THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

Date September 15, 2015

WP # 0003MSS-253-2014

Score test for a separable covariance structure with the first component as compound symmetric correlation matrix

Katarzyna Filipiak
Department of Mathematical and Statistical Methods
Poznan University of Life Sciences
Wojska Polskiego 28, PL-60637 Poznan, Poland
E-mail: kas_l@up.poznan.pl

Daniel Klein
Institute of Mathematics, Faculty of Science
P. J. Safarik University
Jesenna 5, 040 01 Kosice, Slovakia
daniel.klein@upjs.sk

Anuradha Roy _
Department of Management Science and Statistics
The University of Texas at San Antonio
One UTSA Circle, San Antonio, TX 78249 USA
Email: aroy@utsa.edu
Phone: +00-210-458-6343, Fax: +00-210-458-6350

ONE UTSA CIRCLE
SAN ANTONIO, TEXAS 78249-0631
210.458.4317 | BUSINESS.UTSA.EDU

Score test for a separable covariance structure with the first component as compound symmetric correlation matrix

Katarzyna Filipiak
Department of Mathematical and Statistical Methods
Poznań University of Life Sciences
Wojska Polskiego 28, PL-60637 Poznań, Poland
E-mail: kasfil@up.poznan.pl

Daniel Klein
Institute of Mathematics, Faculty of Science
P. J. Šafárik University
Jesenná 5, 040 01 Košice, Slovakia
daniel.klein@upjs.sk

Anuradha Roy *
Department of Management Science and Statistics
The University of Texas at San Antonio
One UTSA Circle, San Antonio, TX 78249 USA
Email: aroy@utsa.edu
Phone: +00-210-458-6343, Fax: +00-210-458-6350

Abstract

Likelihood ratio tests (LRTs) for separability of a covariance structure for doubly multivariate data are widely studied in the literature. There are three types of LRT: biased tests based on an asymptotic chi-square null distribution; unbiased/unmodified tests based on an empirical null distribution; and unbiased/modified tests with a test statistic modified to follow a theoretical chi-square null distribution. The Rao's score test (RST) statistic, an alternative for both biased and unbiased/unmodified versions of the corresponding LRT test statistics are derived for a common case. The separability of a covariance structure with the first component as a compound symmetric correlation matrix under the assumption of multivariate normality is tested. Simulation studies compare the biased LRT to biased RST, and unbiased/unmodified LRT to unbiased/unmodified RST. The RSTs outperform their corresponding LRTs in general. Three examples are presented. Since the RST does not require estimation of a general variance-covariance matrix (the alternative hypothesis), this test can be performed for small sample sizes, where the variance-covariance matrix could not be estimated for the corresponding LRT, making the LRT infeasible. In cases where both LRT and RST are feasible, the RST outperforms a comparable LRT.

Keywords: Empirical null distribution; Likelihood ratio test; Maximum likelihood estimates; Rao's score test; Separable covariance structure

2000 Mathematics Subject Classification: Primary 62H15; Secondary 62H12

*Correspondence to: Anuradha Roy, Department of Management Science and Statistics, The University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, USA

1 Introduction

This article is concerned with a very important hypothesis testing problem of a $2-$separable covariance structure (defined in Section 2) as found in two-level or doubly multivariate data. Modern experimental techniques allow to collect and store *multi-level* multivariate data (Leiva and Roy, 2011) in almost all fields such as agriculture, biology, biomedical, medical, environmental and engineering research, where the observations are collected on more than one response variable (q) at different locations (p) repeatedly over time (t) and at different depths (d) etc. These multi-level multivariate observations may have variances that differ across locations, time and depths, and developing efficient techniques for accounting these variations is of great importance for any statistical analysis.

In many practical problems, where the repeated measures occur, the covariance matrix of these repeated measures is found to have some structure. For measurements of the same type made in the same way it is usual to assume variance homogeneity too. Crowder and Hand (1990, p.60) say "While it is robust not to assume knowledge of the covariance structure, this can result in rather weak inference in the sense that too many degrees of freedom are used up in estimating the covariance parameters, leaving too few for the parameters of interest." The unstructured (UN) covariance matrix does not require stationarity, but is overparametrized since correlation should decay as the space or time points become more widely separated and estimating parameters which are close to zero only adds extra variability due to estimation of excessive parameters and thus losing degrees of freedom. Thus, for example, we assume stationarity as a consequence of the assumption of equicorrelated covariance structure - compound symmetry (CS) - which may be appropriate where the repeated measurements are all made at about the same time, as in the often used 'split-plots' set-up. The CS structure is also plausible where the measurements are made at unequally spaced times over a longer period. The advantages of using CS structure over repeated measures include flexibility in using the structured covariance matrices for the repeated measures and savings in degrees of freedom for testing of hypothesis. In other cases, there might be some strict temporal sequence where the covariance matrix has AR(1) structure, as often seen in medical data.

For doubly multivariate data, separable structure can additionally be used to model data without losing many degrees of freedom and still avoid an over-constrained model. Consider an example of a medical data set where the detection of a cancerous region from surrounding tissues (skeletonization) of patients suffering from breast cancer is the focus. Pinto Pereira et al. (2009) divided each breast image into 48 regions and then estimated the percent density (PD) for each one of its regions. However, they only used one marker, the PD, in their analysis. A better result with a high reliability may be achieved if joint analysis of the PD and a measure of microcalcifications, which are often the only detectable sign of breast cancer, can be done together. These two measurements ($q = 2$), the PD and a measure of microcalcifications, are not only correlated among themselves, but also exploit the strong regional covariance over the 48 regions ($p = 48$). In this example equicorrelated covariance structure could be one of the plausible structures over 48 regions. Besides CS, a few other plausible correlation structures over repeated measures among many are autoregressive of order one $\big(\mathrm{AR}(1)\big)$, circular and Toeplitz. Non-stationary unstructured (UN) and antedependent variance-covariance matrices are other possibilities. All structures on the repeated measures are tentative; so before any statistical analysis of doubly multivariate data one needs to perform tests for the most suitable separable structures with the first component (structure on repeated measurements) as one of the above plausible structures, i.e., (CS $\otimes$ UN), or (AR(1) $\otimes$ UN) or (UN $\otimes$ UN) etc.

1.1 Existing Tests

The most common hypotheses testing procedures for large samples are the likelihood ratio (Wilks, 1938), the Wald (Wald, 1943), and the Rao's score (Rao, 1948) tests. These were all developed using one-level multivariate models. These tests have earned the status of default methods, with a neat and unified asymptotic theory. They are widely used in almost all areas from agriculture to engineering research among many others even for the smallest possible sample size (n). The likelihood ratio test criterion Λ (Anderson, 1984) or a function of it, $\mathcal{L} = -2\ln\Lambda$ (Wald, 1943), is the most commonly used test statistic. The quantity $\mathcal{L}$ under the null hypothesis is asymptotically distributed as a χ^2 under normality assumption and is used as the test statistic with large sample size. When the data are not large enough, χ^2 distribution is generally an inadequate approximation thus resulting in erroneous conclusions. When the sample size is small or moderate, Korin (1968) studied the accuracy of the approximation and expressed the null distribution of $\mathcal{L}$ in the form of an asymptotic series of central χ^2 distribution and then derived the distribution of $\mathcal{L}$ using this series.

All the above-mentioned tests have been established for traditional multivariate data (say with q response variables); in other words, just for 'one-level multivariate data' in a large sample setting. Hypothesis testing of a $2-$separable covariance structure with both unstructured components has been widely studied by many authors $\big($Dutilleul, 1999; Roy and Khattree, 2003; Lu and Zimmerman, 2005; Roy, 2007; Srivastava *et al.,* 2008; Werner *et al.,* 2008$\big)$. Roy and Khattree (2005a, b) have also studied this $2-$separable covariance structure by assuming a compound symmetry (CS) or autoregressive of order one $\big(\mathrm{AR}(1)\big)$ correlation structures on the first component just to avoid the identifiability problem. Roy and Khattree (2007a) have shown that the choice of appropriate covariance structure is crucial for two-level multivariate data in the context of classification, and it almost always affects the misclassification error rate, in a major way. Thus, it is vital to test the appropriate covariance structure on the multi-level multivariate observations before any statistical analysis. Roy and Leiva (2008) further studied the 2−separable covariance structure by assuming both the components as structured (CS or AR(1)) which is useful for spatio-temporal repeated measurements. For example, for modeling the covariance of multivariate environmental monitoring data obtained repeatedly over time and space, or for modeling covariance structure of glucose measurement at 15 different regions ($p = 15$) in both the hemispheres ($q = 2$) of the brain (Worsley et al., 1991). All these authors used likelihood ratio test (LRT) statistic for testing various permutations of patterns of 2−separable covariance structures. Among these authors Lu and Zimmerman (2005) and Roy and Leiva (2008) have used unbiased/unmodified LRT, and simulations are used to build its sampling distribution and find quantiles. Others worked on biased LRT, based on the theoretical chi-square null distribution; in this case the rejection rate of null hypothesis is not equal to the nominal Type I error when the null hypothesis is true. It is worthwhile to mention here that using biased LRT, *MIXED* procedure of SAS Software (SAS Institute Inc., 2009) can test the hypotheses for 2−separable covariance structure with the first component as CS or AR(1) correlation or unstructured covariance structures. Therefore, we see that hypotheses tests for separable structures are a well developed area, and biased and unbiased/unmodified LRTs are available.

Several authors also proposed unbiased/modified LRT statistic in which the test statistic is modified in order to match the theoretical chi-square distribution to test the separability of variance-covariance structure. Mitchell *et al.* (2006) derived a modified LRT statistic to test the separability of a covariance matrix using the ratio of the mean of the LRT to the asymptotic mean to estimate the critical value of the distribution of the LRT statistic for two-level data. Simpson (2010) proposed an adjusted likelihood ratio test of separability for unbalanced two-level multivariate data using the technique proposed by Mitchell *et al.* (2006). He also suggested another less conservative and more straightforward adjustment in his paper. Simpson (2010) as well addressed the particular case where the within subject correlation decreases exponentially in both levels. Very recently Manceur and Dutilleul (2013) presented an unbiased/modified LRT, based on penalty-based homothetic transformation of the LRT statistic, for separability of a variance-covariance structure, by multiplying the test statistic by a constant. This constant is estimated by simulation so that the distribution of the test statistic approaches chi-square even for small samples. At the core of their work was the finding that a simple homothetic transformation based on an optimal penalty was sufficient to modify the biased LRT statistics for separability, so the distributions of LRT statistics thus modified are χ^2 already for moderate sample sizes. Manceur and Dutilleul (2013) also calculated standardized empirical bias. However, since the estimation of unstructured $\mathbf{\Omega}$ is necessary in all these modified tests, the sample size n bigger than pq is required.

The LRT statistic is reliable with very large samples. Nevertheless, in the real-life applications we have only finite samples; small sample sizes are the norm because of limited measurement opportunities. One way of overcoming the problem of the accuracy of the asymptotic approximation under the null distribution of the unmodifed LRT statistic for testing 2-separable covariance structure for small or moderate sample sizes is to exploit the empirical null distribution (END) of the LRT statistic. Lu and Zimmerman (2005) and Roy and Leiva (2008) derived ENDs of the LRT statistics for testing $2-$separable covariance structure with both unstructured components and both structured components respectively. One can clearly see from these two articles that the ENDs of the LRT statistics are quite different from their limiting χ^2 distributions for small sample size with $n > pq$. Therefore, the LRT fails as a matter of practical use because its distribution is very different from its limiting χ^2 distribution for small samples; in addition, the LRT cannot be used for $n \leq pq$ for the general unstructured variance-covariance matrix as alternative hypothesis, a common problem for LRT. Nevertheless, researchers still use the theoretical chi square distribution even for small samples as exact tests are not available in such cases.

In many two-level multivariate data applications it is possible to model a dataset with $n \leq pq$ without testing for separability of the variance-covariance matrix by postulating the separability. For example, *MIXED* procedure of SAS can fit linear models, and Roy and Khattree's (2007a, b) classification rules can classify individuals with separable covariance structure when $n \leq pq$. It is commonly done in practice. However, before applying *MIXED* procedure of SAS, or Roy and Khattree's classification rules one must test whether the data have separable covariance structure. Unfortunately, all the above mentioned available unmodified LRT based tests or the modified LRT based tests need the assumption $n > pq$, which is often not possible in applied setting given the limitations on data collection. So, even if the methods are available for modeling data using separable structure when $n \leq pq$, the testing is not, which is the limiting factor of any statistical analysis for two-level data. However, Simpson et al. (2014) very recently provides a method in this context which avoids this limitation. We propose a different approach.

1.2 Proposed Tests

Rao's score test (RST) is an alternative or competitor to LRT; in this article we propose a new approach, an unmodified RST procedure, to test a $2-$separable covariance structure with the first component as a CS correlation matrix, which essentially means that all measurements for any characteristic within the same subject are equicorrelated. The biggest advantage of RST is that it only exploits the null hypothesis, and thus does not need the assumption $n > pq$ as LRT does. We compare the performances of this new RST procedure with unmodifed LRT procedure. When both components of the 2−separable covariance structure are unstructured, the RST requires a sample size $n > \max(p, q)$, which can be large for many repeated measures (p). However, when the first separable component is the CS correlation structure, RST only requires a sample size $n > q$, which is independent of the number of repeated measures. Given the increasing collection of multi-level data on which separability could be assessed, we develop a new method of testing separability of a covariance structure using RST when n is just greater than q, which is a substantial improvement over the LRT. This method will give the opportunity to many statistical practitioners and researchers to test the separability in small sample situation before applying the separability structure to their applications.

We perform simulation experiments to check the finite sample performance of both the RST and the LRT statistics, comparing a biased LRT to a biased RST, and an unbiased/unmodified LRT to an unbiased/unmodified RST. Both LRT and RST are equivalent to the first order of asymptotics, but differ to some extent in the second order properties; neither is uniformly superior to the other. Thus, empirical type I error is determined for both LRT and RST statistics to show that the biasedness of RST is much smaller than LRT for nominal significance level 0.01 as well as 0.05. Moreover, we derive the ENDs of the RST and LRT statistics, compare an unbiased/unmodified LRT to an unbiased/unmodified RST, and show that for small samples the END of the RST statistic gives much more reliable inference than the END of the LRT statistic. In other words, we show that the difference between the END of RST statistic and its limiting χ^2 distribution is much smaller than the difference between the END of LRT statistic and its limiting χ^2 distribution for any small or moderate sample size. We also derive ENDs of RST statistics for $q < n \leq pq$, the computation of which is not even possible for LRT statistics. The simplicity of the standard χ^2 test is convenient, but comes at potentially considerable cost because it differs substantially from the END especially for large number of repeated measurements (p). To show the performance of the ENDs for both RST and LRT statistics we perform simulation studies up to 15 repeated measurements with the number of variables as two and three.

This article is organized as follows. In Section 2 separability hypothesis of a covariance matrix for two-level data is introduced, and the formulation of the test statistics is presented. RST is defined in Section 3. Simulation studies are performed in Section 4 to calculate the observed Type I error rates, the ENDs of RST and LRT statistics and the power of the tests. Three real data examples to show the performance of our new proposed method, comparing the biased LRT to the biased RST and the unbiased/unmodified LRT to the unbiased/unmodified RST, are given in Section 5, and finally Section 6 summarizes with several comments along with the scope for the future research. Proofs of some basic results of matrix algebra which are needed in deriving the maximum likelihood estimates (MLEs) of the matrix parameters and the RST statistic in Section 3 are presented in A. Empirical percentiles of the null distributions of LRT and RST statistics for several combinations of p, q and n are presented in B.

