



# Permutation based testing on covariance separability

Seongoh Park<sup>1</sup> · Johan Lim<sup>1</sup> · Xinlei Wang<sup>2</sup> · Sanghan Lee<sup>3</sup>

Received: 19 August 2017 / Accepted: 8 September 2018 / Published online: 27 September 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

Separability is an attractive feature of covariance matrices or matrix variate data, which can improve and simplify many multivariate procedures. Due to its importance, testing separability has attracted much attention in the past. The procedures in the literature are of two types, likelihood ratio test (LRT) and Rao's score test (RST). Both are based on the normality assumption or the large-sample asymptotic properties of the test statistics. In this paper, we develop a new approach that is very different from existing ones. We propose to reformulate the null hypothesis (the separability of a covariance matrix of interest) into many sub-hypotheses (the separability of the sub-matrices of the covariance matrix), which are testable using a permutation based procedure. We then combine the testing results of sub-hypotheses using the Bonferroni and two-stage additive procedures. Our permutation based procedures are inherently distribution free; thus it is robust to non-normality of the data. In addition, unlike the LRT, they are applicable to situations when the sample size is smaller than the number of unknown parameters in the covariance matrix. Our numerical study and data examples show the advantages of our procedures over the existing LRT and RST.

**Keywords** Bonferroni test · Covariance matrix · Multiple multivariate data · Non-normal data · Permutation test · Separability

## 1 Introduction

Covariance separability is an attractive property of covariance matrices, which can improve and simplify many multivariate procedures. A separable covariance matrix is defined by its representation as the Kronecker product of two covariance matrices. The

---

✉ Johan Lim  
johanlim@snu.ac.kr

<sup>1</sup> Department of Statistics, Seoul National University, Seoul, Korea

<sup>2</sup> Department of Statistical Science, Southern Methodist University, Dallas, TX, USA

<sup>3</sup> Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA

most common multivariate procedures for a separable covariance matrix are those with the matrix normal distribution or transposable data (Dawid 1981; Gupta and Nagar 1999; Wang and West 2009). For example, Viroli (2010) and Glanz and Carvalho (2013) consider the mixture model and an EM procedure to estimate it. Yin and Li (2012) study the sparse Gaussian graphical model under the matrix normal assumption. Allen and Tibshirani (2012) and Tan and Witten (2014) study various inferential issues on transposable data, where transposability implies both rows and columns of the matrix data are correlated. Extensions to three-level multivariate data are also presented in the literature (e.g., Roy and Leiva 2008, 2011).

Due to its importance in inferential procedures, testing covariance separability has received considerable attention from previous researchers. Lu and Zimmerman (2005) and Mitchell et al. (2006) consider repeatedly measured multivariate data. They, independently, study the likelihood ratio (LR) statistic under the normality assumption, and propose an approximation to its quantiles under the null hypothesis. Fuentes (2006) studies the separability in spatio-temporal processes. She uses a spectral representation of the process, and reformulates the test problem to a simple two-way ANOVA problem. Li et al. (2007) work with stationary spatio-temporal random fields. Using the asymptotic normality of the estimated covariance functions, they build an unified framework for various testing problems. Recently, Filipiak et al. (2016, 2017) propose Rao's score test (RST) under normality for the repeatedly measured multivariate data, and use the asymptotic chi-square distribution as its null distribution.

The normality is a common assumption in most of previous works on testing covariance separability, which is often not true in practice. In this paper, we propose permutation based testing procedures to resolve this difficulty. More specifically, we rewrite the null hypothesis on a separable covariance matrix  $\Sigma$  as an intersection of many individual sub-hypotheses. The sub-hypotheses are on separability of specially structured sub-matrices of  $\Sigma$ , where the LR statistic is invariant to the permutation of (groups of) variables. Thus, the  $p$  value for each sub-hypothesis can be approximated numerically. The final decision is obtained by combining  $p$  values of individual sub-hypotheses using the Bonferroni and multi-stage additive procedures.

The remainder of this paper is organized as follows. In Sect. 2, we briefly introduce the LRT under normality assumption by Mitchell et al. (2006) and the RST by Filipiak et al. (2016, 2017). In Sect. 3, we propose permutation based procedures. We then apply our methods and also two existing procedures (the LRT and RST) to simulated data, and compare their sizes and powers in Sect. 4. We analyze two real data examples in Sect. 5. Concluding remarks and discussions are given in Sect. 6.

## 2 Existing procedures for testing separability

We consider a multivariate random variable  $\mathbf{X}_{p \times q}$ ,

$$\mathbf{X}_{p \times q} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1q} \\ X_{21} & X_{22} & \cdots & X_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pq} \end{pmatrix},$$

and its vectorized variable

$$\mathbf{vX} = (X_{11}, X_{12}, \dots, X_{1q}, X_{21}, X_{22}, \dots, X_{p1}, X_{p2}, \dots, X_{pq})^T$$

which has mean  $\mu = (\mu_{11}, \mu_{12}, \dots, \mu_{1q}, \mu_{21}, \mu_{22}, \dots, \mu_{p1}, \mu_{p2}, \dots, \mu_{pq})^T$  and covariance matrix  $\Sigma$  (a  $pq \times pq$  matrix). Suppose we have  $n$  repeatedly measured observations (independently and identically distributed copies) of  $\mathbf{X}$  and  $\mathbf{vX}$ , which are  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  and  $\mathbf{vX}_1, \mathbf{vX}_2, \dots, \mathbf{vX}_n$ , respectively. The goal of this paper is to test the separability of  $\Sigma$ , i.e.,  $\Sigma = U \otimes V$  for two covariance matrices  $U = (u_{ij}, 1 \leq i, j \leq p)$  (a  $p \times p$  matrix) and  $V = (v_{kl}, 1 \leq k, l \leq q)$  (a  $q \times q$  matrix) using these observations.

A popular procedure to test the separability of  $\Sigma$  is the likelihood ratio (LR) test under normality. Suppose  $\mathbf{vX}$  is normal and has the covariance matrix  $\Sigma = U \otimes V$ . The maximum likelihood estimators (MLEs) of  $U$  and  $V$  are the solution to the following equations (Dutilleul 1999),

$$\begin{aligned} \widehat{V} &= \frac{1}{pn} \sum_{h=1}^n (\mathbf{X}_h - \widehat{\mathbf{M}})^T \widehat{U}^{-1} (\mathbf{X}_h - \widehat{\mathbf{M}}) \\ \widehat{U} &= \frac{1}{qn} \sum_{h=1}^n (\mathbf{X}_h - \widehat{\mathbf{M}}) \widehat{V}^{-1} (\mathbf{X}_h - \widehat{\mathbf{M}})^T, \end{aligned}$$

where  $\mathbf{M}$  is the  $p \times q$  matrix formed from  $\mu$  and  $\widehat{\mathbf{M}}$  is the sample average of  $\mathbf{X}_1, \dots, \mathbf{X}_n$ . The LR statistic is then found to be

$$\text{lrt} = nq \log |\widehat{U}| + np \log |\widehat{V}| - n \log |\mathbf{S}|, \tag{1}$$

where  $\mathbf{S}$  is the sample covariance matrix

$$\mathbf{S} = \frac{1}{n} \sum_{h=1}^n (\mathbf{vX}_h - \widehat{\mu})(\mathbf{vX}_h - \widehat{\mu})^T$$

and  $\widehat{\mu}$  is the vectorized estimator of  $\widehat{\mathbf{M}}$ .

For the null distribution of the LR statistic in (1), Mitchell et al. (2006) propose to match its first moment to that of a scaled chi-square distribution. They first show that the null distribution of the LR statistic only depends on  $p, q$ , and  $n$ , not the specific form of  $U$  and  $V$ . Thus, all separable covariance models  $\Sigma = U \otimes V$  yield the same null distribution given  $p, q, n$ . For each combination of  $(p, q, n)$ , they approximate the null distribution with a scaled chi-square distribution:

$$\text{lrt} \approx k \cdot \chi^2(\xi), \tag{2}$$

where  $\xi = pq(pq + 1)/2 - p(p + 1)/2 - q(q + 1)/2 + 1$  is the number of degrees of freedom of the asymptotic null distribution; and  $k$  is determined by matching the first moment of  $\text{lrt}$  to that of the  $k \cdot \chi^2(\xi)$ , given by

$$k = (1/\xi) \left\{ -n \left( pq \log 2 + \sum_{j=1}^{pq} \psi(0.5(n - j)) - pq \log n \right) - (n/(n - 1))(p(p + 1)/2 + q(q + 1)/2 + pq - 1) \right\},$$

where  $\psi$  is the digamma function. They further approximate the critical value (denoted by  $\text{lrt}_\alpha$ ) with that of the approximate scaled chi-square distribution as

$$\text{lrt}_\alpha \simeq k \cdot \chi_\alpha^2(\xi), \tag{3}$$

where  $\text{lrt}_\alpha$  is the  $100 \times (1 - \alpha)$ -th quantile of the null distribution of  $\text{lrt}$  and  $\chi_\alpha^2(\xi)$  is that of the chi-square distribution  $\chi^2(\xi)$ .

