**[2]

Received: 1 January 2017 / Accepted: 2 August 2017

Abstract A new hypothesis testing of equality of mean vectors in two populations using D^2 statistic for multivariate repeated measures data on q response variables at p sites or time points in a block exchangeable covariance matrix setting is proposed. The minimum sample size needed for our new test is only $q + 1$, unlike $pq + 1$ in Hotelling's T^2 test. The new test is very efficient in small sample scenario, when the number of observations is not adequate to estimate the $pq \times pq$ dimensional unknown unstructured variance–covariance matrix. Some simulation studies are performed to compare the power of the new D^2 test and the existing BT^2 test. The performance of the proposed D^2 test is demonstrated with the two medical data sets.

Keywords BT^2 statistic · D^2 statistic · Hotelling's T^2 statistic · Lawley–Hotelling trace distribution

Mathematics Subject Classification 62H12 · 62H15 · 62H10 · 62E17

Žežula's and Klein's research were supported by grants VEGA MŠ SR 1/0344/14 and 1/0073/15.

Electronic supplementary material The online version of this article (doi:10.1007/s11749-017-0549-z) contains supplementary material, which is available to authorized users.

✉ Anuradha Roy
Anuradha.Roy@utsa.edu; anuroy150@gmail.com

[1] Faculty of Science, Institute of Mathematics, P. J. Šafárik University, 040 01 Kosice, Slovakia

[2] Department of Management Science and Statistics, The University of Texas at San Antonio, San Antonio, TX 78249, USA

Published online: 04 September 2017

1 Introduction

Multivariate hierarchical data, multivariate multilevel data, multivariate repeated measures data are data where the observations are frequently correlated. Uncorrelated error is often a violated assumption of the statistical procedures in these kinds of data. Violations often occur when the error terms are not independent, but correlated. For example, analysis of multivariate repeated measures data needs to take into account the correlations among the measurements of q different variables as well as the correlations among measurements taken at p different sites or time points.

The classical linear models for the multivariate data with a vector valued random variable (response vector) can be extended to a multivariate repeated measures data or doubly multivariate data with a matrix valued random variable $\boldsymbol{X}$. For example, a simple location model for doubly multivariate data can be presented as

$$\boldsymbol{X} = \boldsymbol{M} + \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{M}$ is a $q \times p$ location center matrix, which may depend on the explaining variables, and $\boldsymbol{E}$ is the $q \times p$ error matrix. The independent and identically distributed $q \times p$ matrix observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ may then come from a multivariate hierarchical model with two hierarchy levels (n classes in upper level, p individuals in the lower level with q response variables measured on each individual). Students in classes, patients in hospitals and repeated measurements on subjects are examples of these models, for more examples see Beacon and Thompson (1996) or Goldstein (2011). In the multivariate repeated measures case, one has n individuals, q variables and p repetitions for each variable (see Roy 2006; Roy and Fonseca 2012; Roy et al. 2015).

An example of multivariate data with structure defined by (1) is a study of cerebral metabolism in epileptic patients by Sperling et al. (1990). In this study, the metabolic rate of glucose at 16 locations in the brain by positron emission tomography (PET) scans was measured. These 16 locations include 8 regions of interest: the first five, frontal, sensorimotor, temporal, parietal and occipital are known as cortical regions, and the last three, caudate nucleus, lenticular nucleus and thalamus are known as subcortical regions. Sperling et al. (1990) measured the right-sided (R) and the left-sided (L) metabolic rates in each region of interest. Clearly, the data are doubly multivariate with $q = 5$ and $p = 2$, and $q = 3$ and $p = 2$ in cortical and subcortical regions, respectively. The sample consisted of 18 Normal control subjects, 8 patients with a Right brain hemisphere focus of the epilepsy and 8 patients with a Left brain hemisphere focus of the epilepsy. We use part of these data as an illustration of our method, namely Normal control and Right hemisphere focus groups. The metabolic rates for cortical region, measured in mg/100g/min, for all three groups are shown in Table 2 in Sperling et al. (1990) and reproduced in Table S1 in the online supplementary material for the Normal control and Right hemisphere focus groups. The metabolic rates for subcortical region, measured in mg/100g/min, for all three groups are shown in Table 3 in Sperling et al. (1990) and reproduced in Table S2 in the online supplementary material for the Normal control and Right hemisphere focus groups.

Classical multivariate analysis assumes both unstructured mean and unstructured variance–covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}(vec(\boldsymbol{E}))$, where the number of unknown parameters to be estimated is $r = pq(pq+1)/2$, which can be large for arbitrary values of p or q. However, for doubly multivariate data one may assume a separable covariance structure, i.e.,

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2, \tag{2}$$

where the $p \times p$ matrix $\boldsymbol{\Sigma}_1$ is the variance–covariance matrix of the p sites or time points (the same for all variables/locations) and the $q \times q$ matrix $\boldsymbol{\Sigma}_2$ is the variance–covariance matrix of the q variables or locations (the same for all sites/time points). Here $\otimes$ represents the Kronecker product. This choice (2) implies the assumption of separability of the total variability into two sources originated by the within- and between-variable variation. There are many articles which work with the models (1), but only few of them consider it in conjunction with the separable error term (e.g., Viroli 2012). The number of unknown parameters in the separable structure is only $s = p(p+1)/2 + q(q+1)/2$, which can be much less than r since the difference $r - s$ is quadratic function in p or q with the other dimension fixed, and biquadratic one for $p = q$.

If the p columns of $\boldsymbol{X}$ are exchangeable, then $E(\boldsymbol{X}) = \boldsymbol{M} = \boldsymbol{\xi}\mathbf{1}_p'$ (constant $q \times 1$ mean vector structure over time), and the p columns of $\boldsymbol{E}$ are also exchangeable [see, e.g., Arnold (1973, 1979) and Roy et al. (2015)]. Thus, the number of free parameters in the mean vector is only q, and the block exchangeable or block compound symmetry (BCS) covariance structure of $vec(\boldsymbol{E})$ is then

$$\begin{aligned}\boldsymbol{\Sigma} &= \begin{pmatrix} \boldsymbol{\Sigma}_0 & \boldsymbol{\Sigma}_1 & \dots & \boldsymbol{\Sigma}_1 \\ \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_0 & \dots & \boldsymbol{\Sigma}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_1 & \dots & \boldsymbol{\Sigma}_0 \end{pmatrix} \\ &= \boldsymbol{I}_p \otimes (\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1) + \boldsymbol{J}_p \otimes \boldsymbol{\Sigma}_1, \end{aligned} \tag{3}$$

where each column of $\boldsymbol{E}$ has the variance–covariance matrix $\boldsymbol{\Sigma}_0$ and any two different columns have the covariance matrix $\boldsymbol{\Sigma}_1$. Here $\boldsymbol{I}_p$ is the $p \times p$ identity matrix, $\boldsymbol{J}_p = \mathbf{1}_p\mathbf{1}_p'$, and $\mathbf{1}_p$ is the $p \times 1$ vector of ones. The pattern of the block exchangeable covariance structure arises from imposing symmetry on blocks of variables. From the definition of $\boldsymbol{\Sigma}$, it is clear that $p \geq 2$ is needed for the block exchangeable covariance structure. We assume that the $q \times q$ matrix $\boldsymbol{\Sigma}_0$ is positive definite (denoted by $\boldsymbol{\Sigma}_0 > 0$), and the $q \times q$ symmetric matrix $\boldsymbol{\Sigma}_1$ must satisfy $\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1 > 0$ and $\boldsymbol{\Sigma}_0 + (p-1)\boldsymbol{\Sigma}_1 > 0$ in order to guarantee the positive definiteness of $\boldsymbol{\Sigma}$ [for a proof, see Lemma 2.1 in Roy and Leiva (2011)].

However, one can also be faced situations when exchangeability of error terms does not go together with exchangeability of means, see Wilks (1946). We can see it at the above-cited example of cerebral metabolism data as well, since one cannot expect the same metabolic activity at different sides of the brain. That is why we choose to assume an unstructured mean $\boldsymbol{M}$ in the basic model. Symmetry is imposed on $\boldsymbol{X}$ for

considerations that are external to the data—imposed by external considerations. So, the covariance matrix is tied to the same invariance at different sides of the brain, that is why block exchangeable covariance structure is a direct consequence.

BCS covariance structure may be a realistic assumption in many doubly multivariate data. It has been studied most extensively by Arnold (1976) and Szatrowski (1976). There are many articles which work with the models (1), but only few of them consider it in conjunction with the block exchangeable error term (Arnold 1979). This structure has the desirable feature that it needs to estimate a smaller set of unknown parameters, which are $q(q+1)$ instead of r. Moreover, this number does not even depend on p (as is the case of both r and s). This means that one can get more information, e.g., by increasing the number of repeated measurements p, without estimating more parameters. Also, for this covariance structure the repeated measurements over time need not be equally spaced.

