



# Testing independence under a block compound symmetry covariance structure

Katarzyna Filipiak<sup>1</sup> · Mateusz John<sup>1</sup> · Daniel Klein<sup>2</sup>

Received: 27 September 2021 / Revised: 23 May 2022 / Accepted: 26 May 2022 /

Published online: 29 June 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

The goal of this article is to test the hypothesis related to the independence of features between any two repeated measures in a block compound symmetry structure under the doubly multivariate normal model. The Rao score and Wald test statistics are determined and the characteristic function of the likelihood ratio test statistic is presented. For all of these test statistics, the asymptotic distributional properties are compared using simulation studies, and the robustness of the empirical distributions is considered. Furthermore, for power analysis purpose, the Kullback-Leibler divergence is proposed to measure discrepancy between hypotheses and the power of each mentioned tests, as well as F-test and Roy's largest root test, is studied. Finally, all mentioned tests are applied to a real data example.

**Keywords** Doubly multivariate model · Block compound symmetry · Rao score test · Wald test · Likelihood ratio test · Roy's largest root test · Independence · Entropy loss function · Power

---

Mateusz John and Daniel Klein have equally contributed to this work.

---

✉ Katarzyna Filipiak  
katarzyna.filipiak@put.poznan.pl

Mateusz John  
mateusz.john@put.poznan.pl

Daniel Klein  
daniel.klein@upjs.sk

<sup>1</sup> Institute of Mathematics, Poznań University of Technology, Piotrowo 3A, Poznań 60-965, Poland

<sup>2</sup> Institute of Mathematics, P. J. Šafárik University, Jesenná 5, Košice 04001, Slovakia

## 1 Introduction

In modern experiments, huge datasets consisting of measurements of many features, observed repeatedly in time, at various locations and depths, taken from many individuals, are usually collected. This paper deals with doubly multivariate data that can be stored in a three-dimensional tensor of observations of size  $n \times m \times u$ , where  $n$  is the number of individuals (sample size) and  $m$  and  $u$  are the numbers of repeated measurements of, for example, different features and locations, respectively. In such experiments, especially in genetics or medicine, the sample size is often too small to estimate all the unknown parameters of the model. To avoid this problem, regularization of estimators is proposed as one of the solutions, while the second one is to consider a patterned covariance matrix.

One of the common patterns for a doubly multivariate model is a block compound symmetry (BCS) structure, which is a direct extension of compound symmetry, common for multivariate models. BCS has been introduced in Rao (1945, 1953), where the problem of discriminating genetically different groups is studied. Note that the BCS structure is common for experiments, where the covariance matrix does not change when vectors from different repeated measures are interchanged, and hence, in the literature, it is also called an exchangeable structure.

A general method to test means and covariance matrices related to models with the BCS covariance structure has been proposed in Arnold (1973, 1979) as a way of reducing the number of unknown parameters for estimation. A similar problem has been also studied in Szatrowski (1976, 1978, 1982). Furthermore, Perlman (1987) found that if the collected data has symmetries, it is possible to obtain a more accurate estimate of the covariance matrix. Recently, Leiva (2007) formulated the generalized Fisher linear discrimination method under the BCS covariance structure and derived the maximum likelihood estimators of BCS. The estimation of BCS, as well as circular Toeplitz structure, was considered in Liang et al. (2012), while the optimal estimators of the BCS structure was proposed in Roy et al. (2016); Koziół et al. (2018). The application of the BCS structure in multivariate interval data problems is shown in Hao et al. (2015). The tests for the mean structure under the model with the BCS covariance matrix were proposed in Zmysłony et al. (2018); Žežula et al. (2018), while the likelihood ratio test (LRT) and the Rao score test (RST) to test the BCS covariance structure were presented in Roy and Leiva (2011); Roy et al. (2018); Filipiak and Klein (2021). The asymptotic normal distribution of the LRT under the assumption of the size of each block and the sample size tending to infinity was presented in Sun and Xie (2020). Very recently, Liang et al. (2021) derived the LRT for testing simultaneously the mean and particular structures of blocks of the BCS covariance matrix.

The aim of this paper is to test independence between features measured repeatedly, e.g., over time or locations, under the normal model with the BCS covariance structure. The LRT for such a hypothesis was studied, for example, in Fonseca et al. (2018), where also a new test statistic being F-distributed, say FT statistic, is proposed. In the same year Tsukada (2018) compared the power of LRT, modified LRT (using Bartlett correction), Wald test (WT) and RST, but only for selected types of alternative hypothesis. It should be noted that the Wald test statistic is given without the proof in Tsukada (2018). In this paper we revise this test and show that the form presented

in Tsukada (2018) is not in line with the definition of Wald test given by Rao (2005). Very recently Koziol et al. (2021) gave a review on testing hypotheses under the BCS structures, and introduced also Roy's test statistics having the largest root distribution.

The main goals of this paper are to determine the RST and WT statistics and to derive the exact distribution of LRT. Moreover, using simulation studies we compare the asymptotical properties of mentioned likelihood ratio based tests and we verify their robustness for non-Gaussian data. Finally, we recall the FT and Roy's largest root tests and we study the power of each considered test. For this purpose, we introduce the entropy loss function (Kullback-Leibler divergence between two distributions) as a measure of discrepancy between the null and alternative hypotheses. We show that such an approach allows to compare the power of the test for various alternatives, in contrast to the approach usually considered in the literature, where particular structures of alternative hypotheses are studied; cf. Fonseca et al. (2018), Tsukada (2018). Note, finally, that the presented results can be applied in many areas of science, e.g. genetics, medicine, dietetics, agriculture, physics, image processing or engineering. In this paper we use horticultural real data example to compare all considered tests.

The paper is organized as follows. In Sect. 2 the model and hypotheses of interest, as well as the maximum likelihood estimators (MLEs) of unknown parameters, are presented. The RST, WT, LRT, FT and Roy's test statistics are formulated in Sect. 3, together with their properties, such as the convergence of the distribution of RST and WT statistics to the limiting chi-square distribution, the exact distribution of LRT statistic, as well as the independence of the distributions of all test statistics on the true values of unknown parameters. The powers of all tests are analyzed in Sect. 4. Finally, to illustrate presented methods, the independence of the petal lengths between any two flowers of Kalanchoe plants is tested in Sect. 5. The article is summarized in the Discussion section.

## 2 Model and hypothesis

We consider an experiment performed on  $n$  individuals in which  $m$  features are repeatedly measured  $u$  times, where these repeated measurements could be time points, locations, depths, etc. Let  $\mathbf{X}_i = (\mathbf{x}'_{i1}, \dots, \mathbf{x}'_{iu})'$ ,  $i = 1, \dots, n$ , be independent and identically distributed  $um$ -dimensional vectors of observations, where  $\mathbf{x}_{ij}$ ,  $j = 1, \dots, u$ , are  $m$ -dimensional vectors of measurements of the  $j$ th feature on the  $i$ th individual (at each of the  $u$  repetitions).

A normal matrix model is assumed here, in which the observation vectors for all individuals are placed in rows one below the other, that is,

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)' \sim N_{n,um}(\mathbf{1}_n \boldsymbol{\mu}', \mathbf{I}_n, \boldsymbol{\Omega}), \quad (1)$$

where  $\mathbf{1}_n$  is an  $n$ -dimensional vector of ones,  $\boldsymbol{\mu}$  is a  $um$ -dimensional general mean (the same for every individual),  $\mathbf{I}_n$  is an identity matrix of order  $n$ , and  $\boldsymbol{\Omega}$  is an unknown

symmetric positive definite covariance matrix of order  $um$ . Model (1) can also be presented in a vectorized form as

$$\text{vec } \mathbf{X} \sim N_{num}((\mathbf{I}_{um} \otimes \mathbf{1}_n)\boldsymbol{\mu}, \boldsymbol{\Omega} \otimes \mathbf{I}_n),$$

where  $\text{vec}$  is the operator stacking the columns of a matrix one below the other and  $\otimes$  is a Kronecker product.

It is known, that if  $\boldsymbol{\Omega}$  is unstructured, its MLE is of the form

$$\mathbf{S} = \frac{1}{n} \mathbf{X}' \mathbf{Q}_n \mathbf{X}, \tag{2}$$

where  $\mathbf{Q}_n = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n'$  is the orthogonal projector onto the orthocomplement of the column space of  $\mathbf{1}_n$ , while an unbiased estimator has the form

$$\mathbf{S}^* = \frac{1}{n-1} \mathbf{X}' \mathbf{Q}_n \mathbf{X}.$$

Note that, if  $um$  is close to the sample size, both estimators are ill-conditioned. Furthermore, if  $um > n$  the estimators are singular. To avoid these problems and to reduce the number of unknown parameters, one may impose the appropriate structure on the covariance matrix, which decreases the number of unknown parameters. In this paper we consider the BCS structure

$$\boldsymbol{\Omega}_{\text{BCS}} = \begin{pmatrix} \Gamma_0 & \Gamma_1 & \dots & \Gamma_1 \\ \Gamma_1 & \Gamma_0 & \dots & \Gamma_1 \\ \vdots & & \ddots & \vdots \\ \Gamma_1 & \Gamma_1 & \dots & \Gamma_0 \end{pmatrix} = \mathbf{I}_u \otimes \Gamma_0 + (\mathbf{J}_u - \mathbf{I}_u) \otimes \Gamma_1 := \boldsymbol{\Gamma} \tag{3}$$

with symmetric positive definite (p.d.) matrix  $\Gamma_0$  of order  $m$ , and with symmetric matrix  $\Gamma_1$  of order  $m$  such that  $\boldsymbol{\Gamma}$  is p.d. Matrix  $\Gamma_0$  is a variance-covariance matrix of  $m$  features at any given repeated measurement, while  $\Gamma_1$  is a covariance matrix of  $m$  features between any two repeated measurements.

After reparameterization one can get an equivalent form of BCS structure of the form

$$\boldsymbol{\Gamma} = \mathbf{Q}_u \otimes \boldsymbol{\Delta}_1 + \mathbf{P}_u \otimes \boldsymbol{\Delta}_2, \tag{4}$$

where  $\mathbf{P}_u = \frac{1}{u} \mathbf{1}_u \mathbf{1}_u'$  is the orthogonal projector onto the column space of  $\mathbf{1}_u$ . This form is more useful from a computational point of view. Since  $\mathbf{P}_u \mathbf{Q}_u = \mathbf{0}$ , to ensure the positive definiteness of  $\boldsymbol{\Gamma}$  it is enough to assume that  $\boldsymbol{\Delta}_i, i = 1, 2$ , are symmetric positive definite matrices. The relationship between (3) and (4) can be represented as

$$\begin{cases} \boldsymbol{\Delta}_1 = \Gamma_0 - \Gamma_1 \\ \boldsymbol{\Delta}_2 = \Gamma_0 + (u - 1)\Gamma_1. \end{cases}$$

Note that the BCS structure is also called exchangeable, since the vector  $\mathbf{x}_{ij}$  can be interchanged with  $\mathbf{x}_{ij'}, j, j' = 1, \dots, u$ , without changing the covariance matrix.

Since the space of BCS structures is a quadratic subspace, that is, the power of BCS also belongs to the same space, cf. Seely (1971), the MLE of  $\Gamma$  is a projection of  $\mathbf{S}$  given in (2) onto the space of BCS structures, that is,

$$\widehat{\Gamma} = \mathbf{Q}_u \otimes \widehat{\Delta}_1 + \mathbf{P}_u \otimes \widehat{\Delta}_2, \tag{5}$$

with

$$\widehat{\Delta}_1 = \frac{1}{u-1} \text{BTr}_m [(\mathbf{Q}_u \otimes \mathbf{I}_m)\mathbf{S}], \quad \widehat{\Delta}_2 = \text{BTr}_m [(\mathbf{P}_u \otimes \mathbf{I}_m)\mathbf{S}]; \tag{6}$$

cf. Filipiak et al. (2020). Alternatively, if  $\Gamma$  is expressed as (3),

$$\widehat{\Gamma}_0 = \frac{1}{u} \text{BTr}_m \mathbf{S}, \quad \widehat{\Gamma}_1 = \frac{1}{u(u-1)} (\text{BSum}_m \mathbf{S} - \text{BTr}_m \mathbf{S}), \tag{7}$$

where  $\text{BTr}_m(\mathbf{A}) = \sum_{i=1}^u \mathbf{A}_{ii}$  is a block trace operator defined on the partitioned matrix  $\mathbf{A} = (\mathbf{A}_{ij})$ ,  $i, j = 1, \dots, u$ , with blocks of order  $m$ ; cf. Filipiak et al. (2018), and  $\text{BSum}_m(\mathbf{A}) = \sum_{i=1}^u \sum_{j=1}^u \mathbf{A}_{ij}$ .

We are interested in testing the hypothesis related to the independence of features between two repeated measurements. This means that we are testing the block diagonality of the covariance matrix, which can be presented as

$$H_0 : \mathbf{\Omega} = \Gamma \text{ and } \Gamma_1 = \mathbf{0} \quad \text{vs} \quad H_1 : \mathbf{\Omega} = \Gamma \tag{8}$$

or, equivalently, using parameterization (4),

$$H_0 : \Delta_1 = \Delta_2 \quad \text{vs} \quad H_1 : \mathbf{\Omega} = \mathbf{Q}_u \otimes \Delta_1 + \mathbf{P}_u \otimes \Delta_2. \tag{9}$$

The spectral form of the BCS structure given in (9) provides simpler algebraic transformations than the previous one, and thus will usually be considered in the forthcoming sections.

