Statistica Sinica Preprint No: SS-2018-0268	
Title	Hypothesis Testing for the Covariance Matrix in High-Dimensional Transposable Data with Kronecker Product Dependence Structure
Manuscript ID	SS-2018-0268
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0268
Complete List of Authors	Anestis Touloumis John C. Marioni and Simon Tavarè
Corresponding Author	Anestis Touloumis
E-mail	A.Touloumis@brighton.ac.uk
Notice: Accepted version subject to English editing.	

Hypothesis Testing for the Covariance Matrix in High-Dimensional Transposable Data with Kronecker Product Dependence Structure

Anestis Touloumis, John C. Marioni and Simon Tavaré

University of Brighton, University of Cambridge and EMBL-EBI, and Columbia University

Abstract: The matrix-variate normal distribution is a popular model for high-dimensional transposable data because it decomposes the dependence structure of the random matrix into the Kronecker product of two covariance matrices, one for each of the row and column variables. However, there is a lack of hypothesis testing procedures for the row or column covariance matrix in high-dimensional settings. Tests for assessing the sphericity, identity and diagonality hypothesis for the row (column) covariance matrix in high-dimensional settings while treating the column (row) dependence structure as a 'nuisance' parameter are introduced. The proposed tests are robust to normality departures provided that the Kronecker product dependence structure holds. In simulations, the proposed tests appear to maintain the nominal level and they tended to be powerful against the alternative hypotheses tested. The utility of the proposed tests is demonstrated by analyzing a microarray and an electroencephalogram study. The proposed testing methodology has been implemented in the R package HDTD.

Key words and phrases: Covariance matrix, high-dimensional settings, hypothesis testing, matrix-valued random variables, transposable data.

1. Introduction

Transposable data (Allen and Tibshirani, 2010) refer to matrix-valued random variables that treat the rows and columns as two distinct sets of variables of interest. To illustrate the term, consider the mouse aging atlas project (Zahn et al., 2007), where gene expression levels were measured in different tissue samples collected from multiple mice. For each mouse, the data can be organized in a $9 \times 8,932$ matrix where the rows index nine different tissues and the columns index $8,932$ genes under study. Biological questions will involve at least one of the two sets of variables, tissues and genes. For instance, one might want to infer the dependence structure among genes and/or among tissues or study the overall mean gene expression relationship across the nine tissues. Besides studies in genetics (Allen and Tibshirani, 2010, 2012; Efron, 2009; Teng and Huang, 2009; Yin and Li, 2012; Ning and Liu, 2013; Touloumis, Tavaré and Marioni, 2015), transposable data arise in electroencephalogram EEG studies (Zhang et al., 1995; Leng and Tang, 2012; Xia and Li, 2017), spatio-temporal studies (Genton, 2007; Mardia and Goodall, 1993), cross-classified multivariate data (Galecki, 1994; Naik and Rao, 2001), functional MRI (Allen and Tibshirani, 2010; Zhu and Li, 2018), financial market targeting (Leng and Tang, 2012) and in time-series (Carvalho and West, 2007; Lee, Daniels and Joo,

2013) among others.

To introduce the notation, consider N independent and identically distributed (i.i.d.) $r \times c$ random matrices $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ such that in each matrix there are r row variables and c column variables. To reflect a high-dimensional setting or equivalently the 'small sample size, large number of parameters' paradigm, assume that the sample size N is smaller than the number of observations $r \times c$ in a single matrix. The challenge with high-dimensional transposable data is to model parsimoniously the covariance structure of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ while respecting the structural information provided by presenting the data in matrix form. For this reason, the matrix-variate normal distribution (Dawid, 1981; Gupta and Nagar, 2000) is a popular choice to model high-dimensional transposable data (Allen and Tibshirani, 2010, 2012; Efron, 2009; Teng and Huang, 2009; Carvalho and West, 2007; Leng and Tang, 2012; Yin and Li, 2012; Tsiligkaridis and Hero, 2013; Zhou, 2014; Zhu and Li, 2018). It is defined by three matrix parameters, the mean matrix $\mathbf{M}$ and two positive-definite matrices $\boldsymbol{\Sigma}_{\mathrm{R}}$ and $\boldsymbol{\Sigma}_{\mathrm{C}}$. These matrices satisfy the relations $E(\boldsymbol{X}_i) = \mathbf{M}$ and $\mathrm{Cov}\left[\mathrm{vec}(\boldsymbol{X}_i)\right] = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\mathrm{C}} \otimes \boldsymbol{\Sigma}_{\mathrm{R}}$, where $\mathrm{vec}(\mathbf{A})$ vectorizes matrix $\mathbf{A}$ by its columns and $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$. In simple terms, the matrix-variate normal distribution allows