2 Separability hypothesis of a covariance matrix

Let $\boldsymbol{X}_i$ for $i = 1, \ldots, n$ be the independent and identically distributed $(q \times p)-$dimensional observation matrices, measurements on q characteristics at p time points on ith individual. We assume $\boldsymbol{X}_i \sim N_{q,p}(\boldsymbol{M}, \boldsymbol{\Omega})$, i.e., $\mathrm{vec}\boldsymbol{X}_i \sim N_{pq}\left(\mathrm{vec}\boldsymbol{M}, \boldsymbol{\Omega}\right)$, where $\mathrm{vec}\boldsymbol{M} \in \mathbb{R}^{pq}$, $\mathrm{vec}(\cdot)$ is the operator stacking the columns of a $q \times p$ matrix into $pq \times 1$ dimensional vector, and $\boldsymbol{\Omega}$ is assumed to be a $pq \times pq$ positive definite matrix. We denote $\mathrm{vec}\boldsymbol{X}_i = \boldsymbol{x}_i$ and $\mathrm{vec}\boldsymbol{M} = \boldsymbol{\mu}$. We define the vector of unknown parameters $\boldsymbol{\theta} = \left(\boldsymbol{\mu}', \mathrm{vech}'\boldsymbol{\Omega}\right)'$, where $\mathrm{vech}(\cdot)$ is the operator stacking the columns of a $pq \times pq$ dimensional symmetric matrix into $(pq(pq+1)/2-$dimensional vector by eliminating all the supradiagonal elements. The number of unknown parameters to be estimated in $\boldsymbol{\Omega}$ is $pq(pq+1)/2$, which increases very rapidly with the increase of the dimension of either the number of characteristics q, or the number of time points p. Estimation of $\boldsymbol{\Omega}$ is impossible when the sample size $n \leq pq$. So, researchers usually rely on structured covariance matrices which depend on a smaller set of unknown parameters. The problem, though, is knowing what the structure is. A form of the covariance matrix $\boldsymbol{\Omega}$ suitable for doubly multivariate data or two-level data is a $2-$separable variance-covariance structure as follows:

$$\underset{pq\times pq}{\boldsymbol{\Omega}} = \underset{p\times p}{\boldsymbol{\Psi}} \otimes \underset{q\times q}{\boldsymbol{\Sigma}}, \tag{1}$$

where both $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are unstructured positive definite matrices and $\otimes$ represents the Kronecker product. The $q \times q$ matrix $\boldsymbol{\Sigma}$ represents the variance-covariance matrix of the q response variables at any given time point. We assume $\boldsymbol{\Sigma}$ does not depend on time and it is the same for all time points. The $p \times p$ matrix $\boldsymbol{\Psi}$ represents the variance-covariance matrix of the repeated measurements on a given characteristic and it is assumed to be the same for all characteristics as well. The number of unknown parameters to be estimated in the separable structure (1) is only $p(p+1)/2 + q(q+1)/2 - 1$ (with the first diagonal element of $\boldsymbol{\Psi}$ as one to circumvent the over-identifiability problem of $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ in $\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$, that is why the total number of parameters gets reduced by one) which is much less than $pq(pq+1)/2$ in an unstructured variance-covariance matrix $\boldsymbol{\Omega}$. Several authors, e.g., Boik (1991), Galecki (1994), Naik and Rao (2001), Chaganty and Naik (2002), and Roy and Khattree (2007a, b) have observed many advantages of using the separable covariance structure over the usual unstructured variance-covariance matrix for analyzing two-level multivariate data. In this article we consider CS correlation structure on $\boldsymbol{\Psi}$, so that the number of unknown parameters to be estimated further reduces to $1 + q(q+1)/2$.

We consider the RST and LRT for testing the separability of the variance-covariance matrix $\boldsymbol{\Omega}$ with half structured and half unstructured matrices, i.e.,

$$H_0 : \boldsymbol{\Omega} = \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma},\ \boldsymbol{\Psi}\ \text{CS} \quad \text{against} \quad H_A : \boldsymbol{\Omega}\ \text{unstructured}. \tag{2}$$

The matrices $\boldsymbol{\Psi}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are positive definite matrices. The variance-covariance matrix $\boldsymbol{\Sigma}$ is assumed to be unstructured. Given that $\boldsymbol{\Psi}$ has a CS correlation structure, it can be written as $\boldsymbol{\Psi} = (1-\rho)\boldsymbol{I}_p + \rho \mathbf{1}_p \mathbf{1}_p'$. Since $\boldsymbol{\Psi}$ is a positive definite matrix, we should have $-1/(p-1) < \rho < 1$. Note that $\boldsymbol{\Psi}$ is a CS correlation structure with only one unknown correlation parameter ρ. We choose $\boldsymbol{\Psi}$ a correlation matrix with all p diagonal elements as one, not a covariance matrix just to circumvent the identifiability problem of the $p \times p$ matrix $\boldsymbol{\Psi}$ and the $q \times q$ matrix $\boldsymbol{\Sigma}$. Thus, the $(p \times p)-$dimensional CS correlation matrix $\boldsymbol{\Psi}$ has only one parameter, and $(q \times q)-$dimensional unstructured variance-covariance matrix $\boldsymbol{\Sigma}$ has $q(q+1)/2$ parameters. Recently, there has been a discussion by Dutilleul and Roy in Lee *et al.* (2010) on the definition of ML estimators and their identifiability for $2-$separable variance-covariance structure. This problem is circumvented in this paper by choosing $\boldsymbol{\Psi}$ as a correlation matrix.

When the log-likelihood function is a smooth curve well approximated by a quadratic function, RST and LRT are identical under null hypothesis $\big($see Lemma 1 in Engle (1984), p. 782$\big)$. Rao (1984, p. 418; 2005) also mentioned that under the normality assumption RST and LRT statistics have the same asymptotic distribution χ^2_ν where the degrees of freedom (df) ν is equal to the number of parameters estimated under H_A minus the number of parameters estimated under H_0. Thus, in our separable structure set up (2) we have

$$\nu = \frac{pq(pq+1)}{2} - \frac{q(q+1)}{2} - 1. \tag{3}$$

One may see the testing of the H_0 in (2) as a following sequence of two hypotheses:

$$H_{01} : \boldsymbol{\Omega} = \boldsymbol{V} \otimes \boldsymbol{\Sigma},\ \boldsymbol{V}\ \text{UN} \quad \text{against} \quad H_{A1} : \boldsymbol{\Omega}\ \text{unstructured}, \tag{4}$$

assuming $\boldsymbol{V}$ as positive definite with df ν_1 as

$$\nu_1 = \frac{pq(pq+1)}{2} - \frac{q(q+1)}{2} - \frac{p(p+1)}{2} + 1,$$

and then, assuming the general separable structure H_{01} in (4) has been accepted, testing separability with one CS factor matrix against general separability as follows:

$$H_{02}: \boldsymbol{\Omega} = \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma},\ \boldsymbol{\Psi}\ \text{CS} \quad \text{against} \quad H_{A2}: \boldsymbol{V} \otimes \boldsymbol{\Sigma},\ \boldsymbol{V}\ \text{UN}, \tag{5}$$

with df ν_2 as

$$\nu_2 = \frac{p(p+1)}{2} - 2.$$

In other words, H_0 in (2), is equivalent to the test sequence

$$H_0 \equiv H_{02} \circ H_{01},$$

where '$\circ$' means 'after'. Thus, we see that H_{01} is the null hypothesis corresponding to the test of general separability and H_{02} is the null hypothesis corresponding to the test of separability with the first component as CS correlation structure. Now, the minimum sample size required to test H_{01} in (4) using LRT is $n_1 = pq + 1$ and the same for H_{02} in (5) using LRT is $n_2 = \max(p, q) + 1$, whereas, using RST the minimum sample sizes required to test the H_{01} in (4) and H_{02} in (5) are $n_1 = \max(p, q) + 1$ and $n_2 = q + 1$ respectively. Therefore, the minimum sample size needed to test $H_{02} \circ H_{01}$ using LRT is $pq + 1$ and the minimum sample size needed to test $H_{02} \circ H_{01}$ using RST is $\big(\max(p, q) + 1\big)$. Nevertheless, the minimum sample size needed to test H_0 using LRT is $pq + 1$ and the minimum sample size needed to test H_0 using RST is $q + 1$. Hence, instead of going straight to H_0, if we test the sequence $H_{02} \circ H_{01}$, the required sample size for RST is $\big(\max(p, q) + 1\big)$. So, if the number of repeated measurements p is large as in the examples in the Introduction and $n < p$, then the sequence of $H_{02} \circ H_{01}$ cannot be tested. However, if $n > p$, one can test the sequence of $H_{02} \circ H_{01}$ in those examples. So, when the sample size is small, testing the sequence $H_{02} \circ H_{01}$ may not be possible. H_{02} in (5) is likely to be used in applications when it is known apriori that the data already has the general separable structure.

The proposed RST in this paper is for the Hypothesis (2), and not for the equivalent test sequence $H_{02} \circ H_{01}$. Testing the equivalent test sequence $H_{02} \circ H_{01}$ would need new theoretical calculations and simulations. Thus, we discuss only the testing of Hypothesis (2) in the following sections.

3 Rao's score test statistic

Let us assume that the log-likelihood function $\ln L(\boldsymbol{\mu}, \boldsymbol{\Omega}; \boldsymbol{X})$ with the data matrix $\boldsymbol{X}$, where $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{pq,n}$, is partially differentiable with respect to each coordinate of the parameter vector $(\boldsymbol{\mu}', \text{vech}'\boldsymbol{\Omega})'$ for every data matrix $\boldsymbol{X}$. Now we derive the expressions of the LRT and the RST statistics for testing the Hypothesis (2).

The LRT is based upon the difference between the maximum of the log-likelihood under the null and under the alternative hypotheses. The likelihood ratio Λ can be written as

$$\Lambda = \frac{\max_{H_0} L}{\max_{H_A} L}.$$

It is well known that, for large sample size and under normality assumption, the LRT statistic $-2\ln\Lambda$ is approximately distributed as χ^2_ν under H_0. The degrees of freedom ν is given in (3). It is to be noted that if any of the covariance parameters fall on the boundary of their parameter space then the asymptotic distribution of $-2\ln\Lambda$ becomes a mixture of χ^2 distributions as discussed in Self and Liang (1987). The Hypothesis (2) using LRT is discussed thoroughly in Roy and Khattree (2005a). In this paper we will derive and discuss the same Hypothesis (2) using RST.

Let $\boldsymbol{s}(\boldsymbol{\theta}) = (\boldsymbol{s}_1(\boldsymbol{\theta})', \boldsymbol{s}_2(\boldsymbol{\theta})')' = \left(\dfrac{\partial \ln L}{\partial \boldsymbol{\mu}'}, \dfrac{\partial \ln L}{\partial \mathrm{vech}'\boldsymbol{\Omega}}\right)'$ be the score vector. Then the Fisher information matrix can be defined as

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}) \quad = \quad -\mathrm{E}\left[\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] \overset{df}{=} \begin{pmatrix} \boldsymbol{\mathcal{F}}_{11} & \boldsymbol{\mathcal{F}}_{12} \\ \boldsymbol{\mathcal{F}}_{12}' & \boldsymbol{\mathcal{F}}_{22} \end{pmatrix}, \tag{6}$$

where $\boldsymbol{\mathcal{F}}_{11}$, $\boldsymbol{\mathcal{F}}_{12}$ and $\boldsymbol{\mathcal{F}}_{22}$ are $pq \times pq$, $pq \times pq(pq+1)/2$ and $pq(pq+1)/2 \times pq(pq+1)/2$ matrices respectively. Let the Fisher information matrix exist and be invertible. The Rao's score (RS)

$$\boldsymbol{s}(\widehat{\boldsymbol{\theta}})'\boldsymbol{\mathcal{F}}^{-1}(\widehat{\boldsymbol{\theta}})\boldsymbol{s}(\widehat{\boldsymbol{\theta}}), \tag{7}$$

where $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\mu}}', \mathrm{vech}'(\widehat{\boldsymbol{\Psi}} \otimes \widehat{\boldsymbol{\Sigma}})\right)'$ is the MLE of $\boldsymbol{\theta}$ under the null hypothesis H_0, is defined as the RST statistic. This statistic is also approximately distributed as χ^2_ν with the same degrees of freedom ν given in (3). We now obtain the expression of the RST statistic to test the null hypothesis H_0 in the following section.

Let us define the centered form of the data matrix $\boldsymbol{Y}(\boldsymbol{\mu})$ as $\boldsymbol{Y}(\boldsymbol{\mu}) = \boldsymbol{X} - \boldsymbol{1}_n' \otimes \boldsymbol{\mu}$. Then

$$\boldsymbol{Y}(\boldsymbol{\mu}) \sim N_{pq,n}(\boldsymbol{0}, \boldsymbol{\Omega}, \boldsymbol{I}_n),$$

which means

$$\mathrm{vec}\boldsymbol{Y}(\boldsymbol{\mu}) \sim N_{npq}(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Omega}).$$

The log-likelihood function in terms of this centered data matrix $\boldsymbol{Y}(\boldsymbol{\mu})$ can be written as

$$\ln L(\boldsymbol{\mu}, \boldsymbol{\Omega}; \boldsymbol{X}) = -\frac{npq}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\mathrm{vec}'\boldsymbol{Y}(\boldsymbol{\mu})(\boldsymbol{I}_n \otimes \boldsymbol{\Omega}^{-1})\mathrm{vec}\boldsymbol{Y}(\boldsymbol{\mu}). \tag{8}$$

In order to determine the score vector, we first differentiate the above log-likelihood function $\ln L$

with respect to $\boldsymbol{\mu}$. We get

$$
\begin{aligned}
\frac{\partial \ln L}{\partial \boldsymbol{\mu}'} &= -\frac{1}{2} \cdot \frac{\partial \mathrm{vec}'\boldsymbol{Y}(\boldsymbol{\mu})(\boldsymbol{I}_n \otimes \boldsymbol{\Omega}^{-1})\mathrm{vec}\boldsymbol{Y}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}'} \\
&= -\frac{1}{2} \cdot \frac{\partial \mathrm{vec}'\boldsymbol{Y}(\boldsymbol{\mu})(\boldsymbol{I}_n \otimes \boldsymbol{\Omega}^{-1})\mathrm{vec}\boldsymbol{Y}(\boldsymbol{\mu})}{\partial \mathrm{vec}'\boldsymbol{Y}(\boldsymbol{\mu})} \cdot \frac{\partial \mathrm{vec}\boldsymbol{Y}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}'} \\
&= -\frac{1}{2} \cdot 2 \cdot \mathrm{vec}'\boldsymbol{Y}(\boldsymbol{\mu})(\boldsymbol{I}_n \otimes \boldsymbol{\Omega}^{-1}) \cdot \left(-\frac{\partial \mathrm{vec}(\boldsymbol{1}_n' \otimes \boldsymbol{\mu})}{\partial \boldsymbol{\mu}'} \right) \\
&= \mathrm{vec}'\boldsymbol{Y}(\boldsymbol{\mu})(\boldsymbol{1}_n \otimes \boldsymbol{\Omega}^{-1}), \qquad (9)
\end{aligned}
$$

which is a $pq-$dimensional row vector. Now, we differentiate $\ln L$ given in (8) with respect to $\mathrm{vech}\boldsymbol{\Omega}$. Using Proposition 3 (ii), Proposition 1 (iii) in A and the symmetricity of $\boldsymbol{\Omega}$ we get

$$
\begin{aligned}
\frac{\partial \ln L}{\partial \mathrm{vech}'\boldsymbol{\Omega}} &= \frac{\partial \ln L}{\partial \mathrm{vec}'\boldsymbol{\Omega}} \cdot \frac{\partial \mathrm{vec}'\boldsymbol{\Omega}}{\partial \mathrm{vech}'\boldsymbol{\Omega}} = \frac{\partial \ln L}{\partial \mathrm{vec}'\boldsymbol{\Omega}} \cdot \boldsymbol{D}_{pq} \\
&= -\frac{n}{2} \frac{\partial \ln |\boldsymbol{\Omega}|}{\partial |\boldsymbol{\Omega}|} \cdot \frac{\partial |\boldsymbol{\Omega}|}{\partial \mathrm{vec}'\boldsymbol{\Omega}} \cdot \boldsymbol{D}_{pq} \\
&\quad -\frac{1}{2} \frac{\partial \mathrm{vec}'\boldsymbol{Y}(\boldsymbol{\mu})(\boldsymbol{I}_n \otimes \boldsymbol{\Omega}^{-1})\mathrm{vec}\boldsymbol{Y}(\boldsymbol{\mu})}{\partial \mathrm{vec}'(\boldsymbol{I}_n \otimes \boldsymbol{\Omega}^{-1})} \cdot \frac{\partial \mathrm{vec}(\boldsymbol{I}_n \otimes \boldsymbol{\Omega}^{-1})}{\partial \mathrm{vec}'\boldsymbol{\Omega}^{-1}} \cdot \frac{\partial \mathrm{vec}\boldsymbol{\Omega}^{-1}}{\partial \mathrm{vec}'\boldsymbol{\Omega}} \cdot \boldsymbol{D}_{pq} \\
&= -\frac{n}{2} \mathrm{vec}'\boldsymbol{\Omega}^{-1}\boldsymbol{D}_{pq} \\
&\quad + \frac{1}{2} (\mathrm{vec}'\boldsymbol{Y}(\boldsymbol{\mu}) \otimes \mathrm{vec}'\boldsymbol{Y}(\boldsymbol{\mu}))(\boldsymbol{I}_n \otimes \boldsymbol{K}_{pq,n} \otimes \boldsymbol{I}_{pq})(\mathrm{vec}\boldsymbol{I}_n \otimes \boldsymbol{I}_{p^2q^2})(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\boldsymbol{D}_{pq} \\
&= -\frac{n}{2} \mathrm{vec}'\boldsymbol{\Omega}^{-1}\boldsymbol{D}_{pq} + \frac{1}{2} \mathrm{vec}'(\boldsymbol{Y}(\boldsymbol{\mu})\boldsymbol{Y}'(\boldsymbol{\mu}))(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\boldsymbol{D}_{pq}, \qquad (10)
\end{aligned}
$$

which is a $pq(pq+1)/2-$dimensional row vector. Now, the log-likelihood $\ln L$ is maximized at a value $\widehat{\boldsymbol{\theta}}$ when $\dfrac{\partial \ln L}{\partial \boldsymbol{\theta}'} = \boldsymbol{0}$. It is easy to see from (9) that the MLE of $\boldsymbol{\mu}$ is

$$
\widehat{\boldsymbol{\mu}} = \frac{1}{n} \boldsymbol{X} \boldsymbol{1}_n.
$$

Now, substituting the value of $\widehat{\boldsymbol{\mu}}$ in $\boldsymbol{Y}(\boldsymbol{\mu})$ we get

$$
\begin{aligned}
\mathrm{vec}\boldsymbol{Y}(\widehat{\boldsymbol{\mu}}) &= \mathrm{vec}\left(\boldsymbol{X} - \boldsymbol{1}_n' \otimes \frac{1}{n}\boldsymbol{X}\boldsymbol{1}_n \right) = \mathrm{vec}\boldsymbol{X} - \left(\boldsymbol{1}_n \otimes \left(\frac{1}{n}\boldsymbol{1}_n' \otimes \boldsymbol{I}_{pq} \right) \mathrm{vec}\boldsymbol{X} \right) \\
&= (\boldsymbol{Q}_{\boldsymbol{1}_n} \otimes \boldsymbol{I}_{pq})\mathrm{vec}\boldsymbol{X} = \mathrm{vec}\left(\boldsymbol{X}\boldsymbol{Q}_{\boldsymbol{1}_n} \right).
\end{aligned}
$$