The above approximation by Mitchell et al. (2006) relies on the multivariate normality assumption in evaluating the expectation of  $\text{lrt}$ . In particular, they evaluate the expectation of the generalized variance  $\log |\mathbf{S}|$  under the multivariate normality. As shown in our numerical study in Sect. 4, if the data are from a non-normal distribution (in our study, we use a multivariate t-distribution with degrees of freedom 5 and multivariate version of the chi-square distribution with degrees of freedom 3), the critical value in (3) is severely biased and accordingly the size of the LRT is much larger than the aimed level.

As an alternative to the LRT, very recently, Filipiak et al. (2016, 2017) consider the RST under normality, whose testing statistic is

$$\begin{aligned} \text{rst} &= \frac{npq}{2} - \text{tr} \left( (\hat{\mathbf{U}}^{-1} \otimes \hat{\mathbf{V}}^{-1}) \mathbf{Z}^T \mathbf{P}^\perp \mathbf{Z} \right) \\ &+ \frac{1}{2n} \text{tr} \left( (\hat{\mathbf{U}}^{-1} \otimes \hat{\mathbf{V}}^{-1}) \mathbf{Z}^T \mathbf{P}^\perp \mathbf{Z} (\hat{\mathbf{U}}^{-1} \otimes \hat{\mathbf{V}}^{-1}) \mathbf{Z}^T \mathbf{P}^\perp \mathbf{Z} \right), \end{aligned} \tag{4}$$

where  $\mathbf{Z} = (\mathbf{vX}_1, \dots, \mathbf{vX}_n)^T$  is a matrix having  $n$  rows by stacking transposed  $\mathbf{vX}_h, h = 1, \dots, n$ , and  $\mathbf{P}^\perp = \mathbf{I}_n - n^{-1} \mathbf{1}_n \mathbf{1}_n^T$  is a projection matrix onto orthogonal complement of the column space of  $n$  dimensional vector  $\mathbf{1}_n = (1, \dots, 1)^T$ , and  $\text{tr}(\mathbf{A})$  is the trace of the matrix  $\mathbf{A}$ .

The RST statistic has an advantage in its applicability to a small sized data, where the required number of samples is smaller ( $n > \max(p, q)$ ) than that of LRT ( $n > pq$ ). Also, it is well-known that the RST statistic is asymptotically distributed as the chi-square distribution with degrees of freedom  $\xi$  (defined above) under the null hypothesis. However, as pointed out in Filipiak et al. (2016, 2017), the finite sample distribution of (4) is unknown under the null hypothesis and also is quite different from the asymptotic. The authors numerically approximate its critical values using Monte Carlo samples from the normal distribution. Nonetheless, like the LRT, the

performance of the RST also depends on the normality assumption as shown later in Fig. 1 and tables in Sect. 4.

### 3 Permutation based procedures

All procedures to test the covariance separability in the literature, which include Mitchell et al. (2006) and Filipiak et al. (2016, 2017), strongly depend on the normality assumption or large sample asymptotic, which is often not true in practice. In this section, we propose permutation based procedures that are free of any distributional assumption on the data. To do so, we first rewrite the null hypothesis (the separability of  $\Sigma$ ) into the intersection of many small individual hypotheses on (a specific form of) sub-matrices of  $\Sigma$  and test the individual hypotheses via permutation.

Suppose, as defined earlier,  $\mathbf{vX}$  is a  $pq$  dimensional random vector of

$$(X_{11}, X_{12}, \dots, X_{1q}, X_{21}, X_{22}, \dots, X_{2q}, \dots, X_{p1}, X_{p2}, \dots, X_{pq})^T$$

with covariance matrix  $\Sigma = (\sigma_{st}, 1 \leq s, t \leq pq)$  (a  $pq \times pq$  matrix). The covariance matrix  $\Sigma$  is defined to be separable, if it can be written as the Kronecker product of two covariance matrices  $U = (u_{ij}, 1 \leq i, j \leq p)$  (a  $p \times p$  matrix) and  $V = (v_{kl}, 1 \leq k, l \leq q)$  (a  $q \times q$  matrix):

$$\Sigma = U \otimes V = \begin{pmatrix} u_{11}V & \cdots & u_{1p}V \\ \vdots & \ddots & \vdots \\ u_{p1}V & \cdots & u_{pp}V \end{pmatrix} \tag{5}$$

Let

$$\begin{aligned} \Sigma_{ij}^r &= \begin{pmatrix} u_{ii} & u_{ij} \\ u_{ji} & u_{jj} \end{pmatrix} \otimes V = \begin{pmatrix} u_{ii}V & u_{ij}V \\ u_{ji}V & u_{jj}V \end{pmatrix} \\ &= \begin{pmatrix} u_{ii} \begin{pmatrix} v_{11} & \cdots & v_{1q} \\ \vdots & \ddots & \vdots \\ v_{q1} & \cdots & v_{qq} \end{pmatrix} & u_{ij} \begin{pmatrix} v_{11} & \cdots & v_{1q} \\ \vdots & \ddots & \vdots \\ v_{q1} & \cdots & v_{qq} \end{pmatrix} \\ u_{ji} \begin{pmatrix} v_{11} & \cdots & v_{1q} \\ \vdots & \ddots & \vdots \\ v_{q1} & \cdots & v_{qq} \end{pmatrix} & u_{jj} \begin{pmatrix} v_{11} & \cdots & v_{1q} \\ \vdots & \ddots & \vdots \\ v_{q1} & \cdots & v_{qq} \end{pmatrix} \end{pmatrix}, \end{aligned}$$

which is the submatrix of  $\Sigma$  corresponding to the subvector

$$\mathbf{X}_{ij}^r = (X_{i1}, X_{i2}, \dots, X_{iq}, X_{j1}, X_{j2}, \dots, X_{jq})^T.$$

Similarly, we define  $\Sigma_{kl}^c$  as

$$\begin{aligned} \Sigma_{kl}^c &= \begin{pmatrix} v_{kk} & v_{kl} \\ v_{lk} & v_{ll} \end{pmatrix} \otimes \mathbf{U} = \begin{pmatrix} \mathbf{U}v_{kk} & v_{kl}\mathbf{U} \\ \mathbf{U}v_{lk} & v_{ll}\mathbf{U} \end{pmatrix} \\ &= \begin{pmatrix} \begin{pmatrix} u_{11} & \cdots & u_{1p} \\ \vdots & \ddots & \vdots \\ u_{p1} & \cdots & u_{pp} \end{pmatrix} v_{kk} & \begin{pmatrix} u_{11} & \cdots & u_{1p} \\ \vdots & \ddots & \vdots \\ u_{p1} & \cdots & u_{pp} \end{pmatrix} v_{kl} \\ \begin{pmatrix} u_{11} & \cdots & u_{1p} \\ \vdots & \ddots & \vdots \\ u_{p1} & \cdots & u_{pp} \end{pmatrix} v_{lk} & \begin{pmatrix} u_{11} & \cdots & u_{1p} \\ \vdots & \ddots & \vdots \\ u_{p1} & \cdots & u_{pp} \end{pmatrix} v_{ll} \end{pmatrix}. \end{aligned}$$

Then, the hypothesis “ $\Sigma$  is separable” is equivalent to “ $\Sigma_{ij}^r$  and  $\Sigma_{kl}^c$  are separable for all  $1 \leq i < j \leq p$  and  $1 \leq k < l \leq q$ .” We let  $\mathcal{H}_{0,ij}^r$  and  $\mathcal{H}_{0,kl}^c$  be the separability hypotheses on  $\Sigma_{ij}^r$  and  $\Sigma_{kl}^c$ , respectively, for each choice of  $(i, j)$  and  $(k, l)$ . Below we propose a permutation based procedure to test the individual hypotheses  $\mathcal{H}_{0,ij}^r$  and  $\mathcal{H}_{0,kl}^c$ .