researchers to decompose the high-dimensional dependence structure into the Kronecker product of two lower-dimensional covariance matrices $\mathbf{\Sigma}_{\mathrm{C}}$ and $\mathbf{\Sigma}_{\mathrm{R}}$, recognized as the covariance matrices of the column and row variables respectively. Consequently, the covariance between two elements of $\boldsymbol{X}_i$, say $X_{ir_1c_1}$ and $X_{ir_2c_2}$, is given by

$$\mathrm{Cov}(X_{ir_1c_1}, X_{ir_2c_2}) = (\mathbf{\Sigma}_{\mathrm{R}})_{r_1r_2} (\mathbf{\Sigma}_{\mathrm{C}})_{c_1c_2}$$

for all $i = 1, \ldots, N$, $r_1, r_2 = 1, \ldots, r$ and $c_1, c_2 = 1, \ldots, c$ and where $(\mathbf{A})_{a_1a_2}$ denotes the (a_1, a_2) element of the matrix $\mathbf{A}$. To exemplify this relation, consider again the mouse aging project. Therein, $\mathbf{\Sigma}_{\mathrm{R}}$ will describe the dependence structure among tissues and $\mathbf{\Sigma}_{\mathrm{C}}$ the dependence structure among genes. Hence, the covariance between the expression levels of gene r_1 in tissue c_1 and of gene r_2 in tissue c_2 will be the product of the covariance between the two genes and of the covariance between the two tissues.

The Kronecker product covariance matrix decomposition is not necessarily an over-simplifying and convenient assumption. In fact, Hafner, Linton and Tang (2016) showed that it can approximate (in the least squares sense) the true high-dimensional covariance matrix well.

This result provides some theoretical justification on the use of the matrix-variate normal distribution (or more precisely of any distribution with a Kronecker product covariance matrix) in high-dimensional settings

with transposable data. In addition, hypothesis testing procedures (Aston, Pigoli and Tavakoli, 2017) and diagnostic plots (Ning and Liu, 2013; Yin and Li, 2012) are also available to evaluate the Kronecker product assumption for a given dataset.

To the best of our knowledge, no formal procedure exists for performing hypothesis testing for $\mathbf{\Sigma}_{\mathrm{R}}$ (or $\mathbf{\Sigma}_{\mathrm{C}}$) in high-dimensional transposable data under the matrix-variate normal distribution, while treating $\mathbf{M}$ and $\mathbf{\Sigma}_{\mathrm{C}}$ (or $\mathbf{\Sigma}_{\mathrm{R}}$) as matrix-valued nuisance parameters. To fill this gap, we consider the following three hypothesis tests: the sphericity hypothesis test

$$\mathrm{H}_0 : \mathbf{\Sigma}_{\mathrm{R}} = \sigma^2 \mathbf{I}_r \text{ vs. } \mathrm{H}_1 : \mathbf{\Sigma}_{\mathrm{R}} \neq \sigma^2 \mathbf{I}_r \tag{1.1}$$

where $\sigma^2 > 0$ is an unknown constant and $\mathbf{I}_p$ is the identity matrix of size p, the identity hypothesis test

$$\mathrm{H}_0 : \mathbf{\Sigma}_{\mathrm{R}} = \mathbf{I}_r \text{ vs. } \mathrm{H}_1 : \mathbf{\Sigma}_{\mathrm{R}} \neq \mathbf{I}_r \tag{1.2}$$

and the diagonality hypothesis test

$$\mathrm{H}_0 : \mathbf{\Sigma}_{\mathrm{R}} = \mathbf{\Delta}_{\mathbf{\Sigma}_{\mathrm{R}}} \text{ vs. } \mathrm{H}_1 : \mathbf{\Sigma}_{\mathrm{R}} \neq \mathbf{\Delta}_{\mathbf{\Sigma}_{\mathrm{R}}} \tag{1.3}$$

where $\mathbf{\Delta}_{\mathbf{A}}$ denotes the diagonal matrix with diagonal elements the corresponding elements of $\mathbf{A}$. This suggests that the diagonality hypothesis test can also be written as:

$$\mathrm{H}_0 : (\mathbf{\Sigma}_{\mathrm{R}})_{r_1 r_2} = 0 \text{ for all } r_1 \neq r_2 \text{ vs. } \mathrm{H}_1 : \text{ not } \mathrm{H}_0 \,. \tag{1.4}$$