We now derive the expression of the RST statistic for H_0. We start with the calculations of the four component matrices in the Fisher information matrix $\boldsymbol{\mathcal{F}}(\boldsymbol{\theta})$ in (6). We have

$$\begin{aligned}
\boldsymbol{\mathcal{F}}_{11} &= -\mathrm{E}\left[\frac{\partial}{\partial\boldsymbol{\mu}'}\left(\frac{\partial \ln L}{\partial\boldsymbol{\mu}'}\right)\right] = -\mathrm{E}\left[\frac{\partial(\mathbf{1}_n'\otimes\boldsymbol{\Omega}^{-1})\mathrm{vec}\boldsymbol{Y}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}'}\right] \\
&= -\mathrm{E}\left[\frac{\partial(\mathbf{1}_n'\otimes\boldsymbol{\Omega}^{-1})\mathrm{vec}\boldsymbol{Y}(\boldsymbol{\mu})}{\partial\mathrm{vec}'\boldsymbol{Y}(\boldsymbol{\mu})}\cdot\frac{\partial\mathrm{vec}\boldsymbol{Y}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}'}\right] \\
&= -(\mathbf{1}_n'\otimes\boldsymbol{\Omega}^{-1})\mathrm{E}\left[\frac{\partial\mathrm{vec}(\boldsymbol{X}-\mathbf{1}_n'\otimes\boldsymbol{\mu})}{\partial\boldsymbol{\mu}'}\right] \\
&= (\mathbf{1}_n'\otimes\boldsymbol{\Omega}^{-1})(\boldsymbol{I}_n\otimes\boldsymbol{K}_{1,1}\otimes\boldsymbol{I}_{pq})(\mathbf{1}_n\otimes\boldsymbol{I}_{qp}) \\
&= n\boldsymbol{\Omega}^{-1}, \qquad (11)
\end{aligned}$$

$$\begin{aligned}
\text{and}\quad \boldsymbol{\mathcal{F}}_{12} &= \boldsymbol{\mathcal{F}}_{12}' = -\mathrm{E}\left[\frac{\partial}{\partial\mathrm{vech}'\boldsymbol{\Omega}}\left(\frac{\partial \ln L}{\partial\boldsymbol{\mu}'}\right)\right] = -\mathrm{E}\left[\frac{\partial(\mathbf{1}_n'\otimes\boldsymbol{\Omega}^{-1})\mathrm{vec}\boldsymbol{Y}(\boldsymbol{\mu})}{\partial\mathrm{vech}'\boldsymbol{\Omega}}\right] \\
&= -\mathrm{E}\left[\frac{\partial(\mathbf{1}_n'\otimes\boldsymbol{\Omega}^{-1})\mathrm{vec}\boldsymbol{Y}(\boldsymbol{\mu})}{\partial\mathrm{vec}'(\mathbf{1}_n'\otimes\boldsymbol{\Omega}^{-1})}\cdot\frac{\partial\mathrm{vec}(\mathbf{1}_n'\otimes\boldsymbol{\Omega}^{-1})}{\partial\mathrm{vec}'\boldsymbol{\Omega}^{-1}}\cdot\frac{\partial\mathrm{vec}\boldsymbol{\Omega}^{-1}}{\partial\mathrm{vec}'\boldsymbol{\Omega}}\cdot\boldsymbol{D}_{pq}\right] \\
&= \mathrm{E}\left[\big(\mathrm{vec}'\boldsymbol{Y}(\boldsymbol{\mu})\otimes\boldsymbol{I}_{pq}\big)(\boldsymbol{I}_n\otimes\boldsymbol{K}_{pq,1}\otimes\boldsymbol{I}_{pq})(\mathbf{1}_n\otimes\boldsymbol{I}_{p^2q^2})(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\right]\boldsymbol{D}_{pq} \\
&= \mathbf{0}. \qquad (12)
\end{aligned}$$

Now by denoting $\boldsymbol{H} = \mathrm{vec}\big(\boldsymbol{Y}(\boldsymbol{\mu})\boldsymbol{Y}'(\boldsymbol{\mu})\big)$ and then using the expression (6.5) of Ghazal and Neudecker (2000, p. 81): $\mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}'] = \mathrm{tr}[\boldsymbol{V}]\boldsymbol{U}+\boldsymbol{M}\boldsymbol{M}'$ for $\mathrm{vec}\boldsymbol{Z}\sim N_{kl}(\mathrm{vec}\boldsymbol{M},\boldsymbol{V}\otimes\boldsymbol{U})$, we get

$$\mathrm{E}[\boldsymbol{H}] = \big(\mathrm{tr}[\boldsymbol{I}_n]\big)(\mathrm{vec}\boldsymbol{\Omega}) = n(\mathrm{vec}\boldsymbol{\Omega}).$$

Using this result, the Equation (10) and the Lemma 1 in A we can write

$$\begin{aligned}
\boldsymbol{\mathcal{F}}_{22} &= -\mathrm{E}\left[\frac{\partial}{\partial\mathrm{vech}'\boldsymbol{\Omega}}\left(\frac{\partial \ln L}{\partial\mathrm{vech}'\boldsymbol{\Omega}}\right)\right] \\
&= \frac{n}{2}\boldsymbol{D}_{pq}'\cdot\frac{\partial\mathrm{vec}\boldsymbol{\Omega}^{-1}}{\partial\mathrm{vec}'\boldsymbol{\Omega}}\cdot\boldsymbol{D}_{pq}-\frac{1}{2}\mathrm{E}\left[\frac{\partial\boldsymbol{D}_{pq}'(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{H}}{\partial\mathrm{vec}'\boldsymbol{\Omega}}\right]\boldsymbol{D}_{pq} \\
&= -\frac{n}{2}\boldsymbol{D}_{pq}'(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{D}_{pq} \\
&\quad +\frac{1}{2}(\mathrm{E}[\boldsymbol{H}]\otimes\boldsymbol{D}_{pq})'(\boldsymbol{I}_{pq}\otimes\boldsymbol{K}_{pq,pq}\otimes\boldsymbol{I}_{pq})(\boldsymbol{I}_{p^2q^2}\otimes\mathrm{vec}\boldsymbol{\Omega}^{-1}+\mathrm{vec}\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{I}_{p^2q^2})(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{D}_{pq} \\
&= \frac{n}{2}\boldsymbol{D}_{pq}'(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{D}_{pq}. \qquad (13)
\end{aligned}$$

Now, substituting the values of $\boldsymbol{\mathcal{F}}_{11}$, $\boldsymbol{\mathcal{F}}_{12}$ and $\boldsymbol{\mathcal{F}}_{22}$ from (11), (12) and (13) in (6) we get the Fisher information matrix as

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}) = \begin{pmatrix} n\boldsymbol{\Omega}^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{n}{2}\boldsymbol{D}_{qp}'(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{D}_{qp} \end{pmatrix}.$$

Then, applying the Proposition 1.3.3. of Kollo and von Rosen (2005) and the Proposition 1 (iv) in

A we get the inverse of the Fisher information matrix as follows:

$$\begin{aligned}
\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}) &= \begin{pmatrix} n\boldsymbol{\Omega}^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{n}{2}\boldsymbol{D}'_{qp}(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{D}_{qp} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \frac{1}{n}\boldsymbol{\Omega} & \mathbf{0} \\ \mathbf{0} & \left(\frac{n}{2}\boldsymbol{D}'_{qp}(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{D}_{qp}\right)^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{n}\boldsymbol{\Omega} & \mathbf{0} \\ \mathbf{0} & \frac{2}{n}\boldsymbol{D}^{+}_{qp}(\boldsymbol{\Omega}\otimes\boldsymbol{\Omega})\boldsymbol{D}^{+'}_{qp} \end{pmatrix}. \qquad (14)
\end{aligned}$$

Since the hypothesis we are interested in this article involves only the variance-covariance matrix, we derive the RST statistic corresponding to vech$\boldsymbol{\Omega}$. Now, the component of the score vector corresponding to vech$\boldsymbol{\Omega}$ is

$$\begin{aligned}
s_2(\boldsymbol{\theta}) &= -\frac{n}{2}\boldsymbol{D}'_{pq}\text{vec}\boldsymbol{\Omega}^{-1} + \frac{1}{2}\boldsymbol{D}'_{pq}(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\text{vec}(\boldsymbol{Y}(\boldsymbol{\mu})\boldsymbol{Y}'(\boldsymbol{\mu})) \\
&\overset{df}{=} s_{21}(\boldsymbol{\theta}) + s_{22}(\boldsymbol{\theta}),
\end{aligned}$$

where

$$\begin{aligned}
s_{21}(\boldsymbol{\theta}) &= -\frac{n}{2}\boldsymbol{D}'_{pq}(\text{vec}\boldsymbol{\Omega}^{-1}), \\
\text{and} \quad s_{22}(\boldsymbol{\theta}) &= \frac{1}{2}\boldsymbol{D}'_{pq}(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\text{vec}(\boldsymbol{Y}(\boldsymbol{\mu})\boldsymbol{Y}'(\boldsymbol{\mu})).
\end{aligned}$$

Let $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\Sigma}}$ be the ML estimators under the null hypothesis. Therefore, the estimator $\widehat{\boldsymbol{\theta}}$ under the null hypothesis H_0 is

$$\widehat{\boldsymbol{\theta}} = \begin{pmatrix} \widehat{\boldsymbol{\mu}} \\ \text{vech}(\widehat{\boldsymbol{\Psi}}\otimes\widehat{\boldsymbol{\Sigma}}) \end{pmatrix}.$$

For MLEs $\widehat{\boldsymbol{\Sigma}}$, and $\widehat{\boldsymbol{\Psi}}$ or $\widehat{\rho}$ see Equations (3) and (6) in Roy and Khattree (2005a). These two equations in Roy and Khattree (2005a) are analytically intractable, and should be solved simultaneously and iteratively to get the MLEs $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Psi}}$; see Roy and Khattree (2005a, p. 301) for the algorithm to solve the Equations (3) and (6) in their paper. The SAS code for the algorithm is available from Roy's website. Now substituting the expression $\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta})$ from (14) in (7) we write the RST statistic or Rao's score (RS) for the null hypothesis H_0 as the sum of four components due to $s_{21}(\widehat{\boldsymbol{\theta}})$, $s_{22}(\widehat{\boldsymbol{\theta}})$ and $\boldsymbol{\mathcal{F}}_{22}^{-1}$ as

$$\begin{aligned}
\text{RS} &= \frac{2}{n}s_2'(\widehat{\boldsymbol{\theta}})\boldsymbol{D}^{+}_{pq}(\widehat{\boldsymbol{\Omega}}\otimes\widehat{\boldsymbol{\Omega}})\boldsymbol{D}^{+'}_{pq}s_2(\widehat{\boldsymbol{\theta}}) \\
&= \text{RS}_{11} + \text{RS}_{12} + \text{RS}_{21} + \text{RS}_{22}, \qquad (15)
\end{aligned}$$

where the notation $\widehat{\boldsymbol{\Omega}}$ is used to represent $\widehat{\boldsymbol{\Psi}}\otimes\widehat{\boldsymbol{\Sigma}}$. Now, from the symmetry of the quadratic form we have $\text{RS}_{21} = \text{RS}_{12}$. We will now evaluate the components in (15) one by one using Proposition 2 in A to get an expression of RS in terms of $\widehat{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\Sigma}}$. We have

$$\begin{aligned}
\mathrm{RS}_{11} &= \frac{2}{n} s_{21}'(\widehat{\boldsymbol{\theta}}) \boldsymbol{D}_{pq}^{+}(\widehat{\boldsymbol{\Omega}} \otimes \widehat{\boldsymbol{\Omega}}) \boldsymbol{D}_{pq}^{+'} s_{21}(\widehat{\boldsymbol{\theta}}) \\
&= \frac{n}{2} \mathrm{vec}' \widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{N}_{pq}(\widehat{\boldsymbol{\Omega}} \otimes \widehat{\boldsymbol{\Omega}}) \boldsymbol{N}_{pq}(\mathrm{vec} \widehat{\boldsymbol{\Omega}}^{-1}) \\
&= \frac{n}{2} \mathrm{vec}' \widehat{\boldsymbol{\Omega}}^{-1} (\widehat{\boldsymbol{\Omega}} \otimes \widehat{\boldsymbol{\Omega}}) \mathrm{vec} \widehat{\boldsymbol{\Omega}}^{-1} \\
&= \frac{npq}{2},
\end{aligned} \tag{16}$$

$$\begin{aligned}
\mathrm{RS}_{12} &= \frac{2}{n} s_{21}'(\widehat{\boldsymbol{\theta}}) \boldsymbol{D}_{pq}^{+}(\widehat{\boldsymbol{\Omega}} \otimes \widehat{\boldsymbol{\Omega}}) \boldsymbol{D}_{pq}^{+'} s_{22}(\widehat{\boldsymbol{\theta}}) \\
&= -\frac{1}{2} \mathrm{vec}' \widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{N}_{pq}(\widehat{\boldsymbol{\Omega}} \otimes \widehat{\boldsymbol{\Omega}}) \boldsymbol{N}_{pq}(\widehat{\boldsymbol{\Omega}}^{-1} \otimes \widehat{\boldsymbol{\Omega}}^{-1}) \mathrm{vec}(\boldsymbol{Y}(\widehat{\boldsymbol{\mu}}) \boldsymbol{Y}'(\widehat{\boldsymbol{\mu}})) \\
&= -\frac{1}{2} \mathrm{vec}' \widehat{\boldsymbol{\Omega}}^{-1} \mathrm{vec}(\boldsymbol{X} \boldsymbol{Q}_{\mathbf{1}_n} \boldsymbol{X}') \\
&= -\frac{1}{2} \mathrm{tr} \left[\widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{X} \boldsymbol{Q}_{\mathbf{1}_n} \boldsymbol{X}' \right] \\
&= -\frac{1}{2} \mathrm{tr} \left[(\widehat{\boldsymbol{\Psi}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) \boldsymbol{X} \boldsymbol{Q}_{\mathbf{1}_n} \boldsymbol{X}' \right],
\end{aligned} \tag{17}$$

and

$$\begin{aligned}
\mathrm{RS}_{22} &= \frac{2}{n} s_{22}'(\widehat{\boldsymbol{\theta}}) \boldsymbol{D}_{pq}^{+}(\widehat{\boldsymbol{\Omega}} \otimes \widehat{\boldsymbol{\Omega}}) \boldsymbol{D}_{pq}^{+'} s_{22}(\widehat{\boldsymbol{\theta}}) \\
&= \frac{1}{2n} \mathrm{vec}'(\boldsymbol{Y}(\widehat{\boldsymbol{\mu}}) \boldsymbol{Y}'(\widehat{\boldsymbol{\mu}}))(\widehat{\boldsymbol{\Omega}}^{-1} \otimes \widehat{\boldsymbol{\Omega}}^{-1}) \boldsymbol{N}_{pq}(\widehat{\boldsymbol{\Omega}} \otimes \widehat{\boldsymbol{\Omega}}) \boldsymbol{N}_{pq}(\widehat{\boldsymbol{\Omega}}^{-1} \otimes \widehat{\boldsymbol{\Omega}}^{-1}) \mathrm{vec}(\boldsymbol{Y}(\widehat{\boldsymbol{\mu}}) \boldsymbol{Y}'(\widehat{\boldsymbol{\mu}})) \\
&= \mathrm{vec}'(\boldsymbol{X} \boldsymbol{Q}_{\mathbf{1}_n} \boldsymbol{X}')(\widehat{\boldsymbol{\Omega}}^{-1} \otimes \widehat{\boldsymbol{\Omega}}^{-1}) \mathrm{vec}(\boldsymbol{X} \boldsymbol{Q}_{\mathbf{1}_n} \boldsymbol{X}') \\
&= \frac{1}{2n} \mathrm{tr} \left[(\widehat{\boldsymbol{\Psi}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) \boldsymbol{X} \boldsymbol{Q}_{\mathbf{1}_n} \boldsymbol{X}' (\widehat{\boldsymbol{\Psi}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) \boldsymbol{X} \boldsymbol{Q}_{\mathbf{1}_n} \boldsymbol{X}' \right].
\end{aligned} \tag{18}$$

Now, after substituting the values of $\mathrm{RS}_{11}, \mathrm{RS}_{12}, \mathrm{RS}_{21}$ and RS_{22} from (16), (17) and (18) in (15), we get the RST statistic or RS to test the null hypothesis H_0 as

$$\begin{aligned}
\mathrm{RS} &= \frac{nqp}{2} - \mathrm{tr} \left[(\widehat{\boldsymbol{\Psi}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) \boldsymbol{X} \boldsymbol{Q}_{\mathbf{1}_n} \boldsymbol{X}' \right] + \\
&\quad + \frac{1}{2n} \mathrm{tr} \left[(\widehat{\boldsymbol{\Psi}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) \boldsymbol{X} \boldsymbol{Q}_{\mathbf{1}_n} \boldsymbol{X}' (\widehat{\boldsymbol{\Psi}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) \boldsymbol{X} \boldsymbol{Q}_{\mathbf{1}_n} \boldsymbol{X}' \right],
\end{aligned}$$

which has an asymptotic χ^2 distribution with ν degrees of freedom, where ν is given in (3).