Let us denote  $\Delta_1 = \Delta_2$  in (9) by  $\Delta$ . Then, the null hypothesis can be written as  $\mathbf{\Omega} = \mathbf{I}_u \otimes \Delta$ . Since the space of block diagonal matrices is a quadratic subspace, the MLE of  $\Delta$  is a projection of  $\mathbf{S}$  onto the space of block diagonal matrices, that is,

$$\widehat{\Delta} = \frac{1}{u} \text{BTr}_m \mathbf{S}. \tag{10}$$

The MLEs given in (6) and (10) will be used for determining the test statistics in the next section.

### 3 Test statistics

In this section we give an overview of the tests for the considered hypothesis, with determination of RST and WT statistics. Note that the form of RST statistic has been stated by Tsukada (2018), however, in this paper we present an alternative proof of its form. Moreover, in Tsukada (2018) the WT statistic has been given without any proof; therefore, in this paper we prove that the WT statistic has a more complex form. We

also verify the convergence of likelihood ratio based tests to the limiting chi-square distribution, using empirical distributions for RST and WT, and the exact distribution of LRT, formulated in Theorem 3 and proved in Appendix C.

Another test for (9), FT, having F-distribution with respective degrees of freedom, has been introduced by Fonseca et al. (2018). Its generalization is Roy’s test, having the largest root distribution; cf. Mardia et al. (1979). In this paper we recall their forms, and we verify the robustness of all tests for non-Gaussian data.

We start with determining the RST statistic. The proof of the following theorem can be found in Appendix A.

**Theorem 1** *Under hypothesis (9) the Rao score test statistic can be expressed as*

$$RS = \frac{n}{2} \operatorname{tr} \left\{ \left[ \mathbf{I}_{um} - \widehat{\Gamma}(\mathbf{I}_u \otimes \widehat{\Delta})^{-1} \right]^2 \right\},$$

where  $\widehat{\Gamma}$  and  $\widehat{\Delta}$  are given in, respectively, (5) and (10).

Denoting the MLEs of covariance matrix under alternative and null hypothesis by, respectively,  $\widehat{\Omega}_{H_1}$  and  $\widehat{\Omega}_{H_0}$ , we may represent the above RS test statistic as

$$RS = \frac{n}{2} \operatorname{tr} \left\{ \left[ \mathbf{I}_{um} - \widehat{\Omega}_{H_1} \widehat{\Omega}_{H_0}^{-1} \right]^2 \right\},$$

which is in line with the RS for testing various covariance structures in Filipiak and Klein (2021).

It is worth noting that under hypothesis (8), which is obviously equivalent to (9), we may formulate the following corollary, that can be proven directly from Theorem 1 by considering the parameterization of  $\widehat{\Gamma}$  and  $\widehat{\Delta}$  through  $\widehat{\Gamma}_0$  and  $\widehat{\Gamma}_1$ .

**Corollary 1** *Under hypothesis (8) the Rao score test statistic can be expressed as*

$$RS = \frac{nu(u-1)}{2} \operatorname{tr} \left[ (\widehat{\Gamma}_0^{-1} \widehat{\Gamma}_1)^2 \right],$$

where  $\widehat{\Gamma}_0$  and  $\widehat{\Gamma}_1$  are given in (7).

Note, that the RST statistic presented in Corollary 1 can be also expressed as formula (3.20) in Tsukada (2018).

Finally, recall that, due to Rao (2005), under considered null hypothesis and if the sample size  $n \rightarrow \infty$ , presented RST statistic is  $\chi^2$  distributed with  $m(m + 1)/2$  degrees of freedom. The same limiting distribution is related to the second well know test - Wald test, presented in the next theorem, with the proof in Appendix B.

**Theorem 2** *Under hypothesis (9) the Wald test statistic can be expressed as*

$$W = \frac{n}{2} \operatorname{vec}'(\widehat{\Delta}_1 - \widehat{\Delta}_2) \left[ \frac{1}{u-1} (\widehat{\Delta}_1 \otimes \widehat{\Delta}_1) + (\widehat{\Delta}_2 \otimes \widehat{\Delta}_2) \right]^{-1} \operatorname{vec}(\widehat{\Delta}_1 - \widehat{\Delta}_2)$$

with  $\widehat{\Delta}_1$  and  $\widehat{\Delta}_2$  given in (6).

We shall note, that to determine the Wald test statistic under hypothesis (8) it is enough to replace matrices  $\Delta_1, \Delta_2$  by respective transformations given in (2), however, the form of  $W$  will be much more complex. It is also possible to determine  $W$  going directly from null hypothesis (8), nevertheless, in such a case the Fisher information matrix given in Appendix A cannot be applied directly.

Finally observe, that plugging the MLE of covariance matrix under null hypothesis instead of alternative into the form of  $WT$ , respective test statistic presented by Tsukada (2018) will be obtained. Note however, that such approach is not in line with the definition of the Wald test given by Rao (2005).

The Rao score test is based on the MLE of vector of parameters under null hypothesis, while the Wald test used the MLE under alternative. The third test of Rao’s “holy trinity” is the likelihood ratio test, based on comparison of the MLEs under the null and alternative hypotheses. When testing (9), the likelihood ratio  $\Lambda$  has the form

$$\Lambda = \left( \frac{|\widehat{\Delta}_1|^{u-1} |\widehat{\Delta}_2|}{|\widehat{\Delta}|^u} \right)^{n/2}. \tag{11}$$

It is well known (Rao 2005) that under the null hypothesis,  $LR = -2 \ln \Lambda$  is approximately distributed as  $\chi^2$  with  $m(m + 1)/2$  degrees of freedom. It should be noted that if the covariance parameters fall on the boundary of their parameter space, then the asymptotic distribution of  $LR$  becomes a mixture of  $\chi^2$  distributions, as discussed in Self and Liang (1987). Instead of an approximate distribution, which works well only for relatively large sample sizes, one can use the exact distribution of the  $LR$  presented in the following theorem, with the proof given in Appendix C.

**Theorem 3** *The characteristic function of  $LR = -2 \ln \Lambda$ , with  $\Lambda$  being the likelihood ratio test statistic given in (11), is of the form*

$$\varphi(t) = \frac{u^{-numit}}{(u-1)^{-n(u-1)mit}} \cdot \prod_{j=1}^m \left[ \frac{\Gamma\left(\frac{(n-1)(u-1)+1-j}{2} - itn(u-1)\right)}{\Gamma\left(\frac{(n-1)(u-1)+1-j}{2}\right)} \cdot \frac{\Gamma\left(\frac{n-j}{2} - itn\right)}{\Gamma\left(\frac{n-j}{2}\right)} \cdot \frac{\Gamma\left(\frac{(n-1)u+1-j}{2}\right)}{\Gamma\left(\frac{(n-1)u+1-j}{2} - itnu\right)} \right].$$

The exact distribution of  $LRT$  statistic with the use of the above characteristic function can be computed with the use of R package `CharFunToolR` developed in Gajdoš (2018) on the basis of Matlab package `CharFunTool` provided in Witkovský (2018).

We shall also mention, that in the literature some modifications of  $LRT$  are studied. One of the example is multiplication of  $LR$  by a constant equal to  $1 - (u^2 - u + 1)(2m^2 + 3m - 1)/(6(n - 1)u(u - 1)(m + 1))$ ; cf. Tsukada (2018). Using the general theory of asymptotic expansions from Anderson (2003), such modified test statistic converges faster to respective  $\chi^2$  distribution. Nevertheless, since in this paper we give the exact distribution of  $LRT$ , we do not consider mentioned modification as a separate test.

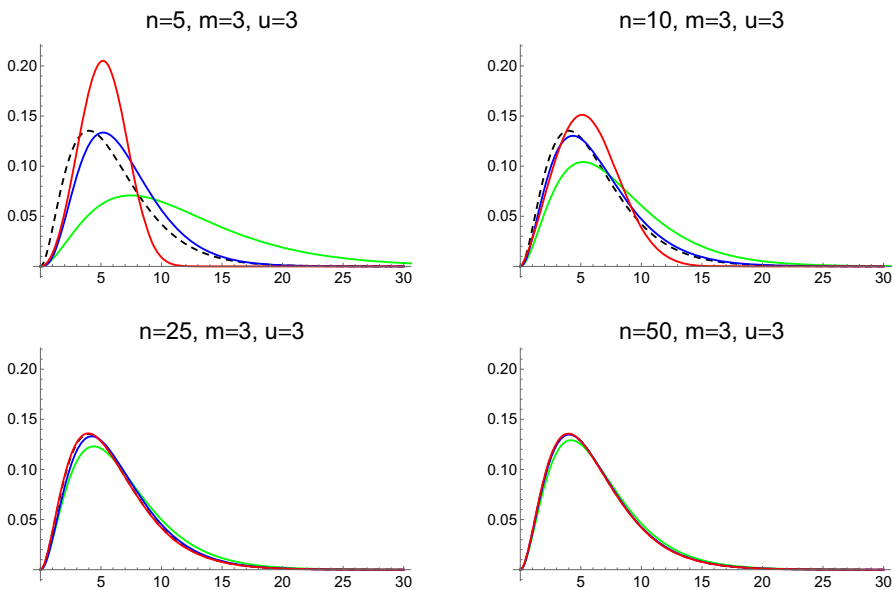
Finally note, that a big advantage of all presented test statistics is the following property, with the proof given in Appendix D.

**Proposition 1** *The distributions of RST, WT and LRT statistics under the null hypothesis in (9) do not depend on the true values of  $\mu$  and  $\Delta$ .*

Using simulations, we compare now the behavior of the empirical distributions of RST and WT statistics with respect to their convergence to the limiting distribution and we collate them with the exact distribution of LRT.

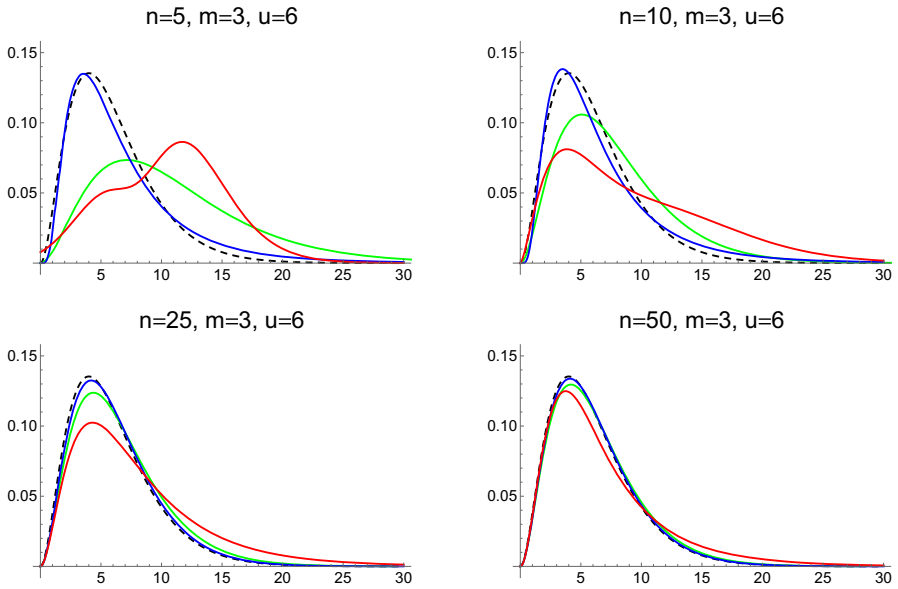
Recall, that all proposed tests can be performed if  $n > m$ . Moreover, for  $n \rightarrow \infty$ , the distribution of RS, W and LR test statistics tends to the  $\chi^2$  distribution with  $m(m+1)/2$  degrees of freedom. Thus, in Figs. 1, 2, 3, 4 and 5 we present the empirical null distributions of RST and WT, exact distribution of LRT and the limiting  $\chi^2$  distribution with respect to the sample size for  $u = 3$  and respectively  $m \in \{3, 6, 9\}$  and for  $m = 3$  and respectively  $u \in \{6, 9\}$ . It can be seen that distributions of all test statistics tend to the limiting distribution with the increase of  $n$ , however, the convergence of RST is the quickest and even for relatively small sample size does not differ significantly from the limiting one, which is not the case for WT nor LRT.