To illustrate the practical importance of testing these three hypotheses, consider first the diagonality test. The null hypothesis implies independence of the row variables in such a way that the transposable data can be written in terms of r independent populations, one for each row. In particular, the r_1-th population consists of N c-variate random vectors with mean vector the r_1-th row of $\mathbf{M}$ and covariance matrix $(\mathbf{\Sigma}_{\mathrm{R}})^{-1}_{r_1 r_1}\,\mathbf{\Sigma}_{\mathrm{C}}$. Therefore, the diagonality hypothesis test under the matrix-variate normal model is equivalent to testing whether the r row random vectors are independently distributed with proportional covariance matrices but not necessarily a common mean vector.

Next, consider the sphericity test. The null hypothesis is more restrictive since it requires the r independent populations to have identical covariance matrix (equal to $\sigma^{-2}\mathbf{\Sigma}_{\mathrm{C}}$) and thus, it can be utilized to explore whether the r rows are independently distributed with common covariance matrix but varying mean vectors. Another use of the sphericity hypothesis test is to assess indirectly whether a known row covariance matrix $\mathbf{\Sigma}_{R0}$ equals the row-wise covariance structure $\mathbf{\Sigma}_R$, that is testing

$$\mathrm{H}_0 : \mathbf{\Sigma}_{\mathrm{R}} = \mathbf{\Sigma}_{\mathrm{R0}} \text{ vs. } \mathrm{H}_1 : \mathbf{\Sigma}_{\mathrm{R}} \neq \mathbf{\Sigma}_{\mathrm{R0}}\,.$$

To accomplish this, one must apply the transformation $\boldsymbol{X}_i \longmapsto \mathbf{\Sigma}_{\mathrm{R0}}^{-1/2}\boldsymbol{X}_i$ and then test the sphericity hypothesis on the transformed random matri-

ces. In this case, the constant σ^2 is the normalizing constant that makes the $\mathbf{\Sigma}_{\mathrm{R}}$ and $\mathbf{\Sigma}_{\mathrm{C}}$ identifiable (see also Section 2).

To this end, consider now the identity test. The null hypothesis implies that all row variances are equal to 1. This test is useful only in studies where transposable data for each subject have been preprocessed in such a way that the measurements across column and/or row variables have sample mean zero and unit variance. Examples of column- and/or doubly standardized data can be found in microarray studies (Efron, 2009).

It is not straightforward to assess hypothesis tests (1.1), (1.2) or (1.3) by applying existing testing procedures for high-dimensional covariance matrices of random vectors such as the testing procedures of Chen, Zhang and Zhong (2010) and Srivastava, Yanagihara and Kubokawa (2014) among others. For more detailed literature on testing the covariance structure with high-dimensional random vectors see, for example, Ahmad and von Rosen (2015). Unfortunately, these methods do not account for the presence of a column-wise dependence structure and/or an unrestricted mean matrix $\mathbf{M}$. In preliminary simulations (see Section 10 in the Supplementary Materials), we have found such tests approximate the nominal size only when the column variables were indeed independent ($\mathbf{\Sigma}_{\mathrm{C}} = \mathbf{I}_c$) and a constant r-variate mean $\boldsymbol{\mu}$ vector holds for the row variables. Otherwise, they led

to inflated sizes, for example always falsely rejecting the null hypothesis in the presence of moderate to strong column-wise correlation pattern and/or more complicated forms of the mean matrix.