Remark 1. *Observe that the above* RST *statistic depends only on the data matrix* $\boldsymbol{X}$*, the ML estimate of* $\boldsymbol{\Psi}$ *with an explicit expression in* ρ*, and the ML estimate of the* $q \times q$ *dimensional variance-covariance matrix* $\boldsymbol{\Sigma}$*. The* RST *statistic does not need the ML estimate of the* $pq \times pq$ *dimensional unstructured variance covariance matrix* $\boldsymbol{\Omega}$*, as does the* LRT *statistic. Thus, the minimum number of observations needed to calculate the* RST *statistic is only* $q + 1$*, a quantity independent of* p*, whereas the minimum number of observations needed to calculate the* LRT *statistic is* $pq + 1$*, as it* *depends on the ML estimate of the $pq \times pq$ dimensional unstructured variance-covariance matrix* $\mathbf{\Omega}$*, which can grow very fast with the increase in* p*. Thus, the* RST *is an huge improvement over the* LRT*: one can test the null hypothesis* H_0 *with only* $q+1$ *observations using RST.*

4 Simulation Study

We perform simulation studies to compare the LRT and RST for the biased and unbiased/unmodified approaches. These simulations use two-level multivariate data assuming the 2−separable covariance structure $\mathbf{\Psi} \otimes \mathbf{\Sigma}$ on each subject, where $\mathbf{\Psi}$ has a CS correlation structure with correlation coefficient ρ on the repeated measurements for each characteristic. First, to compare the biased LRT with biased RST, we compute the observed Type I error rates, $\widehat{\alpha}$ to measure their biasedness. In other words, we would like to see for which n the observed $\widehat{\alpha}$ approaches the nominal α level for different values of p, the number of repeated measurements, and for different values of ρ, when the nominal Type I error rate $\alpha = 0.01$. Second, we compare the unbiased/unmodified LRT and RST using empirical null distributions (ENDs). Suppose the RST to contrast the Hypothesis (2) has the rejection region $\{RST > \kappa_\alpha\}$, and κ_α is chosen so that the test has significance level $\alpha = 0.01$ or 0.05. The significance level (or Type I error) for the RST for Hypothesis (2) is defined as

$$\alpha = P(\text{reject } \ H_0 \ \text{ when } \ H_0 \ \text{ is true}) = P_{\substack{\mathbf{\Psi} \otimes \mathbf{\Sigma} \\ \mathbf{\Psi}\text{C.S.}}}(\text{RST statistic} > \kappa_\alpha), \tag{19}$$

so, κ_α is the $100(1-\alpha)\%$ quantile of the RST under null hypothesis. Note that the distribution of RST is not known either under H_0 or any model under H_A. But we can compute this test (i.e., compute κ_α) and its properties using Monte Carlo simulation (Rizzo, 2008). To reduce the Monte Carlo error one needs to increase the simulation size. Let $\widehat{\kappa}_\alpha$ is an estimate of κ_α. We denote the observed Type I error rates for LRT and RST statistics as $\widehat{\alpha}_{\text{LRT}}$ and $\widehat{\alpha}_{\text{RST}}$ respectively, and calculate them when the nominal significance level $\alpha = 0.01$ as well as $\alpha = 0.05$, for different values of n, p and ρ in the following section.

4.1 Observed Type I error rates for biased LRT and RST statistics

Samples of various sizes from small to large, e.g., $n = 10, 15, 20, 25, 30, 50, 75$ and 100 are generated from a pq−variate normal population $\text{N}_{pq}(\mathbf{0}, \mathbf{\Psi} \otimes \mathbf{\Sigma})$. The number of repeated measurements p is chosen as 3, 4, 5 and 7, and the number of characteristics q as 3. The (3×3)−dimensional variance-covariance matrix $\mathbf{\Sigma}$ is taken as

$$\mathbf{\Sigma} = \begin{pmatrix} 2 & 1 & 2 \\ 1 & 4 & 3 \\ 2 & 3 & 5 \end{pmatrix},$$

and the correlation coefficient ρ of repeated measurements in the CS correlation structure $\mathbf{\Psi}$ is chosen as -0.4, -0.2, -0.1, 0.3, 0.5, 0.7 and 0.9 for each p such that $-\frac{1}{p-1} < \rho < 1$. We generate 50,000 samples for each combination of the parameters under H_0. Table 1 shows the empirical Type I error rates for both LRT and RST statistics, with all combinations of n, p and ρ for the nominal Type I error rate $\alpha = 0.01$ for 50,000 simulations. For $p = 3, 4, 5$ and 7 both LRT and RST statistics are approximately distributed as χ^2_ν under H_0, with degrees of freedom $\nu = 38, 71, 113$ and 224 respectively using (3). The observed Type I error rates for all the statistics appear to increase with p for a fixed ρ, which is manifested by the uniformly larger Type I error rates for $p = 7$. We also notice that Type I errors do not substantially change with ρ for LRT as well as for RST statistics and it does not depend on $\mathbf{\Sigma}$.

As expected the Type I error rates do decrease with the sample size n. Notice that RST performs much better for small and moderate n than its counterpart LRT. For $p = 4$ the Type I error rate 0.01 is achieved only for sample size 25 in the case of RST. For each combination of n, p and ρ we see that $\widehat{\alpha}_{\mathrm{RST}}$ is always much less than $\widehat{\alpha}_{\mathrm{LRT}}$. It is clear from Table 1 that $\widehat{\alpha}_{\mathrm{RST}}$ is approximately equal to the nominal significance level 0.01 for small and moderate sample sizes, which is not the case for $\widehat{\alpha}_{\mathrm{LRT}}$. A sample size of about 75 is required for $p = 5$ so that the empirical Type I error rate is approximately equal to the nominal significance level 0.01 in the case of RST, but a sample size of about 200 is required for $p = 5$ in the case of LRT for the same scenario. For $p = 7$, the empirical significance level decreases very slowly and even at $n = 200$ it is not close to the nominal significance level $\alpha = 0.01$ in the case of LRT; see Roy and Khattree (2005a) for detail. This suggests that LRT may not perform well when p is large, even for the large sample sizes. In contrast, a sample size of about 100 is required for $p = 7$, so that the empirical Type I error rate is approximately equal to the nominal significance level 0.01 in the case of RST. Thus, we see that RST performs much better than LRT in the small to moderate sample cases.

Since Type I errors do not substantially change with ρ for LRT as well as for RST statistics as noticed in Table 1, we also calculate $\widehat{\alpha}_{\mathrm{LRT}}$ and $\widehat{\alpha}_{\mathrm{RST}}$ when the nominal significance level $\alpha = 0.05$, for $\rho = 0.05$ and for various values of n and p for 50,000 simulations. The results are presented in Table 2.

Table 1: Observed Type I error rates $\widehat{\alpha}_{\text{LRT}}$ and $\widehat{\alpha}_{\text{RST}}$ when the nominal significance level $\alpha = 0.01$ for different values of n, p, and ρ based on 50,000 simulations

$p \rightarrow$		3		4		5		7	
n	$\rho \downarrow$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$
10	-0.4	0.913	0.021	—	—	—	—	—	—
	-0.2	0.914	0.021	—	—	—	—	—	—
	-0.1	0.914	0.021	—	—	—	—	—	—
	0.1	0.913	0.021	—	—	—	—	—	—
	0.3	0.913	0.021	—	—	—	—	—	—
	0.5	0.913	0.021	—	—	—	—	—	—
	0.7	0.913	0.021	—	—	—	—	—	—
	0.9	0.913	0.021	—	—	—	—	—	—
15	-0.4	0.365	0.016	—	—	—	—	—	—
	-0.2	0.365	0.016	0.876	0.023	—	—	—	—
	-0.1	0.364	0.016	0.876	0.023	—	—	—	—
	0.1	0.364	0.016	0.876	0.023	—	—	—	—
	0.3	0.365	0.016	0.876	0.023	—	—	—	—
	0.5	0.365	0.016	0.876	0.023	—	—	—	—
	0.7	0.365	0.016	0.876	0.023	—	—	—	—
	0.9	0.364	0.015	0.876	0.023	—	—	—	—
20	-0.4	0.174	0.014	—	—	—	—	—	—
	-0.2	0.174	0.014	0.500	0.018	0.902	0.024	—	—
	-0.1	0.174	0.014	0.500	0.018	0.901	0.024	—	—
	0.1	0.174	0.014	0.500	0.018	0.901	0.025	—	—
	0.3	0.174	0.014	0.500	0.018	0.901	0.024	—	—
	0.5	0.174	0.014	0.500	0.018	0.901	0.024	—	—
	0.7	0.174	0.014	0.500	0.018	0.901	0.024	—	—
	0.9	0.174	0.014	0.500	0.019	0.901	0.024	—	—
25	-0.4	0.104	0.014	—	—	—	—	—	—
	-0.2	0.104	0.014	0.283	0.018	0.634	0.021	—	—
	-0.1	0.105	0.014	0.283	0.018	0.634	0.021	0.999	0.029
	0.1	0.104	0.014	0.284	0.018	0.633	0.021	0.999	0.029
	0.3	0.104	0.014	0.284	0.018	0.633	0.021	0.999	0.029
	0.5	0.104	0.014	0.284	0.017	0.634	0.021	0.999	0.029
	0.7	0.105	0.014	0.283	0.017	0.634	0.021	0.999	0.029
	0.9	0.105	0.014	0.283	0.017	0.634	0.021	0.999	0.029

$p \rightarrow$		3		4		5		7	
n	$\rho \downarrow$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$
30	-0.4	0.076	0.012	—	—	—	—	—	—
	-0.2	0.075	0.012	0.181	0.016	0.420	0.019	—	—
	-0.1	0.075	0.012	0.181	0.016	0.419	0.019	0.965	0.025
	0.1	0.075	0.013	0.181	0.016	0.419	0.018	0.965	0.025
	0.3	0.076	0.012	0.181	0.016	0.419	0.019	0.965	0.025
	0.5	0.076	0.012	0.182	0.016	0.420	0.019	0.965	0.025
	0.7	0.076	0.012	0.181	0.016	0.419	0.019	0.965	0.025
	0.9	0.075	0.013	0.181	0.016	0.420	0.019	0.965	0.025
50	-0.4	0.034	0.011	—	—	—	—	—	—
	-0.2	0.034	0.011	0.063	0.013	0.127	0.016	—	—
	-0.1	0.034	0.011	0.063	0.013	0.127	0.015	0.446	0.018
	0.1	0.034	0.011	0.063	0.013	0.127	0.015	0.445	0.018
	0.3	0.035	0.011	0.063	0.013	0.127	0.015	0.445	0.017
	0.5	0.034	0.011	0.063	0.013	0.126	0.015	0.446	0.018
	0.7	0.034	0.011	0.063	0.013	0.126	0.015	0.446	0.018
	0.9	0.035	0.011	0.063	0.013	0.126	0.015	0.445	0.018
75	-0.4	0.024	0.011	—	—	—	—	—	—
	-0.2	0.024	0.011	0.035	0.012	0.059	0.014	—	—
	-0.1	0.024	0.011	0.035	0.012	0.059	0.014	0.172	0.014
	0.1	0.024	0.011	0.035	0.012	0.060	0.014	0.172	0.014
	0.3	0.024	0.011	0.035	0.012	0.060	0.013	0.172	0.014
	0.5	0.024	0.011	0.035	0.012	0.060	0.013	0.172	0.014
	0.7	0.024	0.011	0.035	0.012	0.060	0.013	0.172	0.014
	0.9	0.024	0.011	0.035	0.012	0.060	0.013	0.172	0.014
100	-0.4	0.019	0.011	—	—	—	—	—	—
	-0.2	0.019	0.011	0.026	0.012	0.040	0.013	—	—
	-0.1	0.019	0.011	0.025	0.011	0.039	0.013	0.096	0.014
	0.1	0.018	0.011	0.025	0.012	0.039	0.013	0.096	0.014
	0.3	0.018	0.011	0.026	0.012	0.039	0.013	0.096	0.014
	0.5	0.019	0.011	0.026	0.012	0.039	0.013	0.096	0.014
	0.7	0.019	0.011	0.026	0.012	0.039	0.013	0.096	0.014
	0.9	0.019	0.011	0.026	0.012	0.039	0.013	0.096	0.014

It is to be noted that the standard error of a statistic is important in evaluating the accuracy of an estimate. Since the empirical Type I error rate, which is a proportion, estimates the nominal significance level α, the standard error of an empirical Type I error rate is $\sqrt{\big(\alpha(1-\alpha)\big)/50000}$. So,

Table 2: Observed Type I error rates $\widehat{\alpha}_{\text{LRT}}$ and $\widehat{\alpha}_{\text{RST}}$ when the nominal significance level $\alpha = 0.05$ for different values of n, p, with $\rho = 0.5$ based on 50,000 simulations

$p \rightarrow$	3		4		5		7	
n	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$
10	0.965	0.083	—	—	—	—	—	—
15	0.593	0.068	0.952	0.088	—	—	—	—
20	0.372	0.062	0.719	0.076	0.967	0.090	—	—
25	0.269	0.061	0.519	0.073	0.826	0.081	1.000	0.103
30	0.208	0.060	0.389	0.067	0.658	0.076	0.992	0.093
50	0.122	0.054	0.192	0.057	0.308	0.065	0.689	0.071
75	0.093	0.054	0.127	0.056	0.184	0.060	0.386	0.065
100	0.078	0.052	0.100	0.054	0.135	0.057	0.258	0.063

in case of nominal significance level 0.01, the standard error of an empirical Type I error rate is 4.45×10^{-4} and in case of nominal significance level 0.05, the standard error of an empirical Type I error rate is 9.75×10^{-4}. The maximum standard errors for $\widehat{\alpha}_{\text{LRT}}$ and $\widehat{\alpha}_{\text{RST}}$ for each p of the simulated type I rates are presented in Table 3. These results give a better sense of how different the simulated type I error distributions are between the LRT and the RST.

Table 3: Maximal empirical standard errors of $\widehat{\alpha}_{\text{LRT}}$ and $\widehat{\alpha}_{\text{RST}}$ when the nominal significance level $\alpha = 0.05$ for different values of p, with $\rho = 0.5$ based on 50,000 simulations

$p \rightarrow$	3		4		5		7	
	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$	$\widehat{\alpha}_{\text{LRT}}$	$\widehat{\alpha}_{\text{RST}}$
	0.00215	0.00064	0.00224	0.00067	0.00221	0.00068	0.00222	0.00075

Therefore, we see that when the number of repeated measurements (p) is not small, our proposed test may have little power for small samples, especially when the repeated measures are correlated, however performance of RST is much better than LRT. This simulation study allows us to assess the relative performance of these two testing procedures by comparing the empirical Type I error rate under various settings. We see that when the number of repeated measurements increases, both the tests lose power for having higher degrees of freedom. Also, as mentioned before, the empirical significance level decreases very slowly with the increased sample size (n) to the nominal significance level, for both $\alpha = 0.01$ and $\alpha = 0.05$ for a fixed p.

Plots of the Type I error rate as a function of the sample size for RST and LRT statistics for several repeated measures p, $q = 3$, $\rho = 0.5$ and $\alpha = 0.01$ are given in Figure 1. Clearly, as the number of repeated measures, p, increases, empirical Type I error increases. Also, Type I error rate decreases as the sample size (n) increases for each combination of p, $q = 3$, and $\rho = 0.5$ for LRT, however, Type I error rate is always very low, almost equal to $\alpha = 0.01$ for all sample sizes for RST. This finding motivated us to compute the empirical percentiles of the null distribution of RST as well as LRT statistics for various values of n, p, q and $\rho = 0.5$ in the following section as we have only finite samples in real data applications. We use these empirical percentiles tables to conduct power analysis in Section 4.3.