The next two presented tests, the FT and Roy's test, are based on unbiased estimators of unknown parameters, instead of MLEs. Roy et al. (2016) presented such unbiased estimators of  $\Delta_1$  and  $\Delta_2$  in terms of multiple sums of vector products. Recall that since the space of BCS structures is a quadratic subspace, these estimators can also be obtained by projection of sample covariance matrix  $\mathbf{S}^*$  onto the space of BCS matrices;

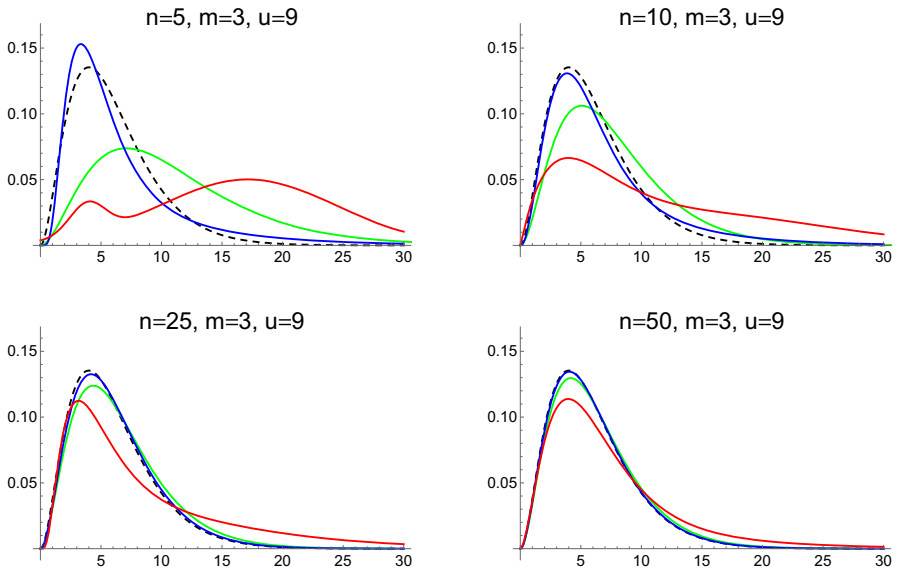


**Fig. 1** Empirical null distribution of RST (blue) and WT (red) and exact distribution of LRT (green) along with the  $\chi^2_6$  distribution (black dashed) for  $m = 3$ ,  $u = 3$ . (Colour figure online)

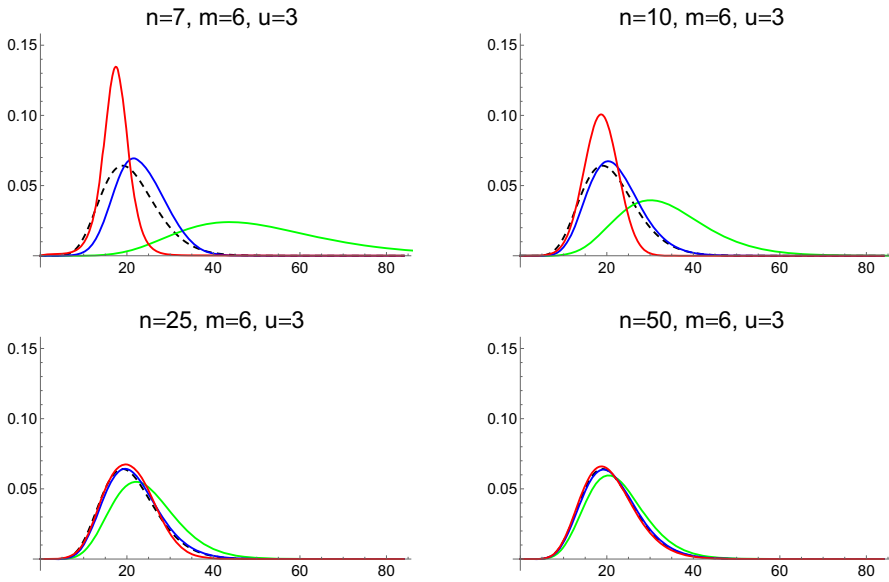




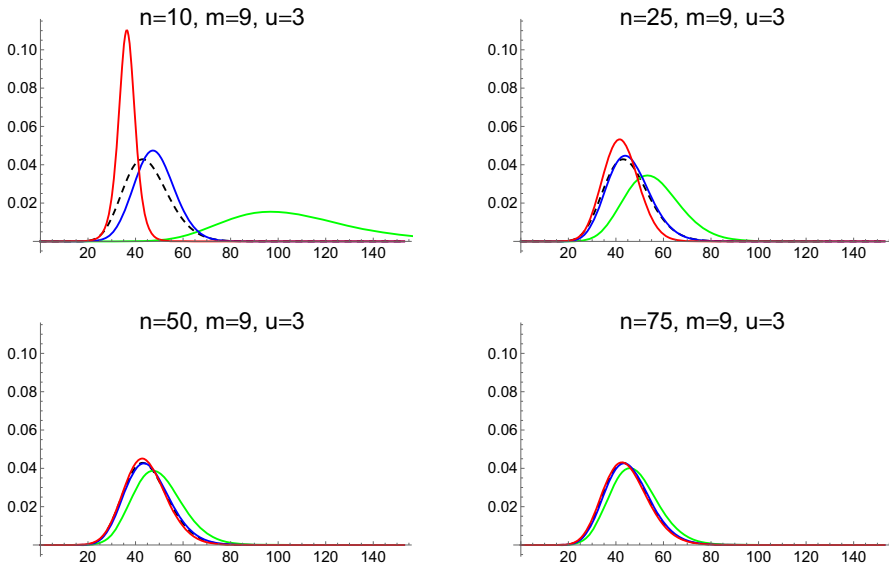
**Fig. 2** Empirical null distribution of RST (blue) and WT (red) and exact distribution of LRT (green) along with the  $\chi^2_6$  distribution (black dashed) for  $m = 3, u = 6$ . (Colour figure online)



**Fig. 3** Empirical null distribution of RST (blue) and WT (red) and exact distribution of LRT (green) along with the  $\chi^2_6$  distribution (black dashed) for  $m = 3, u = 9$ . (Colour figure online)



**Fig. 4** Empirical null distribution of RST (blue) and WT (red) and exact distribution of LRT (green) along with the  $\chi^2_{21}$  distribution (black dashed) for  $m = 6, u = 3$ . (Colour figure online)



**Fig. 5** Empirical null distribution of RST (blue) and WT (red) statistics and exact distribution of LRT (green) statistic along with the  $\chi^2_{45}$  distribution (black dashed) for  $m = 9, u = 3$ . (Colour figure online)

cf. Filipiak et al. (2020). Thus, the unbiased estimators given in Roy et al. (2016) can also be represented as

$$\tilde{\Delta}_1 = \frac{1}{u-1} \text{BTr}_m [(\mathbf{Q}_u \otimes \mathbf{I}_m)\mathbf{S}^*], \quad \tilde{\Delta}_2 = \text{BTr}_m [(\mathbf{P}_u \otimes \mathbf{I}_m)\mathbf{S}^*].$$

The FT statistic, introduced by Fonseca et al. (2018), has the form

$$F = \frac{\mathbf{v}'\tilde{\Delta}_2\mathbf{v}}{\mathbf{v}'\tilde{\Delta}_1\mathbf{v}},$$

and is F distributed with  $n - 1$  and  $(n - 1)(u - 1)$  degrees of freedom; cf. Fonseca et al. (2018)[Lemma 3.1]. Since the unbiased estimators of  $\Delta_1$  and  $\Delta_2$  differ from respective MLEs only by a constant, we can also represent FT statistic in terms of MLEs, that is

$$F = \frac{\mathbf{v}'\hat{\Delta}_2\mathbf{v}}{\mathbf{v}'\hat{\Delta}_1\mathbf{v}}.$$

Noting that

$$\hat{\Delta}_1 \sim W_m \left( \frac{1}{n(u-1)}\Delta_1, (n - 1)(u - 1) \right), \quad \hat{\Delta}_2 \sim W_m \left( \frac{1}{n}\Delta_2, n - 1 \right)$$

are independent; cf. Roy et al. (2015), the F distribution of the above FT statistic also follows.

Observe, that according to (D.1), the FT statistic can be expressed as

$$F = \frac{\mathbf{v}'\Delta^{1/2}\hat{\Upsilon}_2\Delta^{1/2}\mathbf{v}}{\mathbf{v}'\Delta^{1/2}\hat{\Upsilon}_1\Delta^{1/2}\mathbf{v}} = \frac{\mathbf{w}'\hat{\Upsilon}_2\mathbf{w}}{\mathbf{w}'\hat{\Upsilon}_1\mathbf{w}},$$

where  $\hat{\Upsilon}_1$  and  $\hat{\Upsilon}_2$  are given in (D.2). Thus, even if under null hypothesis the distribution of the above FT statistic does not depend on the true value of  $\Delta$ , the choice of vector  $\mathbf{v}$  should be appropriate. Note, that if  $\mathbf{v}$  is equal to the column of identity matrix, in fact the hypothesis about specific entry of covariance matrix being equal to zero is tested, which is the same as (9) only if  $m = 1$ . Furthermore, if  $\mathbf{v} = \mathbf{1}_m$  (as it was assumed in, e.g., Fonseca et al. (2018)) then the hypothesis about the sum of all elements of  $\Gamma_1$  being equal to zero is tested. Thus, the proposed test statistic is appropriate to test (9) if all the entries of  $\Gamma_1$  are of the same sign (or some of them, but not all, are zeros). In fact, the choice of vector  $\mathbf{v}$  as the vector of nonnegative (nonpositive) components corresponds to testing the value of weighted sum of the elements of  $\Gamma_1$ . Concluding, for testing (9), it would be natural not to fix a single  $\mathbf{v}$ , but to choose some optimal vector of quadratic forms in FT.

To achieve higher power of the test it is natural to choose a vector  $\mathbf{v}$  maximizing the value of the test statistic. Thus, maximizing  $F$  for all  $\mathbf{v} \in \mathbb{R}^m$  we obtain the test statistic  $F_m = \lambda_{\max}(\widehat{\Delta}_2 \widehat{\Delta}_1^{-1})$ ; cf. Koziol et al. (2021), where  $\lambda_{\max}(\cdot)$  is the largest root of the matrix in parenthesis. The inference can be made using Roy's test of the form

$$R = \frac{\frac{1}{u-1} F_m}{1 + \frac{1}{u-1} F_m},$$

which has the largest root distribution with parameters  $m$ ,  $(n-1)(u-1)$ , and  $n-1$ ; for more details see e.g. Mardia et al. (1979). For simplicity we will abbreviate this distribution as RLR (Roy's largest root).

Notice, that vector  $\mathbf{v}$  in  $F_m$  is the eigenvector corresponding to the maximal eigenvalue of  $\widehat{\Delta}_2 \widehat{\Delta}_1^{-1}$ , thus it is not fixed anymore, but depends on the data. As a result, as it is mentioned in Koziol et al. (2021), Roy's test does not necessarily have higher power than the  $F$ -test, it is advantageous only when the largest eigenvalue is substantially larger than the remaining ones.

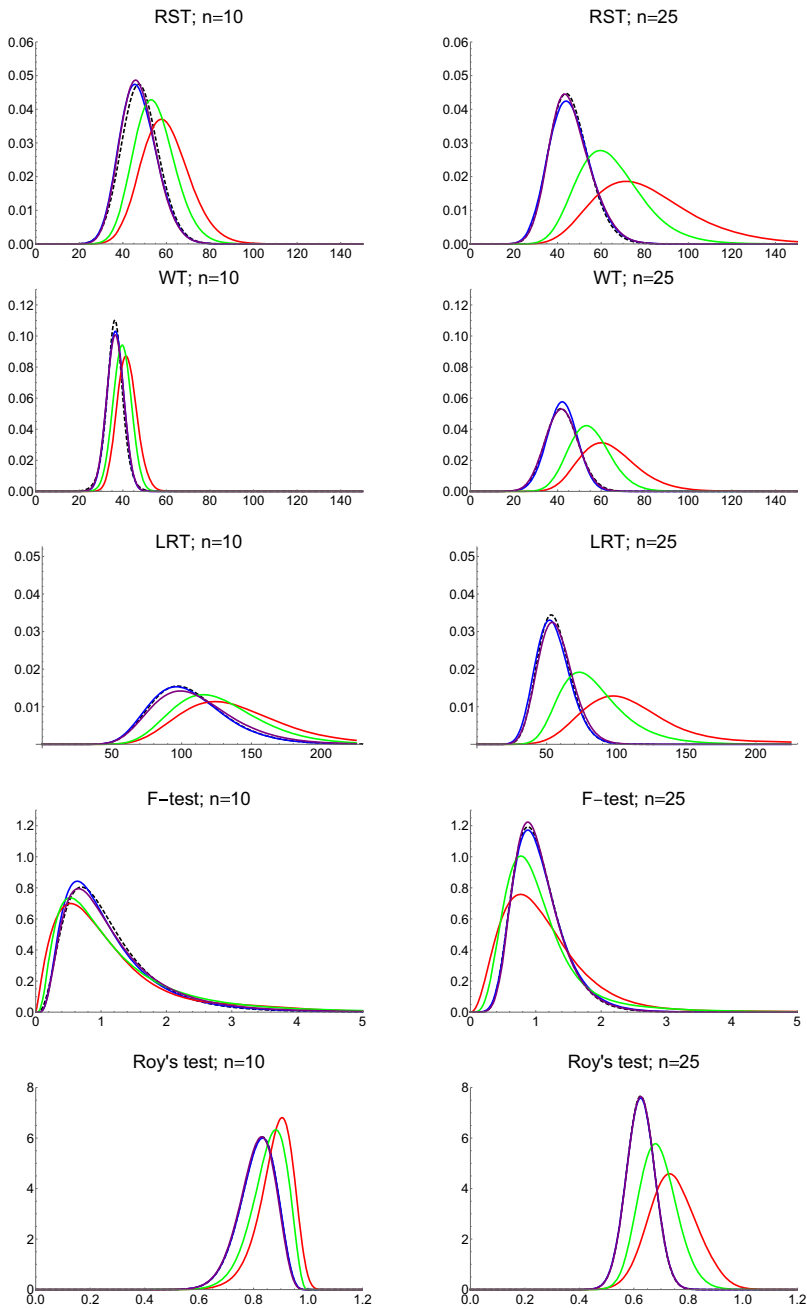
In order to check the robustness of considered tests with respect to some perturbations from normality, for various combinations of  $m$ ,  $u$  and  $n$ , we generated the data from the following non-Gaussian distributions: multivariate  $t_3$ ,  $t_5$ , gamma distribution with parameters (2, 1.5), and uniform on the interval (0, 1). The results for  $m = 9$ ,  $u = 3$  and  $n \in \{10, 25\}$  are given in Fig. 6. It can be seen that for small sample sizes, the distributions of all test statistics under non-normality are quite close to the relevant empirical null distributions, however, if the sample size increases, for considered  $t$  distributions, the test statistics are appearing to tend to some other distribution than chi-square. Similar observation was noticed for other sets of parameters (results not presented in this paper). Therefore, the robustness of test statistics, especially for multivariate  $t$  distributions, will be the topic of future research.