We consider the test  $\mathcal{H}_{0,ij}^r$  which is the hypothesis on the covariance matrix of the sub-vector

$$\mathbf{X}_{ij}^r = (X_{i1}, X_{i2}, \dots, X_{iq}, X_{j1}, X_{j2}, \dots, X_{jq})^T.$$

Here, we assume  $u_{ii} = u_{jj}$  (say they are equal to 1) and, before implementing the procedure below for the data, we standardize the data to take this assumption into account. Under this assumption,  $\mathbf{X}_{ij}^r$  has the covariance matrix

$$\Sigma_{ij}^r = \begin{pmatrix} \mathbf{V} & u_{ij}\mathbf{V} \\ u_{ji}\mathbf{V} & \mathbf{V} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} v_{11} & \cdots & v_{1q} \\ \vdots & \ddots & \vdots \\ v_{q1} & \cdots & v_{qq} \end{pmatrix} & \begin{pmatrix} v_{11} & \cdots & v_{1q} \\ \vdots & \ddots & \vdots \\ v_{q1} & \cdots & v_{qq} \end{pmatrix} \\ u_{ji} \begin{pmatrix} v_{11} & \cdots & v_{1q} \\ \vdots & \ddots & \vdots \\ v_{q1} & \cdots & v_{qq} \end{pmatrix} & \begin{pmatrix} v_{11} & \cdots & v_{1q} \\ \vdots & \ddots & \vdots \\ v_{q1} & \cdots & v_{qq} \end{pmatrix} \end{pmatrix}. \tag{6}$$

In addition,

$$\mathbf{X}_{ji}^r = (X_{j1}, X_{j2}, \dots, X_{jq}, X_{i1}, X_{i2}, \dots, X_{iq})^T$$

has the same covariance matrix with that of  $\mathbf{X}_{ij}^r$ . Thus, if the distribution of  $\mathbf{X}$  is specified only based on its mean and covariance matrix (for example, the elliptical distribution in Anderson 2003), the distributions of  $\mathbf{X}_{ij}^r$  and  $\mathbf{X}_{ji}^r$  are equal if they have common means. The detail about permutation of multivariate data can be found in Li et al. (2010, 2012) and Klingenberg et al. (2009).

The above allows us to construct a permutation based testing procedure for the sub-hypotheses  $\mathcal{H}_{0,ij}^r$  and also for  $\mathcal{H}_{0,kl}^c$ . To be specific, let  $\mathbf{Y}^h, h = 1, 2, \dots, n$ , be

independent copies of  $(\mathbf{X}_{ij}^r)^T$  and let the likelihood ratio test (LRT) statistic for  $\mathcal{H}_{0,ij}^r$  using  $\mathbf{Y}^h$ 's be  $\mathbf{T}$ , that is,

$$\mathbf{T} = nq \log |\widehat{\mathbf{U}}_{ij}| + 2n \log |\widehat{\mathbf{V}}| - n \log |\mathbf{S}_{[i,j]}|, \tag{7}$$

where  $\widehat{\mathbf{U}}_{ij}$  and  $\widehat{\mathbf{V}}$  are the MLE of  $\mathbf{U}_{ij} = [1, u_{ij}; u_{ji}, 1]$ ,  $\mathbf{V}$ , and  $\mathbf{S}_{[i,j]}$  is the sample covariance matrix obtained from  $\mathbf{Y}^h$ 's. Note that we assume  $p = 2$  for the time being, so that  $\mathbf{Y}^h = (\mathbf{Y}_1^h, \mathbf{Y}_2^h)$ , where  $\mathbf{Y}_1^h = (X_{i1}^h, \dots, X_{iq}^h)$  and  $\mathbf{Y}_2^h = (X_{j1}^h, \dots, X_{jq}^h)$ . Suppose  $\pi = (\pi(h), h = 1, 2, \dots, n)$  is a vector of i.i.d. random numbers having values 0 or 1 with probability 1/2. The permutation of  $\{\mathbf{Y}^h = (\mathbf{Y}_1^h, \mathbf{Y}_2^h), h = 1, 2, \dots, n\}$  for  $\pi$  is defined as  $\{\mathbf{Y}^h(\pi) = (\mathbf{Y}_1^h(\pi), \mathbf{Y}_2^h(\pi)), h = 1, 2, \dots, n\}$ , where

$$\mathbf{Y}^h(\pi) = (\mathbf{Y}_1^h(\pi), \mathbf{Y}_2^h(\pi)) = \begin{cases} (\mathbf{Y}_1^h, \mathbf{Y}_2^h) & \text{if } \pi(h) = 0 \\ (\mathbf{Y}_2^h, \mathbf{Y}_1^h) & \text{if } \pi(h) = 1. \end{cases}$$

In the sequel, the permuted LR statistic corresponding to  $\pi$  is computed as

$$\mathbf{T}(\pi) = nq \log |\widehat{\mathbf{U}}_{ij}(\pi)| + 2n \log |\widehat{\mathbf{V}}(\pi)| - n \log |\mathbf{S}_{[i,j]}(\pi)|, \tag{8}$$

where  $\widehat{\mathbf{U}}_{ij}(\pi)$ ,  $\widehat{\mathbf{V}}(\pi)$ , and  $\mathbf{S}_{[i,j]}(\pi)$  are the estimators with the  $1 \times 2q$  permuted samples  $\{\mathbf{Y}^h(\pi) = (\mathbf{Y}_1^h(\pi), \mathbf{Y}_2^h(\pi)), h = 1, 2, \dots, n\}$  defined above. We approximate the null distribution of  $\mathbf{T}$  with the empirical distribution function of the permuted statistics  $\{\mathbf{T}(\pi), \pi \in \Pi\}$ , where  $\Pi$  is the collection of all possible permutations, and then evaluate the  $p$  value for the hypothesis  $\mathcal{H}_{0,ij}^r$ . The above permutation algorithm is summarized as follows. For the column-wise permutation algorithm, we replace “rows” with “columns”; and “p” with “q”.

---

**Algorithm 1** “Row”-wise Permutation

---

**Input:**  $\{\mathbf{X}_h\}_{h=1}^n$ : standardized  $p \times q$  dimensional data  
 1: **for**  $(i, j) \in \{(i, j) : 1 \leq i < j \leq p\}$  **do**  
 2:      $\mathbf{Y}^h \leftarrow 1 \times 2q$  subvector from  $i, j$ -th “rows” of  $\mathbf{X}_h, h = 1, \dots, n$   
 3:      $\text{LRT}_{\text{obs}} \leftarrow$  the LR statistic computed with  $\{\mathbf{Y}^h\}_{h=1}^n$   
 4:     **for**  $b = 1, \dots, B$  **do**  
 5:         Generate  $\pi_h \stackrel{i.i.d.}{\sim} \text{Ber}(1/2), h = 1, \dots, n$   
 6:          $\text{LRT}_b \leftarrow$  the LR statistic from permuted data  $\{\mathbf{Y}^h(\pi)\}_{h=1}^n$   
 7:     **end for**  
 8:      $p_{ij} \leftarrow B^{-1} \sum_{b=1}^B \mathbf{I}(\text{LRT}_b > \text{LRT}_{\text{obs}})$   
 9: **end for**  
 10:  $p_0 \leftarrow \min \left\{ \binom{p}{2} \min_{i < j} p_{ij}, 1 \right\}$   
**Output:**  $p_0$ : the overall  $p$  value

---

Finally, we do the same procedure for each individual hypothesis  $\mathcal{H}_{0,ij}^r$  and  $\mathcal{H}_{0,kl}^c$  and obtain their  $p$  values  $p_{ij}^r$  and  $p_{kl}^c$ .