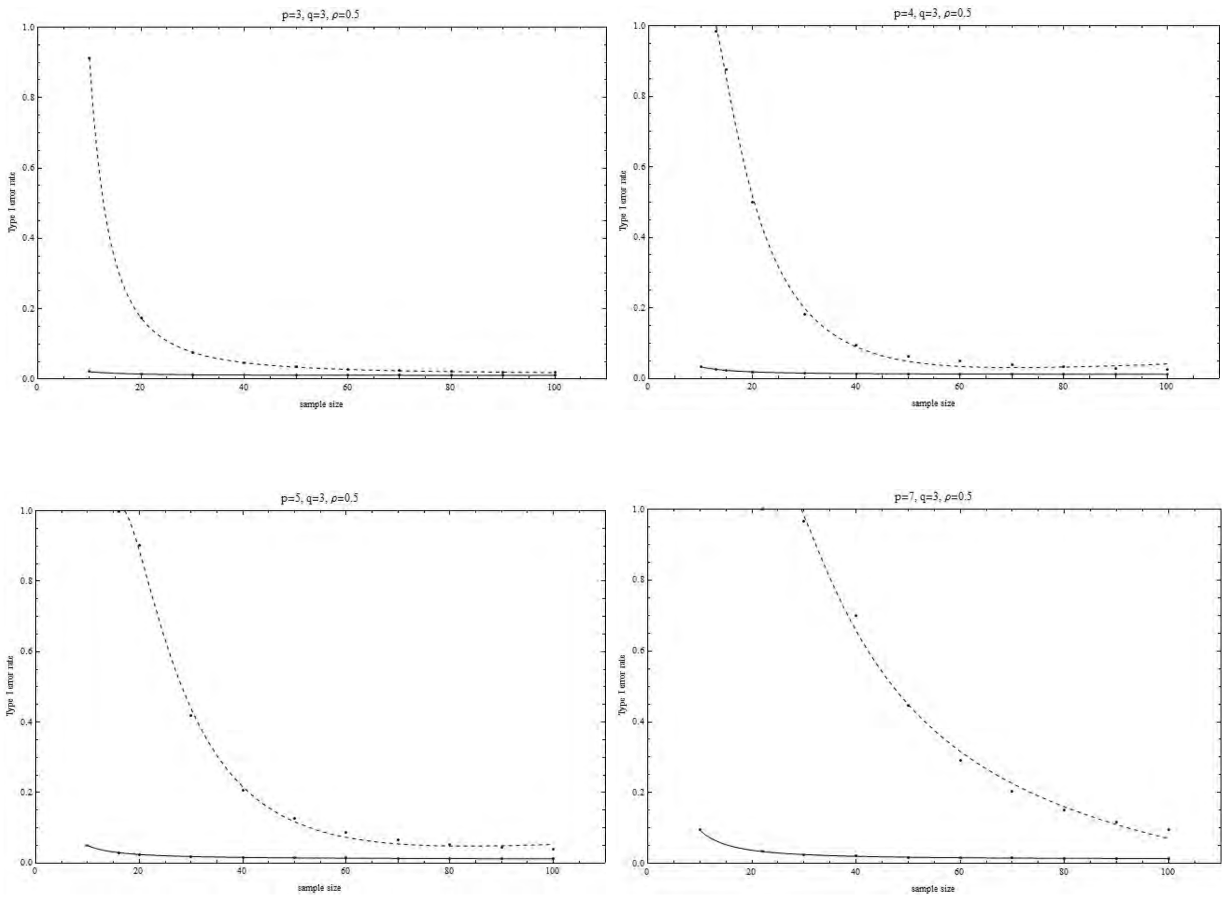

Figure 1: Plots of the Type I error rate as a function of the sample size for RST and LRT statistics for several repeated measures p, $q = 3$, $\rho = 0.5$ and for $\alpha = 0.01$. Plot lines: Dashed – LRT; Solid – RST.

4.2 Empirical 90th, 95th and 99th percentiles of the null distribution of unbiased/unmodified LRT and RST statistics

In this section we conduct some simulation experiments to study the finite sample performance by estimating the percentiles of the END of RST as well as LRT statistics. The number of repeated measurements p is chosen as 2, 3, 4, 5, 7, 10 and 15, and the number of characteristic q is taken as 2 and 3. Samples of various sizes are drawn from $\mathrm{N}_{pq}(\mathbf{0}, \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma})$. We assume $\boldsymbol{\Psi}$ has CS correlation structure with $\rho = 0.5$ only (since previous results showed little sensitivity to ρ). The $(3 \times 3)-$dimensional variance-covariance matrix $\boldsymbol{\Sigma}$ is taken as in the previous section. Tables 4 – 6 and Tables 11 – 16 present the estimates of the empirical 90^{th}, 95^{th} and 99^{th} percentiles of the END of LRT along with the END of RST statistics based on 50,000 simulations for various values of n, p and q. We have compared simulations for 10,000, 50,000 and 100,000 runs for various choices of p and q, and we have found that simulated results are stable for 50,000. After some preliminary studies we decided to use 50,000 runs.

The empirical percentiles allow us to assess the relative performance of the two testing procedures LRT and RST in small to moderate sample size set-up by comparing the percentiles of the END with that of its limiting χ^2 distribution under various settings of n, p and q. We also see in our simulation studies that with different correlation coefficients ρ in the CS correlation structure $\boldsymbol{\Psi}$, the percentiles of the ENDs of both LRT and RST statistics change minutely with ρ for various values of n, p and q. Thus, it appears that the empirical percentiles of the null distributions in Tables 4 – 6 and Tables 11 – 16 will work reasonably well in practice for approximating the limiting χ^2 in small to moderate sample size set-up. It is to be noted that the computation of ENDs is time consuming; for example, a medium category computer takes about 17 hours to compute the results presented in Table 4, while it takes about 55 hours for the Table 16.

We see from Tables 4 – 6 and Tables 11 – 16 that after certain n the empirical percentiles converge very slowly to χ^2 percentiles. It is clear that the bias from the limiting χ^2 percentile decreases as sample size increases. It appears that the percentiles of RST statistic provide better approximation and work well for approximating the limiting χ^2 distribution than that of the LRT statistic. We observe that, when n is small, both the ENDs of LRT and RST statistics are to some extent different from the limiting χ^2 distribution. From Figure 2 we see that just for $n = 4$ and $n = 9$ ENDs of RST statistics are very close to its limiting χ^2 distribution for $p = q = 3$. We also see from Figure 2 that for $n = 4$ ENDs of RST statistics is not close to its limiting χ^2 distribution for $p = 5$ and $q = 3$, however for $n = 15$ it is fairly close to its limiting χ^2 distribution. Note that computations of the ENDs of LRT statistics are not even possible for $n = 4$ and $n = 9$ for $p = q = 3$; also computations of the ENDs of LRT statistics are not even possible for $n = 4$ and $n = 15$ for $p = 5$ and $q = 3$. From Figures 3 it is clear that just for $n = 20$ ENDs of the RST statistics are very close to its limiting χ^2 distribution for $p = 5$ and $q = 3$, whereas for its counterpart LRT statistics it is not the case.

Table 4: Empirical 90th, 95th and 99th percentiles of the null distribution of LRT and RST statistics based on 50,000 simulations for $p = 3$, $q = 2$ and different values of n

n	$Q_{\text{LRT}}(90)$	$Q_{\text{LRT}}(95)$	$Q_{\text{LRT}}(99)$	$Q_{\text{RST}}(90)$	$Q_{\text{RST}}(95)$	$Q_{\text{RST}}(99)$
3	—	—	—	30.227	33.030	38.047
4	—	—	—	28.321	31.408	37.331
5	—	—	—	27.336	30.322	36.940
6	—	—	—	26.766	29.598	36.149
7	68.933	79.094	101.535	26.342	29.319	36.053
8	51.331	58.133	72.047	26.202	29.064	35.482
9	44.399	50.033	61.456	25.831	28.680	35.196
10	40.748	45.725	55.948	25.680	28.529	34.688
15	32.903	36.515	44.132	25.427	28.155	34.164
20	30.421	33.822	40.883	25.263	28.060	33.840
25	29.022	32.338	39.168	25.075	28.055	33.853
27	28.798	31.995	38.568	25.161	27.905	34.214
30	28.181	31.277	37.733	25.039	27.743	33.630
40	27.232	30.456	36.949	25.020	27.902	33.815
50	26.750	29.747	35.892	24.925	27.709	33.557
75	26.134	28.939	35.040	24.982	27.750	33.582
100	25.599	28.543	34.368	24.789	27.585	33.340
125	25.523	28.465	34.350	24.837	27.687	33.493
150	25.378	28.191	34.253	24.814	27.504	33.324
200	25.161	27.978	33.891	24.750	27.481	33.449
∞	24.769	27.587	33.409	24.769	27.587	33.409

From the results of the simulation studies we see that the ENDs of RST statistics present significant set of nice features. They not only have a good asymptotic behavior for increasing sample sizes, but also have very good performance for very small sample sizes, e.g., for sample sizes exceeding only by one or two the number of variables q. We see that there is an associated error when we use percentile of RST statistic instead of true χ^2 percentile. For example, we see that 90th percentile of the RST statistic is 21.112 for $p = 2, q = 3$ and $n = 100$, whereas the true χ^2 percentile is 21.064. Therefore, the relative error is (21.112 - 21.064)/21.064 =0.226%. This shows that END introduces some error, however it is acceptable when p is not too large and n is not too small. Table 7 presents the relative errors between the RST statistics and their ENDs for 90th percentile for different values of n and p, along with $q = 3$. The same for the LRT statistics for 90th percentile are also given in Table 7 in the second row in both *italics* and parenthesis for each combination of n and p. We see that the percent errors between the LRT statistics and their ENDs are much more higher than that of the RST statistics. Also, note that for small sample size n, LRT fails to calculate the percentiles under H_0. The errors associated with ENDs for both RST and LRT statistics increase with p, which is expected though, for a fixed sample size n. Also, the errors decrease with n for a fixed p, as large n means more information and thus the errors get reduced. The same pattern of behavior is observed for 95th as well as for 99th percentiles (results are not shown here). All these characteristics add up to make the ENDs of the RST statistics the best choice for practical applications of the test studied for small as well as for moderate sample sizes. We thus have the following remark.

Remark 2. *From Table 7, as well as from Figures 2 and 3 we observe that for small and moderate sample sizes the END of the RST statistic converges to the limiting* χ^2 *distribution much faster than the corresponding END of the LRT statistic. Thus, we conclude that the END of RST statistic performs much better than the END of LRT statistic for both small and moderate sample studies, and it is then prudent to use the END of RST statistic as opposed to the END of LRT statistic for any real-life applications assuming a stationary model for one of the two matrices.*

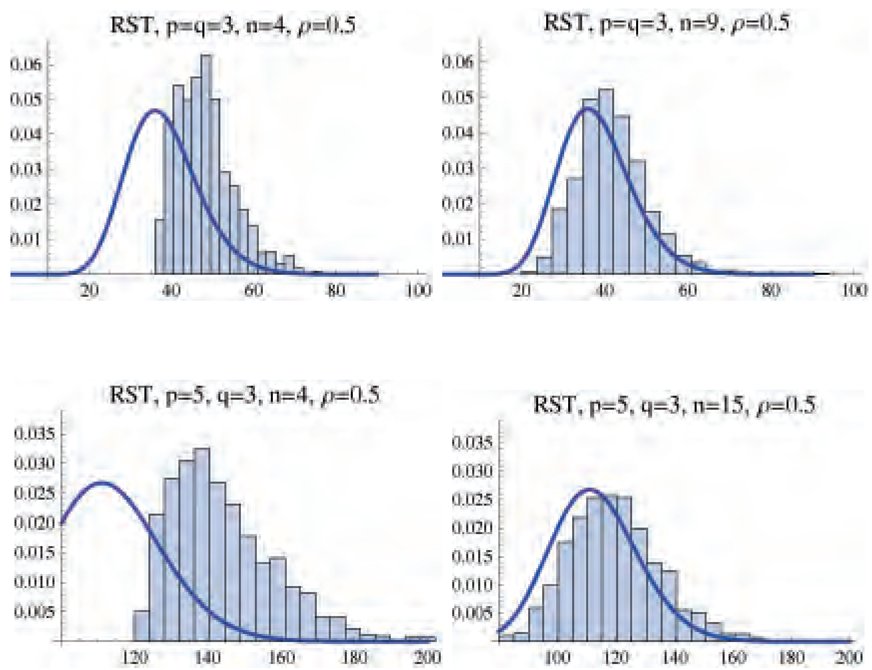

Figure 2: Plots of the empirical histogram and the limiting χ^2 distribution for RST statistics for sample sizes 4 and 9 for $p = 3$ (up). The same for sample sizes 4 and 15 for $p = 5$ (down).

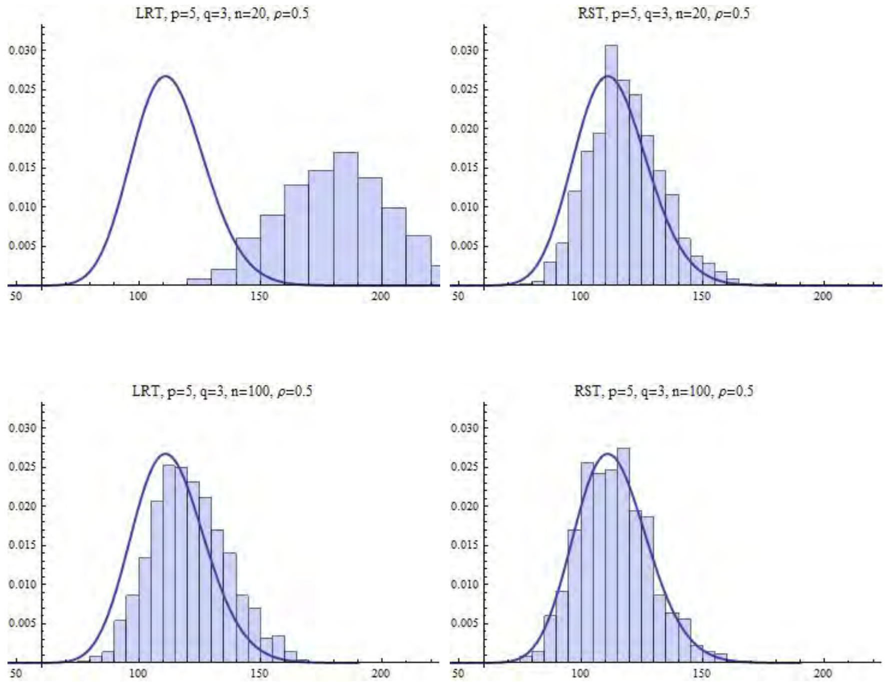

Figure 3: Plots of the empirical histogram and the limiting χ^2 distribution for LRT and RST statistics for sample sizes 20 and 100 for $p = 5$.

Table 5: Empirical 90th, 95th and 99th percentiles of the null distribution of LRT and RST statistics based on 50,000 simulations for $p = 2$, $q = 3$ and different values of n

n	$Q_{\text{LRT}}(90)$	$Q_{\text{LRT}}(95)$	$Q_{\text{LRT}}(99)$	$Q_{\text{RST}}(90)$	$Q_{\text{RST}}(95)$	$Q_{\text{RST}}(99)$
4	—	—	—	22.344	23.458	25.425
5	—	—	—	22.592	24.206	27.077
6	—	—	—	22.522	24.407	27.954
7	64.211	74.505	97.059	22.321	24.491	28.501
8	47.083	53.649	67.636	22.216	24.465	28.768
9	40.000	45.297	56.885	21.885	24.180	29.010
10	36.430	41.250	51.250	22.006	24.275	29.115
15	28.772	32.388	39.738	21.593	24.023	28.885
20	26.366	29.695	36.434	21.477	23.981	28.958
25	25.072	28.210	34.654	21.410	23.928	29.121
27	24.846	27.897	34.381	21.402	23.941	29.005
30	24.218	27.174	33.295	21.238	23.669	29.016
40	23.419	26.358	32.381	21.253	23.854	28.960
50	22.885	25.746	31.678	21.194	23.787	29.220
75	22.252	24.967	30.469	21.208	23.728	29.191
100	21.922	24.509	30.232	21.112	23.595	28.870
125	21.744	24.409	30.202	21.095	23.669	29.213
150	21.612	24.273	29.812	21.084	23.655	29.002
200	21.465	24.188	29.782	21.097	23.667	29.201
∞	21.064	23.685	29.141	21.064	23.685	29.141

Table 6: Empirical 90th, 95th and 99th percentiles of the null distribution of LRT and RST statistics based on 50,000 simulations for $p = 3$, $q = 3$ and different values of n

n	$Q_{\mathrm{LRT}}(90)$	$Q_{\mathrm{LRT}}(95)$	$Q_{\mathrm{LRT}}(99)$	$Q_{\mathrm{RST}}(90)$	$Q_{\mathrm{RST}}(95)$	$Q_{\mathrm{RST}}(99)$
4	—	—	—	58.318	62.261	72.774
5	—	—	—	56.072	60.108	69.189
6	—	—	—	54.589	58.741	67.418
7	—	—	—	53.957	58.087	67.135
8	—	—	—	53.314	57.342	66.195
9	—	—	—	52.568	56.538	65.284
10	130.348	145.199	178.340	52.306	56.270	65.028
15	75.239	81.388	93.834	51.238	55.108	63.240
20	66.025	71.293	82.013	50.694	54.552	62.882
25	61.507	66.411	76.058	50.549	54.501	62.657
27	60.521	65.215	74.874	50.482	54.331	62.389
30	59.103	63.834	73.027	50.370	54.332	62.294
40	56.291	60.628	69.513	50.193	54.037	61.810
50	54.679	59.157	67.366	49.927	53.776	61.636
75	52.975	57.117	65.265	49.926	53.844	61.660
100	51.833	55.932	64.165	49.614	53.553	61.496
125	51.358	55.326	63.733	49.619	53.454	61.776
150	51.212	55.040	63.079	49.755	53.490	61.361
200	50.647	54.494	62.158	49.596	53.395	60.849
∞	49.513	53.384	61.162	49.513	53.384	61.162

Table 7: Relative error between the RST statistics and their ENDs for 90th percentile for various values of p and n.