To address this issue, we extend the work of Chen, Zhang and Zhong (2010) to matrix-variate distributions. In all cases, we estimate a (scaled) squared Frobenius norm that measures the distance between the corresponding null and alternative hypotheses for $\mathbf{\Sigma}_{\mathrm{R}}$ while treating $\mathbf{M}$ and $\mathbf{\Sigma}_{\mathrm{C}}$ as 'nuisance' matrix parameters. This is reasonable because the squared Frobenius norm of the difference of the Kronecker product $\mathbf{\Sigma}_{\mathrm{C}} \otimes \mathbf{\Sigma}_{\mathrm{R}}$ under the sphericity, identity or diagonality hypothesis and the corresponding alternative hypothesis depends only on the squared Frobenius norm for $\mathbf{\Sigma}_{\mathrm{R}}$. Next, the unknown parameters in the squared Frobenius norms will be replaced by unbiased and/or consistent estimators. This allows us to derive the general asymptotic distributions of the proposed test statistics and hence, to explore their asymptotic power. In addition, we show that the proposed tests are nonparametric, meaning that under suitable conditions they can account for some departures from the matrix-variate normal distribution.

It is important to emphasize that the methods developed here can manage the high-dimensional setting in a very parsimonious and efficient way.

The proposed test statistics are computationally cheap since their construction relies on estimating just four parameters: $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}})$, $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}}^2)$, $\mathrm{tr}(\boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}}^2)$ and $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{C}}^2)$. Explicit estimation of $r(r-1)/2 + c(c-1)/2$ non-redundant elements in $\boldsymbol{\Sigma}_{\mathrm{R}}$ and $\boldsymbol{\Sigma}_{\mathrm{C}}$ is avoided, a cumbersome task for a large number of rows and/or columns. To appreciate the computational gains, assume that we want to test the dependence structure for the tissues in the mouse aging example. Full estimation of the mean matrix and the gene covariance matrix requires estimation of $1,140$ non-redundant nuisance parameters while the proposed methods need to account only for the gene-covariance matrix via a single parameter $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{C}}^2)$, which can be consistently estimated.

We want to underline that the role of the row and column variables can be interchanged, which implies that if the interest lies in applying the sphericity, identity or diagonality hypothesis test to the column covariance matrix, then the transformation $\boldsymbol{X}_i \longmapsto \boldsymbol{X}_i'$ should be performed prior to carrying out the test on the transformed data. In other words, the proposed tests can be applied to $\boldsymbol{\Sigma}_{\mathrm{C}}$ by simply transposing the data matrices.

This paper is organized as follows. In Section 2, we present the working framework that allows us to handle and develop test statistics with high-dimensional transposable data under a Kronecker product patterned covariance matrix in a nonparametric manner. In Section 3, we propose

tests for assessing the sphericity, identity and diagonality hypotheses of the row (or column) covariance matrix. For each of the tests proposed, we derive the general asymptotic distribution, indicate the rejection region and provide a lower bound for the asymptotic analysis. Further, we point to a software implementation of our methods. In Section 4, we demonstrate the good performance of the proposed tests in simulation studies. In Section 5, we apply the test statistics to the mouse aging dataset and an EEG dataset. We summarize our findings and discuss future research in Section 6. Technical details can be found in the Supplementary Materials.

2. Notation and Assumptions

Suppose there are r row variables and c column variables and assume that $r \times c$ random matrices $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ are generated by the matrix-valued nonparametric model

$$\boldsymbol{X}_i = \boldsymbol{\Sigma}_{\mathrm{R}}^{1/2} \boldsymbol{Z}_i \boldsymbol{\Sigma}_{\mathrm{C}}^{1/2} + \mathbf{M}\,, \tag{2.1}$$

where

- $\boldsymbol{\Sigma}_{\mathrm{R}} = \boldsymbol{\Sigma}_{\mathrm{R}}^{1/2} \boldsymbol{\Sigma}_{\mathrm{R}}^{1/2}$ is the $r \times r$ row covariance matrix.
- $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N$ are $r \times c$ i.i.d. random matrices and $Z_{ir_1c_1}$ denotes the (r_1, c_1) element of $\boldsymbol{Z}_i$ for $r_1 = 1, \ldots, r$ and $c_1 = 1, \ldots, c$.
- $E\left(Z_{ir_1c_1}\right) = 0$, $\mathrm{Var}\left(Z_{ir_1c_1}\right) = 1$, $E(Z_{ir_1c_1}^4) = 3 + B$ for a finite constant

$B \geq -2$, $E(Z_{ir_1c_1}^8) < \infty$ and for any positive integers $l_1, \ldots, l_q$ with $\sum_{v=1}^{q} l_v \leq 8$

$$E\left(\prod_{k=1}^{q} Z_{ir_kc_k}\right) = \prod_{k=1}^{q} E(Z_{ir_kc_k})$$

for $(r_1, c_1) \neq \ldots \neq (r_q, c_q)$. Thus, the elements of $\mathbf{Z}_i$ can be viewed as white noise that can also accommodate weak dependence patterns.