## 4 Power study of considered tests

For power comparison purpose Fonseca et al. (2018) considered the covariance structure under alternative constructed by choosing a block diagonal matrix with  $\Gamma_0$  on the diagonal and stating a scaled randomly generated matrix  $\Gamma_1$  as off-diagonal blocks, that is,

$$\Gamma = \mathbf{I}_u \otimes \Gamma_0 + (\mathbf{J}_u - \mathbf{I}_u) \otimes \lambda \Gamma_1,$$

where  $\lambda$  is a parameter ensuring positive definiteness of  $\Gamma$ . Note, that for various matrices  $\Gamma_1$ , parameter  $\lambda$  belongs to different domains (ensuring positive definiteness), and hence the discrepancies  $|\lambda|$  are not comparable. Moreover, such approach allows to consider only very specific types of alternatives and also null hypothesis, choosing the same  $\Gamma_0$  in both, null and alternative. Thus, in this paper, as a measure of discrepancy between given alternative,  $\Gamma$ , and a set of block-diagonal matrices  $\mathbf{I}_u \otimes \Delta$ , we mini-



**Fig. 6** Empirical null distributions of test statistics (presented in rows) under normality (black dashed), the multivariate  $t_3$  (red),  $t_5$  (green), gamma  $G(2, 1.5)$  (blue), and uniform  $U(0, 1)$  (purple) distributions, for  $m = 9, u = 3, n = 10$  (left panel),  $n = 25$  (right panel). (Colour figure online)

mize Kullback-Leibler divergence between two distributions that differ in covariance matrix, that is

$$\zeta = \min_{\Delta} \left\{ \text{tr} \left[ \Gamma^{-1} (\mathbf{I}_u \otimes \Delta) \right] - \ln \left| \Gamma^{-1} (\mathbf{I}_u \otimes \Delta) \right| - um \right\}, \tag{12}$$

where the symmetric p.d. matrix  $\Gamma$  has BCS structure with some given symmetric matrices  $\Gamma_0$  (p.d.) and  $\Gamma_1$ , while  $\Delta$  is a symmetric p.d. matrix for which the minimum is attained. Using the same differentiation rules as in Appendix A, it can be shown that the minimum (12) is obtained for

$$\Delta = \left( \frac{u-1}{u} \Delta_1^{-1} + \frac{1}{u} \Delta_2^{-1} \right)^{-1} \tag{13}$$

with  $\Delta_1$  and  $\Delta_2$  defined in (2). It should be noted that (13) determines the block diagonal structure which is the closest one in the sense of (12), and does not need to be the same as the diagonal blocks used in the alternative hypothesis. Observe moreover that since the value of  $\zeta$  is not upper bounded, we use the transformation  $\eta = 1 - \frac{1}{1+\zeta}$  that shrinks  $\zeta$  into the  $[0, 1)$  interval. Note that, in contrast to the method used by Fonseca et al. (2018), for arbitrary randomly generated  $\Gamma_0$  and  $\Gamma_1$ , the minimum  $\Delta$  and the discrepancy  $\eta$  can be determined and compared to each other.

Summing up, for various values of  $u$  and  $m$ , we first randomly generate matrices  $\Gamma_0$  and  $\Gamma_1$ , for which we determine the discrepancy  $\eta$ . For example, for  $u = m = 3$ , we choose the matrices

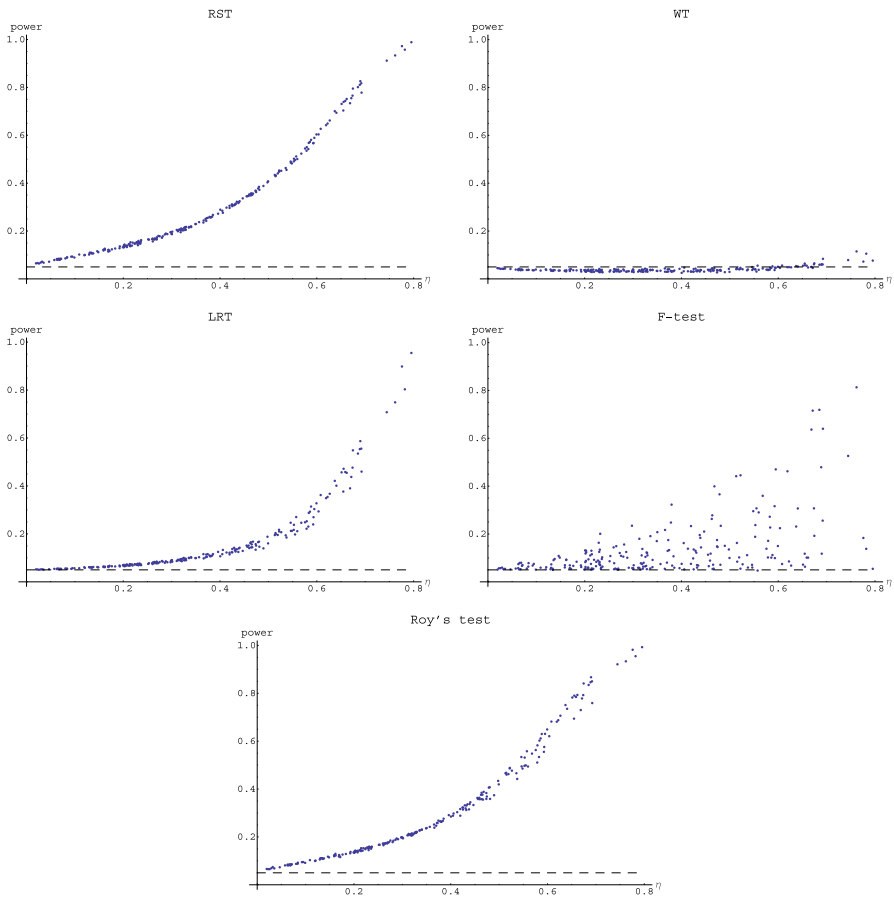
$$\Gamma_0 = \begin{pmatrix} 88.910 & -13.002 & 14.855 \\ -13.002 & 84.921 & 5.285 \\ 14.855 & 5.285 & 120.934 \end{pmatrix}, \quad \Gamma_1 = \begin{pmatrix} 26.195 & -0.231 & -4.579 \\ -0.231 & 2.357 & -1.647 \\ -4.579 & -1.647 & 3.495 \end{pmatrix}, \tag{14}$$

for which the minimum in (12) is attained at

$$\Delta = \begin{pmatrix} 75.995 & -13.2497 & 17.5422 \\ -13.2497 & 84.7433 & 5.49956 \\ 17.5422 & 5.49956 & 120.196 \end{pmatrix}$$

giving  $\eta = 0.2012$ .

In the study on power, we start by verifying the power of all the tests mentioned for  $m = u = 3$  and  $n = 5$ , such that  $\Gamma_0$  is given in (14), and  $\Gamma_1$  is randomly generated 300 times. For each case that gives a positive definite  $\Gamma$  (exactly for 237 cases), we generate the data matrix  $\mathbf{X} \sim N_{n,um}(\mathbf{0}, \mathbf{I}_n, \Gamma)$ , for which all tests are then performed. For rejection of null hypothesis we use respectively the empirical null distribution of RST and WT, the exact distribution of LRT (presented in Theorem 3 and computed using CharFunToolR package), F distribution with  $n - 1$  and  $(n - 1)(u - 1)$  degrees of freedom for FT, and RLR distribution with  $m$ ,  $(n - 1)(u - 1)$ , and  $n - 1$  degrees of freedom (computed with the use of the algorithm of Chiani (2016)) for Roy’s test. Similarly to Fonseca et al. (2018) and Koziol et al. (2021), to perform FT we choose  $\mathbf{v} = \mathbf{1}_m$ . In all comparisons, the significance level 0.05 is used. The empirical power

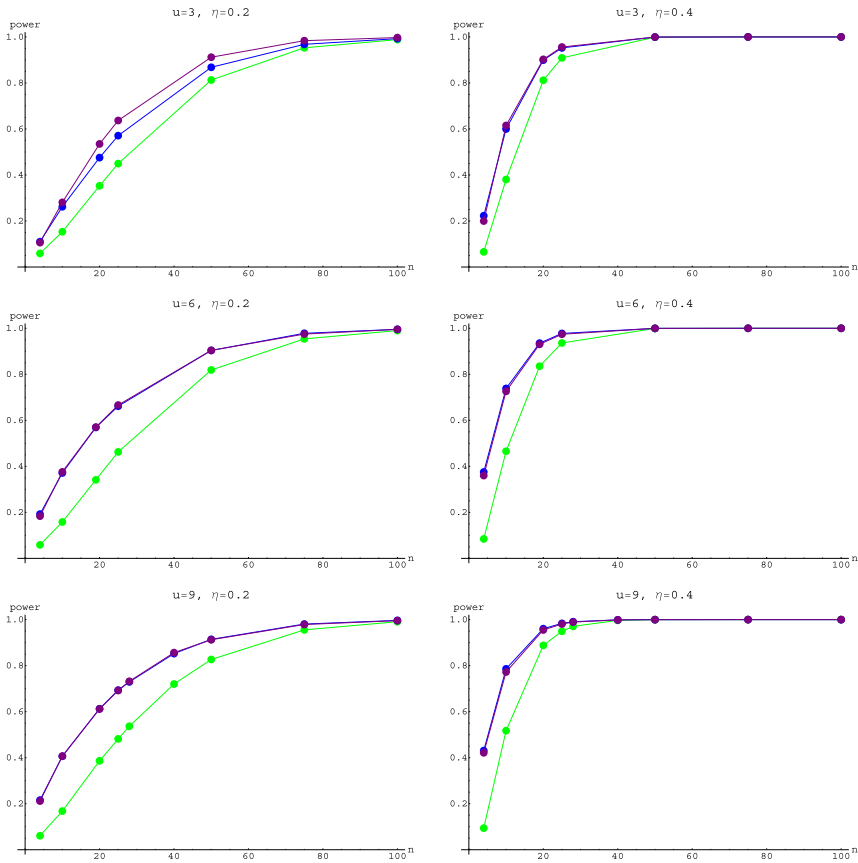


**Fig. 7** Empirical powers of RST, WT, LRT, FT, and Roy's test statistics for  $m = u = 3, n = 5$ , with respect to discrepancy  $\eta$ . (Colour figure online)

is calculated as the ratio between the number of rejections and the number of all trials performed. The results of the simulations are presented in Fig. 7.

It can be seen that RST, LRT, and Roy's test statistics increase with the discrepancy, in contrast to FT, for which for two equally distant BCS structures the power differs significantly, and to WT, for which the power is often below nominal significance level (Wald test is biased). The different behavior of these latter tests may also be caused by inappropriate choice of discrepancy; however, this topic will be studied in future research. Concluding, in the following part of this chapter, we compare power of RST, LRT and Roy's test only.