$p \rightarrow$ $n \downarrow$	2	3	4	5	7	10	15
4	6.076 —	17.784 —	21.657 —	23.889 —	27.246 —	29.662 —	31.157 —
6	6.922 —	10.253 —	13.076 —	14.971 —	16.863 —	18.065 —	19.022 —
8	5.466 (*123.522*)	7.677 —	9.716 —	10.707 —	12.167 —	13.091 —	13.625 —
10	4.474 (*72.949*)	5.642 (*163.263*)	7.170 —	8.191 —	9.461 —	10.173 —	10.621 —
20	1.958 (*25.168*)	2.385 (*33.350*)	3.400 (*45.282*)	3.885 (*63.792*)	4.455 —	4.851 —	5.073 —
30	0.824 (*14.971*)	1.732 (*19.370*)	2.310 (*24.541*)	2.583 (*31.212*)	2.902 (*49.192*)	3.228 —	3.300 —
40	0.895 (*11.180*)	1.374 (*13.691*)	1.625 (*16.691*)	1.861 (*20.799*)	2.086 (*30.286*)	2.321 (*52.174*)	2.458 —
50	0.615 (*8.645*)	0.836 (*10.434*)	1.121 (*12.737*)	1.560 (*15.827*)	1.642 (*22.031*)	1.901 (*34.860*)	1.947 (*76.922*)
75	0.683 (*5.638*)	0.834 (*6.993*)	0.815 (*8.127*)	1.071 (*9.903*)	1.200 (*13.192*)	1.266 (*19.492*)	1.289 (*33.036*)
100	0.226 (*4.074*)	0.205 (*4.686*)	0.637 (*5.804*)	0.703 (*7.072*)	1.023 (*9.563*)	0.947 (*13.603*)	0.980 (*21.648*)

Note: the values in the parenthesis and *italics* are the relative error between the LRT statistics and their ENDs for 90th percentile for various values of p and n.

Computations for calculating the END for fixed n, p, q, ρ and $\mathbf{\Sigma}$ are carried out by the algorithm presented below. The *Mathematica* code to compute END are available from the authors on request. Algorithm Outline:

Step 1 Fix n, p, q, ρ and $\mathbf{\Sigma}$. Calculate $\mathbf{\Psi} = (1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{J}_p$.

Step 2 Set the seed to 123213789.

Step 3 Generate the observation matrix $\mathbf{X}$ from the matrix normal distribution $N_{pq,n}(\mathbf{0}, \mathbf{\Psi} \otimes \mathbf{\Sigma}, \boldsymbol{I}_n)$.

Step 4 Get the pooled sample variance-covariance matrix for repeated measures, say $\mathbf{G}$.

Step 5 Obtain an initial estimate of ρ as $\widehat{\rho}_0 = (\mathbf{1}_p'\mathbf{G}\,\mathbf{1}_p - \mathrm{tr}(\mathbf{G}))/p(p-1)$. Take $\widehat{\mathbf{\Psi}}_0 = (1-\widehat{\rho}_0)\boldsymbol{I}_p + \widehat{\boldsymbol{\rho}}_0\boldsymbol{J}_p$ as an initial estimate of $\mathbf{\Psi}$.

Step 6 Compute $\widehat{\mathbf{\Sigma}} = \frac{1}{np}\sum_{i=1}^{p}(\boldsymbol{e}_i \otimes \boldsymbol{I}_q)'(\widehat{\mathbf{\Psi}} \otimes \boldsymbol{I}_q)\mathbf{X}\,\boldsymbol{Q}_{\mathbf{1}_n}\boldsymbol{X}'(\boldsymbol{e}_i \otimes \boldsymbol{I}_q)$, where $\boldsymbol{e}_i$ is p-th column of $\boldsymbol{I}_p$.

Step 7 Compute $k_0 = nqp(p-1)$, $a = \mathrm{tr}((\boldsymbol{I}_p \otimes \widehat{\mathbf{\Sigma}})\mathbf{X}\,\boldsymbol{Q}_{\mathbf{1}_n}\boldsymbol{X}')$, and $b = \mathrm{tr}((\boldsymbol{J}_p \otimes \widehat{\mathbf{\Sigma}})\mathbf{X}\,\boldsymbol{Q}_{\mathbf{1}_n}\boldsymbol{X}')$.

Step 8 Compute the value of $\widehat{\rho}$ by solving the cubic equation $k_0(p-1)\widehat{\rho}^3 + (k_0 - k_0(p-1) + (p-1)^2a - (p-1)b)\widehat{\rho}^2 + (2(p-1)a - k_0)\widehat{\rho} + a - b = 0$. Ensure that $-1/(p-1) < \widehat{\rho} < 1$. Truncate $\widehat{\boldsymbol{\rho}}$ to $-1/(p-1)$ or 1, if it is outside this range.

Step 9 Compute the revised estimate of $\widehat{\mathbf{\Psi}}$ from $\widehat{\rho}$.

Step 10 Compute the revised estimate of $\widehat{\mathbf{\Sigma}}$ using $\widehat{\mathbf{\Psi}}$ obtained in Step 9.

Step 11 Repeat Steps 7, 8, 9 and 10 until convergence is attained. This is ensured by verifying that the maximum of the absolute difference between two successive values of $\widehat{\rho}$ and the absolute difference between two successive values of $\mathrm{tr}(\widehat{\mathbf{\Sigma}})$ is less than a pre-determine number $\varepsilon(= 10^{-6}$, say).

Step 12 Calculate the Rao's score test statistic $RS = nqp/2 - \mathrm{tr}(\boldsymbol{Z}) + (1/2n)\mathrm{tr}(\boldsymbol{Z}^2)$, where

$$\mathbf{Z} = (\widehat{\mathbf{\Psi}} \otimes \widehat{\mathbf{\Sigma}})\mathbf{X}\,\boldsymbol{Q}_{\mathbf{1}_n}\boldsymbol{X}'.$$

Step 13 Repeat Steps 3-12 50,000 times.

4.3 Power Simulations

Power analysis is very important for applications to real data. Thus, we carry out some power simulations to study the finite sample performance of the tests comparing the LRT and the RST approaches. The power of a statistical test (Lehmann and Romano, 2005), e.g., for RST, is a function and is defined as

$$\beta(\mathbf{\Omega}) = P_{\mathbf{\Omega}}(\text{reject } \; H_0 \; \text{ when using RST given } \; H_0 \; \text{ is false}) = P_{\mathbf{\Omega}}(RST > \widehat{\kappa}_\alpha),$$

where κ_α is defined in (19), and $\widehat{\kappa}_\alpha$ is an estimate of κ_α This function cannot be computed exactly, but can be approximated using Monte Carlo technique (Rizzo, 2008). Now, both RST and LRT statistics depend on n, p and q. So, both RST and LRT statistics are functions of n, p and q as shown in Tables 8 and 9. It seems RST is more powerful than the LRT for the alternative in our study.

Like observed Type I error rates $\widehat{\alpha}_{\text{LRT}}$ and $\widehat{\alpha}_{\text{RST}}$ here also samples of various sizes from small to large, e.g., $n = 4, 6, 8, 10, 15, 20, 25, 30, 50, 75$ and 100 are generated from a $pq-$variate normal population $\text{N}_{pq}(\mathbf{0}, \mathbf{\Omega})$, where $\mathbf{\Omega}$ is an unstructured positive definite matrix. The number of repeated measurements p is chosen as 3, 4, 5 and 7, and the number of characteristics q as 3. We generate 50,000 samples for each combination of parameters p, q and n. The empirical powers of LRT and RST for $p = 3, 4, 5$ and $p = 7$ and for different values of n are given in Tables 8 and 9 for $\alpha = 0.01$ and 0.05 respectively.

Table 8: Empirical powers of LRT and RST for different values of n and p for $\alpha = 0.01$ based on 50,000 simulations

$p \rightarrow$	3		4		5		7	
n	LRT	RST	LRT	RST	LRT	RST	LRT	RST
4	—	0.022	—	0.021	—	0.018	—	0.014
6	—	0.044	—	0.034	—	0.028	—	0.021
8	—	0.067	—	0.053	—	0.037	—	0.027
10	0.020	0.097	—	0.078	—	0.055	—	0.037
15	0.134	0.229	0.058	0.174	—	0.117	—	0.070
20	0.316	0.397	0.205	0.330	0.099	0.228	—	0.130
25	0.536	0.583	0.372	0.479	0.226	0.346	0.073	0.209
30	0.713	0.750	0.580	0.662	0.391	0.504	0.158	0.295
50	0.985	0.989	0.961	0.976	0.887	0.927	0.682	0.773
75	1.000	1.000	1.000	1.000	0.997	0.999	0.973	0.986
100	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000

While there is some correspondence between Type I error and power in these tests, it is not a strong linkage. This is because neither the null nor alternative hypotheses are point hypotheses; both describe sets of covariance structures. This makes the concept of a statistical distance between hypotheses very difficult to measure. The measure of discrepancy between hypotheses is based on Genton (2007). It is to be noted that this measure is used in nearest Kronecker product for a space time covariance matrix problem, and it is not sufficient to measure the distance between our structured covariance matrix and the alternative. For our studies we assumed nonseparable covariance matrix as true hypothesis. The results presented in the tables are for large distances from separability. For determining empirical power of the LRT and the RST we use the corresponding empirical null distributions.

Table 9: Empirical powers of LRT and RST for different values of n and p for $\alpha = 0.05$ based on 50,000 simulations

$p \rightarrow$	3		4		5		7	
n	LRT	RST	LRT	RST	LRT	RST	LRT	RST
4	—	0.092	—	0.086	—	0.073	—	0.065
6	—	0.142	—	0.125	—	0.102	—	0.084
8	—	0.199	—	0.166	—	0.137	—	0.107
10	0.094	0.269	—	0.233	—	0.176	—	0.132
15	0.350	0.469	0.202	0.391	—	0.304	—	0.216
20	0.583	0.672	0.441	0.591	0.275	0.469	—	0.327
25	0.773	0.817	0.642	0.743	0.474	0.624	0.231	0.448
30	0.882	0.909	0.798	0.863	0.645	0.758	0.391	0.583
50	0.997	0.998	0.992	0.995	0.967	0.983	0.875	0.926
75	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.998
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

As expected, empirical power increases steadily with n for a fixed p, as increase in n provides more information. We observe this phenomenon for both $\alpha = 0.01$ and 0.05 for LRT as well as for RST. Now, since the number of parameters in the $(pq \times pq)-$dimensional unstructured variance-covariance matrix $\mathbf{\Omega}$ is $pq(pq+1)/2$, increase in p for a fixed n means more parameters to estimate. So, too many degrees of freedom are used up in estimating too many parameters in the $(pq \times pq)-$dimensional covariance matrix $\mathbf{\Omega}$, and consequently power decreases with the increase of p for LRT as well as for RST. The most important factor to be noticed from Tables 8 and 9 is that the power of LRT is always smaller than the power of RST, especially for small samples for both $\alpha = 0.01$ and 0.05. If we compare Tables 8 and 9, as expected we see that the increase in the Type I error (α), increases the power of a test. We notice that as the number of repeated measures, p, increases empirical power decreases. It is to be noted, that a bigger power study is needed, since our study is only for one unstructured $\mathbf{\Omega}$ as the alternative hypothesis.

5 Three real data examples

To illustrate our proposed testing method, in this section we test the Hypothesis (2) on three data sets. The first one is of relatively smaller in size, and the second and third ones are of moderately larger sizes. We use the biased along with the unbiased and unmodified RST in these examples to see the performance of our new method and evaluate the performance in comparison to the biased along with the unbiased and unmodified LRT method.

Example 1 (*Dental Data*). The data set is from Timm (1980, Table 7.2). The data were originally collected by T. Zullo of the School of Dental Medicine at the University of Pittsburgh. There are nine subjects in the data set. Measurements at three different time points ($p = 3$) were made on each of $q = 3$ characteristics. Note that the null hypothesis cannot be tested using the LRT as the number of subjects $n = 9$ is not greater than $pq = 9$. Therefore, if we take all three measurements LRT cannot be performed, nonetheless RST can be performed as it only requires $n > q$. The calculated value of RST statistic is 54.8215 (see Table 10) with 38 df $\left(\frac{9\cdot 10}{2} - \frac{3\cdot 4}{2} - 1 = 38\right)$. Now, $\chi^2_{38,0.05} = 53.384$ and $\chi^2_{38,0.01} = 61.162$. Therefore, we reject the null hypothesis at 5% level of significance and fail to reject it at 1% level of significance. From Table 6 we notice that our calculated RST statistic is less than the corresponding critical values, the empirical 95th percentiles for $n = 9$: $\text{END}^{\text{RST}}_{9,0.1} = 52.568$ and $\text{END}^{\text{RST}}_{9,0.05} = 56.538$. So, we reject the null hypothesis at 10% level of significance for RST ($p-$value < 0.1). Nevertheless, we fail to reject the null hypothesis at 5% level of significance for RST (p–value > 0.05). Thus, we draw a little different conclusion when we use END of RST statistic in place of the limiting χ^2 distribution.

For the purpose of comparison of the LRT and the RST, we now consider only two measurements ($q = 2$). The test statistics values are given in Table 10. We will first consider measurements 1 and 2. The calculated value of the LRT statistic with 17 df is 27.0035. Now, $\chi^2_{17,0.1} = 24.769$ and $\chi^2_{17,0.05} = 27.587$. Therefore, we fail to reject the null hypothesis marginally at 5% level of significance, but reject it at 10% level of significance. However, RS= 22.1995 with the same 17 df. Therefore, in this case we fail to reject the null hypothesis at 10% level of significance (p–value > 0.1). Nevertheless, these inferences are not accurate or correct, as from Table 4 we notice that both our calculated LRT and RST statistics are less than the corresponding empirical 90th percentiles for $n = 9$: $\text{END}^{\text{LRT}}_{9,0.1} = 44.399$ and $\text{END}^{\text{RST}}_{9,0.1} = 25.831$. So, we fail to reject the null hypothesis at 10% level of significance for both LRT and RST ($p-$value > 0.1).

Table 10: Calculated values of LRT, RST statistics and their $p-$values along with the $p-$values of the limiting χ^2_ν distribution for different data sets

Data	n	p	q	ν	LRT (END $p-$value)	RST (END $p-$value)	LRT (χ^2_ν $p-$value)	RST (χ^2_ν $p-$value)
Dental	9	3	1,2,3	38	—	54.8215 (> 0.05 & < 0.10)	—	(> 0.01 & < 0.05)
Dental	9	3	1,2	17	27.0035 (> 0.1)	22.1995 (> 0.1)	(> 0.05 & < 0.10)	> 0.1
Dental	9	3	1,3	17	40.4858 (> 0.1)	23.3546 (> 0.1)	< 0.01	> 0.1
AIDS	27	3	1,2,3	38	134.8540 (< 0.01)	114.6980 (< 0.01)	< 0.01	< 0.01
AIDS	27	3	1,2	17	80.4133 (< 0.01)	61.9511 (< 0.01)	< 0.01	< 0.01
Mineral	25	2	1,2,3	14	22.7675 (> 0.1)	17.7937 (> 0.1)	(> 0.05 & < 0.10)	> 0.1

We further consider this data set with measurements 1 and 3. In this case the calculated values of LRT and RST are 40.4858 and 23.3546 respectively with 17 df. Since again $\chi^2_{17,0.1} = 24.769$ and $\chi^2_{17,0.01} = 33.409$, we reject the null hypothesis with $p-$value < 0.01 using LRT, but fail to reject the null hypothesis with $p-$value > 0.1 using RST. So, we see very different conclusions for both the tests. However, when we consider empirical null distributions, we see from Table 4 that again our calculated LRT and RST statistics are less than the respective 90th percentile values, and therefore we fail to reject the null hypothesis at 10% level of significance for both LRT and RST ($p-$values > 0.1).

Thus, we draw very different conclusions on the null Hypothesis (2) when we use ENDs in place of the limiting χ^2 distributions. Nonetheless, the conclusions using LRT and RST are the same if we use the respective empirical distributions when $n > pq$.