- $\mathbf{\Sigma}_{\mathrm{C}} = \mathbf{\Sigma}_{\mathrm{C}}^{1/2}\mathbf{\Sigma}_{\mathrm{C}}^{1/2}$ is the $c \times c$ column covariance matrix such that $\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}) = c$, where $\mathrm{tr}(\mathbf{A})$ denotes the trace of the matrix $\mathbf{A}$.

- $\mathbf{M} = E(\boldsymbol{X}_i)$ is the $r \times c$ mean matrix.

Model (2.1) is a special case of the nonparametric matrix-valued model for transposable data employed in Touloumis, Tavaré and Marioni (2015) with $\mathbf{\Sigma} = \mathbf{\Sigma}_{\mathrm{C}} \otimes \mathbf{\Sigma}_{\mathrm{R}}$, where $\mathbf{\Sigma}$ is the covariance matrix of $\boldsymbol{x}_i = \mathrm{vec}(\boldsymbol{X}_i)$, the vectorized form of $\boldsymbol{X}_i$. Hence, it contains the matrix-variate normal distribution as a member ($B = 0$), preserves the interpretation of $\mathbf{\Sigma}_{\mathrm{R}}$ and $\mathbf{\Sigma}_{\mathrm{C}}$ as row and column covariance matrices respectively, and allows consideration of some non-normal distributions, such as the members of the elliptically contoured family of distributions and of the independent component model (Oja, 2010) subject to a Kronecker product covariance decomposition.

The trace restriction on $\mathbf{\Sigma}_{\mathrm{C}}$ makes the two covariance matrices identifiable since otherwise we have $\mathbf{\Sigma} = (t\mathbf{\Sigma}_{\mathrm{C}}) \otimes (\mathbf{\Sigma}_{\mathrm{R}}/t)$ for any $t > 0$. In

the context of the matrix-variate normal distribution, this issue has been resolved by setting either $\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}) = c$ (Mardia and Goodall, 1993; Theobald and Wuttke, 2006) or a diagonal element of $\mathbf{\Sigma}_{\mathrm{C}}$ equal to 1 (Naik and Rao, 2001; Srivastava, von Rosen and von Rosen, 2008; Yin and Li, 2012). Although none of these constraints affects the row and column correlation patterns, we adopt the former because it eases the construction of unbiased and/or consistent estimators of the parameters that we base the proposed test statistics upon.

To manage high-dimensional settings, we assume that as $N \to \infty$,

$$rc = r(N)c(N) \to \infty \ , \ N = O(rc) \ , \ \frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{m}}^4)}{\mathrm{tr}^2(\mathbf{\Sigma}_{\mathrm{m}}^2)} \to 0 \text{ for } \mathrm{m} \in \{\mathrm{R}, \mathrm{C}\} \, . \quad (2.2)$$

Assumption (2.2) specifies neither the pairwise limiting ratios of the triplet (N, r, c) nor the rate at which $r \to \infty$ and $c \to \infty$. Thus it covers applications in which: i) the sample size might not be expected to increase proportionally to the dimension of the transposable data matrices and ii) r and/or c tend to ∞ a lot faster than N. These situations were tested in the simulation study, where the proposed tests appeared to behave well. Assumption (2.2) does not seriously limit the scope of the row and column covariance structures under consideration. Covariance matrices with eigenvalues bounded away from 0 and ∞ (Chen, Zhang and Zhong, 2010), that satisfy a first-order autoregressive correlation pattern with bounded

variances (Chen, Zhang and Zhong, 2010), or that have a few divergent eigenvalues as long as they diverge slowly (Chen and Qin, 2010) all satisfy the trace ratio restrictions in (2.2). Therefore, model (2.1) and assumption (2.2) constitute a flexible working framework that allows us to handle a wide range of studies with high-dimensional transposable data.