For power comparison, we consider  $m \in \{3, 6\}$ ,  $u \in \{3, 6, 9\}$ , and  $m = 9, u \in \{3, 6\}$ . For each pair  $(u, m)$  we choose  $\mathbf{\Gamma}_0$  and  $\mathbf{\Gamma}_1$  in such a way that the discrepancy  $\eta$  equals 0.2 and 0.4. The forms of all chosen matrices, except those given in (14), are available from the authors on request. Then, similarly as in the previous case, for



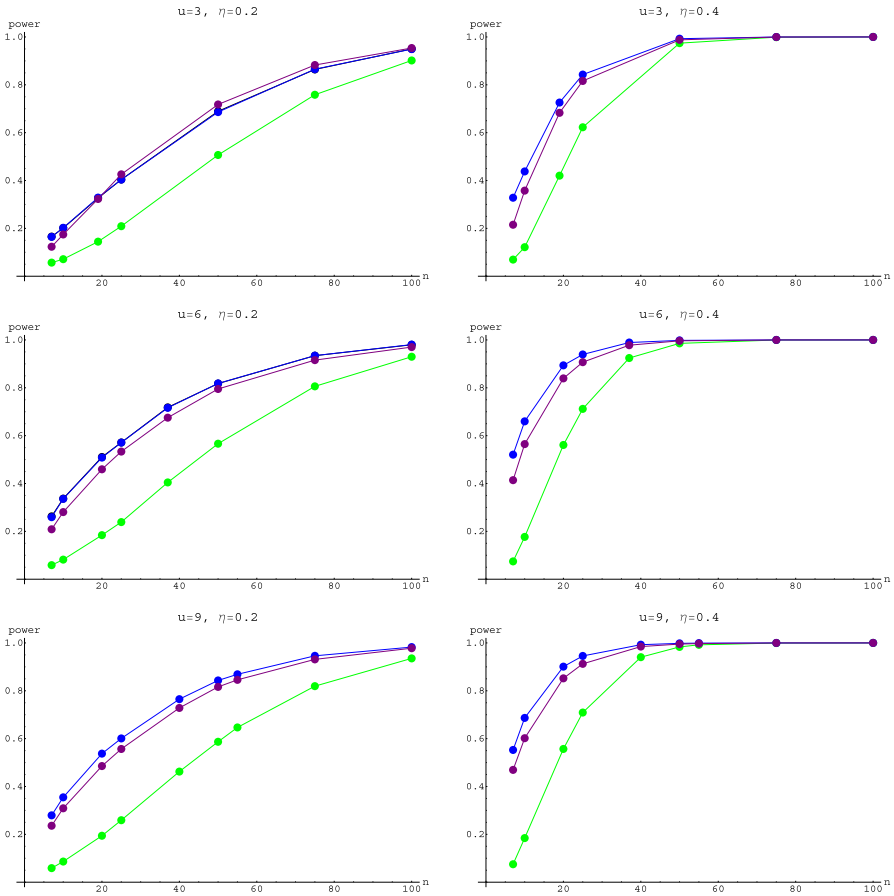
**Fig. 8** Empirical powers of RST (blue), LRT (green) and Roy's test (purple) statistics depending on  $n$  for  $m = 3$ ,  $u \in \{3, 6, 9\}$  and for  $\eta = 0.2$  (left panel) and  $\eta = 0.4$  (right panel). (Colour figure online)

selected sample sizes that ensure the existence of all three tests ( $n > m$ ), we generate 50,000 data matrices, for which the empirical powers are computed. The simulation results are presented in Figs. 8, 9 and 10. It can be seen that for both  $\eta$  the power of the RST (blue line) and Roy's test (purple line) exceed the power of the LRT (green line) for each considered sample size. Moreover, for  $\eta = 0.4$ , the power of RST and Roy's test is indistinguishable.

## 5 Real data example

In this section we consider an example originally presented in Liang et al. (2015)[Table 1], where the hierarchical model with block circular structure (in particular BCS structure) has been studied.

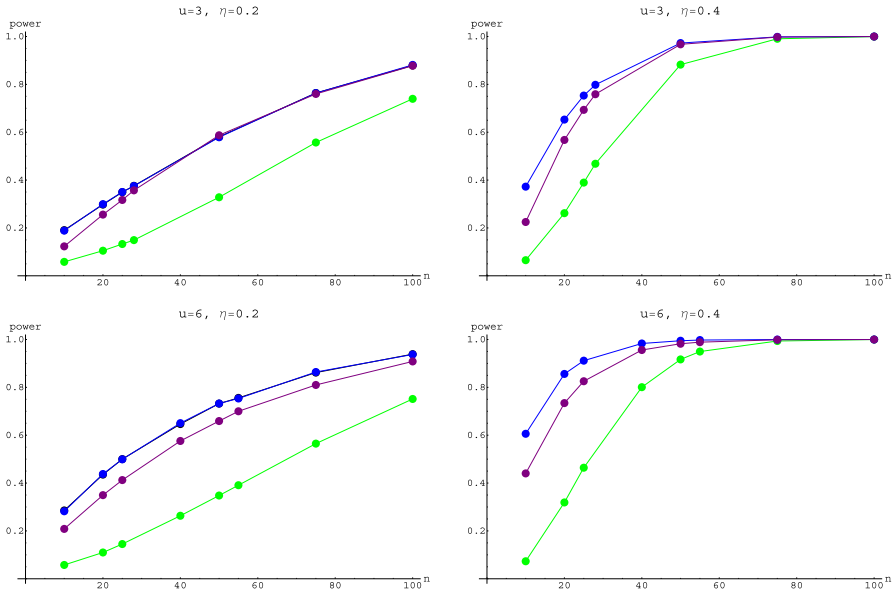




**Fig. 9** Empirical powers of RST (blue), LRT (green) and Roy’s test (purple) statistics depending on  $n$  for  $m = 6$ ,  $u \in \{3, 6, 9\}$  and for  $\eta = 0.2$  (left panel) and  $\eta = 0.4$  (right panel). (Colour figure online)

Data consisting of measurements of petal length made in 11 specific *Kalanchoe* plants from the same greenhouse are analyzed. From each plant 3 flowers have been randomly chosen. Note that each flower has 4 petals. We assume that the covariance between every two flowers is the same, which follows the BCS structure of the dispersion matrix. It is worth noting that since in each flower the arrangement of the petals is circular, Liang et al. (2015) additionally assumed a circular structure of covariance between the petals in each flower. This assumption is not required in this paper. For clarity, in this experiment we have  $n = 11$  individuals,  $m = 4$  petals on each of  $u = 3$  flowers.

Our aim is to verify the hypothesis related to the independence of petal lengths between any two flowers; hence, the hypothesis (9) is suitable here. We use RST, WT, and LRT statistics and exact and approximate quantiles of their distributions to make the decision. Note that in the case of RST and WT the empirical null



**Fig. 10** Empirical powers of RST (blue), LRT (green) and Roy’s test (purple) statistics depending on  $n$  for  $m = 9, u \in \{3, 6\}$  and for  $\eta = 0.2$  (left panel) and  $\eta = 0.4$  (right panel). (Colour figure online)

distributions are used as the exact distributions, while the quantiles (and thus also the  $p$ -value) of the exact LRT distribution are computed using the R package `CharFunToolR`. In all three cases, the  $\chi^2_{10}$  distribution is used as the limiting one. We also compute F test statistics for three different choices of  $\mathbf{v}$ :  $\mathbf{1}_m, \mathbf{v}_{\max}$  being the eigenvector related to maximal eigenvalue of  $\widehat{\Delta}_2 \widehat{\Delta}_1^{-1}$ , and some randomly generated  $\mathbf{v}_g = (0.859853, 0.175291, 0.011513, 0.405039)'$ , as well as Roy’s test statistic and we determine the  $p$ -values based on, respectively, F and RLR distributions with respective degrees of freedom. To calculate the  $p$ -value of Roy’s test, we use the algorithm presented in Chiani (2016). The values of the test statistics together with the respective  $p$ -values are given in Table 1.

The  $p$ -values computed from the exact distributions of all the tests suggest the same decision: at the significance level 0.05, the hypothesis of independence is not rejected. Furthermore, the decision made on the basis of limiting distribution of RST, WT and

**Table 1** Values of RST, LRT, WT,  $F(\mathbf{1}_m)$ ,  $F(\mathbf{v}_{\max})$ ,  $F(\mathbf{v}_g)$ , and Roy’s test statistics together with respective  $p$ -values for real data example

	RST	LRT	WT	$F(\mathbf{1}_m)$	$F(\mathbf{v}_{\max})$	$F(\mathbf{v}_g)$	R
test statistics	5.974	11.532	8.600	0.700	0.978	0.144	0.601
exact $p$ -value	0.879	0.565	0.569	0.714	0.491	0.998	0.402
$\chi^2_{10}$ $p$ -value	0.817	0.318	0.570	–	–	–	–

LRT remains the same. However, in the case of RST and WT, the exact and chi-square  $p$ -values are close to each other, which is not the case for LRT. This observation is very general, as can be seen from the significant discrepancies between the exact and limiting distributions of LRT, especially for small  $n$ , given in Figs. 1, 2, 3, 4 and 5.

Note that for different choice of  $\mathbf{v}$  in the F test we obtain various values of test statistic and, obviously, different  $p$ -values. In the example considered, the decision remains the same; however, for other datasets, different decisions can be made. Thus, from a practical point of view, the Roy's test should be preferable to the F-test, as it does not depend on the choice of  $\mathbf{v}$ . However, since the determination of the RLR distribution, as well as the exact distribution of LRT, involves complex computations with the use of special packages, while the (empirical) exact distributions of RST and WT do not differ significantly from limiting  $\chi^2$  distribution, even if the normality assumption is not fulfilled, the RST or WT procedure with  $p$ -value taken from the limiting distribution seems to be the most useful for practitioners. Finally, because of biasedness of WT shown in the previous section, the type II error can be much higher than one can accept. Therefore, the RST is suggested to be used by practitioners.

## 6 Discussion

In this article we determined the RST and WT statistics as well as we showed the characteristic function of LRT statistic for testing the independence of features between repeated measurements in the BCS covariance structure. For all of these test statistics we proved that their null distributions do not depend on the true parameters. For FT and Roy's test this conclusion is obvious. The robustness analysis performed for selected distributions showed, that all mentioned tests are relatively consistent, however, some future research must be done under the non-normality assumption, especially for multivariate  $t$  distribution of the data. Nevertheless, since WT is biased, and since the values of F test statistic strongly depend on the choice of vector  $\mathbf{v}$ , it is difficult to verify the power of these test, and thus they are not taken into consideration in power comparison. In the power analysis we showed that the powers of RST and Roy's tests do not differ significantly and usually exceed the power of LRT.

Summing up, the F test would be good for testing the hypothesis about the values of specific elements of covariance matrix, in which case vector  $\mathbf{v}$  should be chosen according to the tested hypothesis. Because of biasedness of WT and since the determination of the exact distribution of LRT and RLR distribution is relatively complex, the RST with its limiting  $\chi^2$  distribution can be recommended for the use by researchers.

**Acknowledgements** The authors thank the Mathematical Research and Conference Center of the Polish Academy of Sciences, Będlewo, Poland, for providing the opportunity and support for this paper. This research is partially supported by Statutory Activities No. 0213/SBAD/0110 (Mateusz John) and by the Slovak Research and Development Agency under the contract no. APVV-17-0568 (Daniel Klein).

### Appendix A: Proof of Theorem 1

Following Rao (2005) the Rao score test statistics is a function of score vector  $\mathbf{s}$  and Fisher information matrix  $\mathbf{F}$ , that is

$$RS = \mathbf{s}'(\hat{\boldsymbol{\theta}})\mathbf{F}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{s}(\hat{\boldsymbol{\theta}}).$$

Note, that the score vector  $\mathbf{s}(\boldsymbol{\theta})$  is a vector of first derivatives of log-likelihood function with respect to the vector of parameters under alternative, that is  $\boldsymbol{\theta} = (\boldsymbol{\mu}', \text{vech}' \boldsymbol{\Delta}_1, \text{vech}' \boldsymbol{\Delta}_2)'$ , where  $\text{vech } \mathbf{A}$  is a vector obtained from  $\text{vec } \mathbf{A}$  by eliminating all elements arranged above the main diagonal of  $\mathbf{A}$ ; cf. Magnus and Neudecker (1986). Observe, that it is easy to recover all the entries of  $\text{vec } \mathbf{A}$  from  $\text{vech } \mathbf{A}$  by duplication of respective elements, that is  $\text{vec } \mathbf{A} = \mathbf{D}_m \text{vech } \mathbf{A}$ , where  $\mathbf{D}_m$  is an  $m^2 \times \frac{1}{2}m(m + 1)$  duplication matrix; cf. Magnus and Neudecker (1986).

Due to Magnus and Neudecker (1986), by the derivative of an arbitrary function  $\mathbf{G}(\mathbf{A})$  with respect to  $\mathbf{A}$  we mean  $\frac{d \text{vec } \mathbf{G}(\mathbf{A})}{d \text{vec}' \mathbf{A}}$ . If  $\mathbf{A}$  is a symmetric matrix, the derivative is computed with respect to  $\text{vech}' \mathbf{A}$ . It is easy to see that  $\frac{d \text{vec } \mathbf{A}}{d \text{vech}' \mathbf{A}} = \mathbf{D}_m$ , and thus, according to the chain rule,  $\frac{d \text{vec } \mathbf{G}(\mathbf{A})}{d \text{vech}' \mathbf{A}} = \frac{d \text{vec } \mathbf{G}(\mathbf{A})}{d \text{vec}' \mathbf{A}} \cdot \mathbf{D}_m$ . Since the log-likelihood function is a scalar function, in considered case resulting score vector is of dimension  $um + m(m + 1)$ .