Our next step is to combine  $p$  values of individual sub-hypotheses to make a final decision on the separability of  $\Sigma$ , given the overall significance level  $\alpha$ . Here, we

consider two combining procedures, both based on the Bonferroni correction. The first procedure, named as “m-perm”, considers only the sub-hypotheses along with the smaller dimension of the data matrix. If  $p$  is smaller than  $q$ , we consider the sub-hypotheses  $\mathcal{H}_{0,ij}^r, i, j = 1, 2, \dots, p$  with  $i < j$ . The row-wise  $\binom{p}{2}$  pairs of hypotheses result in  $p$  values  $\{p_{12}^r, p_{13}^r, \dots, p_{p-1,p}^r\}$ , and the Bonferroni correction compares all individual  $p$  values  $p_{ij}^r$  with the adjusted significance level  $\alpha/\binom{p}{2}$ . Equivalently, the  $p$  value is set as  $\binom{p}{2} \min_{i < j} p_{ij}^r$ . The second procedure, named as “two-s”, is the combination of the row and column-wise Bonferroni procedures with the idea of multi-stage additive testing (Sheng and Qiu 2007). Here, we first test the separability by testing the row-wise sub-hypotheses (here,  $p < q$  is assumed) using the Bonferroni procedure at level  $\gamma_1$ . If the separability is not rejected at the first stage, we further test the column-wise sub-hypotheses using the Bonferroni procedure at level  $\gamma_2$ . The  $p$  value of the two stage procedure becomes  $\binom{p}{2} \min_{i < j} p_{ij}^r$ , if  $\binom{p}{2} \min_{i < j} p_{ij}^r \leq \gamma_1$ ;  $\gamma_1 + (1 - \gamma_1) \binom{q}{2} \min_{k < l} p_{kl}^c$ , otherwise. In this paper, we set the significance levels  $\gamma_1$  and  $\gamma_2$  to be equal, i.e.,  $\gamma = 1 - \sqrt{1 - \alpha}$ , following Sheng and Qiu (2007).

Despite the additional efforts on combining  $p$  values, the proposed procedures have at least two advantages over the existing LRT and RST. First, the LRT and RST strongly depend on the normality assumption, which is often not true in practice. In addition, even under the normality, the null distributions of the LR statistic and RST statistic are still not fully characterized, and several approximate formulas or Monte Carlo approximation is used in the literature. Unlike the existing procedures, our procedures are distribution free and can be used with minimal distributional assumptions. Second, our procedures are applicable to small sized data. For example, the required sample size for our procedures is  $n > 2 \min(p, q)$ , while that of the LRT and RST is  $n > pq$  and  $n > \max(p, q)$ , respectively.

## 4 A numerical study

We numerically investigate the sizes and powers of the proposed permutation based procedures and compare their performances to those of the LRT and the RST. Here, we use the linear model as in Mitchell et al. (2006); that is

$$\mathbf{X} = \mathbf{M} + \mathbf{E}, \quad (9)$$

where  $\mathbf{M}$  is the  $p \times q$  matrix corresponding to the mean vector  $\mu$ ;  $\mathbf{E}$  is the  $p \times q$  error matrix, whose  $(i, j)$ -th element is  $e_{ij}$ , and let

$$\mathbf{vE} = (e_{11}, e_{12}, \dots, e_{1q}, e_{21}, \dots, e_{2q}, \dots, e_{p1}, \dots, e_{pq})^T.$$

We consider three distributions for  $\mathbf{vE}$  with mean 0 and covariance matrix  $\Sigma$ ; (i) the multivariate normal distribution, (ii) the multivariate t-distribution with degrees of freedom 5, and (iii) the “multivariate” chi-square distribution with degrees of freedom 3. Contrary to the first two, the last distribution is asymmetrical, which is generated by



$$\mathbf{vE} = \Sigma^{1/2}(\mathbf{vF} - \mu_F),$$

where each component of  $\mathbf{vF}$  independently follows a univariate chi-square distribution with degrees of freedom 3, and a mean vector  $\mu_F$  of  $\mathbf{vF}$  is used to center the error at the mean zero. It is worthwhile to point out that these multivariate distributions can be characterized only by mean and covariance matrix when degrees of freedom, if relevant, are fixed, which ensures the applicability of our procedures for testing the covariance matrix. We also assume that  $\mathbf{M}$  is the zero matrix for simplicity. Following Mitchell et al. (2006), we set  $p = 4$  (the row size),  $q = 3, 5, 10$  (the column size), and  $n = 20, 25, 50, 75$  (the number of replicated samples). The covariance matrix is assumed to have the form of

$$C[(i, t + k), (j, t)] := \text{cov}(X_{i(t+k)}, X_{jt}) = \sigma^2[\gamma I(i \neq j) + I(i = j)] \frac{\rho_i^k}{1 - \rho_i \rho_j}, \tag{10}$$

where  $1 \leq i, j \leq 4, 1 \leq t \leq t + k \leq q$ , and  $I(A)$  is an indicator function for the event  $A$ . For instance, the model (10) with  $p = 4, q = 3$  is written by

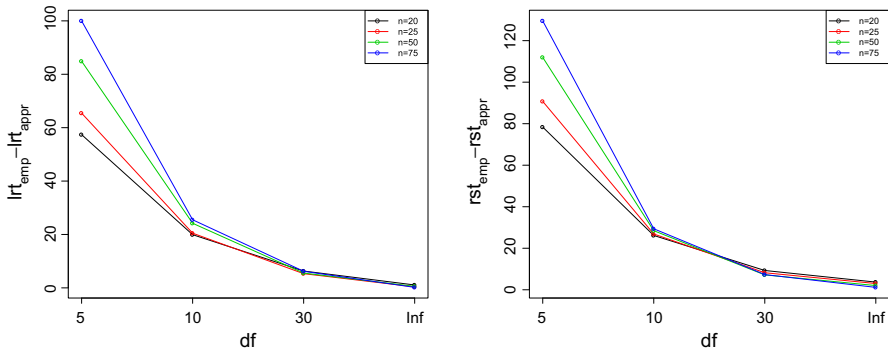
$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44} \end{pmatrix},$$

where each block matrix is defined by

$$\Sigma_{ij} = \sigma^2 * \gamma^{I(i \neq j)} * \frac{1}{1 - \rho_i \rho_j} * \begin{pmatrix} 1 & \rho_j & \rho_j^2 \\ \rho_i & 1 & \rho_j \\ \rho_i^2 & \rho_i & 1 \end{pmatrix}, \tag{11}$$

where  $*$  indicates elementwise multiplication. In the study below, we set  $\sigma = 1$ , and  $\gamma = 0.7$ .

We first examine the magnitude of biases of Mitchell et al. (2006)’s approximation for the LRT and the RST (with the asymptotic chi-square distribution) to the critical value when the normality is violated. To do it, we generate samples from the multivariate t-distribution where the covariance is assumed to be separable  $((\rho_1, \rho_2, \rho_3, \rho_4) = (0.6, 0.6, 0.6, 0.6))$ . We fix row and column sizes as  $p = 4, q = 3$  and compute 10,000 LRT and RST statistics with different sample sizes  $n = 20, 25, 50, 75$  and degrees of freedom  $df = 5, 10, 30, \infty$  ( $\infty$  corresponds to the multivariate normal distribution). To understand the approximation error by Mitchell et al. (2006) and Filipiak et al. (2016, 2017), we consider the differences between the empirical 95-th percentile and the approximation by (3) for the LRT and that by the asymptotic chi-square distribution for the RST. Figure 1 shows that the approximate critical values by both methods are biased, and the magnitude of the biases increases as either the number of samples increases (decreases) when the degrees of freedom are small (large) or the non-normality grows (the degrees of freedom decreases).



**Fig. 1** The difference between the true critical value and its approximation from the multivariate t-distribution under separability

Next, we compare the empirical sizes and powers of the permutation based procedures to the LRT based on (3) and RST. To evaluate the empirical sizes and powers, we consider three hypotheses: (i) null (“N”) hypothesis:  $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.6, 0.6, 0.6, 0.6)$ , (ii) the first alternative (“A1”) hypothesis:  $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.6, 0.65, 0.7, 0.75)$ , the second alternative (“A2”) hypothesis:  $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.9, 0.7, 0.7, 0.45)$ . We generate 500 data sets from the model (9) for each combination of  $p, q$ , and  $n$ . In each data set, we use 2000 permuted samples to calculate the  $p$  value. The significance level  $\alpha$  is set to 0.05. The empirical sizes and powers are reported in Table 1.

Table 1 first shows that the proposed permutation based procedures work better than both the LRT with the approximation in (3) and RST in controlling the size at the aimed level 0.05. The sizes of both the LRT with (3) and RST become larger than 0.05 when the underlying distribution has heavier tails (e.g. the t-distribution with smaller degrees of freedom). In addition, when the data are from an asymmetric distribution (e.g. the chi-square distribution), the same type of upward bias occurs in the size.