Example 2 (*Aids Data*). The data set taken from Thompson (1991) corresponds to a sample of 27 patients involved in a pilot study for a new treatment for AIDS. Three different variables: TMHR scores, Karofsky scores, and T-4 cell counts, were measured at three time points at an interval of 90 days during the study. Thus, for this data set $p = 3$ and $q = 3$. The test statistics values are also given in Table 10. We will first consider all three variables. The calculated test statistic values for LRT and RST methods are equal to 134.85402 and 114.698 respectively with 38 df. Now, $\chi^2_{38,0.01} = 61.162$. Therefore, we reject the null hypothesis at 1% level of significance ($p-$value < 0.01) for both LRT and RST. Now, from Table 6 we see that the critical values, the empirical 99th percentiles for $n = 27$ are $\text{END}^{\text{LRT}}_{27,0.01} = 74.874$ and $\text{END}^{\text{RST}}_{27,0.01} = 62.389$. Thus, if we consider END we also reject the null hypothesis at 1% level of significance (p–value < 0.01) for LRT as well as for RST.

We further consider this data set with the 1st and the 2nd variables, i.e., with TMHR scores and Karofsky scores. In this case calculated test statistic values for LRT and RST methods are equal to

80.413 and 61.951 respectively with 17 df. Since $\chi^2_{17,0.01} = 33.409$, we reject the null hypothesis at 1% level of significance (p–value < 0.01) for both LRT and RST. However, from Table 4 we see that the empirical 99th percentiles for $n = 27$ are $\text{END}^{\text{LRT}}_{27,0.01} = 38.568$ and $\text{END}^{\text{RST}}_{27,0.01} = 34.214$. Thus, if we consider END we also reject the null hypothesis at 1% level of significance ($p-$value < 0.01) for both LRT and RST.

From this example we see that if we use ENDs, we come to the same conclusions for both LRT and RST. Moreover, if we use limiting χ^2 distribution we reach the same conclusion as that of the ENDs.

Example 3 (*Mineral Data*). This data set is taken from Johnson and Wichern (2007, p. 43). An investigator measured the mineral content of bones (radius, humerus and ulna) by photon absorptiometry to examine whether dietary supplements would slow bone loss in 25 older women. Measurements were recorded for three bones on the dominant and non-dominant sides. Clearly, for this data set we have $p = 2$ and $q = 3$. The calculated test statistics values for LRT and RST are 22.7675 and 17.7937 respectively with 14 df and are given in Table 10. Observe that $\chi^2_{14,0.1} = 21.064$ and $\chi^2_{14,0.05} = 23.685$. Therefore, basing on LRT, we reject the null hypothesis at 10% level of significance ($p-$value < 0.1), but fail to reject the null hypothesis at 5% level of significance ($p-$value > 0.05). In case of RST we fail to reject the null hypothesis at 10% level of significance ($p-$values > 0.1). Now, from Table 5 we see that the empirical 90th percentiles for $n = 25$ are $\text{END}^{\text{LRT}}_{25,0.1} = 25.072$ and $\text{END}^{\text{RST}}_{25,0.1} = 21.410$. Thus, if we consider ENDs we fail to reject the null hypothesis at 10% level of significance ($p-$value > 0.1) for LRT as well as for RST.

From this example we see that if we use ENDs, we come to the same conclusions for both the tests, whereas if we use limiting χ^2 distribution we reach different conclusions.

From the above examples we have the following suggestions for the researchers and statistical practitioners.

Remark 3. *From the above examples we see that the inference changes most of the time if we use END as opposed to the limiting χ^2 distribution which is very conservative, especially if the test statistic value lies in the close neighborhood of the critical value of the χ^2 distribution. However, the conclusions remain the same for LRT and RST if we use END, which is more desirable. But most importantly, we see that the conclusions drawn from END using RST and the limiting χ^2 distribution are the same all but one time (in a marginal case with too small sample size) in a small sample example above. These observations suggest us to use RST instead of LRT for testing* *separability of the variance-covariance matrix with first component as CS correlation matrix for small and moderate sample sizes, and especially in small sample sizes. From these studies it can be seen that for precise conclusion it is always better to use END of RST if available instead of χ^2. However, the above examples show that if END of RST is not available, the decision based on the limiting χ^2 distribution would not differ much. Remark 2 also reinforces Remark 3.*

6 Summary and scope for the future

Two-level or doubly multivariate data are thriving in all disciplines in the 21st century, so this topic is of wide interest to many researchers and statistical practitioners in many industries. Here we develop a new hypothesis testing procedure to test the separability of a covariance matrix for two-level multivariate data using RST, which is no longer just an alternative or competitor to LRT, but is much superior to LRT in small and in moderate-sized data sets. A first advantage of the RST over the LRT is that it does not require an estimate of the information matrix under the alternative hypothesis. A second advantage of the RST over the LRT is that it converges to a Chi-square distribution much faster according to our simulation study. From the theoretical point of view the drawback of the RST is that at the beginning the RST needs more calculations connected with the Fisher information matrix $\boldsymbol{\mathcal{F}}$, requiring second derivatives of likelihood functions and the inverse of the Fisher information matrix $\boldsymbol{\mathcal{F}}$. However, it is enough to calculate it one time to obtain a simple form of the RST statistic. So, from computational point of view, RST is faster to obtain, because it does not need to find MLE's under H_A, which are more time-consuming. We see from Remark 1 that one may increase p for more information, and still can get stable estimate of the RST statistic with permissible minimum sample size $q+1$, a quantity independent of p. Nevertheless, the condition $n > pq$ has some advantages, by providing smaller bias and higher precision to ML estimates that help for the behavior of both LRT and RST statistics and the characterization of their distributions - compared to when $n \leq pq$ and n is equal to or just about the permissible minimum sample size.

In this article we have taken the correlation matrix $\boldsymbol{\Psi}$ as CS. First, it is well known (see Naik and Rao, 2001; Jones, 1993) that the correlation matrix $\boldsymbol{\Psi}$ of the repeated measures usually has a simpler structure such as CS, AR(1), circular or Toeplitz as opposed to a general structure. In our formulation, it is easier to accommodate different structures for the correlation matrix of repeated measures (via $\boldsymbol{\Psi}$). Thus, it may be worthwhile to develop tests of the null hypothesis H_0 with $\boldsymbol{\Psi}$ as AR(1), circular or Toeplitz for various types of two-level multivariate data sets. Likewise, if one prefers a non-stationary covariance structure, one can develop an RST statistic with $\boldsymbol{\Psi}$ as an unstructured or antedependent covariance matrix. All these studies would surely help in providing an improved statistical analysis for two-level multivariate data.

The modeling of the mean may have an effect on the performance of separability tests for variance-covariance structures. Since the hypothesis we are interested in this article involves only the variance-covariance matrix, we derive the RST statistic corresponding to vech$\boldsymbol{\Omega}$ by calculating the component of the score vector corresponding to vech$\boldsymbol{\Omega}$. If one wants to see the effect of the mean vector one needs to derive the RST statistic corresponding to $\boldsymbol{\mu}$ too, i.e., derive the RST statistic corresponding to the score vector $\boldsymbol{s}(\boldsymbol{\theta})$. We would like to solve this problem in near future and publish it in a future correspondence.

A relatively new criterion for testing hypothesis, referred to as the gradient test, has been proposed by Terrell (2002). Its statistic shares the same first order asymptotic properties with the three classical tests, the likelihood ratio, the Wald and the Rao's score statistics, and is very simple when compared with the same three classical tests. We will explore Terrell's method to develop a test statistic for the null hypothesis H_0 in near future, and will report it in a future correspondence.

Acknowledgement

The authors thank Professor Dietrich von Rosen at the Swedish University of Agricultural Sciences, Sweden, for the idea and inspiration of the method in the paper. The authors also thank Mr. Michael Anderson at the University of Texas at San Antonio for the careful reading of this paper. This research is partially supported by Ministry of Science and Higher Education in Iuventus Plus programme, grant No. 0123/IP3/2011/71 (Katarzyna Filipiak) and by grant VEGA MŠ SR 1/0832/12 (Daniel Klein). The authors would like to thank the associate editor for his thoughtful comments, and the three reviewers for their constructive comments of an earlier version of this article, and for providing valuable recommendations that led to the substantially improved version of the article.

A Some algebraic definitions and results

Following Magnus and Neudecker (1986) let $\boldsymbol{N}_m$ be the symmetric idempotent $m^2 \times m^2$ matrix defined as $\boldsymbol{N}_m = \frac{1}{2}(\boldsymbol{I}_{m^2} + \boldsymbol{K}_{m,m})$, where $m^2 \times m^2$ matrices $\boldsymbol{I}_{m^2}$ and $\boldsymbol{K}_{m,m}$ represent the identity matrix and the commutation matrix (c.f. Kollo and von Rosen, 2005) respectively. Then a unique $m^2 \times m(m+1)/2-$dimensional transformation matrix $\boldsymbol{D}_m$ is called *a duplication matrix* if

$$\boldsymbol{D}_m \text{vech} \boldsymbol{A} = \text{vec} \boldsymbol{A};$$

see Magnus and Neudecker (1986). Using the above definitions we have the following propositions.

Proposition 1. *The following equalities hold:*

(i) $(\boldsymbol{I}_n \otimes \boldsymbol{K}_{n,n})(\boldsymbol{K}_{n,n} \otimes \boldsymbol{I}_n) = \boldsymbol{K}_{n^2,n}$*;*

(ii) $\boldsymbol{K}_{m,k}(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{K}_{l,n} = \boldsymbol{B} \otimes \boldsymbol{A}$ *for any* $k \times l$ *matrix* $\boldsymbol{A}$ *and* $m \times n$ *matrix* $\boldsymbol{B}$*;*

(iii) $\text{vec}(\boldsymbol{A} \otimes \boldsymbol{B}) = (\boldsymbol{I}_l \otimes \boldsymbol{K}_{n,k} \otimes \boldsymbol{I}_m)(\text{vec}\boldsymbol{A} \otimes \text{vec}\boldsymbol{B})$ *for any* $k \times l$ *matrix* $\boldsymbol{A}$ *and* $m \times n$ *matrix* $\boldsymbol{B}$*;*

(iv) $\left(\boldsymbol{D}'_m(\boldsymbol{A}^{-1} \otimes \boldsymbol{A}^{-1})\boldsymbol{D}_m\right)^{-1} = \boldsymbol{D}^{+}_m(\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{D}^{+\prime}_m$ *for any* $m \times m$ *nonsingular matrix* $\boldsymbol{A}$*, where* $\boldsymbol{D}^{+}_m$ *is a Moore-Penrose inverse of* $\boldsymbol{D}_m$*.*

Proposition 2. *For any* $m \times m$ *symmetric matrix* $\boldsymbol{A}$ *the following equalities hold:*

(i) $\boldsymbol{N}_m \text{vec}\boldsymbol{A} = \text{vec}\boldsymbol{A}$*;*

(ii) $\boldsymbol{K}_{m,m}\boldsymbol{D}_m = \boldsymbol{D}_m$*;*

(iii) $\boldsymbol{N}_m(\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{N}_m = \boldsymbol{N}_m(\boldsymbol{A} \otimes \boldsymbol{A}) = (\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{N}_m$*;*

(iv) $\boldsymbol{D}_m\boldsymbol{D}^{+}_m = \boldsymbol{N}_m$*.*

The statements in the above propositions can be found in Magnus and Neudecker (1986) or Ghazal and Neudecker (2000).

Proposition 3. *Let* $\boldsymbol{F}(\boldsymbol{Z})$ *be a* $k \times l$ *matrix function of* $\boldsymbol{Z}$*.*

(i) *If* $\boldsymbol{Z}$ *is an* $m \times n$ *matrix, then* $\dfrac{\partial \text{vec}\boldsymbol{F}(\boldsymbol{Z})}{\partial \text{vec}'\boldsymbol{Z}}$ *is a* $kl \times mn$ *matrix such that its* (i,j)*th element is the derivative of the* i*th element of* $\text{vec}\boldsymbol{F}(\boldsymbol{Z})$ *with respect to the* j*th element of* $\text{vec}\boldsymbol{Z}$*.*

(ii) *If* $\boldsymbol{Z}$ *is an* $m \times m$ *symmetric matrix then*

$$\frac{\partial \text{vec}\boldsymbol{F}(\boldsymbol{Z})}{\partial \text{vech}'\boldsymbol{Z}} = \frac{\partial \text{vec}\boldsymbol{F}(\boldsymbol{Z})}{\partial \text{vec}'\boldsymbol{Z}} \cdot \frac{\partial \text{vec}\boldsymbol{Z}}{\partial \text{vech}'\boldsymbol{Z}} = \frac{\partial \text{vec}\boldsymbol{F}(\boldsymbol{Z})}{\partial \text{vec}'\boldsymbol{Z}} \cdot \boldsymbol{D}_n,$$

where the derivative of the first term in the multiplicaton is calculated treating Z as non-symmetric. Proposition 3 (ii) follows from the chain rule as described in Magnus and Neudecker (1986). Using the above propositions we now have the following lemma.

Lemma 1. *For any* $m \times m$ *symmetric matrix* $\boldsymbol{A}$

(i) $(\text{vec}'\boldsymbol{A} \otimes \boldsymbol{I}_m)(\boldsymbol{I}_m \otimes \text{vec}\boldsymbol{A}^{-1}) = \boldsymbol{I}_m$*;*

(ii) $(\mathrm{vec}\boldsymbol{A} \otimes \boldsymbol{D}_m)'(\boldsymbol{I}_m \otimes \boldsymbol{K}_{m,m} \otimes \boldsymbol{I}_m)(\boldsymbol{I}_{m^2} \otimes \mathrm{vec}\boldsymbol{A}^{-1}) = \boldsymbol{D}_m'$;

(iii) $(\mathrm{vec}\boldsymbol{A} \otimes \boldsymbol{D}_m)'(\boldsymbol{I}_m \otimes \boldsymbol{K}_{m,m} \otimes \boldsymbol{I}_m)(\mathrm{vec}\boldsymbol{A}^{-1} \otimes \boldsymbol{I}_{m^2}) = \boldsymbol{D}_m'$;

Proof. **(i)** Let $\boldsymbol{A}_{i\bullet}$ denote the i-th row and $\boldsymbol{A}_{\bullet j}$ denote the j-th column of matrix $\boldsymbol{A}$. Then clearly

$$\boldsymbol{A}_{i\bullet}\boldsymbol{A}_{\bullet j}^{-1} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Since

$$(\mathrm{vec}'\boldsymbol{A} \otimes \boldsymbol{I}_n)(\boldsymbol{I}_n \otimes \mathrm{vec}\boldsymbol{A}^{-1}) = \left\{\boldsymbol{A}_{i\bullet}\boldsymbol{A}_{\bullet j}^{-1}\right\}_{ij},$$

we obtain (i).

(ii) From Proposition 2 (ii) and Proposition 1 (i), (iii) and Lemma 1 (i) we can write

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

(iii) We have

$$\begin{aligned}
& (\mathrm{vec}\boldsymbol{A} \otimes \boldsymbol{D}_m)'(\boldsymbol{I}_m \otimes \boldsymbol{K}_{m,m} \otimes \boldsymbol{I}_m)(\mathrm{vec}\boldsymbol{A}^{-1} \otimes \boldsymbol{I}_{m^2}) \\
= \;& \boldsymbol{D}_m' \sum_{i=1}^{m}\sum_{j=1}^{m}(\mathrm{vec}\boldsymbol{A} \otimes \boldsymbol{I}_{m^2})'(\boldsymbol{I}_m \otimes \boldsymbol{E}_{ij} \otimes \boldsymbol{E}_{ij}' \otimes \boldsymbol{I}_m)(\mathrm{vec}\boldsymbol{A}^{-1} \otimes \boldsymbol{I}_{m^2}) \\
= \;& \boldsymbol{D}_m' \sum_{i=1}^{m}\sum_{j=1}^{m}(\mathrm{vec}\boldsymbol{A} \otimes \boldsymbol{I}_{m^2})'(\mathrm{vec}(\boldsymbol{E}_{ij}\boldsymbol{A}^{-1}) \otimes \boldsymbol{I}_{m^2}) \\
= \;& \boldsymbol{D}_m' \sum_{i=1}^{m}\sum_{j=1}^{m}\mathrm{tr}\left[\boldsymbol{A}\boldsymbol{E}_{ij}\boldsymbol{A}^{-1}\right](\boldsymbol{E}_{ij}' \otimes \boldsymbol{I}_m) \\
= \;& \boldsymbol{D}_m' \sum_{i=1}^{m}(\boldsymbol{E}_{ii} \otimes \boldsymbol{I}_m) \\
= \;& \boldsymbol{D}_m',
\end{aligned}$$

where $\boldsymbol{E}_{ij} = \boldsymbol{e}_i\boldsymbol{e}_j'$ and $\boldsymbol{e}_i$ is the i-th column of $\boldsymbol{I}_m$.