Noting, that  $\hat{\boldsymbol{\theta}}$  is the MLE of  $\boldsymbol{\theta}$  under the null hypothesis, for (9) we obtain  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\mu}}', \text{vech}' \hat{\boldsymbol{\Delta}}, \text{vech}' \hat{\boldsymbol{\Delta}})'$  with  $\hat{\boldsymbol{\mu}} = \frac{1}{n}\mathbf{X}'\mathbf{1}_n$  and  $\hat{\boldsymbol{\Delta}}$  presented in (10). Observe, however, that since the considered hypothesis does not contain any restrictions on  $\boldsymbol{\mu}$ , the first entry of the score vector (first derivative with respect to  $\boldsymbol{\mu}$ ) will reduce to  $\mathbf{0}$  when  $\boldsymbol{\mu}$  is replaced by its MLE. Thus, without loss of generality, we consider the score vector of the dimension  $m(m + 1)$ . Similarly, the Fisher information matrix is of order  $m(m + 1)$ .

Considering  $\boldsymbol{\Gamma}$  given by (4) and since  $|\boldsymbol{\Gamma}| = |\boldsymbol{\Delta}_1|^{u-1} |\boldsymbol{\Delta}_2|$ , the log-likelihood function under  $H_1$  in (9) can be presented as

$$\ln L = -\frac{num}{2} \ln(2\pi) - \frac{n(u - 1)}{2} \ln |\boldsymbol{\Delta}_1| - \frac{n}{2} \ln |\boldsymbol{\Delta}_2| - \frac{1}{2} \text{tr}(\mathbf{Y}\boldsymbol{\Gamma}^{-1}\mathbf{Y}')$$

with  $\mathbf{Y} := \mathbf{Y}(\boldsymbol{\mu}) = \mathbf{X} - \mathbf{1}_n\boldsymbol{\mu}'$ . In order to obtain the score vector, we differentiate the above log-likelihood function with respect to  $\boldsymbol{\theta}$ . Using the chain rule as described in Magnus and Neudecker (1986), the differentiation formulas given in Fackler (2005), and Corollary 2.10 of Filipiak et al. (2018), we obtain

$$\begin{aligned} \frac{\partial \ln L}{\partial \boldsymbol{\Delta}_1} &= -\frac{n}{2} \left[ (u - 1) \text{vec}' \boldsymbol{\Delta}_1^{-1} \right. \\ &\quad \left. - \text{vec}' \left( \frac{1}{n}\mathbf{Y}'\mathbf{Y} \right) (\mathbf{I}_u \otimes \mathbf{K}_{m,u} \otimes \mathbf{I}_m) (\text{vec } \mathbf{Q}_u \otimes \mathbf{I}_{m^2}) (\boldsymbol{\Delta}_1^{-1} \otimes \boldsymbol{\Delta}_1^{-1}) \right] \mathbf{D}_m \\ &= -\frac{n}{2} \left[ (u - 1) \text{vec}' \boldsymbol{\Delta}_1^{-1} \right. \\ &\quad \left. - \text{vec}' \{ \text{BTr}_m [ (\mathbf{Q}_u \otimes \mathbf{I}_m) \frac{1}{n}\mathbf{Y}'\mathbf{Y} ] \} (\boldsymbol{\Delta}_1^{-1} \otimes \boldsymbol{\Delta}_1^{-1}) \right] \mathbf{D}_m, \end{aligned}$$

where  $\mathbf{K}_{m,u}$  is an  $um \times um$  commutation matrix that transforms an  $m \times u$  matrix  $\mathbf{A}$  as  $\mathbf{K}_{m,u} \text{vec } \mathbf{A} = \text{vec } \mathbf{A}'$ . Similar result can be obtained for  $\frac{\partial \ln L}{\partial \Delta_2}$  with  $\Delta_1$  replaced by  $\Delta_2$ , with the projection matrix  $\mathbf{Q}_u$  replaced by  $\mathbf{P}_u$ , and with  $\text{tr } \mathbf{Q}_u = u - 1$  replaced by  $\text{tr } \mathbf{P}_u = 1$ . Plugging in  $\hat{\theta}$  under  $H_0$  into the above formulas, and observing that

$$\frac{1}{n} \mathbf{Y}'\mathbf{Y} = \frac{1}{n} \mathbf{Y}'(\hat{\mu})\mathbf{Y}(\hat{\mu}) = \frac{1}{n} \mathbf{X}'\mathbf{Q}_n\mathbf{X} = \mathbf{S},$$

we get

$$\mathbf{s}(\hat{\theta}) = -\frac{n}{2}(\mathbf{I}_2 \otimes \mathbf{D}'_m) \cdot \begin{pmatrix} (u-1) \text{vec } \hat{\Delta}^{-1} - (\hat{\Delta}^{-1} \otimes \hat{\Delta}^{-1}) \text{vec}\{\text{BTr}_m[(\mathbf{Q}_u \otimes \mathbf{I}_m)\mathbf{S}]\} \\ \text{vec } \hat{\Delta}^{-1} - (\hat{\Delta}^{-1} \otimes \hat{\Delta}^{-1}) \text{vec}\{\text{BTr}_m[(\mathbf{P}_u \otimes \mathbf{I}_m)\mathbf{S}]\} \end{pmatrix},$$

which, due to the formulas (6) for MLEs of BCS structure, can be simplified to

$$\mathbf{s}(\hat{\theta}) = -\frac{n}{2}(\mathbf{I}_2 \otimes \mathbf{D}'_m) \begin{pmatrix} (u-1)(\hat{\Delta}^{-1} \otimes \hat{\Delta}^{-1}) \text{vec}(\hat{\Delta} - \hat{\Delta}_1) \\ (\hat{\Delta}^{-1} \otimes \hat{\Delta}^{-1}) \text{vec}(\hat{\Delta} - \hat{\Delta}_2) \end{pmatrix}. \tag{A.1}$$

To compute Fisher information matrix, second order partial derivatives and their expected values are calculated. The detailed computations only for parameter  $\Delta_1$  are presented here, as the derivatives with respect to  $\Delta_2$  result from the same arguments.

Using derivatives from Fackler (2005) and formula (1.4.23) from Kollo and von Rosen (2005) we get

$$\frac{\partial^2 \ln L}{\partial \Delta_1^2} = \frac{n}{2} \mathbf{D}'_m \left\{ (u-1)(\Delta_1^{-1} \otimes \Delta_1^{-1}) - (\text{vec}'\{\text{BTr}_m[(\mathbf{Q}_u \otimes \mathbf{I}_m)\frac{1}{n}\mathbf{Y}'\mathbf{Y}]\} \otimes \mathbf{I}_{m^2})(\mathbf{I}_m \otimes \mathbf{K}_{m,m} \otimes \mathbf{I}_m) \cdot (\mathbf{I}_{m^2} \otimes \text{vec}' \Delta_1^{-1} + \text{vec } \Delta_1^{-1} \otimes \mathbf{I}_{m^2})(\Delta_1^{-1} \otimes \Delta_1^{-1}) \right\} \mathbf{D}_m.$$

To compute the expectation of the above, we use the notation  $E(\mathbf{Y}'\mathbf{Y}) = \mathbf{\Gamma}$  (cf. Kollo and von Rosen (2005)[Th. 2.2.9(i)]), and hence from orthogonality of  $\mathbf{P}_u$  and  $\mathbf{Q}_u$ , we obtain

$$\mathbf{F}_{11} = -\frac{n}{2} \mathbf{D}'_m \left( (u-1)(\Delta_1^{-1} \otimes \Delta_1^{-1}) - \{\text{vec}'[\text{BTr}_m(\mathbf{Q}_u \otimes \Delta_1)] \otimes \mathbf{I}_{m^2}\}(\mathbf{I}_m \otimes \mathbf{K}_{m,m} \otimes \mathbf{I}_m) \cdot (\mathbf{I}_{m^2} \otimes \text{vec}' \Delta_1^{-1} + \text{vec } \Delta_1^{-1} \otimes \mathbf{I}_{m^2})(\Delta_1^{-1} \otimes \Delta_1^{-1}) \right) \mathbf{D}_m.$$

Noting, that the block-diagonal elements of  $\mathbf{Q}_u \otimes \Delta_1$  are equal to  $\frac{u-1}{u} \Delta_1$  it is easy to see that  $\text{BTr}_m(\mathbf{Q}_u \otimes \Delta_1) = (u-1)\Delta_1$ . Furthermore, since for symmetric matrix  $\mathbf{A}$

$$\begin{aligned} & (\text{vec}' \mathbf{A} \otimes \mathbf{D}'_m)(\mathbf{I}_m \otimes \mathbf{K}_{m,m} \otimes \mathbf{I}_m)(\text{vec } \mathbf{A}^{-1} \otimes \mathbf{I}_{m^2}) \\ &= (\text{vec}' \mathbf{A} \otimes \mathbf{D}'_m)(\mathbf{I}_m \otimes \mathbf{K}_{m,m} \otimes \mathbf{I}_m)(\mathbf{I}_{m^2} \otimes \text{vec } \mathbf{A}^{-1}) = \mathbf{D}'_m \end{aligned}$$

cf. Filipiak et al. (2016)[Lemma 1], we obtain

$$\mathbf{F}_{11} = \frac{n(u-1)}{2} \mathbf{D}'_m (\mathbf{\Delta}_1^{-1} \otimes \mathbf{\Delta}_1^{-1}) \mathbf{D}_m. \tag{A.2}$$

Plugging in the MLE's of unknown parameters under  $\mathbf{H}_0$ , we finally obtain

$$\mathbf{F}(\hat{\boldsymbol{\theta}}) = \frac{n}{2} \begin{pmatrix} (u-1) \mathbf{D}'_m (\hat{\mathbf{\Delta}}^{-1} \otimes \hat{\mathbf{\Delta}}^{-1}) \mathbf{D}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{D}'_m (\hat{\mathbf{\Delta}}^{-1} \otimes \hat{\mathbf{\Delta}}^{-1}) \mathbf{D}_m \end{pmatrix}$$

and, from Filipiak et al. (2016)[Prop. 1(iv)],

$$\mathbf{F}^{-1}(\hat{\boldsymbol{\theta}}) = \frac{2}{n} (\mathbf{I}_2 \otimes \mathbf{D}_m^+) \begin{pmatrix} \frac{1}{u-1} (\hat{\mathbf{\Delta}} \otimes \hat{\mathbf{\Delta}}) & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{\Delta}} \otimes \hat{\mathbf{\Delta}} \end{pmatrix} (\mathbf{I}_2 \otimes \mathbf{D}_m^{+'}),$$

where  $\mathbf{D}_m^+$  is a Moore-Penrose inverse of  $\mathbf{D}_m$ . Denoting the score vector (A.1) as

$$\mathbf{s}(\hat{\boldsymbol{\theta}}) = -\frac{n}{2} (\mathbf{I}_2 \otimes \mathbf{D}'_m) \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{pmatrix},$$

we get

$$\text{RS} = \frac{n}{2} (\mathbf{s}'_1, \mathbf{s}'_2) (\mathbf{I}_2 \otimes \mathbf{D}_m \mathbf{D}_m^+) \begin{pmatrix} \frac{1}{u-1} (\hat{\mathbf{\Delta}} \otimes \hat{\mathbf{\Delta}}) & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{\Delta}} \otimes \hat{\mathbf{\Delta}} \end{pmatrix} (\mathbf{I}_2 \otimes \mathbf{D}_m^{+'} \mathbf{D}'_m) \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{pmatrix}.$$

Using Magnus and Neudecker (1986)[formulas (54) and (36)] we may further simplify the above RS to

$$\begin{aligned} \text{RS} &= \frac{n}{2} \left[ (u-1) \text{vec}'(\hat{\mathbf{\Delta}} - \hat{\mathbf{\Delta}}_1) (\hat{\mathbf{\Delta}}^{-1} \otimes \hat{\mathbf{\Delta}}^{-1}) \text{vec}(\hat{\mathbf{\Delta}} - \hat{\mathbf{\Delta}}_1) \right. \\ &\quad \left. + \text{vec}'(\hat{\mathbf{\Delta}} - \hat{\mathbf{\Delta}}_2) (\hat{\mathbf{\Delta}}^{-1} \otimes \hat{\mathbf{\Delta}}^{-1}) \text{vec}(\hat{\mathbf{\Delta}} - \hat{\mathbf{\Delta}}_2) \right] \\ &= \frac{n}{2} \left\{ (u-1) \text{tr}[(\hat{\mathbf{\Delta}} - \hat{\mathbf{\Delta}}_1) \hat{\mathbf{\Delta}}^{-1} (\hat{\mathbf{\Delta}} - \hat{\mathbf{\Delta}}_1) \hat{\mathbf{\Delta}}^{-1}] \right. \\ &\quad \left. + \text{tr}[(\hat{\mathbf{\Delta}} - \hat{\mathbf{\Delta}}_2) \hat{\mathbf{\Delta}}^{-1} (\hat{\mathbf{\Delta}} - \hat{\mathbf{\Delta}}_2) \hat{\mathbf{\Delta}}^{-1}] \right\} \\ &= \frac{n}{2} \left\{ (u-1) \text{tr}[\mathbf{I}_m - 2\hat{\mathbf{\Delta}}_1 \hat{\mathbf{\Delta}}^{-1} + \hat{\mathbf{\Delta}}_1 \hat{\mathbf{\Delta}}^{-1} \hat{\mathbf{\Delta}}_1 \hat{\mathbf{\Delta}}^{-1}] \right. \\ &\quad \left. + \text{tr}[\mathbf{I}_m - 2\hat{\mathbf{\Delta}}_2 \hat{\mathbf{\Delta}}^{-1} + \hat{\mathbf{\Delta}}_2 \hat{\mathbf{\Delta}}^{-1} \hat{\mathbf{\Delta}}_2 \hat{\mathbf{\Delta}}^{-1}] \right\} \\ &= \frac{n}{2} \left[ \text{tr} \mathbf{I}_{um} - 2 \text{tr}(\mathbf{Q}_u \otimes \hat{\mathbf{\Delta}}_1 \hat{\mathbf{\Delta}}^{-1}) + \text{tr}(\mathbf{Q}_u \otimes \hat{\mathbf{\Delta}}_1 \hat{\mathbf{\Delta}}^{-1} \hat{\mathbf{\Delta}}_1 \hat{\mathbf{\Delta}}^{-1}) \right. \\ &\quad \left. - 2 \text{tr}(\mathbf{P}_u \otimes \hat{\mathbf{\Delta}}_2 \hat{\mathbf{\Delta}}^{-1}) + \text{tr}(\mathbf{P}_u \otimes \hat{\mathbf{\Delta}}_2 \hat{\mathbf{\Delta}}^{-1} \hat{\mathbf{\Delta}}_2 \hat{\mathbf{\Delta}}^{-1}) \right] \end{aligned}$$