The aim of the paper is to derive the basic tests for the mean: one-sample as well as for two-sample cases under the assumption of block exchangeable covariance matrix for the data. For two-sample cases, we assume the equality of block exchangeable covariance matrices in the two populations. Hypothesis testing of block exchangeable structure for one population was developed by Roy and Leiva (2011) and then very recently by Coelho and Roy (2017). There are few methods proposed in the literature for testing either single mean vector or equality of two (or more) mean vectors for multivariate repeated measures populations with block exchangeable covariance structure. However, they are based on different principles (see later) and lead to different statistics which are either approximate or still difficult to handle (e.g., product of beta random variables). Since our method provides an exact (and effectively computable) distribution of the test statistic under normality, it turns out to be more effective than the previous ones, especially for small sample sizes.

One may also use the random effects models instead of introducing new mean parameters for each site or time point. This would limit the number of parameters, and similar hypotheses could be tested based on them, however, under other assumptions. Lin and Wang (2013) proposed a multivariate skew-normal linear mixed model by assuming a multivariate skew-normal distribution for the random effects and a multivariate normal distribution for the random errors for skewed distribution. They used a continuous-time damped exponential correlation structure to address the within-subject autocorrelation among irregularly observed measures. Lin and Wang (2013) developed a computationally tractable alternative expectation–conditional maximization algorithm for carrying out the maximum likelihood estimation. Wang et al (2015) extended multivariate-t linear mixed model with left- and/or right-censored responses embodied within multivariate repeated measurements collected at (possibly) irregularly occasions in the presence of potential outliers or heavy-tailed noises simultaneously.

Problems using the doubly multivariate data in paired sample (unstructured) mean testing using the BCS covariance structure was considered by Roy et al. (2015) in small sample case. However, they did not discuss the testing of mean vectors in two independent populations. Testing procedures for both structured and unstructured mean are developed in Sect. 2. They are applied to two medical data sets in Sect. 3. Some concluding remarks are given in Sect. 4.

2 Tests for the mean

Hypothesis testing of mean vector in a doubly multivariate framework is much more difficult than in a multivariate framework as the number of parameters increases with the increase of p. In this article, we develop a two-sample test procedure for the mean vector $\boldsymbol{\mu}$ in a doubly multivariate setup using BCS covariance structure as defined in (3). However, before developing two-sample test, it is necessary to develop a one-sample test of mean vector in the same setup. One-sample test is developed in Sect. 2.1 and two-sample test in Sect. 2.3, respectively.

We assume $\boldsymbol{x} = vec(\boldsymbol{X}) = \left(\boldsymbol{x}_1^{*\prime}, \ldots, \boldsymbol{x}_p^{*\prime}\right)' \sim N_{pq}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\mathrm{E}(\boldsymbol{x}) = \boldsymbol{\mu} = vec(\boldsymbol{M}) = \left(\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_p'\right)'$ and $\mathrm{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}$ as defined in (3).

2.1 One-sample test

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be random samples from $N_{pq}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; $\boldsymbol{x}_i = \left(\boldsymbol{x}_{i,1}^{*\prime}, \ldots, \boldsymbol{x}_{i,p}^{*\prime}\right)' \ \forall\, i = 1, \ldots, n$, and $\overline{\boldsymbol{x}}$ be the sample mean of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. We test the following hypothesis

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{against} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0. \tag{4}$$

Let $\underset{\sim}{\boldsymbol{X}} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \left(\underbrace{\boldsymbol{X}_{\bullet 1}^{*\prime}}, \ldots, \underbrace{\boldsymbol{X}_{\bullet p}^{*\prime}}\right)'$ be the $(pq \times n)$-dimensional data matrix, where $\underbrace{\boldsymbol{X}_{\bullet j}^{*}}$, $i = 1, \ldots, p$ is the data matrix at the ith site or time point.

Let $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{+}\boldsymbol{A}'$ be the orthogonal projector onto the column space of any matrix $\boldsymbol{A}$, and $\boldsymbol{Q}_{\boldsymbol{A}} = \boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{A}}$ be the orthogonal projector on its orthogonal complement, where $(\boldsymbol{A}'\boldsymbol{A})^{+}$ is the Moore–Penrose inverse of $(\boldsymbol{A}'\boldsymbol{A})$. For the sake of simplification, the matrix $\boldsymbol{P}_{\boldsymbol{J}_n}$ and $\boldsymbol{Q}_{\boldsymbol{J}_n}$ will be denoted by $\boldsymbol{P}_n$ and $\boldsymbol{Q}_n$, respectively.

We can define the $(pq \times pq)$-dimensional sample variance–covariance matrix $\boldsymbol{S}$ as follows

$$\boldsymbol{S} = \frac{1}{n-1}\underset{\sim}{\boldsymbol{X}}\boldsymbol{Q}_n\underset{\sim}{\boldsymbol{X}}' = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} & \ldots & \boldsymbol{S}_{1p} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} & \ldots & \boldsymbol{S}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{S}_{p1} & \boldsymbol{S}_{p2} & \ldots & \boldsymbol{S}_{pp} \end{pmatrix},$$

where $\boldsymbol{S}_{ij} = \frac{1}{n-1}\underbrace{\boldsymbol{X}_{\bullet i}^{*}}\boldsymbol{Q}_n\underbrace{\boldsymbol{X}_{\bullet j}^{*\prime}}$; $i, j = 1, \ldots, p$. The matrix $\boldsymbol{S}$ is an unbiased estimator of $\boldsymbol{\Sigma}$, and it is known that $\boldsymbol{S}$ has the Wishart distribution with $n-1$ degrees of freedom and covariance matrix $\frac{1}{n-1}\boldsymbol{\Sigma}$, i.e., $\boldsymbol{S} \sim W_{pq}\left(n-1, \frac{1}{n-1}\boldsymbol{\Sigma}\right)$. Therefore, $\mathrm{E}\left[\boldsymbol{S}_{ij}\right] = \boldsymbol{\Sigma}_{1-\delta_{ij}}$, where $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$. It is natural to use the following unbiased estimators of the variance and covariance matrices $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$:

$$\widehat{\boldsymbol{\Sigma}}_0 = \frac{1}{p}\sum_{i=1}^{p}\boldsymbol{S}_{ii}, \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_1 = \frac{1}{p(p-1)}\sum_{\substack{i=1 \\ i\neq j}}^{p}\sum_{j=1}^{p}\boldsymbol{S}_{ij}.$$

Then the estimator $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{I}_p \otimes \left(\widehat{\boldsymbol{\Sigma}}_0 - \widehat{\boldsymbol{\Sigma}}_1\right) + \boldsymbol{J}_p \otimes \widehat{\boldsymbol{\Sigma}}_1$ could be written as the sum of two orthogonal parts

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{P}_p \otimes \widehat{\boldsymbol{\Delta}}_2 + \boldsymbol{Q}_p \otimes \widehat{\boldsymbol{\Delta}}_1,$$

where $\widehat{\boldsymbol{\Delta}}_1 = \widehat{\boldsymbol{\Sigma}}_0 - \widehat{\boldsymbol{\Sigma}}_1$ and $\widehat{\boldsymbol{\Delta}}_2 = \widehat{\boldsymbol{\Sigma}}_0 + (p-1)\widehat{\boldsymbol{\Sigma}}_1$. Since $\boldsymbol{P}_p \boldsymbol{Q}_p = \boldsymbol{0}$, the inverse of $\widehat{\boldsymbol{\Sigma}}$ can be written as $\widehat{\boldsymbol{\Sigma}}^{-1} = \boldsymbol{P}_p \otimes \widehat{\boldsymbol{\Delta}}_2^{-1} + \boldsymbol{Q}_p \otimes \widehat{\boldsymbol{\Delta}}_1^{-1}$. Then for testing Hypothesis (4), we define the test statistic

$$D^2 = n\,(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)' \left[\boldsymbol{P}_p \otimes \widehat{\boldsymbol{\Delta}}_2^{-1} + \boldsymbol{Q}_p \otimes \widehat{\boldsymbol{\Delta}}_1^{-1}\right](\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0). \tag{5}$$

The distribution of this D^2 test statistic is not Hotelling's T^2, since the estimator $\widehat{\boldsymbol{\Sigma}}$ does not follow the Wishart distribution. We derive the distribution of this test statistic in Sect. 2.1.2. In the next section, we present the test statistic presented in Roy et al. (2015) and show that their test statistic has smaller power than the one presented in this paper.

2.1.1 Similar solution of the problem via orthogonalization

Let $\boldsymbol{Z} = \boldsymbol{H}_p \otimes \boldsymbol{I}_q$ be any matrix such that $\boldsymbol{H}_p$ is a $(p \times p)$ orthogonal matrix with the first row proportional to a vector of 1's. Then, the canonical transformation

$$\boldsymbol{y} = \left(\boldsymbol{H}_p \otimes \boldsymbol{I}_q\right)\boldsymbol{x}$$

of the data is used.