3. Test Statistics

To construct the proposed test statistics, we need to estimate $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}})$, $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}}^2)$, $\mathrm{tr}(\boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}}^2) = \mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}} \circ \boldsymbol{\Sigma}_{\mathrm{R}})$ and $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{C}}^2)$, where $\mathbf{A} \circ \mathbf{B}$ is the Hadamard product of the matrices $\mathbf{A}$ and $\mathbf{B}$. Before introducing the test statistics, we present unbiased and/or consistent estimators of these parameters and we discuss some computational aspects.

3.1 Parameter estimators

The parameters $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}})$, $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}}^2)$ and $\mathrm{tr}(\boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}}^2)$ can be estimated by

$$
\begin{aligned}
T_{1N} =& Y_{1N} - Y_{4N} = \frac{1}{cN}\sum_{i=1}^{N} \mathrm{tr}(\boldsymbol{X}_i\boldsymbol{X}_i') - \frac{1}{cP_2^N}\sum_{i,j}^{*} \mathrm{tr}(\boldsymbol{X}_i\boldsymbol{X}_j')\,,\\
T_{2N} =& Y_{2N} - 2Y_{5N} + Y_{6N}\\
=& \frac{1}{c^2P_2^N}\sum_{i,j}^{*} \mathrm{tr}(\boldsymbol{X}_i\boldsymbol{X}_i'\boldsymbol{X}_j\boldsymbol{X}_j') - 2\frac{1}{c^2P_3^N}\sum_{i,j,k}^{*} \mathrm{tr}(\boldsymbol{X}_i\boldsymbol{X}_i'\boldsymbol{X}_j\boldsymbol{X}_k')\\
&+ \frac{1}{c^2P_4^N}\sum_{i,j,k,l}^{*} \mathrm{tr}(\boldsymbol{X}_i\boldsymbol{X}_j'\boldsymbol{X}_k\boldsymbol{X}_l')\,,
\end{aligned}
$$

and

$$\begin{aligned} T_{3N} =& Y_{6N} - 2Y_{7N} + Y_{8N} \\ =& \frac{1}{c^2 P_2^N} \sum_{i,j}^{*} \mathrm{tr}[(\boldsymbol{X}_i \boldsymbol{X}_i') \circ (\boldsymbol{X}_j \boldsymbol{X}_j')] - 2 \frac{1}{c^2 P_3^N} \sum_{i,j,k}^{*} \mathrm{tr}[(\boldsymbol{X}_i \boldsymbol{X}_i') \circ (\boldsymbol{X}_j \boldsymbol{X}_k')] \\ &+ \frac{1}{c^2 P_4^N} \sum_{i,j,k,l}^{*} \mathrm{tr}[(\boldsymbol{X}_i \boldsymbol{X}_j') \circ (\boldsymbol{X}_k \boldsymbol{X}_l')] \end{aligned}$$

respectively, where $P_t^s = \prod_{k=0}^{t}(s-k)$ and $\sum^*$ denotes summation over mutually distinct indices. The terms Y_{1N}, Y_{2N}, Y_{3N} in T_{1N}, T_{2N} and T_{3N} are the unbiased estimators of the targeted parameters when $\mathbf{M} = \mathbf{0}$ while the terms Y_{4N}, Y_{5N}, Y_{6N}, Y_{7N} and Y_{8N} are U-statistics of order two, three and four that are subtracted so that T_{1N}, T_{2N} and T_{3N} remain unbiased even when $\mathbf{M} \neq \mathbf{0}$. To the best of our knowledge, Chen, Zhang and Zhong (2010) first exploited this usage of U-statistics for constructing test statistics.

To estimate $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{C}}^2)$, we utilize the vectorized form of model (2.1) and write $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{C}}^2) = \mathrm{tr}(\boldsymbol{\Sigma}^2)/\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}}^2)$. To estimate $\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{C}}^2)$ we will use $T_{5N} = T_{4N}/T_{2N}$, that is the ratio of an unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$

$$T_{4N} = \frac{1}{P_2^N} \sum_{i,j}^{*} (\boldsymbol{x}_i' \boldsymbol{x}_j)^2 - 2 \frac{1}{P_3^N} \sum_{i,j,k}^{*} \boldsymbol{x}_i' \boldsymbol{x}_j \boldsymbol{x}_i' \boldsymbol{x}_k + \frac{1}{P_4^N} \sum_{i,j,k,l}^{*} \boldsymbol{x}_i' \boldsymbol{x}_j \boldsymbol{x}_k' \boldsymbol{x}_l \,,$$

to T_{2N}, an unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$. Theorem 1 establishes that T_{1N}, T_{2N}, T_{4N} and T_{5N} are all ratio-consistent estimators of the targeted parameters (a general statistic $\widehat{\theta}_N$ is a ratio-consistent estimator to the parameter