Finally, from idempotency and orthogonality of  $\mathbf{Q}_u$  and  $\mathbf{P}_u$  we obtain

$$\begin{aligned} \text{RS} &= \frac{n}{2} \text{tr} \left[ \mathbf{I}_{um} - 2(\mathbf{Q}_u \otimes \hat{\mathbf{\Delta}}_1 + \mathbf{P}_u \otimes \hat{\mathbf{\Delta}}_2) (\mathbf{I}_u \otimes \hat{\mathbf{\Delta}}^{-1}) \right. \\ &\quad \left. + (\mathbf{Q}_u \otimes \hat{\mathbf{\Delta}}_1 + \mathbf{P}_u \otimes \hat{\mathbf{\Delta}}_2) (\mathbf{I}_u \otimes \hat{\mathbf{\Delta}}^{-1}) \right] \end{aligned}$$

$$\begin{aligned} & \cdot (\mathbf{Q}_u \otimes \widehat{\Delta}_1 + \mathbf{P}_u \otimes \widehat{\Delta}_2)(\mathbf{I}_u \otimes \widehat{\Delta}^{-1}) \Big] \\ & = \frac{n}{2} \operatorname{tr} \left\{ \left[ \mathbf{I}_{um} - \widehat{\Gamma}(\mathbf{I}_u \otimes \widehat{\Delta})^{-1} \right]^2 \right\}. \end{aligned}$$

### Appendix B: Proof of Theorem 2

Due to Rao (2005) the Wald test statistics for testing composite null hypothesis  $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{c}$  can be represented as

$$\mathbf{W} = (\mathbf{h}(\widehat{\boldsymbol{\theta}}) - \mathbf{c})' \mathbf{A}^{-1}(\widehat{\boldsymbol{\theta}})(\mathbf{h}(\widehat{\boldsymbol{\theta}}) - \mathbf{c}),$$

where  $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\mathbf{F}^{-1}(\boldsymbol{\theta})\mathbf{H}'(\boldsymbol{\theta})$ , with  $\mathbf{H}(\boldsymbol{\theta})$  being the matrix of derivatives of  $\mathbf{h}$  with respect to the components of  $\boldsymbol{\theta}$ ,  $\mathbf{F}(\boldsymbol{\theta})$  is the Fisher information matrix and  $\widehat{\boldsymbol{\theta}}$  is the MLE of  $\boldsymbol{\theta}$  under alternative. Considering hypothesis (9) we may note that the null hypothesis can be written as  $\operatorname{vech}(\Delta_1 - \Delta_2) = \mathbf{0}_{m(m+1)/2}$  which is equivalent to  $(\mathbf{I}_{m(m+1)/2}, -\mathbf{I}_{m(m+1)/2})\boldsymbol{\theta} = \mathbf{0}_{m(m+1)/2}$ , where  $\boldsymbol{\theta} = (\operatorname{vech}' \Delta_1, \operatorname{vech}' \Delta_2)'$ . It follows that  $\mathbf{h}(\boldsymbol{\theta}) = (\mathbf{I}, -\mathbf{I})\boldsymbol{\theta}$  with identity matrices of order  $m(m+1)/2$ , and hence  $\mathbf{H}(\boldsymbol{\theta}) = (\mathbf{I}, -\mathbf{I})$ , while  $\mathbf{F}(\boldsymbol{\theta})$  is a block diagonal matrix with diagonal blocks given in (A.2) and equal to  $\frac{n}{2}\mathbf{D}'_m(\Delta_2^{-1} \otimes \Delta_2^{-1})\mathbf{D}_m$ . We then obtain

$$\begin{aligned} \mathbf{A}(\boldsymbol{\theta}) &= \frac{2}{n} (\mathbf{I}, -\mathbf{I}) \begin{pmatrix} \frac{1}{u-1}\mathbf{D}_m^+(\Delta_1 \otimes \Delta_1)\mathbf{D}_m^{+'} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_m^+(\Delta_2 \otimes \Delta_2)\mathbf{D}_m^{+'} \end{pmatrix} \begin{pmatrix} \mathbf{I} \\ -\mathbf{I} \end{pmatrix} \\ &= \frac{2}{n}\mathbf{D}_m^+ \left[ \frac{1}{u-1}(\Delta_1 \otimes \Delta_1) + (\Delta_2 \otimes \Delta_2) \right] \mathbf{D}_m^{+'} \end{aligned}$$

and, plugging the inverse of  $\mathbf{A}(\widehat{\boldsymbol{\theta}})$  into the formula for Wald test statistic, we get

$$\begin{aligned} \mathbf{W} &= \\ & \frac{n}{2} \operatorname{vech}'(\widehat{\Delta}_1 - \widehat{\Delta}_2) \mathbf{D}'_m \left[ \frac{1}{u-1}(\widehat{\Delta}_1 \otimes \widehat{\Delta}_1) + (\widehat{\Delta}_2 \otimes \widehat{\Delta}_2) \right]^{-1} \mathbf{D}_m \operatorname{vech}(\widehat{\Delta}_1 - \widehat{\Delta}_2). \end{aligned}$$

From the definition of duplication matrix we have  $\mathbf{D}_m \operatorname{vech}(\widehat{\Delta}_1 - \widehat{\Delta}_2) = \operatorname{vec}(\widehat{\Delta}_1 - \widehat{\Delta}_2)$  and the thesis follows.

### Appendix C: Proof of Theorem 3

Denoting  $\mathbf{A}_1 = n(u-1)\widehat{\Delta}_1$ ,  $\mathbf{A}_2 = n\widehat{\Delta}_2$  and observing that

$$nu\widehat{\Delta} = nu(u-1)\widehat{\Delta}_1 + nu\widehat{\Delta}_2 = \mathbf{A}_1 + \mathbf{A}_2,$$

we may write (11) as

$$\Lambda_* = \Lambda^{2/n} = \frac{u^{um}}{(u-1)^{(u-1)m}} \cdot \frac{|\mathbf{A}_1|^{u-1} |\mathbf{A}_2|}{|\mathbf{A}_1 + \mathbf{A}_2|^u}.$$

where  $\mathbf{A}_1 \sim W_m(\mathbf{\Delta}_1, (n - 1)(u - 1))$ ,  $\mathbf{A}_2 \sim W_m(\mathbf{\Delta}_2, n - 1)$  and  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are independent; cf. Roy et al. (2015); Filipiak and Klein (2021).

Since the probability density function of any  $m \times m$  matrix  $\mathbf{W} \sim W_m(\mathbf{\Sigma}, \nu)$  can be expressed as

$$f_{\mathbf{W}}(\mathcal{W}) = \frac{e^{-\frac{1}{2} \text{tr}(\mathbf{\Theta}^{-1}\mathcal{W})} |\mathcal{W}|^{\frac{\nu-m-1}{2}}}{2^{\frac{m\nu}{2}} \Gamma_m(\frac{\nu}{2}) |\mathbf{\Sigma}|^{\frac{\nu}{2}}}, \quad \mathcal{W} > 0,$$

with multivariate gamma function of order  $m$  given as

$$\Gamma_m(\frac{\nu}{2}) = \pi^{m(m-1)/4} \prod_{i=1}^m \Gamma(\frac{\nu-i+1}{2}),$$

cf. Kollo and von Rosen (2005)[Th. 2.4.6], we may express the  $h$ -th moment of  $|\mathbf{W}|$  by the following formula

$$\begin{aligned} E(|\mathbf{W}|^h) &= \int_{\mathcal{W}} |\mathcal{W}|^h \cdot f_{\mathbf{W}}(\mathcal{W}) d\mathcal{W} \\ &= \int_{\mathcal{W}} |\mathcal{W}|^h \frac{e^{-\frac{1}{2} \text{tr}(\mathbf{\Theta}^{-1}\mathcal{W})} |\mathcal{W}|^{\frac{\nu-m-1}{2}}}{2^{\frac{m\nu}{2}} \Gamma_m(\frac{\nu}{2}) |\mathbf{\Sigma}|^{\frac{\nu}{2}}} d\mathcal{W} \\ &= \frac{\Gamma_m(\frac{\nu}{2} + h)}{2^{-mh} \Gamma_m(\frac{\nu}{2}) |\mathbf{\Sigma}|^{-h}} \int_{\mathcal{W}} \frac{e^{-\frac{1}{2} \text{tr}(\mathbf{\Theta}^{-1}\mathcal{W})} |\mathcal{W}|^{\frac{\nu-m-1+2h}{2}}}{2^{\frac{m(\nu+2h)}{2}} \Gamma_m(\frac{\nu}{2} + h) |\mathbf{\Sigma}|^{\frac{\nu+2h}{2}}} d\mathcal{W} \\ &= 2^{mh} \cdot |\mathbf{\Sigma}|^h \cdot \frac{\Gamma_m(\frac{\nu}{2} + h)}{\Gamma_m(\frac{\nu}{2})}, \end{aligned} \tag{C.1}$$

where  $\mathcal{W}$  is the symmetric positive-definite matrices space. Hence, the  $h$ -th moment of  $\Lambda_*$  can be written as

$$\begin{aligned} E(\Lambda_*^h) &= \frac{u^{umh}}{(u - 1)^{(u-1)mh}} \\ &\cdot \int_{\mathcal{A}_1} \int_{\mathcal{A}_2} \frac{|\mathcal{A}_1|^{(u-1)h} |\mathcal{A}_2|^h}{|\mathcal{A}_1 + \mathcal{A}_2|^{uh}} \\ &\times \frac{e^{-\text{tr}(\mathbf{I}^{-1}\mathcal{A}_1)/2} |\mathcal{A}_1|^{[(n-1)(u-1)-m-1]/2}}{2^{m(n-1)(u-1)/2} \Gamma_m\left(\frac{(n-1)(u-1)}{2}\right) |\mathbf{1}|^{(n-1)(u-1)/2}} \\ &\times \frac{e^{-\text{tr}(\mathbf{I}^{-1}\mathcal{A}_2)/2} |\mathcal{A}_2|^{(n-2-m)/2}}{2^{m(n-1)/2} \Gamma_m\left(\frac{n-1}{2}\right) |\mathbf{1}|^{(n-1)/2}} d\mathcal{A}_1 d\mathcal{A}_2 \end{aligned}$$



$$\begin{aligned}
 &= \frac{u^{umh}}{(u-1)^{(u-1)mh}} |\mathbf{1}|^{uh} \cdot 2^{umh} \cdot \frac{\Gamma_m\left(\frac{(n-1+2h)(u-1)}{2}\right)}{\Gamma_m\left(\frac{(n-1)(u-1)}{2}\right)} \cdot \frac{\Gamma_m\left(\frac{n-1+2h}{2}\right)}{\Gamma_m\left(\frac{n-1}{2}\right)} \\
 &\quad \times \int_{\mathcal{A}_1} \int_{\mathcal{A}_2} \frac{1}{|\mathcal{A}_1 + \mathcal{A}_2|^{uh}} \\
 &\quad \quad \frac{e^{-\text{tr}(\mathbf{1}^{-1}\mathcal{A}_1)/2}}{|\mathcal{A}_1|^{[(u-1)(n-1+2h)-m-1]/2}} \\
 &\quad \times \frac{2^{m(n-1+2h)(u-1)/2} \Gamma_m\left(\frac{(n-1+2h)(u-1)}{2}\right) |\Delta|^{(n-1+2h)(u-1)/2}}{e^{-\text{tr}(\mathbf{1}^{-1}\mathcal{A}_2)/2} |\mathcal{A}_2|^{(n-1+2h-m-1)/2}} \\
 &\quad \times \frac{2^{m(n-1+2h)/2} \Gamma_m\left(\frac{n-1+2h}{2}\right) |\Delta|^{(n-1+2h)/2}}{(2u)^{umh}} d\mathcal{A}_1 d\mathcal{A}_2 \\
 &= \frac{(u-1)^{(u-1)mh}}{\Gamma_m\left(\frac{(n-1+2h)(u-1)}{2}\right) \Gamma_m\left(\frac{n-1+2h}{2}\right)} \\
 &\quad \times \frac{\Gamma_m\left(\frac{(n-1)(u-1)}{2}\right) \Gamma_m\left(\frac{n-1}{2}\right)}{\Gamma_m\left(\frac{(n-1)(u-1)}{2}\right) \Gamma_m\left(\frac{n-1}{2}\right)} \cdot |\Delta|^{uh} \cdot E(|\mathbf{A}|^{-uh}),
 \end{aligned}$$