Therefore, $\boldsymbol{y} = \left(\boldsymbol{y}_1{}', \ldots, \boldsymbol{y}_p{}'\right)' = \boldsymbol{Z}\boldsymbol{x} \sim N_{pq}\,(\boldsymbol{\nu}, \boldsymbol{\Omega})$ where $\boldsymbol{\nu} = \boldsymbol{Z}\boldsymbol{\mu}$, and according to Roy and Fonseca (2012)

$$\boldsymbol{\Omega} = \boldsymbol{Z}\boldsymbol{\Sigma}\boldsymbol{Z}' = \begin{pmatrix} \boldsymbol{\Delta}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{p-1} \otimes \boldsymbol{\Delta}_1 \end{pmatrix}.$$

It follows that the $q \times 1$ component vectors $\boldsymbol{y}_i,\ i = 1, \ldots, p$, are independent. Even though the estimator $\widehat{\boldsymbol{\Omega}} = \begin{pmatrix} \widehat{\boldsymbol{\Delta}}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{p-1} \otimes \widehat{\boldsymbol{\Delta}}_1 \end{pmatrix}$ does not have a Wishart distribution, the following theorem holds:

Theorem 1 *Distributions of* $(n-1)(p-1)\widehat{\boldsymbol{\Delta}}_1$ *and* $(n-1)\widehat{\boldsymbol{\Delta}}_2$ *are independent, and*

$$\begin{aligned} (n-1)(p-1)\widehat{\boldsymbol{\Delta}}_1 &\sim W_q\left((n-1)(p-1), \boldsymbol{\Delta}_1\right), \quad \textit{and} \\ (n-1)\widehat{\boldsymbol{\Delta}}_2 &\sim W_q\left(n-1, \boldsymbol{\Delta}_2\right). \end{aligned}$$

For the proof of the above theorem, see Theorem 1 in Roy et al. (2015). For testing Hypothesis (4), they derived a similar to Hotelling's T^2 test statistic named Block T^2

$$BT^2 = n\,(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\,\boldsymbol{Z}' \begin{pmatrix} \widehat{\boldsymbol{\Delta}}_2^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_{p-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1} \end{pmatrix} \boldsymbol{Z}\,(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) \sim T^2_{q,n-1} \oplus T^2_{q,(n-1)(p-1)}, \tag{6}$$

where $\oplus$ denotes the convolution operation and $T^2_{\cdot,\cdot}$ denotes Hotelling T^2 distribution with appropriate parameters.

However, such a test statistic is not independent of the choice of the orthogonal matrix $\boldsymbol{H}_p$ which is involved in $\boldsymbol{Z}$. Even if for a given data matrix $\boldsymbol{X}$ the distribution of BT^2 is the same for two different $\boldsymbol{H}_p^1$ and $\boldsymbol{H}_p^2$, it can be easily verified that the particular value of the test statistic is not the same with the exception of $p = 2$. As a consequence, different power can be achieved by using different $\boldsymbol{H}_p$'s. In order to achieve the highest power, one should use the matrix $\boldsymbol{H}_p$ which gives the highest value of BT^2. Unfortunately, such $\boldsymbol{H}_p$ depends on the data observed. However, the following lemma shows a connection between the maximum of BT^2 and D^2.

Lemma 1 *It holds*

$$\max_{H_p} BT^2 \leq D^2,$$

where the maximization is over all orthogonal matrices $\boldsymbol{H}_p$ with the first row proportional to a vector of 1's.

Proof Let us consider the given fixed data matrix $\boldsymbol{X}$ and let us write $\boldsymbol{H}_p = \left(\boldsymbol{h}_1, \boldsymbol{H}'_{2,p}\right)'$, where $\boldsymbol{h}_1 = \frac{1}{\sqrt{p}}\boldsymbol{1}_p$ and $\boldsymbol{H}'_{2,p} = \left(\boldsymbol{h}_2, \ldots, \boldsymbol{h}_p\right)$. The orthogonality of $\boldsymbol{H}_p$ immediately implies

$$\boldsymbol{H}_{2,p}\boldsymbol{H}'_{2,p} = \boldsymbol{I}_{p-1}, \tag{7a}$$

$$\boldsymbol{H}'_{2,p}\boldsymbol{H}_{2,p} = \boldsymbol{Q}_p, \quad \text{and} \tag{7b}$$

$$\boldsymbol{h}'_1\boldsymbol{h}_j = 0, \quad j = 2, \ldots, p. \tag{7c}$$

Since $\boldsymbol{h}_1\boldsymbol{h}'_1 = \boldsymbol{P}_p$, observe that

$$\begin{aligned}
\boldsymbol{Z}'\begin{pmatrix} \widehat{\boldsymbol{\Delta}}_2^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_{p-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1} \end{pmatrix}\boldsymbol{Z} &= (\boldsymbol{H}'_p \otimes \boldsymbol{I}_q)\begin{pmatrix} \widehat{\boldsymbol{\Delta}}_2^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_{p-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1} \end{pmatrix}(\boldsymbol{H}_p \otimes \boldsymbol{I}_q) \\
&= \left(\boldsymbol{h}_1 \otimes \boldsymbol{I}_q; \boldsymbol{H}'_{2,p} \otimes \boldsymbol{I}_q\right)\begin{pmatrix} \widehat{\boldsymbol{\Delta}}_2^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_{p-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1} \end{pmatrix}\begin{pmatrix} \boldsymbol{h}'_1 \otimes \boldsymbol{I}_q \\ \boldsymbol{H}_{2,p} \otimes \boldsymbol{I}_q \end{pmatrix} \\
&= \left(\boldsymbol{h}_1 \otimes \widehat{\boldsymbol{\Delta}}_2^{-1}; \boldsymbol{H}'_{2,p}\boldsymbol{P}_{p-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1}\right)\begin{pmatrix} \boldsymbol{h}'_1 \otimes \boldsymbol{I}_q \\ \boldsymbol{H}_{2,p} \otimes \boldsymbol{I}_q \end{pmatrix} \\
&= \boldsymbol{P}_p \otimes \widehat{\boldsymbol{\Delta}}_2^{-1} + \boldsymbol{H}'_{2,p}\boldsymbol{P}_{p-1}\boldsymbol{H}_{2,p} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1},
\end{aligned}$$

and using (7b) the difference between the matrices in the quadratics forms (5) and (6) is given by

$$\begin{aligned}
\left(\boldsymbol{P}_p \otimes \widehat{\boldsymbol{\Delta}}_2^{-1} + \boldsymbol{Q}_p \otimes \widehat{\boldsymbol{\Delta}}_1^{-1}\right) - \boldsymbol{Z}'\begin{pmatrix} \widehat{\boldsymbol{\Delta}}_2^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_{p-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1} \end{pmatrix}\boldsymbol{Z} \\
= \left(\boldsymbol{H}'_{2,p}\boldsymbol{H}_{2,p} - \boldsymbol{H}'_{2,p}\boldsymbol{P}_{p-1}\boldsymbol{H}_{2,p}\right) \otimes \widehat{\boldsymbol{\Delta}}_1^{-1} \\
= \left(\boldsymbol{H}'_{2,p}\boldsymbol{Q}_{p-1}\boldsymbol{H}_{2,p}\right) \otimes \widehat{\boldsymbol{\Delta}}_1^{-1}.
\end{aligned}$$

Therefore,

$$D^2 - BT^2 = n\,(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)' \left[\left(\boldsymbol{H}'_{2,p}\, \boldsymbol{Q}_{p-1} \boldsymbol{H}_{2,p} \right) \otimes \widehat{\boldsymbol{\Delta}}_1^{-1} \right] (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0). \tag{8}$$

Since $\boldsymbol{Q}_{p-1}$ is symmetric and idempotent, and $\widehat{\boldsymbol{\Delta}}_1^{-1}$ is positive definite with probability one, it follows that $\left(\boldsymbol{H}'_{2,p}\, \boldsymbol{Q}_{p-1} \boldsymbol{H}_{2,p} \right) \otimes \widehat{\boldsymbol{\Delta}}_1^{-1}$ is a positive semidefinite matrix with probability one for any $\boldsymbol{H}_{2,p}$ satisfying (7a)–(7c), which means that $D^2 - BT^2 \geq 0$. □

Lemma 2 *There exists* $\boldsymbol{H}_p$ *satisfying* (7a)–(7c) *for which* $\max_{H_p} BT^2 = D^2$ *only if* $p = 2$ *(for any* q*) or* $q = 1$ *(for any* p*).*

Proof If $p = 2$ then obviously $\boldsymbol{Q}_{p-1} = \boldsymbol{Q}_1 = 0$, therefore $D^2 - BT^2 = 0$ in (8). Now, let $p > 2$ and let $\boldsymbol{T}$ be any $(p-1) \times (p-2)$ matrix which spans the orthogonal complement of the space generated by $\frac{1}{\sqrt{p-1}} \mathbf{1}_{p-1}$ such that $\boldsymbol{Q}_{p-1} = \boldsymbol{T}\boldsymbol{T}'$. Observe that $\boldsymbol{H}'_{2,p}\boldsymbol{T} \neq \mathbf{0}$, since it contains nontrivial linear combinations of linearly independent vectors $\boldsymbol{h}_2, \ldots, \boldsymbol{h}_p$. Clearly, (8) is zero iff $\left[\boldsymbol{T}'\boldsymbol{H}_{2,p} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1/2} \right] (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) = \mathbf{0}$. Since $\widehat{\boldsymbol{\Delta}}_1$ is positive definite with probability one (and thus nonsingular), this is equivalent to

$$\left[\boldsymbol{T}'\boldsymbol{H}_{2,p} \otimes \boldsymbol{I}_q \right] (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) = \mathbf{0}. \tag{9}$$

Using the fact that $\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0 \overset{\text{df}}{=} \text{vec}\left(\overline{\boldsymbol{X}} - \boldsymbol{M}_0\right)$, where $\overline{\boldsymbol{X}}$ and $\boldsymbol{M}_0$ are $q \times p$ matrices, and properties of vec operator we can rewrite last Eq. (9) as

$$\boldsymbol{T}'\boldsymbol{H}_{2,p} \left(\overline{\boldsymbol{X}} - \boldsymbol{M}_0\right)' = \mathbf{0}. \tag{10}$$

Let $\boldsymbol{t}_i$, $i = 1, \ldots, p-2$, denote ith column of the matrix $\boldsymbol{H}'_{2,p}\boldsymbol{T}$, i.e., $\boldsymbol{t}_i$ is linear combination of p-dimensional vectors $\boldsymbol{h}_2, \ldots, \boldsymbol{h}_p$, and since $\boldsymbol{T}$ is of full column rank, $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_{p-2}$ are also linearly independent and they form a subspace $\mathcal{C}_1$ of $\mathbb{R}^p$ of dimension $p-2$.