Second, the performance of the m-perm depends on the permutation direction (either row-wise or column-wise) as well as the true covariance  $\Sigma$ . In both alternatives “A1” and “A2”, the covariance matrix of the column-wise selection (for example,  $\mathbf{X}_{12}^c$ ) is conjectured to be less separable than that of the row-wise selection (for example,  $\mathbf{X}_{12}^r$ ) and, thus, the choice of the column-wise permutation (the cases  $p > q$ ) shows more power than the row-wise permutation. Table 2 reports the empirical powers (for the cases of  $p = 4$  and  $q = 3$ ) of the Bonferroni procedure (at level  $\alpha$ ) based on the row-wise permutation together with those based on the column-wise permutation. This shows the Bonferroni test for sub-hypotheses along with the row-wise direction (p-perm) now has lower power than the column-wise permutation (q-perm); the p-perm often has lower power than both the size-corrected LRT and RST.

Third, the two-s, our second permutation based procedure, considers both row-wise and column-wise permutations and has higher empirical power than the LRT in all but one case (normal distribution and with  $q = 10, r = 75$  under A1). Compared to the RST, it tends to have higher power when  $(p, q) = (4, 3)$  and  $(4, 5)$ , but lower power when  $(p, q) = (4, 10)$  (the case  $p$  and  $q$  are very unequal). It is interesting to see that

**Table 1** Empirical sizes and powers: the “m-perm” indicates the Bonferroni test along with smaller dimension, “two-s” implies the two-stage additive test having the Bonferroni correction at each stage, “lrt” indicates the LRT under normality given by Mitchell et al. (2006), and “rst” presents the RST

p	q	n	Normal				t(5)				$\chi^2(3)$										
			m-perm	two-s	lrt	rst	m-perm	two-s	lrt	rst	m-perm	two-s	lrt	rst							
N	4	3	20	0.024	0.020	0.064	0.088	0.038	0.046	0.546	0.750	0.010	0.012	0.368	0.252						
			25	0.016	0.026	0.054	0.084	0.022	0.028	0.626	0.732	0.010	0.016	0.434	0.250						
	5	20	50	0.016	0.018	0.052	0.078	0.038	0.044	0.766	0.842	0.000	0.008	0.484	0.248						
			75	0.026	0.036	0.034	0.042	0.034	0.048	0.840	0.862	0.002	0.014	0.520	0.286						
			20	0.028	0.020	—	0.134	0.060	0.058	—	0.928	0.014	0.016	—	0.338						
AI	4	3	25	0.024	0.028	0.072	0.084	0.062	0.052	0.738	0.918	0.020	0.022	0.342	0.342						
			50	0.026	0.028	0.052	0.068	0.058	0.054	0.946	0.992	0.016	0.014	0.552	0.300						
	5	20	75	0.036	0.038	0.046	0.050	0.038	0.034	0.988	0.990	0.002	0.002	0.554	0.314						
			10	20	0.034	0.046	—	0.208	0.032	0.042	—	0.998	0.024	0.018	—	0.530					
			25	0.030	0.036	—	0.194	0.052	0.060	—	0.998	0.018	0.026	—	0.476						
AI	4	3	50	0.026	0.032	0.040	0.086	0.066	0.058	0.996	1.000	0.008	0.008	0.408	0.418						
			75	0.024	0.026	0.048	0.082	0.070	0.068	1.000	1.000	0.004	0.000	0.552	0.404						
	5	20	0.040	<b>0.096</b>	0.070	(0.070)	0.130	(0.094)	0.062	<b>0.158</b>	0.584	(0.072)	0.768	(0.064)	0.046	<b>0.118</b>	0.426	(0.072)	0.314	(0.068)	
			25	0.046	<b>0.120</b>	0.114	(0.114)	0.154	(0.110)	0.072	<b>0.152</b>	0.692	(0.066)	0.814	(0.050)	0.060	<b>0.116</b>	0.530	(0.090)	0.364	(0.100)
			50	0.046	<b>0.266</b>	0.190	(0.184)	0.222	(0.196)	0.108	<b>0.272</b>	0.870	(0.074)	0.912	(0.056)	0.072	<b>0.220</b>	0.706	(0.138)	0.466	(0.146)
5	20	75	0.080	<b>0.464</b>	0.312	(0.298)	0.366	(0.334)	0.108	<b>0.364</b>	0.960	(0.070)	0.966	(0.064)	0.096	<b>0.336</b>	0.830	(0.222)	0.640	(0.282)	
		25	0.084	0.082	—	0.180	<b>(0.092)</b>	<b>0.196</b>	0.160	—	0.946	(0.046)	0.108	—	0.108	—	0.480	<b>(0.110)</b>	0.462	(0.112)	
5	25	0.104	0.062	(0.052)	0.174	(0.078)	<b>0.194</b>	0.176	0.818	(0.072)	0.952	(0.064)	<b>0.136</b>	0.134	0.426	(0.070)	0.462	(0.112)			

Table 1 continued

p	q	n	Normal			t(5)			$\chi^2(3)$						
			m-perm	two-s	lrt	rst	m-perm	two-s	lrt	rst	m-perm	two-s	lrt	rst	
50	75	20	0.288	0.248	0.182 (0.180)	0.264 (0.228)	<b>0.330</b>	0.290	0.982 (0.068)	0.994 (0.050)	<b>0.226</b>	0.190	0.736 (0.128)	0.586 (0.140)	
			<b>0.442</b>	0.364	0.258 (0.264)	0.326 (0.260)	<b>0.410</b>	0.372	0.996 (0.066)	0.996 (0.050)	<b>0.332</b>	0.278	0.876 (0.242)	0.724 (0.260)	
			0.048	0.064	—	0.342 ( <b>0.090</b> )	0.066	<b>0.094</b>	—	0.996 (0.054)	0.068	0.068	—	0.634 ( <b>0.092</b> )	—
			0.082	0.076	—	0.332 ( <b>0.116</b> )	0.100	<b>0.142</b>	—	1.000 (0.050)	0.072	0.072	—	0.624 ( <b>0.114</b> )	—
A2	4	3	0.136	0.146	0.138 (0.122)	0.322 ( <b>0.222</b> )	<b>0.234</b>	0.216	0.996 (0.054)	1.000 (0.056)	0.130	0.134	0.620 (0.112)	0.698 ( <b>0.196</b> )	
			0.278	0.236	0.270 (0.268)	0.430 ( <b>0.326</b> )	0.350	<b>0.360</b>	1.000 (0.076)	1.000 (0.062)	0.212	0.178	0.792 (0.230)	0.750 ( <b>0.250</b> )	
			0.074	<b>0.718</b>	0.576 (0.566)	0.704 (0.608)	0.124	<b>0.684</b>	0.928 (0.238)	0.944 (0.064)	0.116	<b>0.700</b>	0.878 (0.472)	0.814 (0.472)	
			0.100	<b>0.870</b>	0.766 (0.766)	0.808 (0.748)	0.150	<b>0.836</b>	0.978 (0.332)	0.984 (0.124)	0.122	<b>0.834</b>	0.962 (0.636)	0.898 (0.560)	
5	20	5	0.316	<b>1.000</b>	1.000 ( <b>1.000</b> )	1.000 ( <b>1.000</b> )	0.272	<b>0.994</b>	1.000 (0.708)	1.000 (0.236)	0.214	<b>0.986</b>	1.000 (0.984)	0.994 (0.962)	
			0.476	<b>1.000</b>	1.000 ( <b>1.000</b> )	1.000 ( <b>1.000</b> )	0.370	<b>0.998</b>	1.000 (0.932)	1.000 (0.480)	0.328	<b>1.000</b>	1.000 ( <b>1.000</b> )	1.000 ( <b>1.000</b> )	
			<b>0.770</b>	0.664	—	0.778 (0.646)	<b>0.734</b>	0.688	—	0.996 (0.098)	<b>0.668</b>	<b>0.668</b>	—	0.926 (0.474)	—
			<b>0.936</b>	0.888	0.658 (0.624)	0.914 (0.840)	<b>0.898</b>	0.856	0.968 (0.218)	1.000 (0.090)	<b>0.900</b>	0.874	0.938 (0.510)	0.954 (0.686)	
10	20	10	<b>1.000</b>	<b>1.000</b>	1.000 ( <b>1.000</b> )	1.000 ( <b>1.000</b> )	<b>0.994</b>	<b>0.994</b>	1.000 (0.580)	1.000 (0.120)	<b>1.000</b>	<b>1.000</b>	1.000 (0.998)	0.998 (0.988)	
			<b>1.000</b>	<b>1.000</b>	1.000 ( <b>1.000</b> )	1.000 ( <b>1.000</b> )	<b>1.000</b>	<b>1.000</b>	1.000 (0.876)	1.000 (0.170)	<b>1.000</b>	<b>1.000</b>	1.000 ( <b>1.000</b> )	1.000 ( <b>1.000</b> )	
			0.098	0.108	—	0.884 ( <b>0.600</b> )	0.102	<b>0.146</b>	—	1.000 (0.058)	0.114	0.126	—	0.962 ( <b>0.510</b> )	
			0.368	0.336	—	0.938 ( <b>0.812</b> )	<b>0.468</b>	0.436	—	1.000 (0.074)	0.398	0.324	—	0.972 ( <b>0.654</b> )	
75	50	75	<b>1.000</b>	<b>1.000</b>	0.990 (0.988)	1.000 ( <b>1.000</b> )	<b>0.994</b>	0.992	1.000 (0.352)	1.000 (0.086)	<b>0.998</b>	1.000 (0.966)	1.000 ( <b>0.998</b> )		
			<b>1.000</b>	<b>1.000</b>	1.000 ( <b>1.000</b> )	1.000 ( <b>1.000</b> )	<b>1.000</b>	<b>1.000</b>	1.000 (0.540)	1.000 (0.082)	<b>1.000</b>	<b>1.000</b>	1.000 ( <b>1.000</b> )	1.000 ( <b>1.000</b> )	