where  $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2 \sim W_m(\Delta, (n-1+2h)u)$ . Now, using (C.1), we can write

$$\begin{aligned}
 E(\Lambda_*^h) &= \frac{u^{umh}}{(u-1)^{(u-1)mh}} \frac{\Gamma_m\left(\frac{(n-1+2h)(u-1)}{2}\right) \Gamma_m\left(\frac{n-1+2h}{2}\right) \Gamma_m\left(\frac{u(n-1)}{2}\right)}{\Gamma_m\left(\frac{(n-1)(u-1)}{2}\right) \Gamma_m\left(\frac{n-1}{2}\right) \Gamma_m\left(\frac{u(n-1+2h)}{2}\right)} \\
 &= \frac{u^{umh}}{(u-1)^{(u-1)mh}} \\
 &\quad \times \prod_{j=1}^m \left[ \frac{\Gamma\left(\frac{(n-1)(u-1)+1-j}{2} + h(u-1)\right)}{\Gamma\left(\frac{(n-1)(u-1)+1-j}{2}\right)} \cdot \frac{\Gamma\left(\frac{n-j}{2} + h\right)}{\Gamma\left(\frac{n-j}{2}\right)} \cdot \frac{\Gamma\left(\frac{(n-1)u+1-j}{2}\right)}{\Gamma\left(\frac{(n-1)u+1-j}{2} + hu\right)} \right].
 \end{aligned}$$

### Appendix D: Proof of Proposition 1

It is enough to show that the distribution of the RS, LR and Wald test statistics presented respectively in Theorem 1, (11) and Theorem 2 under the null hypothesis given in (9) does not depend on the true value of  $\Delta$ .

The observation matrix  $\mathbf{X}$  can be written as  $\mathbf{X} = \mathbf{1}_n^{-'} + \mathbf{E}\Omega^{1/2}$ , where  $\Omega^{1/2} = \mathbf{I}_u \otimes \Delta^{1/2}$  with  $\Delta^{1/2} \Delta^{1/2} = \Delta$  and  $\mathbf{E} \sim N_{n,um}(\mathbf{0}, \mathbf{I}_n, \mathbf{I}_{um})$ , therefore

$$\mathbf{X}'\mathbf{Q}_n\mathbf{X} = \Omega^{1/2}\mathbf{E}'\mathbf{Q}_n\mathbf{E}\Omega^{1/2}.$$

Hence, due to (10) and using Filipiak et al. (2018)[Lemma 2.11], we have

$$\begin{aligned}
 \widehat{\Delta} &= \frac{1}{u} \text{BTr}_m[\mathbf{S}] = \frac{1}{nu} \text{BTr}_m[\mathbf{X}'\mathbf{Q}_n\mathbf{X}] = \\
 &= \frac{1}{nu} \text{BTr}_m[\Omega^{1/2}\mathbf{E}'\mathbf{Q}_n\mathbf{E}\Omega^{1/2}] = \frac{1}{nu} \text{BTr}_m[(\mathbf{I}_u \otimes \Delta^{1/2})\mathbf{E}'\mathbf{Q}_n\mathbf{E}(\mathbf{I}_u \otimes \Delta^{1/2})] = \\
 &= \Delta^{1/2} \frac{1}{nu} \text{BTr}_m[\mathbf{E}'\mathbf{Q}_n\mathbf{E}]\Delta^{1/2} = \Delta^{1/2} \widehat{\Upsilon}_0 \Delta^{1/2},
 \end{aligned}$$

where  $\widehat{\mathbf{Y}}_0$  does not depend on the true values of the unknown parameters. Similarly we find that the estimators in (6) can be expressed as

$$\widehat{\mathbf{\Delta}}_1 = \mathbf{\Delta}^{1/2} \widehat{\mathbf{Y}}_1 \mathbf{\Delta}^{1/2}, \quad \text{and} \quad \widehat{\mathbf{\Delta}}_2 = \mathbf{\Delta}^{1/2} \widehat{\mathbf{Y}}_2 \mathbf{\Delta}^{1/2}, \quad (\text{D.1})$$

where both  $\widehat{\mathbf{Y}}_1$  and  $\widehat{\mathbf{Y}}_2$  do not depend on the true values of the unknown parameters, since

$$\begin{aligned} \widehat{\mathbf{Y}}_1 &= \frac{1}{n(u-1)} \text{BTr}_m[(\mathbf{Q}_u \otimes \mathbf{I}_m) \mathbf{E}' \mathbf{Q}_n \mathbf{E} (\mathbf{Q}_u \otimes \mathbf{I}_m)], \\ \widehat{\mathbf{Y}}_2 &= \frac{1}{n} \text{BTr}_m[(\mathbf{P}_u \otimes \mathbf{I}_m) \mathbf{E}' \mathbf{Q}_n \mathbf{E} (\mathbf{P}_u \otimes \mathbf{I}_m)]. \end{aligned} \quad (\text{D.2})$$

Therefore for estimator  $\widehat{\mathbf{\Gamma}}$  given in (5) it holds

$$\widehat{\mathbf{\Gamma}} = (\mathbf{I}_u \otimes \mathbf{\Delta}^{1/2}) \left( \mathbf{Q}_u \otimes \widehat{\mathbf{Y}}_1 + \mathbf{P}_u \otimes \widehat{\mathbf{Y}}_2 \right) (\mathbf{I}_u \otimes \mathbf{\Delta}^{1/2}),$$

and after substituting into RST statistic given in Theorem 1 it can be presented as

$$\text{RS} = \frac{n}{2} \text{tr} \left\{ \left[ \mathbf{I}_{um} - (\mathbf{Q}_u \otimes \widehat{\mathbf{Y}}_1 + \mathbf{P}_u \otimes \widehat{\mathbf{Y}}_2) (\mathbf{I}_u \otimes \widehat{\mathbf{Y}}_0)^{-1} \right]^2 \right\}.$$

The LRT statistic (11) and Wald test statistic given in Theorem 2 can be expressed as

$$\Lambda = \left( \frac{|\widehat{\mathbf{Y}}_1|^{u-1} |\widehat{\mathbf{Y}}_2|}{|\widehat{\mathbf{Y}}_0|^u} \right)^{n/2},$$

$$\mathbf{W} = \frac{n}{2} \text{vec}'(\widehat{\mathbf{Y}}_1 - \widehat{\mathbf{Y}}_2) \left[ \frac{1}{u-1} (\widehat{\mathbf{Y}}_1 \otimes \widehat{\mathbf{Y}}_1) + (\widehat{\mathbf{Y}}_2 \otimes \widehat{\mathbf{Y}}_2) \right]^{-1} \text{vec}(\widehat{\mathbf{Y}}_1 - \widehat{\mathbf{Y}}_2),$$

respectively.

## References

- Anderson TW (2003) An introduction to multivariate statistical analysis. Wiley, Hoboken
- Arnold SF (1973) Application of the theory of products of problems to certain patterned covariance matrices. *Ann Stat* 1(4):682–699
- Arnold SF (1979) Linear models with exchangeably distributed errors. *J Am Stat Assoc* 74:194–199
- Chiani M (2016) Distribution of the largest root of a matrix for Roy's test in multivariate analysis of variance. *J Multivar Anal* 143:467–471
- Fackler PL (2005) Notes on matrix calculus. <https://media.gradebuddy.com/documents/1897145/1ad5e235-824d-4bbf-81d6-ffd2040e37ec.pdf>
- Filipiak K, John M, Markiewicz A (2020) Comments on maximum likelihood estimation and projections under multivariate statistical models. In: Holgersson T, Singull M (eds) *Recent developments in multivariate and random matrix analysis*. Springer, Berlin, pp 51–66
- Filipiak K, Klein D (2021) Estimation and testing of the covariance structure of doubly multivariate data. In: Filipiak K, Markiewicz A, von Rosen D (eds) *Multivariate, multilinear and mixed linear models*. Springer, Berlin
- Filipiak K, Klein D, Roy A (2016) Score test for a separable covariance structure with the first component as compound symmetric correlation matrix. *J Multivar Anal* 150:105–124

- Filipiak K, Klein D, Vojtková E (2018) The properties of partial trace and block trace operators of partitioned matrices. *Electron J Linear Algebra* 33:3–15
- Fonseca M, Kozioł A, Zmysłony R (2018) Testing hypotheses of covariance structure in multivariate data. *Electron J Linear Algebra* 33:53–62
- Gajdoš A (2018) CharFunToolR: the characteristic functions toolbox (R). <https://github.com/gajdosandrej/CharFunToolR>
- Ghazal AG, Neudecker H (2000) On second-order and fourth-order moments of jointly distributed random matrices: a survey. *Linear Algebra Appl* 321:61–93
- Hao C, Liang Y, Roy A (2015) Equivalency between vertices and centers-coupled-with-radii principal component analyses for interval data. *Stat Probab Lett* 106:113–120
- Kollo T, von Rosen D (2005) *Advanced multivariate statistics with matrices*. Springer, Dordrecht
- Kozioł A, Roy A, Zmysłony R, Leiva R, Fonseca M (2018) Free coordinate estimation for doubly multivariate data. *Linear Algebra Appl* 547:217–239
- Kozioł A, Roy A, Zmysłony R, Žežula I, Fonseca M (2021) Estimation and testing hypotheses in two-level and three-level multivariate data with block compound symmetric covariance structure. In: Filipiak K, Markiewicz A, von Rosen D (eds) *Multivariate, multilinear and mixed linear models*. Springer, Berlin, pp 203–232
- Leiva R (2007) Linear discrimination with equicorrelated training vectors. *J Multivar Anal* 98:384–409
- Liang Y, Coelho CA, von Rosen T (2021) Hypothesis testing in multivariate normal models with block circular covariance structures. *Biom J* 64:557–576
- Liang Y, von Rosen D, von Rosen T (2012) On estimation in multilevel models with block circular symmetric covariance structure. *Acta Comment Univ Tartu Math* 16(1):83–96
- Liang Y, von Rosen D, von Rosen T (2015) On estimation in hierarchical models with block circular covariance structures. *Ann Inst Stat Math* 67:773–791
- Magnus J, Neudecker H (1986) Symmetry, 0–1 matrices and Jacobians, a review. *Econ. Theory* 2:157–190
- Mardia KV, Kent T, Bibby M (1979) *Multivariate analysis*. Academic Press, Cambridge
- Perlman MD (1987) Group symmetry covariance models. *Stat Sci* 2:421–425
- Rao CR (1945) Familial correlations or the multivariate generalizations of the intraclass correlation. *Curr. Sci.* 14:66–67
- Rao CR (1953) Discriminant functions for genetic differentiation and selection. *Sankhyā* 12:229–246
- Rao CR (2005) Score test: historical review and recent developments. *Advances in ranking and selection, multiple comparisons, and reliability*. Springer, Berlin, pp 3–20
- Roy A, Leiva R (2011) Estimating and testing a structured covariance matrix for three-level multivariate data. *Commun Stat Theory Methods* 40:1945–1963
- Roy A, Zmysłony R, Fonseca M, Leiva R (2016) Optimal estimation for doubly multivariate data in blocked compound symmetric covariance structure. *J Multivar Anal* 144:81–90
- Roy A, Filipiak K, Klein D (2018) Testing a block exchangeable covariance matrix. *Statistics* 52:393–408
- 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
- Seely J (1971) Quadratic subspaces and completeness. *Ann Math Stat* 42:710–721
- Self SG, Liang K (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 82:605–610
- Sun G, Xie J (2020) Asymptotic normality and moderate deviation principle for high-dimensional likelihood ratio statistic on block compound symmetry covariance structure. *Statistics* 54:114–134
- 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, Stanford
- Szatrowski TH (1978) Explicit solutions, one iteration convergence and averaging in the multivariate normal estimation problem for patterned means and covariance. *Ann Inst Stat Math* 30A:81–88
- Szatrowski TH (1982) Testing and estimation in the block compound symmetry problem. *J Educ Stat* 7(1):3–18
- Tsukada S (2018) Hypothesis testing for independence under blocked compound symmetric covariance structure. *Commun Math Stat* 6:163–184
- Witkovský V (2018) CharFunTool: the characteristic functions tool-box (MATLAB). <https://github.com/witkovsky/CharFunTool>

- Zmyślony R, Żeżula I, Koziol A (2018) Application of Jordan algebra for testing hypotheses about structure of mean vector in model with block compound symmetric covariance structure. *Electron J Linear Algebra* 33:41–52
- Żeżula I, Klein D, Roy A (2018) Testing of multivariate repeated measures data with block exchangeable covariance structure. *TEST* 27:360–378

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.