Let $\boldsymbol{k}_j$, $j = 1, \ldots, q$, denote jth column of matrix $\left(\overline{\boldsymbol{X}} - \boldsymbol{M}_0\right)'$, i.e., $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_q$ are p-dimensional stochastic vectors, hence linear independent with probability one. Thus, vectors $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_q, \mathbf{1}_p$ form a subspace $\mathcal{C}_2$ of $\mathbb{R}^p$ of dimension $q+1$ with probability 1.

Relations (10) and (7c) imply that subspaces $\mathcal{C}_1$ and $\mathcal{C}_2$ have to be orthogonal. In $\mathbb{R}^p$, it is possible only when $p - 2 + q + 1 \leq p$, i.e., $q \leq 1$. □

2.1.2 Distribution of the test statistic D^2 under H_0

Let us denote $\boldsymbol{b} = \left(\boldsymbol{H}_p \otimes \boldsymbol{I}_q\right)(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)$. We consider vector $\boldsymbol{b}$ be partitioned in p subvectors as $\boldsymbol{b} = (\boldsymbol{b}'_1, \ldots, \boldsymbol{b}'_p)'$, where $\boldsymbol{b}_i$, $i = 1, \ldots, p$, is q-dimensional subvector. Then $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p$ are independently normally distributed with

$$\boldsymbol{b}_1 \sim N_q\left(\mathbf{0}; \frac{1}{n}\boldsymbol{\Delta}_2\right),$$
$$\boldsymbol{b}_i \sim N_q\left(\mathbf{0}; \frac{1}{n}\boldsymbol{\Delta}_1\right) \quad \text{for } i = 2, \ldots, p,$$

under null hypothesis H_0. Since

$$\boldsymbol{H}_p \boldsymbol{P}_p \boldsymbol{H}_p' = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} = \boldsymbol{e}_{1:p}\boldsymbol{e}_{1:p}',$$

and

$$\boldsymbol{H}_p \boldsymbol{Q}_p \boldsymbol{H}_p' = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \sum_{i=2}^{p} \boldsymbol{e}_{i:p}\boldsymbol{e}_{i:p}',$$

where $\boldsymbol{e}_{i:p}$ is the ith column of $\boldsymbol{I}_p$, the statistic D^2 can be written in the form

$$\begin{aligned} D^2 &= n\,(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)' \left[\boldsymbol{P}_p \otimes \widehat{\boldsymbol{\Delta}}_2^{-1} + \boldsymbol{Q}_p \otimes \widehat{\boldsymbol{\Delta}}_1^{-1}\right](\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) \\ &= n\boldsymbol{b}' \left(\boldsymbol{H}_p \otimes \boldsymbol{I}_q\right)\left[\boldsymbol{P}_p \otimes \widehat{\boldsymbol{\Delta}}_2^{-1} + \boldsymbol{Q}_p \otimes \widehat{\boldsymbol{\Delta}}_1^{-1}\right]\left(\boldsymbol{H}_p \otimes \boldsymbol{I}_q\right)' \boldsymbol{b} \\ &= n\boldsymbol{b}_1' \widehat{\boldsymbol{\Delta}}_2^{-1} \boldsymbol{b}_1 + n\sum_{i=2}^{p} \boldsymbol{b}_i' \widehat{\boldsymbol{\Delta}}_1^{-1} \boldsymbol{b}_i \\ &= \operatorname{tr}\left[n\boldsymbol{b}_1\boldsymbol{b}_1' \widehat{\boldsymbol{\Delta}}_2^{-1}\right] + \operatorname{tr}\left[\sum_{i=2}^{p} n\boldsymbol{b}_i\boldsymbol{b}_i' \widehat{\boldsymbol{\Delta}}_1^{-1}\right] \overset{df}{=} T_{01}^2 + T_{02}^2. \end{aligned} \tag{11}$$

Clearly, T_{01}^2 and T_{02}^2 are independent. Since $n\boldsymbol{b}_1\boldsymbol{b}_1' \sim W_q\,(1, \boldsymbol{\Delta}_2)$ and $(n-1)\widehat{\boldsymbol{\Delta}}_2 \sim W_q\,(n-1, \boldsymbol{\Delta}_2)$, T_{01}^2 has Lawley–Hotelling trace (LH-trace) distribution $T_0^2(q; 1, n-1)$ if $n - 1 \geq q$.

Similarly, $\sum_{i=2}^{p} n\boldsymbol{b}_i\boldsymbol{b}_i' \sim W_q\,(p-1, \boldsymbol{\Delta}_1)$ and $(n-1)(p-1)\widehat{\boldsymbol{\Delta}}_1 \sim W_q\,((n-1)(p-1), \boldsymbol{\Delta}_1)$ imply that T_{02}^2 has LH-trace distribution $T_0^2(q; p-1, (n-1)(p-1))$ if $(n-1)(p-1) \geq q$.

Therefore, from (11) we see that the distribution of D^2 is the convolution of two LH-trace distributions, i.e., $T_0^2(q; 1, n-1) \oplus T_0^2(q; p-1, (n-1)(p-1))$. Unfortunately, there is no simple way for obtaining critical values of this convolution, so that we have to use simulations. However, LH-trace distribution is usually approximated by F-distribution (see McKeon 1974). In fact, $T_0^2(q; 1, n-1)$ is equal to usual Hotelling $T_{q,n-1}^2$, which is equivalent to $\frac{(n-1)q}{n-q} F(q, n-q)$. The second term, $T_0^2(q; p-1, (n-1)(p-1))$, can be approximated by

$$\frac{(n-1)(p-1)^2 q}{np - n - p - q} \cdot \frac{b-2}{b} F((p-1)q, b),$$

where

$$b = 4 + \frac{(pq - q + 2)(np - n - p - q - 2)(np - n - p - q + 1)}{(np - n - p)(p + q) - (q - 1)(q + 2)}.$$

Therefore, we can approximate the distribution of D^2 by the convolution of the two (one exact and other approximating) F-distributions, where its critical values can be obtained by the method of Dyer (1982). It is interesting, from practical point of view, to compare these critical values with the simulated ones.

Thus, we see that the first convolution terms in the distributions of BT^2 and D^2 are the same. The difference is made by the second term, where $T^2_{q,(n-1)(p-1)}$ changes to $T^2_0(q;\, p-1, (n-1)(p-1))$.

Remark 1 The statistic T^2_{01} has LH-trace distribution $T^2_0(q;\, 1, n-1)$ if $n-1 \geq q$, and T^2_{02} has LH-trace distribution $T^2_0(q;\, p-1, (n-1)(p-1))$ if $(n-1)(p-1) \geq q$. Since $p \geq 2$ for BCS structure as mentioned in the Sect. 1, we have $(n-1)(p-1) \geq q$ when $n - 1 \geq q$. Therefore, the only condition needed on sample size in order to have T^2_{01} and T^2_{02} LH-trace distributions is $n - 1 \geq q$, i.e., $n \geq q + 1$; regardless of p. In other words, the minimum sample size needed to compute the D^2 test statistic is $q + 1$. The minimum sample size needed to compute the BT^2 test statistic is also $q + 1$ (see Roy et al. 2015), whereas the minimum sample size needed to compute the Hotelling's T^2 test statistic is $pq + 1$. Thus, we see that for a small sample data set where $n \leq pq$ one cannot compute Hotelling's T^2 test statistic, which is possible for the BT^2 and D^2 test statistics.