The cells with “hyphen” are the cases, where  $n \leq pq$  and the LR test is not defined well. The numbers in parentheses are the empirically size-corrected powers using empirical null distribution of LRT (or RST). In the table, when generating samples for the reference distribution, we set covariance factors U, V to be the identity matrix, following Filipiak et al. (2017), Mitchell et al. (2006), because the LRT and RST do not depend on their specific forms

The most powerful methods excluding the LRT, RST without size correction are shown in bold for each scenario

**Table 2** The case of  $p = 4, q = 3$ : “p-perm” and “q-perm” denote empirical powers using row-wise and column-wise permutation, respectively

$n$	A1			A2					
	p-perm	q-perm	lrt	rst	p-perm	lrt	rst		
Normal	20	0.040	<b>0.130</b>	0.070 (0.070)	0.130 (0.094)	0.074	<b>0.812</b>	0.576 (0.566)	0.704 (0.608)
	25	0.046	<b>0.152</b>	0.114 (0.114)	0.154 (0.110)	0.100	<b>0.908</b>	0.766 (0.766)	0.808 (0.748)
	50	0.046	<b>0.320</b>	0.190 (0.184)	0.222 (0.196)	0.316	<b>1.000</b>	<b>1.000 (1.000)</b>	<b>1.000 (1.000)</b>
t(5)	75	0.080	<b>0.538</b>	0.312 (0.298)	0.366 (0.334)	0.476	<b>1.000</b>	<b>1.000 (1.000)</b>	<b>1.000 (1.000)</b>
	20	0.062	<b>0.170</b>	0.584 (0.072)	0.768 (0.064)	0.124	<b>0.742</b>	0.928 (0.238)	0.944 (0.064)
	25	0.072	<b>0.192</b>	0.692 (0.066)	0.814 (0.050)	0.150	<b>0.852</b>	0.978 (0.332)	0.984 (0.124)
50	0.108	<b>0.296</b>	0.870 (0.074)	0.912 (0.056)	0.272	<b>0.990</b>	<b>1.000 (0.708)</b>	<b>1.000 (0.236)</b>	
75	0.108	<b>0.396</b>	0.960 (0.070)	0.966 (0.064)	0.370	<b>0.998</b>	<b>1.000 (0.932)</b>	<b>1.000 (0.480)</b>	

The most powerful methods excluding the LRT, RST without size correction are shown in bold for each scenario

it has higher empirical power than both the LRT and RST even for some cases with the data from the normal distribution. We conjecture this is because the sample size  $n$  is relatively small compared to the dimension of the data  $pq$ . Here, we remark that both the theoretical reference distributions of the LRT and RST are not available in practice.

Finally, we conclude the section with a short report on the computation, which was done with R ver. 3.4.2 and carried out on a PC with Intel Core i7 3.0 GHz processor. The average CPU time without parallel computing for the case  $p = 4$ ,  $q = 5$ ,  $n = 50$  using 2000 permutation steps is 32.04 seconds with standard deviation 1.89 seconds based on 100 repetitions.

## 5 Data examples

### 5.1 Tooth size data

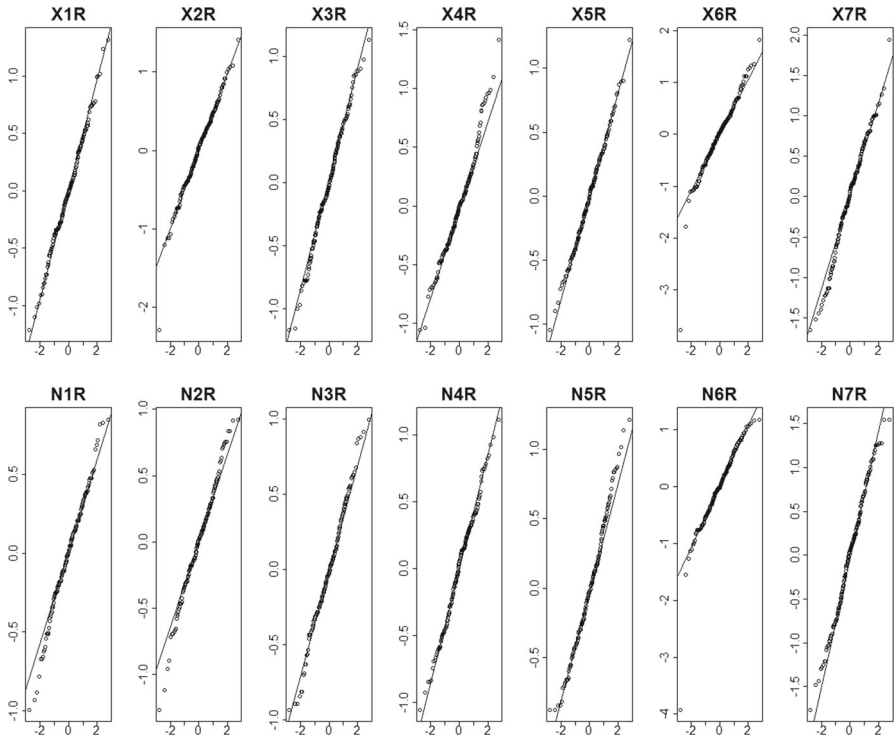
We now apply our method to testing the covariance separability of the tooth size data, which were obtained as a part of Korean National Occlusion Study, conducted from 1997 to 2005 (Wang et al. 2006; Lee et al. 2007). Here, we use the tooth sizes of 179 young Korean men who passed predefined selection criteria among 15,836 respondents recorded in this dataset, to test separability in the analysis. The observation of each subject has a  $2 \times 14$  matrix form, where the first row consists of sizes of 14 teeth in maxilla and the second row consists of those in mandible. We write the size matrix of the  $h$ -th subject as

$$\mathbf{X}_h = \begin{pmatrix} X_{h1}^X, \dots, X_{h7}^X, X_{h8}^X, \dots, X_{h14}^X \\ X_{h1}^N, \dots, X_{h7}^N, X_{h8}^N, \dots, X_{h14}^N \end{pmatrix}$$

where “X” in the upper-script is short for maxilla and “N” for mandible. In this section, we are interested in testing the covariance matrix of  $\mathbf{X}$  (its vectorized version), say  $\Sigma$ , is separable as the Kronecker product of  $U_{2 \times 2}$  (the common covariance matrix between sizes of the upper and lower teeth at the same location) and  $V_{14 \times 14}$  (the covariance matrix among sizes of the teeth within maxillary (or mandible)). In short, we test the hypothesis  $\mathcal{H}_0 : \Sigma = U_{2 \times 2} \otimes V_{14 \times 14}$ .