θ if $\widehat{\theta}_N/\theta$ converges in probability to one) and that T_{3N} is a ratio-consistent estimator of $\mathrm{tr}(\boldsymbol{\Delta}^2_{\boldsymbol{\Sigma}_\mathrm{R}})$ under H_0 in the diagonality hypothesis test (1.3).

Theorem 1. *Under model (2.1) and assumption (2.2):*

$$\frac{T_{1N}}{\mathrm{tr}(\boldsymbol{\Sigma}_\mathrm{R})} \xrightarrow{P} 1, \frac{T_{2N}}{\mathrm{tr}(\boldsymbol{\Sigma}^2_\mathrm{R})} \xrightarrow{P} 1, \frac{T_{4N}}{\mathrm{tr}(\boldsymbol{\Sigma}^2)} \xrightarrow{P} 1, \frac{T_{5N}}{\mathrm{tr}(\boldsymbol{\Sigma}^2_\mathrm{C})} \xrightarrow{P} 1\,,$$

where $\xrightarrow{P}$ *denotes convergence in probability and*

$$\frac{\mathrm{Var}\,(T_{3N})}{\mathrm{tr}^2(\boldsymbol{\Sigma}^2_\mathrm{R})} \to 0\,.$$

Thus, when $\boldsymbol{\Sigma}_\mathrm{R} = \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_\mathrm{R}}$ *we have that*

$$\frac{T_{3N}}{\mathrm{tr}(\boldsymbol{\Delta}^2_{\boldsymbol{\Sigma}_\mathrm{R}})} \xrightarrow{P} 1\,.$$

From a computational perspective, it is worth noting that equivalent formulae for T_{2N}, T_{3N} and T_{4N} available in the Supplementary Material and the cyclic property applied on the trace operators when $r > c$ can significantly reduce the order of calculations of T_{2N}, T_{3N} and T_{4N} from $N^4r^2(r+2c)$ to $N^2\min\{r,c\}^2(\min\{r,c\}+2\max\{r,c\})$. In the special case of centered transposable data matrices ($\mathbf{M} = \mathbf{0}$), further reductions in the computational time can be gained by employing only the first terms in T_{1N}, T_{2N}, T_{3N} and T_{4N}.

3.2 Sphericity test

The proposed test relies on the general limiting distribution of

$$U_N = r\frac{T_{2N}}{T_{1N}^2} - 1\,,$$

a ratio-consistent estimator to the scaled Frobenius norm

$$\frac{1}{r}\mathrm{tr}\left[\frac{\mathbf{\Sigma}_{\mathrm{R}}}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}})/r} - \mathbf{I}_r\right]^2 = r\frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}}^2)}{\mathrm{tr}^2(\mathbf{\Sigma}_{\mathrm{R}})} - 1\,,$$

which measures the distance between the null and alternative hypothesis in the sphericity hypothesis test (1.1), which equals zero if and only if the null hypothesis is true. Let

$$\begin{aligned}\sigma_{U_N}^2 =& \frac{4}{N^2}\left[\frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}{c^2}\right]^2 + \frac{8}{N}\frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}{c^2}\mathrm{tr}\left\{\left[\frac{\mathbf{\Sigma}_{\mathrm{R}}^2}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}}^2)} - \frac{\mathbf{\Sigma}_{\mathrm{R}}}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}})}\right]^2\right\} \\ &+ \frac{4B}{N}\frac{\mathrm{tr}(\mathbf{\Delta}_{\mathbf{\Sigma}_{\mathrm{C}}}^2)}{c^2}\mathrm{tr}\left\{\left[\frac{\mathbf{\Sigma}_{\mathrm{R}}^2}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}}^2)} - \frac{\mathbf{\Sigma}_{\mathrm{R}}}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}})}\right] \circ \left[\frac{\mathbf{\Sigma}_{\mathrm{R}}^2}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}}^2)} - \frac{\mathbf{\Sigma}_{\mathrm{R}}}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}})}\right]\right\}.\end{aligned}$$

Since $-2 \leq B$, $\mathrm{tr}\left(\mathbf{\Delta}_{\mathbf{A}}^2\right) = \mathrm{tr}(\mathbf{A} \circ \mathbf{A}) \leq \mathrm{tr}(\mathbf{A}^2)$ for any symmetric matrix $\mathbf{A}$ and $\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2) \leq c^2$, it follows that $\sigma_{U_N}^2 > 0$.