2.1.3 Exchangeable mean structure

Let us now assume that $\boldsymbol{\mu} = \mathbf{1}_p \otimes \boldsymbol{\xi}_q$. Using similar method as above one can easily arrive at the estimator

$$\widehat{\boldsymbol{\xi}} = \frac{1}{np}\left(\mathbf{1}'_p \otimes \boldsymbol{I}_q\right)\underset{\sim}{\boldsymbol{X}}\mathbf{1}_n \sim N_q\left(\boldsymbol{\xi}, \frac{1}{np}\boldsymbol{\Delta}_2\right).$$

Since $\boldsymbol{Q}_n\mathbf{1}_n = \mathbf{0}$, $\underset{\sim}{\boldsymbol{X}}\boldsymbol{Q}_n\underset{\sim}{\boldsymbol{X}}'$ and $\underset{\sim}{\boldsymbol{X}}\mathbf{1}_n$ are independent. It follows that $\boldsymbol{S}$ and $\widehat{\boldsymbol{\xi}}$ are independent, and also $\widehat{\boldsymbol{\Delta}}_2$ and $\widehat{\boldsymbol{\xi}}$. Thus, Hotelling T^2 statistic for testing $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi}_0$ against $H_1 : \boldsymbol{\xi} \neq \boldsymbol{\xi}_0$ is

$$T^2 = np\left(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0\right)'\widehat{\boldsymbol{\Delta}}_2^{-1}\left(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0\right) \sim T^2_{q,n-1}.$$

This is very similar to the first component of BT^2 or D^2 statistic.

2.2 Power comparison

As it was already explained, choosing different transformation matrices in BT^2 statistic can lead to different power of the test. However, the maximum $\max_{H_p} BT^2$ depends

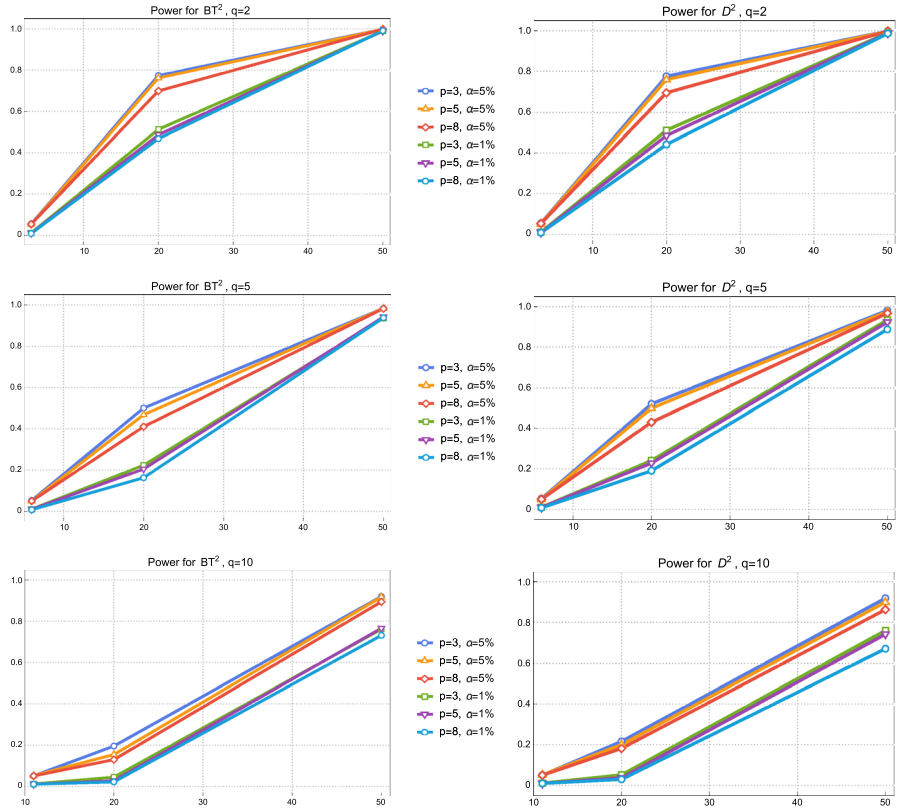

Fig. 1 Power comparison of BT^2 and D^2 for different values of q and p at 5% and 1% level of significance α

on the data; it is not obtained for fixed H_p. On the other hand, for $q > 1$ there is a gap between this maximum and D^2, which also has another distribution, so that the tests are not directly comparable. That is why we did some simulations to compare the powers of the two alternative tests. The maximization of BT^2 was done numerically using software Mathematica. Critical values of the test are computed using R. Implementation details and the R code to compute the critical values of D^2, BT^2 and Hotelling's T^2 tests are freely available as online supplementary material. In the R code, the functions are written in such a way that the user needs to ensure that the data are in the form of $pq \times n$ matrix, where each column is composed of $p\,q$-dimensional vectors one below other (for detail see the online supplementary material). Results of empirical power of BT^2 and D^2 based on 10,000 samples of sizes $n = q + 1$, 20 and 50 (we used only 3 different sample sizes since finding numerically H_p for each data set is very time consuming) are presented in Fig. 1, where horizontal axis represents the sample size and vertical axis represents the power. Figure 1 portrays the power comparison of BT^2 and D^2 for different values of $q = 2$, 5 and 10, and for different values of $p = 3$, 5 and 8 at 5% as well as 1% level of significance α. As expected the power of both BT^2 and D^2 increases with sample size n for all q, p for both 5% and

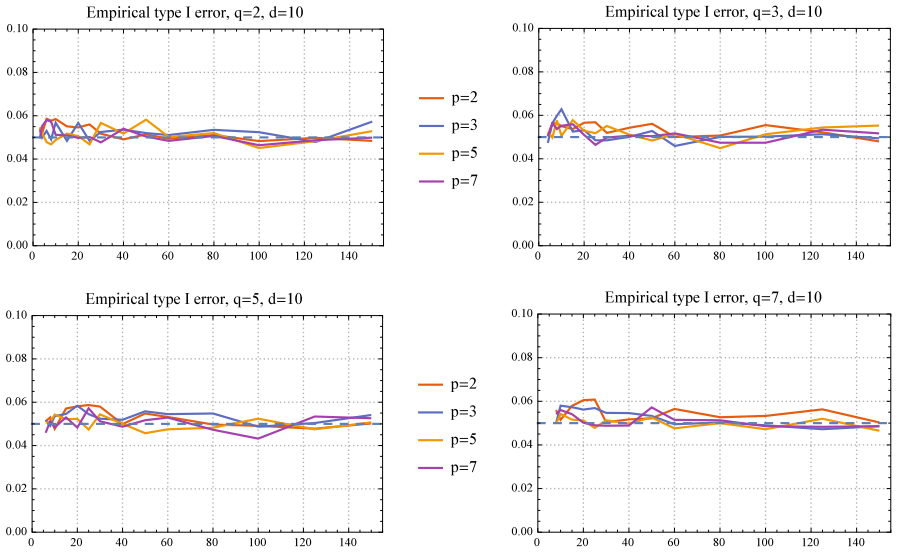

Fig. 2 Empirical Type I errors for different values of q and p at 5% level of significance α with skew parameter $d = 10$

1% significance levels. Figure 1 shows that the differences of the empirical power of BT^2 and D^2 are negligible in all investigated cases.

As suggested by a referee, we also study the performance (empirical Type I errors and powers) of the test statistic D^2 under the violation of the underlying normality assumption. Particularly, we generate the elements of the error term from the skew-normal distribution with the skew parameter $d = 0, 5, 10, 100$. The case $d = 0$ corresponds to the normality assumption, and it is included for the comparison purpose. The parameters are assumed as $q = 2, 3, 5, 7$ and $p = 2, 3, 5, 7$. Results of empirical Type I errors at 5% level of significance α for D^2 based on 10,000 samples of sizes $n = 3, 4, 6, 8, 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 125$ and 150 (the smallest n is always taken as $q + 1$) are presented in Fig. 2. Only the results for the skew parameter $d = 10$ are presented in this figure; the results for other skew parameters are similar. It is seen that the skew parameter does not have an influence on the coverage probability.

The power comparison of D^2 at 5% level of significance α for parameters $q = 2$ and $p = 2$ and different values of skew parameter d is presented in Fig. 3. The curves for different skew parameters are almost identical, and this is true also for all other combinations of parameters q and p. Therefore, in Fig. 4 we present the results only for the skew parameter $d = 10$; it presents the power comparison for different combinations of parameters q and p. As expected, the power decreases with the increase in either of these two parameters.