Before testing the hypothesis on separability, we check the normality assumption which existing procedures rely on. We apply three popular testing procedures for the normality which are available from “MVN” R package by Korkmaz et al. (2014); they are Mardia’s (1970), Henze–Zirkler’s (1990), and Royston’s test (Shapiro and Wilk 1964; Royston 1983, 1992). All of the results clearly indicate non-normality of tooth sizes ( $p$  value = 0), and univariate Q–Q plots (Fig. 2) confirm it again, especially with “X4R” and “N2R”.

To test the separability, we use the centered data by subtracting the mean vector. The cross-covariance matrix between maxillary and mandible regions is estimated from the unstructured sample covariance matrix  $\mathbf{S}$  and the maximum likelihood estimator under separability  $\hat{U} \otimes \hat{V}$ , respectively, as:



**Fig. 2** Univariate Q–Q plots for men’s right-side tooth sizes that are centered. The vertical, horizontal axes represent sample, theoretical quantiles, respectively

$$\mathbf{S}[1 : 7, 15 : 21] = \begin{pmatrix} 0.082 & 0.106 & 0.085 & 0.067 & 0.058 & 0.083 & 0.107 \\ 0.075 & 0.104 & 0.102 & 0.087 & 0.091 & 0.067 & 0.107 \\ 0.050 & 0.088 & 0.124 & 0.103 & 0.097 & 0.092 & 0.113 \\ 0.046 & 0.069 & 0.086 & 0.115 & 0.106 & 0.104 & 0.115 \\ 0.047 & 0.060 & 0.061 & 0.107 & 0.118 & 0.116 & 0.122 \\ 0.061 & 0.067 & 0.072 & 0.106 & 0.126 & 0.202 & 0.180 \\ 0.061 & 0.069 & 0.089 & 0.129 & 0.125 & 0.197 & 0.269 \end{pmatrix}$$

and

$$(\widehat{\mathbf{U}} \otimes \widehat{\mathbf{V}})[1 : 7, 15 : 21] = \begin{pmatrix} 0.033 & 0.017 & 0.015 & 0.011 & 0.010 & 0.013 & 0.014 \\ 0.017 & 0.042 & 0.017 & 0.016 & 0.012 & 0.012 & 0.017 \\ 0.015 & 0.017 & 0.033 & 0.018 & 0.015 & 0.014 & 0.017 \\ 0.011 & 0.016 & 0.018 & 0.034 & 0.022 & 0.018 & 0.023 \\ 0.010 & 0.012 & 0.015 & 0.022 & 0.034 & 0.023 & 0.028 \\ 0.013 & 0.012 & 0.014 & 0.018 & 0.023 & 0.082 & 0.038 \\ 0.014 & 0.017 & 0.017 & 0.023 & 0.028 & 0.038 & 0.090 \end{pmatrix}$$

where  $M[a : b, c : d]$  denotes a submatrix of  $M$  from  $a$ -th row to  $b$ -th row and from  $c$ -th column to  $d$ -th column, producing a  $(b - a + 1) \times (d - c + 1)$  matrix. From the above, we find a significant difference between  $\mathbf{S}[1 : 7, 15 : 21]$  and  $(\widehat{\mathbf{U}} \otimes \widehat{\mathbf{V}})[1 : 7, 15 : 21]$ .

We approximate the  $p$  value using the procedure in Sect. 3, which is approximated to 0.025 by the two-s procedure. Here, the number of permuted data sets to approximate the  $p$  value is set as 10,000. On the other hand, the LRT statistic in (1) is evaluated as 921.91, and the critical value under normality approximated by Mitchell et al. (2006) is 368.26 at the significant level  $\alpha = 0.01$ . The RST statistic is evaluated as 741.45 with the asymptotic critical value  $\chi_{01}^2(299) = 358.81$  and empirical critical value 358.27 at  $\alpha = 0.01$ .

## 5.2 Corpus callosum thickness

Our second example is about two-year longitudinal MRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI). According to Lee et al. (2016), the corpus callosum (CC) thickness profile is calculated based on CC segmentation at equally spaced intervals. To be specific, the CC thicknesses of 135 subjects are measured at 99 points for each year. The separability hypothesis to be tested is  $\mathcal{H}_0 : \Sigma = U_{2 \times 2} \otimes V_{99 \times 99}$ , where we expect  $U$  and  $V$  to explain the covariance structure in CC thickness of the repeated measurements and the measurements within a subject, respectively.

The LRT based on the normality is not applicable to this dataset for two reasons. First, the multivariate normal tests provided by the MVN R-package reveal that the data do not satisfy the multivariate normality. This can also be observed from Fig. 3, in which each shown variable has a heavy right tail. Second, the sample size ( $n = 135$ ) is less than the number of measure points ( $pq = 99 \times 2 = 198$ ), making the LR statistic undefined.

We apply the proposed permutation procedures to the data. More precisely, we apply the permutation test to the sub-hypotheses of all  $\binom{99}{2}$  column-wise pairs. The permutation test for each sub-hypothesis is for the bivariate paired data with the size of  $135 \times 4$ ; for example, if the column pair  $(k, l)$  is chosen, the bivariate paired data are  $\{(X_{hk}^1, X_{hl}^1, X_{hk}^2, X_{hl}^2), h = 1, 2, \dots, 135\}$ . We use 100,000 random permutations to evaluate the  $p$  value of each sub-hypothesis, and the Bonferroni adjusted  $p$  value is given by  $\binom{99}{2} \min_{k < l} p_{kl}^c$ , where  $p_{kl}^c$  is the  $p$  value for the sub-hypothesis  $\mathcal{H}_{kl}^c$  for the  $(k, l)$ -th column pair. The  $p$  value evaluated is less than 0.0001. The RST statistic is evaluated as 16,346.75 with the asymptotic and empirical critical value as  $\chi_{01}^2(14749) = 15,151.49$  and 15,095.09, respectively.

## 6 Discussion

In this paper, we propose permutation based procedures to test the separability of covariance matrices. The procedure divides the null hypothesis on a separable covariance matrix into many sub-hypotheses, which are testable via a permutation method. Compared to the existing LRT and RST under normality, the proposed procedures are distribution free and robust to non-normality of the data. In addition, it is applicable to small sized data, whose size is smaller than dimension of the covariance matrix under test. The numerical study and data examples show that the proposed permuta-



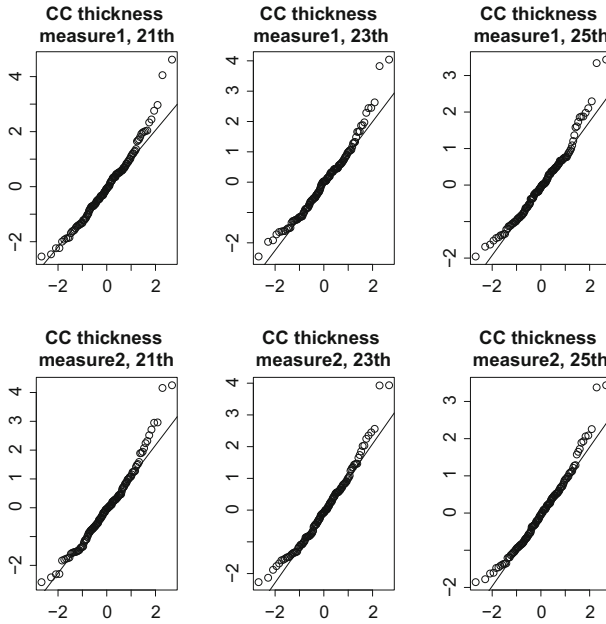


Fig. 3 Univariate Q–Q plots for MRI data, where each columns are centered

tion procedures are more powerful when the data are non-normal and the dimension is high.

Theory on permutation procedures has been well developed for linear permutation test statistics (Strasser and Weber 1999; Finos and Salmaso 2005; Pesarin and Salmaso 2010; Bertoluzzo et al. 2013) and its computational tool is publicly available (the R package “coin” by Hothorn et al. 2017). In our procedure, permutation is applied to testing the sub-hypotheses  $\mathcal{H}_{0,ij}^t$  and  $\mathcal{H}_{0,kl}^c$  for each choice of  $(i, j)$  and  $(k, l)$ ; the sub-hypotheses are on the separability of  $\Sigma_{ij}^t$ s and  $\Sigma_{kl}^c$ s. Here, we use the LR statistic, one of a few known statistics for testing covariance separability. The LR statistic is non-linear and, thus the CRAN package and the asymptotic results of Strasser and Weber (1999) can not be directly applied to it. However, we conjecture that, for an appropriately chosen permutation test statistic, we may encapsulate our problem into the existing conditional inference framework and achieve the proven optimality.