□

B Empirical percentiles of the null distribution of LRT and RST statistics for several combinations of p, q and n

Table 11: Empirical 90th, 95th and 99th percentiles of the null distribution of LRT and RST statistics based on 50,000 simulations for $p = 2$, $q = 2$ and different values of n

n	$Q_{\mathrm{LRT}}(90)$	$Q_{\mathrm{LRT}}(95)$	$Q_{\mathrm{LRT}}(99)$	$Q_{\mathrm{RST}}(90)$	$Q_{\mathrm{RST}}(95)$	$Q_{\mathrm{RST}}(99)$
3	—	—	—	14.870	16.124	17.446
4	—	—	—	13.239	15.367	19.063
5	34.557	41.986	57.578	12.239	14.435	18.953
6	23.919	28.418	38.353	11.804	13.940	18.422
7	20.071	23.857	32.368	11.553	13.524	18.301
8	17.985	21.262	28.342	11.402	13.349	17.797
9	16.586	19.552	26.262	11.209	13.065	17.590
10	15.735	18.695	25.160	11.224	13.155	17.607
15	13.522	15.915	21.253	11.013	12.829	17.068
20	12.746	15.049	20.063	10.995	12.800	16.712
25	12.216	14.431	19.282	10.889	12.716	17.028
27	12.090	14.266	19.097	10.879	12.734	16.839
30	11.923	14.184	18.703	10.828	12.697	16.722
40	11.477	13.621	18.099	10.743	12.616	16.770
50	11.323	13.368	17.951	10.730	12.635	16.792
75	11.167	13.194	17.497	10.740	12.642	16.705
100	10.998	12.969	17.315	10.691	12.568	16.948
125	10.825	12.832	17.102	10.613	12.528	16.627
150	10.811	12.791	17.222	10.628	12.543	16.763
200	10.784	12.757	16.923	10.672	12.558	16.644
∞	10.645	12.592	16.812	10.645	12.592	16.812

Table 12: Empirical 90th, 95th and 99th percentiles of the null distribution of LRT and RST statistics based on 50,000 simulations for $p = 4$, $q = 3$ and different values of n

n	$Q_{\mathrm{LRT}}(90)$	$Q_{\mathrm{LRT}}(95)$	$Q_{\mathrm{LRT}}(99)$	$Q_{\mathrm{RST}}(90)$	$Q_{\mathrm{RST}}(95)$	$Q_{\mathrm{RST}}(99)$
4	—	—	—	105.398	111.562	125.228
5	—	—	—	100.983	106.867	120.120
6	—	—	—	97.964	103.873	116.879
8	—	—	—	95.053	100.818	112.848
10	—	—	—	92.847	98.290	110.686
12	—	—	—	91.950	97.500	108.948
15	158.181	168.873	190.700	90.590	96.112	107.197
20	125.865	133.295	147.892	89.581	94.622	105.384
25	114.058	121.070	134.729	89.054	94.413	105.428
30	107.897	114.259	126.383	88.637	93.804	104.401
35	103.882	109.989	122.271	88.254	93.508	103.897
40	101.096	106.988	118.456	88.043	92.927	103.237
50	97.670	103.604	115.208	87.607	92.746	103.319
75	93.676	99.072	110.081	87.342	92.456	102.729
100	91.664	97.054	107.680	87.187	92.232	102.469
125	90.823	96.117	106.665	87.116	92.223	102.523
150	90.148	95.518	105.823	87.138	92.390	102.466
200	89.178	94.445	104.764	86.967	92.131	102.391
∞	86.635	91.670	101.621	86.635	91.670	101.621

Table 13: Empirical 90th, 95th and 99th percentiles of the null distribution of LRT and RST statistics based on 50,000 simulations for $p = 5$, $q = 3$ and different values of n

n	$Q_{\mathrm{LRT}}(90)$	$Q_{\mathrm{LRT}}(95)$	$Q_{\mathrm{LRT}}(99)$	$Q_{\mathrm{RST}}(90)$	$Q_{\mathrm{RST}}(95)$	$Q_{\mathrm{RST}}(99)$
4	—	—	—	164.330	173.128	191.403
5	—	—	—	156.797	164.600	182.030
6	—	—	—	152.502	160.348	176.634
8	—	—	—	146.845	154.184	170.504
10	—	—	—	143.508	150.908	166.040
12	—	—	—	141.755	148.932	163.625
15	—	—	—	139.811	146.810	161.214
20	217.259	228.517	249.515	137.797	144.415	157.764
25	187.491	196.404	213.762	136.733	143.274	157.020
30	174.044	182.113	197.796	136.070	142.630	155.900
35	165.742	173.532	188.688	135.731	142.095	155.404
40	160.232	167.613	181.731	135.112	141.327	154.295
50	153.637	161.152	175.388	134.712	141.223	153.951
75	145.779	152.569	165.676	134.064	140.456	152.839
100	142.024	148.673	161.851	133.576	140.000	152.877
125	140.005	146.360	159.271	133.635	139.788	151.979
150	138.671	145.233	158.134	133.442	139.624	152.296
200	137.180	143.617	156.191	133.175	139.508	152.130
∞	132.643	138.811	150.882	132.643	138.811	150.882

Table 14: Empirical 90th, 95th and 99th percentiles of the null distribution of LRT and RST statistics based on 50,000 simulations for $p = 7$, $q = 3$ and different values of n

n	$Q_{\mathrm{LRT}}(90)$	$Q_{\mathrm{LRT}}(95)$	$Q_{\mathrm{LRT}}(99)$	$Q_{\mathrm{RST}}(90)$	$Q_{\mathrm{RST}}(95)$	$Q_{\mathrm{RST}}(99)$
4	—	—	—	320.045	333.168	361.732
5	—	—	—	303.298	315.493	342.473
6	—	—	—	293.932	305.697	330.462
8	—	—	—	282.119	292.984	316.203
10	—	—	—	275.313	285.590	307.403
12	—	—	—	270.973	280.887	302.721
15	—	—	—	266.898	276.713	296.705
20	—	—	—	262.723	271.876	290.728
25	439.586	456.694	489.763	260.331	269.477	287.586
30	375.245	388.996	415.646	258.815	267.726	287.194
35	345.770	357.629	380.706	257.898	267.085	284.771
40	327.691	338.589	360.184	256.763	265.778	283.790
50	306.928	317.389	337.764	255.648	264.475	282.015
75	284.697	294.300	312.812	254.536	263.070	279.643
100	275.569	284.881	302.935	254.091	262.666	279.355
125	270.313	279.169	296.581	253.392	261.914	278.743
150	266.792	275.679	292.738	253.124	261.774	277.989
200	262.866	271.316	288.614	252.587	260.956	277.771
∞	251.517	259.914	276.159	251.517	259.914	276.159

Table 15: Empirical 90th, 95th and 99th percentiles of the null distribution of LRT and RST statistics based on 50,000 simulations for $p = 10$, $q = 3$ and different values of n

n	$Q_{\mathrm{LRT}}(90)$	$Q_{\mathrm{LRT}}(95)$	$Q_{\mathrm{LRT}}(99)$	$Q_{\mathrm{RST}}(90)$	$Q_{\mathrm{RST}}(95)$	$Q_{\mathrm{RST}}(99)$
4	—	—	—	644.667	664.529	707.946
5	—	—	—	608.276	626.401	665.933
6	—	—	—	587.009	604.763	640.058
8	—	—	—	562.278	577.856	612.186
10	—	—	—	547.768	562.611	594.212
12	—	—	—	538.109	552.198	582.809
15	—	—	—	529.319	543.350	572.790
20	—	—	—	521.311	534.671	561.423
25	—	—	—	516.178	529.025	554.500
30	—	—	—	513.238	525.937	551.812
35	854.716	877.584	923.150	510.528	523.092	548.606
40	756.593	774.862	811.536	508.731	521.305	545.550
50	670.511	86.437	717.042	506.644	519.276	542.509
75	594.103	607.839	635.022	503.483	516.007	539.454
100	564.821	578.598	604.252	501.898	514.261	537.148
125	548.963	562.197	587.149	501.093	512.856	536.589
150	539.778	552.828	576.845	501.004	512.910	535.901
200	528.065	540.549	563.514	499.988	512.262	534.177
∞	497.190	508.893	531.335	497.190	508.893	531.335

Table 16: Empirical 90th, 95th and 99th percentiles of the null distribution of LRT and RST statistics based on 50,000 simulations for $p = 15$, $q = 3$ and different values of n

n	$Q_{\mathrm{LRT}}(90)$	$Q_{\mathrm{LRT}}(95)$	$Q_{\mathrm{LRT}}(99)$	$Q_{\mathrm{RST}}(90)$	$Q_{\mathrm{RST}}(95)$	$Q_{\mathrm{RST}}(99)$
4	—	—	—	1425.043	1457.008	1524.644
5	—	—	—	1343.057	1370.775	1433.269
6	—	—	—	1293.202	1319.587	1378.464
8	—	—	—	1234.561	1258.438	1309.237
10	—	—	—	1201.916	1224.987	1272.263
12	—	—	—	1181.405	1202.980	1248.336
15	—	—	—	1160.459	1180.732	1223.385
20	—	—	—	1141.643	1161.045	1199.219
25	—	—	—	1129.622	1149.006	1186.455
30	—	—	—	1122.381	1141.361	1180.518
35	—	—	—	1117.097	1135.311	1172.995
40	—	—	—	1113.225	1130.898	1167.173
50	1922.299	1957.103	2026.640	1107.674	1125.905	1161.561
75	1445.463	1468.783	1514.400	1100.529	1118.660	1152.947
100	1321.733	1342.645	1383.925	1097.170	1115.496	1148.483
125	1262.594	1282.593	1319.962	1095.268	1112.801	1146.797
150	1227.295	1246.608	1283.002	1093.888	1111.188	1143.588
200	1186.980	1205.457	1242.884	1091.553	1109.649	1143.251
∞	1086.521	1103.702	1136.416	1086.521	1103.702	1136.416

References

[1] Anderson, T.W., 2003. *An Introduction to Multivariate Statistical Analysis* (3rd Ed.). New Jersey: Wiley.

[2] Boik, J.B., 1991. Scheffe's mixed model for multivariate repeated measures: A relative efficiency evaluation. *Comm. Statist. Theory Methods* 20, 1233–1255.

[3] Chaganty, N.R., Naik, D.N., 2002. Analysis of multivariate longitudinal data using quasi-least squares. *J. Statist. Plann. Inference* 103, 421–436.

[4] Crowder, M.J., Hand, D.J., 1990. *Analysis of Repeated Measures (Monographs on Statistics & Applied Probability)*. Boca Raton, Florida: Chapman & Hall/CRC.

[5] Dutilleul, P., 1999. The MLE algorithm for the matrix normal distribution. *J. Stat. Comput. Simul.* 64, 105–123.

[6] Engle, R., 1984. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. In: Engle, R. and Intriligator, M.D. (Eds.), *Handbook of Econometrics*, Vol. II. North Holland, Amsterdam.

[7] Galecki, A.T., 1994. General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Comm. Statist. Theory Methods* 22, 3105–3120.

[8] Genton, M.G., 2007. Separable approximations of space-time covariance matrices, *Environmetrics* 18, 681–695.

[9] Ghazal, A.G., Neudecker, H., 2000. On second-order and fourth-order moments of jointly distributed random matrices: a survey. *Linear Algebra Appl.* 321, 61–93.

[10] Johnson, R.A., Wichern, D.W., 2007. *Applied Multivariate Statistical Analysis* (6th Ed.). New Jersey: Pearson Prentice Hall.

[11] Jones, R.H. 1993. *Longitudinal Data with Serial Correlation: A State-Space Approach.* London: Chapman Hall.

[12] Kollo, T., von Rosen, D., 2005. *Advanced Multivariate Statistics with Matrices*. Dordrecht: Springer.

[13] Korin, B.P., 1968. On the distribution of a statistics used for testing a covariance matrix. *Biometrika* 55, 171–178.

[14] Lee, C.H., Dutilleul, P., Roy, A., 2010. Comment on "Models with a Kronecker product covariance structure: estimation and testing" by M. S. Srivastava, T. von Rosen and D. von Rosen. *Math. Methods of Statist.* 19, 88–90.

[15] Lehmann, E.L., Romano, J.P., 2005. *Testing Statistical Hypotheses (3rd Ed.)*. New York, USA: Springer.

[16] Leiva, R., Roy, A., 2011. Linear discrimination for multi-level multivariate data with separable means and jointly equicorrelated covariance structure. *J. Statist. Plann. Inference* 141, 1910–1924.

[17] Lu, N., Zimmerman, D., 2005. The likelihood ratio test for a separable covariance matrix. *Statist. Probab. Lett.* 73, 449–457.

[18] Magnus, J., Neudecker, H., 1986. Symmetry, 0-1 matrices and Jacobians, a review. *Econom. Theory* 2, 157–190.

[19] Manceur, A.M., and Dutilleul, P., 2013. Unbiased modified likelihood ratio tests for simple and double separability of a variance-covariance structure. *Statist. Probab. Lett.* 83, 631–636.

[20] Mitchell, M., Genton, M., Gumpertz, M., 2006. A likelihood ratio test for separability of covariances. *J. Multivariate Anal.* 97, 1025–1043.

[21] Naik, D.N., Rao, S.S., 2001. Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *J. Appl. Stat.* 28, 91–105.

[22] Pinto Pereira, S.M., McCormack, V.A., Moss, S.M., dos Santos, S.I. 2009. The spatial distribution of radiodense breast tissue: a longitudinal study. Breast Cancer Res. 11:R33R44.

[23] Rao, C.R., 1948. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Math. Proc. Cambridge Philos. Soc.* 44, 50–57.

[24] Rao, C.R., 1984. *Linear Statistical Inference and its Applications* (2nd Ed.). New Delhi: Wiley Easrtern Ltd.

[25] Rao, C.R., 2005. Score test: Historical review and recent developments. *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*, 3–20.

[26] Rizzo, M., 2008. *Statistical Computing With R.* Chapman & Hall/CRC.

[27] Roy, A., Khattree, R., 2003. Tests for mean and covariance structures relevant in repeated measures based discriminant analysis. *J. Appl. Statist. Sci.* 12, 91–104.

[28] Roy, A., Khattree, R., 2005a. On implementation of a test for Kronecker product covariance structure for multivariate repeated measures data. *Stat. Methodol.* 2, 297–306.

[29] Roy, A., Khattree, R., 2005b. Testing the hypothesis of a Kroneckar product covariance matrix in multivariate repeated measures data. *Proceedings of the 30th Annual SAS Users Group International Conference (SUGI 30), Philadelphia.*

[30] Roy, A., Khattree, R., 2007a. Classification of multivariate repeated measures data with temporal autocorrelation. *J. Appl. Statist. Sci.* 15, 283–294.

[31] Roy, A., Khattree, R., 2007b. Classification rules for repeated measures data from biomedical research. In: Khattree, R. and Naik, D. N. (Eds.), *Computational Methods in Biomedical Research*, 323–370.

[32] Roy, A., 2007. A note on testing of Kronecker product covariance structures for doubly multivariate data. *Proc. Amer. Statist. Assoc., Statistical Computing Section*, 2157–2162.

[33] Roy, A., Leiva, R., 2008. Likelihood ratio tests for triply multivariate data with structured correlation on spatial repeated measurements. *Statist. Probab. Lett.* 78, 1971–1980.

[34] SAS Institute Inc. 2009. SAS/STAT User's Guide Version 9.3, Cary, NC: SAS Institute Inc.

[35] Self, S.G., Liang, K., 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* 82, 605–610.

[36] Simpson, S.L., 2010. An adjusted likelihood ratio test for separability in unbalanced multivariate repeated measures data. *Stat. Methodol.* 7, 511–519.

[37] Simpson, S.L., Edwards, L.J., Muller, K.E., Styner, M.A., 2014. Separability Tests for High-Dimensional, Low Sample Size Multivariate Repeated Measures Data. *Journal of Applied Stat.* DOI: 10.1080/02664763.2014.919251.

[38] Srivastava, M., von Rosen, T., von Rosen, D., 2008. Models with a Kronecker product covariance structure: Estimation and testing. *Math. Methods of Statist.* 17, 357–370.

[39] Terrell, G.R., 2002. The gradient statistic. *Computing Science and Statistics* 34, 206–215.

[40] Thompson, G.L., 1991. A unified approach to rank tests for multivariate and repeated measures designs. *J. Amer. Statist. Assoc.* 86, 410–419.

[41] Timm, N.H., 1980. Multivariate analysis of variance of repeated measurements. In: Krishnaiah, P. R. (Ed.), *Handbook of Statistics*, Vol. 1, North-Holland, 41–87.

[42] Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* 5, 426–482.

[43] Werner, K., Jansson, M., Stoica, P., 2008. On estimation of covariance matrices with Kronecker product structure. *IEEE Trans. Signal Process.* 56, 478–491.

[44] Wilks, S., 1938. The large sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* 9, 60–62.

[45] Worsley, K.J., Evans, A.C., Strother, S.C., Tyler, J.L., 1991. A linear spatial correlation model, with applications to positron emission tomography. *Journal of the American Statistical Association* 86 (413), 55-67.