2.3 Two-sample test

Using the results of the previous sections, we now derive the two-sample test. Let us consider the case where pq dimensional random sample $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ are from

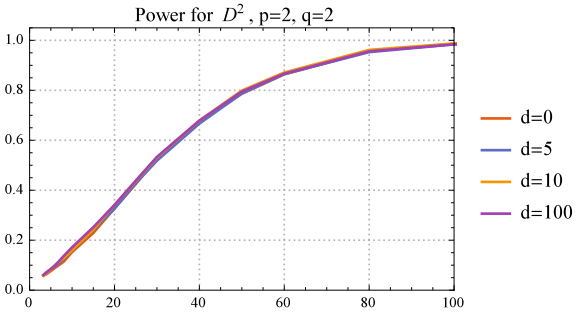

Fig. 3 Power comparison of D^2 for $q = 2$ and $p = 2$ at 5% level of significance α with different skew parameters d

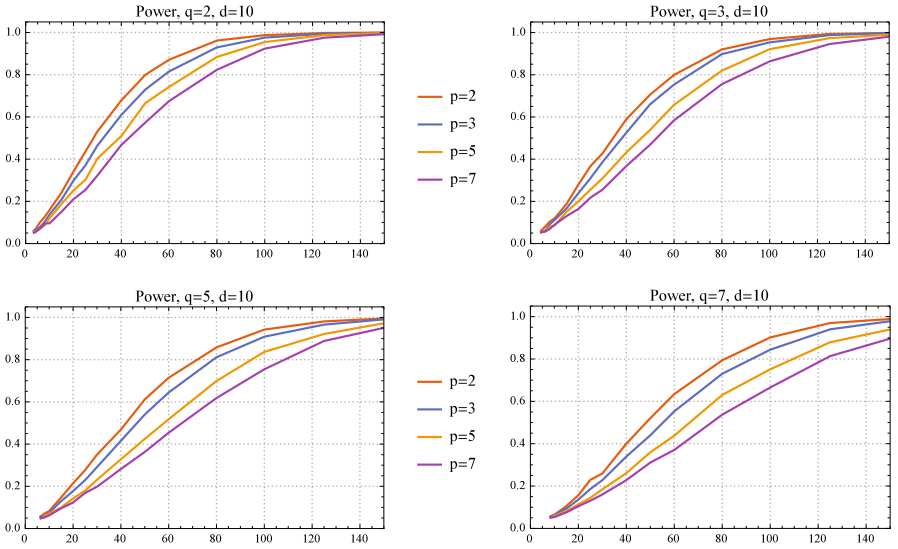

Fig. 4 Power comparison of D^2 for different values of q and p at 5% level of significance α with skew parameter $d = 10$

$N_{pq}\left(\boldsymbol{\mu}_U, \boldsymbol{\Sigma}^1\right)$ and a second random sample $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ are from $N_{pq}\left(\boldsymbol{\mu}_V, \boldsymbol{\Sigma}^2\right)$. We want to test

$$H_0 : \boldsymbol{\mu}_U = \boldsymbol{\mu}_V \quad \text{against} \quad H_1 : \boldsymbol{\mu}_U \neq \boldsymbol{\mu}_V. \tag{12}$$

We assume that the two samples are independent and that $\boldsymbol{\Sigma}^1 = \boldsymbol{\Sigma}^2 = \boldsymbol{\Sigma}$, say, with $\boldsymbol{\Sigma}$ unknown, but assuming to have BCS structure. Let $\underset{\sim}{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ and $\underset{\sim}{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m)$. Let $\overline{\boldsymbol{u}} = \frac{1}{n}\sum_{r=1}^{n} \boldsymbol{u}_r$ and $\overline{\boldsymbol{v}} = \frac{1}{m}\sum_{r=1}^{m} \boldsymbol{v}_r$ be the sample means of the two populations. We know that the sample means $\overline{\boldsymbol{u}}$ and $\overline{\boldsymbol{v}}$ are independent of the covariance matrix estimators $\boldsymbol{S}_1 = \frac{1}{n-1}\underset{\sim}{U}\boldsymbol{Q}_{\boldsymbol{J}_n}\underset{\sim}{U}'$ and $\boldsymbol{S}_2 = \frac{1}{m-1}\underset{\sim}{V}\boldsymbol{Q}_{\boldsymbol{J}_m}\underset{\sim}{V}'$, respectively (Mardia et al. 1979), and therefore they are also independent of the pooled estimator

$$\boldsymbol{S}^{\text{pl}} = \frac{1}{n+m-2}\Big((n-1)\boldsymbol{S}_1 + (m-1)\boldsymbol{S}_2\Big).$$

Thus, $\overline{\boldsymbol{u}} - \overline{\boldsymbol{v}} \sim N_{pq}\left(\boldsymbol{\mu}_U - \boldsymbol{\mu}_V, \frac{n+m}{nm}\boldsymbol{\Sigma}\right) \stackrel{H_0}{=} N_{pq}\left(\boldsymbol{0}, \frac{n+m}{nm}\boldsymbol{\Sigma}\right)$,

and $\boldsymbol{S}^{\text{pl}} \sim W_{pq}\left(n+m-2, \frac{1}{n+m-2}\boldsymbol{\Sigma}\right)$. We can use the following unbiased estimators of the variance and covariance matrices:

$$\widehat{\boldsymbol{\Sigma}}_0 = \frac{1}{p}\sum_{i=1}^{p} \boldsymbol{S}_{ii}^{\text{pl}}, \quad \widehat{\boldsymbol{\Sigma}}_1 = \frac{1}{p(p-1)}\sum_{\substack{i=1\\ i\neq j}}^{p}\sum_{j=1}^{p} \boldsymbol{S}_{ij}^{\text{pl}},$$

$$\widehat{\boldsymbol{\Delta}}_1^{\text{pl}} = \widehat{\boldsymbol{\Sigma}}_0 - \widehat{\boldsymbol{\Sigma}}_1 \quad \text{and} \quad \widehat{\boldsymbol{\Delta}}_2^{\text{pl}} = \widehat{\boldsymbol{\Sigma}}_0 + (p-1)\widehat{\boldsymbol{\Sigma}}_1.$$

Applying Theorem 1, we get two independent Wishart matrices

$$(n+m-2)(p-1)\widehat{\boldsymbol{\Delta}}_1^{\text{pl}} \sim W_q\left((n+m-2)(p-1), \boldsymbol{\Delta}_1\right),$$
$$(n+m-2)\widehat{\boldsymbol{\Delta}}_2^{\text{pl}} \sim W_q\left(n+m-2, \boldsymbol{\Delta}_2\right).$$

Since the estimators $\widehat{\boldsymbol{\Delta}}_1^{\text{pl}}$ and $\widehat{\boldsymbol{\Delta}}_2^{\text{pl}}$ are based on $\boldsymbol{S}^{\text{pl}}$, they are independent of $\overline{\boldsymbol{u}} - \overline{\boldsymbol{v}}$. Therefore, using analogous procedure as in the one-sample case, we arrive to Block T^2 statistic to test Hypothesis (12) in the form

$$\begin{aligned} BT^2 &= \frac{nm}{n+m}\left(\overline{\boldsymbol{u}} - \overline{\boldsymbol{v}}\right)' \boldsymbol{Z}' \begin{pmatrix} \left(\widehat{\boldsymbol{\Delta}}_2^{\text{pl}}\right)^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{p-1}\boldsymbol{J}_{p-1} \otimes \left(\widehat{\boldsymbol{\Delta}}_1^{\text{pl}}\right)^{-1} \end{pmatrix} \boldsymbol{Z}\left(\overline{\boldsymbol{u}} - \overline{\boldsymbol{v}}\right) \qquad (13) \\ &\sim T^2_{q,n+m-2} \oplus T^2_{q,(n+m-2)(p-1)}. \end{aligned}$$

Similarly, denoting $\boldsymbol{d} = \left(\boldsymbol{H}_p \otimes \boldsymbol{I}_q\right)\left(\overline{\boldsymbol{u}} - \overline{\boldsymbol{v}}\right) = (\boldsymbol{d}_1', \ldots, \boldsymbol{d}_p')'$, we get

$$\begin{aligned} D^2 &= \frac{nm}{n+m}\left(\overline{\boldsymbol{u}} - \overline{\boldsymbol{v}}\right)' \left[\boldsymbol{P}_p \otimes \left(\widehat{\boldsymbol{\Delta}}_2^{\text{pl}}\right)^{-1} + \boldsymbol{Q}_p \otimes \left(\widehat{\boldsymbol{\Delta}}_1^{\text{pl}}\right)^{-1}\right]\left(\overline{\boldsymbol{u}} - \overline{\boldsymbol{v}}\right) \\ &= \frac{nm}{n+m}\boldsymbol{d}'\left(\boldsymbol{H}_p \otimes \boldsymbol{I}_q\right)\left[\boldsymbol{P}_p \otimes \left(\widehat{\boldsymbol{\Delta}}_2^{\text{pl}}\right)^{-1} + \boldsymbol{Q}_p \otimes \left(\widehat{\boldsymbol{\Delta}}_1^{\text{pl}}\right)^{-1}\right]\left(\boldsymbol{H}_p \otimes \boldsymbol{I}_q\right)'\boldsymbol{d} \\ &= \operatorname{tr}\left[\frac{nm}{n+m}\boldsymbol{d}_1\boldsymbol{d}_1'\left(\widehat{\boldsymbol{\Delta}}_2^{\text{pl}}\right)^{-1}\right] + \operatorname{tr}\left[\sum_{i=2}^{p}\frac{nm}{n+m}\boldsymbol{d}_i\boldsymbol{d}_i'\left(\widehat{\boldsymbol{\Delta}}_1^{\text{pl}}\right)^{-1}\right] \stackrel{df}{=} T_{03}^2 + T_{04}^2, \end{aligned}$$

where $T_{03}^2 \sim T_0^2(q; 1, n+m-2) = T^2_{q,n+m-2}$ and $T_{04}^2 \sim T_0^2(q; p-1, (n+m-2)(p-1))$ are independent. Thus, in this case we also see that the distribution of D^2 is the convolution of two LH-trace distributions, i.e., $T_0^2(q; 1, n+m-2) \oplus T_0^2(q; p-1, (n+m-2)(p-1))$. As before, we can use the fact that $T^2_{q,n+m-2}$ is equivalent to $\frac{n+m-q-1}{q(n+m-2)}F(q, n+m-q-1)$, and $T_0^2(q; p-1, (n+m-2)(p-1))$ can be approximated by

$$\frac{(n+m-2)(p-1)^2 q}{(n+m-2)(p-1)-q-1} \cdot \frac{c-2}{c}F((p-1)q, c),$$

where

$$c = 4 + \frac{(pq - q + 2)\,[(n + m - 2)(p - 1) - q - 3]\,[(n + m - 2)(p - 1) - q]}{(np + mp - 2p - n - m + 1)\,(p + q) - (q - 1)(q + 2)}\,.$$

The two-sample Hypothesis (12) is tested using both BT^2 and D^2 in the PET data sets discussed in the Sect. 1.