Our procedures in this paper use the Bonferroni rule to combine  $p$  values from testing of many sub-hypotheses. In our problem, the  $p$  values of individual sub-hypotheses are conjectured to be strongly dependent to each other by its nature. For this reason, we adopt the Bonferroni rule to ensure the size of the combined test to be less than the aimed level despite of its conservativeness. In addition, the additional numerical study not reported here shows that the well-known Fisher’s omnibus and Lipták’s rules (under the assumption of independent  $p$  values) are severely biased in their sizes for non-normal data. The same reasoning would be applied to the direct aggregation of individual LR statistics. We could not specify the null distribution of the aggregated

statistic due to the dependency among individual statistics. We thus have difficulty in proceeding with this.

We finally conclude the paper with the remark that the procedure of this paper can easily be applied to testing more complexly structured covariance matrix. Suppose we consider repeatedly measured spatial data (on the lattice system) or image data. The observation of a single subject has the form of a three-way array  $\mathbf{X}_{ijk}$ ,  $i = 1, 2, \dots, a$ ,  $j = 1, 2, \dots, b$ , and  $k = 1, 2, \dots, c$  (with the dimension of  $a \times b \times c$ ), and its separable covariance matrix has the form of  $\Sigma = \mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}$  (here,  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  are  $a \times a$ ,  $b \times b$ , and  $c \times c$  covariance matrix, respectively). To test the hypothesis  $\Sigma = \mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}$ , we read the data as a matrix form as  $\mathbf{Y}_{s,k} = \mathbf{X}_{[ij]k}$  with  $s = 1, 2, \dots, ab$ ,  $k = 1, 2, \dots, c$ , and test the sub-hypothesis  $\mathcal{H}_{0,\{[12],3\}} : \Sigma = \mathbf{U}_1 \otimes \mathbf{C} \left( = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} \right)$  at the level  $\alpha_1$ . We repeat the same procedure for the other two sub-hypotheses  $\mathcal{H}_{0,\{1,[23]\}} : \Sigma = \mathbf{A} \otimes \mathbf{U}_2 \left( = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) \right)$  and  $\mathcal{H}_{0,\{2,[13]\}} : \Sigma = \mathbf{B} \otimes \mathbf{U}_3 \left( = \mathbf{B} \otimes (\mathbf{A} \otimes \mathbf{C}) \right)$  at the level  $\alpha_2$  and  $\alpha_3$ , respectively. Finally, we combine the results of three sub-hypotheses with the Bonferroni or multi-stage additive procedure discussed in Sect. 3.

**Acknowledgements** We are grateful to the Associate Editor and two anonymous reviewers for their many helpful comments. The R-package “NPCovSepTest” of the proposed procedure is online available from <https://sites.google.com/view/seongohpark/software>. J. Lim’s research is supported by National Research Foundation of Korea (Nos. NRF-2017R1A2B2012264; MIST-2011-0030810).

## References

- Anderson TW (2003) An introduction to multivariate statistical analysis, 3rd edn. Wiley, New York
- Allen GI, Tibshirani R (2012) Inference with transposable data: modelling the effects of row and column correlations. *J R Stat Soc Ser B Stat Methodol* 74(4):721–743
- Bertoluzzo F, Pesarin F, Salmaso L (2013) On multi-sided permutation tests. *Commun Stat Simul Comput* 42(6):1380–1390
- Dawid AP (1981) Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika* 68(1):265–274
- Dutilleul P (1999) The mle algorithm for the matrix normal distribution. *J Stat Comput Simul* 64(2):105–123
- Filipiak K, Klein D, Roy A (2016) Score test for a separable covariance structure with the first component as compound symmetric correlation matrix. *J Multivar Anal* 150:105–124
- Filipiak K, Klein D, Roy A (2017) A comparison of likelihood ratio tests and Rao’s score test for three separable covariance matrix structures. *Biom J* 59:192–215
- Finos L, Salmaso L (2005) A new nonparametric approach for multiplicity control: optimal subset procedures. *Comput Stat* 20(4):643–654
- Fuentes M (2006) Testing for separability of spatial-temporal covariance functions. *J Stat Plan Inference* 136(2):447–466
- Glanz H, Carvalho L (2013) An expectation-maximization algorithm for the matrix normal distribution. arXiv preprint [arXiv:1309.6609](https://arxiv.org/abs/1309.6609)
- Gupta AK, Nagar DK (1999) Matrix variate distribution. Chapman Hall/CRC, New York
- Henze N, Zirkler B (1990) A class of invariant consistent tests for multivariate normality. *Commun Stat Theory Methods* 19(10):3595–3617
- Hothorn T, Hornik K, van de Wiel MA, Zeileis A (2017) Package “coin”. Conditional inference procedures in a permutation test framework. ver. 1.2-2. 2017. <https://cran.r-project.org/web/packages/coin/index.html>. Accessed 02 July 2018
- Klingenberg B, Solari A, Salmaso L, Pesarin F (2009) Testing marginal homogeneity against stochastic order in multivariate ordinal data. *Biometrics* 65(2):452–462

- Korkmaz S, Goksuluk D, Zararsiz D (2014) MVN: an R package for assessing multivariate normality. *R J* 6(2):151–162
- Lee SJ, Lee S, Lim J, Ahn SJ, Kim TW (2007) Cluster analysis of tooth size in subjects with normal occlusion. *Am J Orthod Dentofac Orthop* 132(6):796–800
- Lee SH, Bachman AH, Yu D, Lim J, Ardekani BA (2016) Predicting progression from mild cognitive impairment to Alzheimers disease using longitudinal callosal atrophy. *Alzheimers Dement Diagn Assess Dis Monit* 2:68–74
- Li B, Genton MG, Sherman M (2007) A nonparametric assessment of properties of space–time covariance functions. *J Am Stat Assoc* 102(478):736–744
- Li E, Lim J, Lee S-J (2010) Likelihood ratio test for correlated multivariate samples. *J Multivar Anal* 101(3):541–554
- Li E, Lim J, Kim K, Lee S-J (2012) Distribution-free tests of mean vectors and covariance matrices for multivariate paired data. *Metrika* 75(6):833–854
- Lu N, Zimmerman DL (2005) The likelihood ratio test for a separable covariance matrix. *Stat Probab Lett* 73(4):449–457
- Mardia KV (1970) Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57(3):519–530
- Mitchell MW, Genton MG, Gumpertz ML (2006) A likelihood ratio test for separability of covariances. *J Multivar Anal* 97(5):1025–1043
- Pesarin F, Salmaso L (2010) *Permutation tests for complex data*. Wiley, New York
- Roy A, Leiva R (2008) Likelihood ratio tests for triply multivariate data with structured correlation on spatial repeated measurements. *Stat Probab Lett* 78(13):1971–1980
- Roy A, Leiva R (2011) Estimating and testing a structured covariance matrix for three-level multivariate data. *Commun Stat Theory Methods* 40(11):1945–1963
- Royston P (1983) Some techniques for assessing multivariate normality based on the Shapiro–Wilk W. *Appl Stat* 32(2):121–133
- Royston P (1992) Approximating the Shapiro–Wilk W test for non-normality. *Stat Comput* 2(3):117–119
- Shapiro SS, Wilk MB (1964) An analysis of variance test for normality (complete samples). *Biometrika* 52:591–611
- Sheng J, Qiu P (2007) p-Value calculation for multi-stage additive tests. *J Stat Comput Simul* 77(12):1057–1064
- Strasser H, Weber C (1999) On the asymptotic theory of permutation statistics. *Math Methods Stat* 8(2):220–250
- Tan KM, Witten D (2014) Sparse biclustering of transposable data. *J Comput Gr Stat* 23(4):985–1008
- Viroli C (2010) Finite mixtures of matrix normal distributions for classifying three-way data. *Stat Comput* 21(4):511–522
- Wang X, Stokes L, Lim J, Chen M (2006) Concomitant of multivariate order statistics with application to judgment post-stratification. *J Am Stat Assoc* 101:1693–1704
- Wang H, West M (2009) Bayesian analysis of matrix normal graphical models. *Biometrika* 96(4):821–834
- Yin J, Li H (2012) Model selection and estimation in the matrix normal graphical model. *J Multivar Anal* 107:119–140