Remark 2 The statistic T^2_{03} has LH-trace distribution $T^2_0(q; 1, n + m - 2)$ if $n + m - 2 \geq q$, and T^2_{04} has LH-trace distribution $T^2_0(q; p - 1, (n + m - 2)(p - 1))$ if $(n+m-2)(p-1) \geq q$. Since $p \geq 2$ for BCS structure, we have $(n+m-2)(p-1) \geq q$ when $n + m - 2 \geq q$. Therefore, the only condition needed on sample size in order to have T^2_{03} and T^2_{04} LH-trace distributions is $n + m - 2 \geq q$, i.e., $n + m \geq q + 2$. Again, the value of p adds no other requirement on the sample size.

If both samples have also exchangeable mean structure, i.e., $\boldsymbol{\mu}_U = \mathbf{1}_p \otimes \boldsymbol{\xi}_U$, $\boldsymbol{\mu}_V = \mathbf{1}_p \otimes \boldsymbol{\xi}_V$, we can test the hypothesis $H_0 : \boldsymbol{\xi}_U = \boldsymbol{\xi}_V$ using the same method as in Sect. 2.1.3. We obtain the test statistic

$$T^2 = \frac{nmp}{n + m} \left(\widehat{\boldsymbol{\xi}}_U - \widehat{\boldsymbol{\xi}}_V\right)' \left(\widehat{\boldsymbol{\Delta}}^{\text{pl}}_2\right)^{-1} \left(\widehat{\boldsymbol{\xi}}_U - \widehat{\boldsymbol{\xi}}_V\right) \sim T^2_{q,n+m-2}\,.$$

3 Real data examples

Our new method of testing the equality of mean vectors in BCS structure setting is applied to two multivariate repeated measures data sets (Sperling et al. 1990) with two populations on cerebral metabolism in epileptic patients (see the online supplementary material). The cerebral cortex has long been recognized as essential in development of seizures and epilepsy.

As described in Sect. 1, the metabolic rates of glucose are measured in 16 locations by positron emission tomography (PET) scans including 8 regions of interest: the first five ($q = 5$), frontal, sensorimotor, temporal, parietal and occipital regions, are known as cortical regions, and the last three ($q = 3$), caudate nucleus, lenticular nucleus and thalamus regions, are known as subcortical regions. Each region of interest includes right-sided (R) and left-sided (L) metabolic rates ($p = 2$). The Hypothesis (12) for equality of mean vectors of Normal control ($n = 18$) and the Right focus groups ($m = 8$) is tested separately for cortical as well as for subcortical regions. Normal control group in the both data sets has BCS covariance structure. Since in both data sets $p = 2$, by Lemma 2 we have $D^2 = \max BT^2$.

3.1 Cortical metabolic rate

We rearrange the variables in the data set by grouping together the right-sided metabolic rates ($i = 1$) at the frontal, sensorimotor, temporal, parietal and occipital regions and then left-sided metabolic rates ($i = 2$) at the same regions. We also compare our findings with the conventional Hotelling's T^2 statistic. The unbiased

estimates of the mean vectors $\boldsymbol{\mu}^N$ and $\boldsymbol{\mu}^R$ for the Normal control and the Right focus groups for the cortical regions are

$$\widehat{\boldsymbol{\mu}}^N = (4.110, 4.035, 3.791, 3.643, 3.864, 4.162, 4.092, 3.754, 3.623, 3.887) \text{ and}$$
$$\widehat{\boldsymbol{\mu}}^R = (3.988, 3.954, 3.736, 3.523, 4.165, 4.103, 4.143, 3.960, 3.806, 4.323),$$

respectively. For convenience, numerical results are displayed to three (rounded) decimal places. The mean vectors for the two groups appear to be slightly different. The estimates $\widehat{\boldsymbol{\Sigma}}_0^N$ and $\widehat{\boldsymbol{\Sigma}}_1^N$ for Normal control group are

$$\widehat{\boldsymbol{\Sigma}}_0^N = \begin{bmatrix} 0.435 & 0.443 & 0.335 & 0.422 & 0.391 \\ 0.443 & 0.495 & 0.354 & 0.471 & 0.428 \\ 0.335 & 0.354 & 0.359 & 0.362 & 0.385 \\ 0.422 & 0.471 & 0.362 & 0.506 & 0.460 \\ 0.391 & 0.428 & 0.385 & 0.460 & 0.507 \end{bmatrix} \text{ and } \widehat{\boldsymbol{\Sigma}}_1^N = \begin{bmatrix} 0.427 & 0.442 & 0.326 & 0.418 & 0.388 \\ 0.442 & 0.484 & 0.353 & 0.459 & 0.419 \\ 0.326 & 0.353 & 0.340 & 0.359 & 0.385 \\ 0.418 & 0.459 & 0.359 & 0.482 & 0.448 \\ 0.388 & 0.419 & 0.385 & 0.448 & 0.495 \end{bmatrix},$$

and the estimates $\widehat{\boldsymbol{\Sigma}}_0^R$ and $\widehat{\boldsymbol{\Sigma}}_1^R$ for the Right focus group are

$$\widehat{\boldsymbol{\Sigma}}_0^R = \begin{bmatrix} 0.523 & 0.415 & 0.273 & 0.346 & 0.258 \\ 0.415 & 0.403 & 0.246 & 0.323 & 0.244 \\ 0.273 & 0.246 & 0.482 & 0.264 & 0.346 \\ 0.346 & 0.323 & 0.264 & 0.304 & 0.267 \\ 0.258 & 0.244 & 0.346 & 0.267 & 0.325 \end{bmatrix} \text{ and } \widehat{\boldsymbol{\Sigma}}_1^R = \begin{bmatrix} 0.514 & 0.407 & 0.257 & 0.348 & 0.263 \\ 0.407 & 0.390 & 0.221 & 0.319 & 0.240 \\ 0.257 & 0.221 & 0.400 & 0.262 & 0.354 \\ 0.348 & 0.319 & 0.262 & 0.290 & 0.243 \\ 0.263 & 0.240 & 0.354 & 0.243 & 0.274 \end{bmatrix}.$$

The 5×5 matrices $\widehat{\boldsymbol{\Sigma}}_0^N$ represent the estimate of the variance–covariance matrix of the five response variables (frontal, sensorimotor, temporal, parietal and occipital) with any metabolic rate, whereas the 5×5 matrices $\widehat{\boldsymbol{\Sigma}}_1^N$ represent the covariance matrix of the five response variables between the two metabolic rates in the Normal control group. Similarly, $\widehat{\boldsymbol{\Sigma}}_0^R$ and $\widehat{\boldsymbol{\Sigma}}_1^R$ for the Right focus group.

The estimates of the variance–covariance matrices for the Normal control and Right focus groups seem to be similar (see Sect. 4). Now, to test the Hypothesis (12) we calculate pooled sample block exchangeable variance–covariance matrices from the above estimates, the components of which are

$$\widehat{\boldsymbol{\Sigma}}_0^{\text{pl}} = \begin{bmatrix} 0.461 & 0.435 & 0.317 & 0.400 & 0.352 \\ 0.435 & 0.468 & 0.322 & 0.428 & 0.374 \\ 0.317 & 0.322 & 0.395 & 0.333 & 0.374 \\ 0.400 & 0.428 & 0.333 & 0.447 & 0.404 \\ 0.352 & 0.374 & 0.374 & 0.404 & 0.454 \end{bmatrix} \text{ and } \widehat{\boldsymbol{\Sigma}}_1^{\text{pl}} = \begin{bmatrix} 0.453 & 0.432 & 0.306 & 0.398 & 0.351 \\ 0.432 & 0.457 & 0.315 & 0.419 & 0.367 \\ 0.306 & 0.315 & 0.357 & 0.331 & 0.376 \\ 0.398 & 0.419 & 0.331 & 0.426 & 0.388 \\ 0.351 & 0.367 & 0.376 & 0.388 & 0.430 \end{bmatrix}.$$

The calculated BT^2 statistic (13) (which is a convolution of two Hotelling's T^2 with degrees of freedoms ($q = 5$; $n + m - 2 = 24$) and ($q = 5$; $(n + m - 2)(p - 1) = 24$, respectively) is 32.49, and the corresponding p value is 0.0198. Therefore, we conclude that the Normal control and the Right focus groups are different. However, when

we use the classical Hotelling's T^2 test, we get the calculated T^2 statistic as 28.08 (with degrees of freedom ($pq = 10; n + m - 2 = 24$)) and the corresponding p value is 0.1574. Thus, Hotelling's T^2 test concludes that the two groups are not different, which contradicts to the conclusion of the BT^2 test. Therefore, we see that neglecting the correct variance–covariance structure of the data by the Hotelling's T^2 test masks the slight difference between the two groups and leads to incorrect conclusion.

3.2 Subcortical metabolic rate

As before, we rearrange the variables in the data set by grouping together the right-sided metabolic rates ($i = 1$) at the caudate nucleus, lenticular nucleus and thalamus regions and then left-sided metabolic rates ($i = 2$) at the caudate nucleus, lenticular nucleus and thalamus regions. Here we also compare our findings with the classical Hotelling's T^2 statistic. The unbiased estimates of the mean vectors $\boldsymbol{\mu}^N$ and $\boldsymbol{\mu}^R$ for the Normal control and the Right focus groups for the subcortical regions are

$$\widehat{\boldsymbol{\mu}}^N = (3.828, 4.294, 3.999, 3.959, 4.400, 3.984) \text{ and}$$
$$\widehat{\boldsymbol{\mu}}^R = (3.698, 4.343, 3.998, 4.076, 4.326, 4.041),$$

respectively. The difference of the two mean vectors is even smaller than in the previous example. The estimates $\widehat{\boldsymbol{\Sigma}}_0^N$ and $\widehat{\boldsymbol{\Sigma}}_1^N$ for Normal control group are

$$\widehat{\boldsymbol{\Sigma}}_0^N = \begin{bmatrix} 0.514 & 0.434 & 0.435 \\ 0.434 & 0.647 & 0.547 \\ 0.435 & 0.547 & 0.522 \end{bmatrix} \text{ and } \widehat{\boldsymbol{\Sigma}}_1^N = \begin{bmatrix} 0.482 & 0.434 & 0.440 \\ 0.434 & 0.631 & 0.543 \\ 0.440 & 0.543 & 0.510 \end{bmatrix}.$$

and the estimates $\widehat{\boldsymbol{\Sigma}}_0^R$ and $\widehat{\boldsymbol{\Sigma}}_1^R$ for the Right focus group are

$$\widehat{\boldsymbol{\Sigma}}_0^R = \begin{bmatrix} 0.569 & 0.543 & 0.317 \\ 0.543 & 0.757 & 0.493 \\ 0.317 & 0.493 & 0.377 \end{bmatrix} \text{ and } \widehat{\boldsymbol{\Sigma}}_1^R = \begin{bmatrix} 0.548 & 0.539 & 0.315 \\ 0.539 & 0.726 & 0.485 \\ 0.315 & 0.485 & 0.367 \end{bmatrix}.$$

The 3×3 matrices $\widehat{\boldsymbol{\Sigma}}_0^N$ represent the estimate of the variance–covariance matrix of the three response variables (caudate nucleus, lenticular nucleus and thalamus) with any metabolic rate, whereas the 3×3 matrices $\widehat{\boldsymbol{\Sigma}}_1^N$ represent the covariance matrix of the three response variables between the two metabolic rates in the Normal control group. Similarly, $\widehat{\boldsymbol{\Sigma}}_0^R$ and $\widehat{\boldsymbol{\Sigma}}_1^R$ for the Right focus group.

The estimates of the variance–covariance matrices for the Normal control and Right focus groups appear to be similar (see Sect. 4). Now, to test Hypothesis (12) we calculate pooled sample block exchangeable variance–covariance matrices from the above estimates, the components of which are

$$\widehat{\boldsymbol{\Sigma}}_0^{\mathrm{pl}} = \begin{bmatrix} 0.530 & 0.466 & 0.401 \\ 0.466 & 0.679 & 0.531 \\ 0.401 & 0.531 & 0.480 \end{bmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_1^{\mathrm{pl}} = \begin{bmatrix} 0.502 & 0.465 & 0.404 \\ 0.465 & 0.659 & 0.526 \\ 0.404 & 0.526 & 0.468 \end{bmatrix}.$$

The resulting BT^2 statistic (13) (which is a convolution of two Hotelling's T^2 with degrees of freedoms ($q = 3; n+m-2 = 14$) and ($q = 3; (n+m-2)(p-1) = 14$), respectively) equals to 12.02, and the corresponding p value is 0.1341. Thus, we conclude that the Normal control group and the Right focus group are not different, as we have suspected. Again, we can compare it with the classical Hotelling's T^2 test, getting Hotelling's T^2 statistic value equal to 18.07 (with degrees of freedom ($pq = 6; n+m-2 = 14$) and the corresponding p value is 0.0691. This is not significant at 5% level, so that in this case we arrive at the same decision.

However, we can see that it can easily happen the opposite to the previous example, i.e., that BT^2 would be insignificant and Hotelling's T^2 significant. Thus, not taking into account the inherent covariance structure that is present in the data set can be dangerous in both directions.

4 Concluding remarks

In this article, we study the hypothesis testing of equality of mean vectors from one population as well as from two independent populations for two-level multivariate data with block exchangeable covariance structure. Such a structure is a realistic assumption in many cases and substantially reduces the number of estimated parameters. In this paper, we assume the equality of the BCS structure in two populations; however, we are working on the problem of different BCS covariance structures in the two populations and publish it in a future correspondence. Also, in future we will investigate the test with product of beta random variables and report it in another correspondence. The proposed methodology can readily be generalized to more than two levels and will be reported in a follow-up paper. Another inspiration for future work is to develop test statistic under the scenarios of missing values (e.g., Lin et al. 2009; Roy 2006) or censoring observations, e.g., Wang et al (2015).

5 Supplementary materials

The 'TestMeanBCS.R' file contains the R code to compute the critical values of D^2, BT^2 and Hotelling's T^2 tests. Data sets used in the illustration of our proposed method in Sect. 3.

Acknowledgements The authors thank the associate editor and the two anonymous reviewers for their careful reading, constructive comments and suggestions giving us the opportunity to significantly improve our manuscript.

References

Arnold SF (1973) Application of the theory of products of problems to certain patterned covariance matrices. Ann Stat 1(4):682–699

Arnold SF (1976) Applications of products to the generalized compound symmetry problem. Ann Stat 3:227–233

Arnold SF (1979) Linear models with exchangeably distributed errors. J Am Stat Assoc 74:194–199

Beacon HJ, Thompson SG (1996) Multi-level models for repeated measurement data: application to quality of life data in clinical trials. Stat Med 15:2717–2732

Coelho CA, Roy A (2017) Testing the hypothesis of a block compound symmetric covariance matrix for elliptically contoured distributions. TEST 26:308–330

Dyer D (1982) The convolution of generalized F distributions. J Am Stat Assoc 77(377):184–189

Goldstein H (2011) Multilevel statistical models, 4th edn. Edward Arnold, New York

Lin TI, Wang WL (2013) Multivariate skew-normal linear mixed models for multi-outcome longitudinal data. Stat Model 13:199–221

Lin TI, Ho HJ, Chen CL (2009) Analysis of multivariate skew normal models with incomplete data. J Multivar Anal 100:2337–2351

Mardia KV, Kent JT, Bibby JM (1979) Multivariate analysis. Academic Press Inc, New York

McKeon JJ (1974) F approximations to the distribution of Hotelling's T_0^2. Biometrika 61(2):381–383

Roy A (2006) A new classification rule for incomplete doubly multivariate data using mixed effects model with performance comparisons on the imputed data. Stat Med 25(10):1715–1728

Roy A, Fonseca M (2012) Linear models with doubly exchangeable distributed errors. Commun Stat Theory Methods 41(13):2545–2569

Roy A, Leiva R (2011) Estimating and testing a structured covariance matrix for three-level multivariate data. Commun Stat Theory Methods 40(10):1945–1963

Roy A, Leiva R, Žežula I, Klein D (2015) Testing the equality of mean vectors for paired doubly multivariate observations in blocked compound symmetric covariance matrix setup. J Multivar Anal 137:50–60

Sperling MR, Gur RC, Alavi A, Gur RE, Resnick S, O'Connor MJ, Reivich M (1990) Subcortical metabolic alterations in partial epilepsy. Epilepsia 31(2):145–155

Szatrowski TH (1976) Estimation and testing for block compound symmetry and other patterned covariance matrices with linear and non-linear structure. Technical Report Number OLK NSF 107, Department of Statistics, Stanford University

Viroli C (2012) On matrix-variate regression analysis. J Multivar Anal 111:296–309

Wang WL, Lin TI, Lachos VH (2015) Extending multivariate-*t* linear mixed models for multiple longitudinal data with censored responses and heavy tails. Stat Methods Med Res. doi:10.1177/0962280215620229

Wilks SS (1946) Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. Ann Math Stat 17(3):257–281