Theorem 2. *Under model (2.1) and assumption (2.2)*

$$\sigma_{U_N}^{-1}\left[\frac{\mathrm{tr}^2(\mathbf{\Sigma}_{\mathrm{R}})}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}}^2)}\frac{U_N+1}{r} - 1\right] \xrightarrow{d} \mathrm{N}(0,1)$$

where $\xrightarrow{d}$ *denotes convergence in distribution and* $\mathrm{N}(0,1)$ *denotes the standard normal distribution.*

Under H_0 in the sphericity hypothesis test (1.1), $\sigma^2_{U_N}$ reduces to

$$\frac{4}{N^2}\left[\frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}{c^2}\right]^2.$$

In most applications, $\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)$ will be unknown but it can be replaced by its ratio-consistent estimator T_{5N}. Hence, Slutsky's Theorem and Theorems 1 and 2 imply that a test with nominal α level of significance rejects H_0 in the sphericity hypothesis test (1.1) when

$$\frac{N-1}{2}\frac{c^2}{T_{5N}}U_N \geq z_{1-\alpha}\,,$$

where z_p is the p quantile of $\mathrm{N}(0,1)$. The scaling factor $(N-1)/N$ serves as a precaution against inflated empirical sizes in finite samples and it is motivated by the work of Mao (2016), who compared U-statistics based testing procedures for assessing the sphericity and identity hypothesis test for the covariance matrix of high-dimensional vector-valued random variables. It is a basically a correction in the asymptotic variance of U_N that accounts for estimating the mean matrix $\mathbf{M}$ by the sample mean matrix in T_{1N} and T_{2N} (see the corresponding alternative formulae available in the Supplementary Materials). In addition, T_{3N} depends on the sample mean matrix and for this reason, we will also apply the $(N-1)/N$ correction to the identity and the diagonality test.

The asymptotic normality of U_N also permits us to investigate the power

of the proposed test. In this direction, let

$$0 \leq \xi_{1N} = 1 - \frac{1}{r}\frac{\mathrm{tr}^2(\mathbf{\Sigma}_{\mathrm{R}})}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}}^2)} < 1$$

and

$$\xi_{2N} = \mathrm{tr}\left\{\left[\frac{\mathbf{\Sigma}_{\mathrm{R}}^2}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}}^2)} - \frac{\mathbf{\Sigma}_{\mathrm{R}}}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}})}\right]^2\right\},$$

and note that for large N

$$\frac{4}{N^2}\left[\frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}{c^2}\right]^2 \leq \sigma_{U_N}^2 \leq \frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}{c^2}\left[\frac{4}{N^2} + \frac{4(2+B)}{N}\xi_{2N}\right].$$

Theorem 3. *Under model (2.1) and assumption (2.2)*

$$\liminf_N \beta_N^S \geq 1 - \Phi\left(z_{1-a} - \frac{1}{2}\liminf_N \sqrt{\frac{c^2}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}\frac{N^2\xi_{1N}}{1+(2+B)N\xi_{2N}}}\right),$$

where β_N^S is the power function of the proposed sphericity test and Φ is the cumulative distribution function of $\mathrm{N}(0,1)$.

Theorem 3 states that the proposed sphericity test is consistent as long as

$$\liminf_N \sqrt{\frac{c^2}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}\frac{N^2\xi_{1N}}{1+(2+B)N\xi_{2N}}} = \infty.$$

This does not impose severe restrictions for the row covariance. For example, the test is consistent provided that ξ_{1N} and ξ_{2N} are both bounded away from 0 and that ξ_{2N} is bounded away from ∞. Theorem 3 also implies that in finite samples and when conditioning on the remaining parameters, the

strength of column-wise correlation might affect the power of the proposed test. Heuristically, we expect weak column-wise correlation patterns to increase the power of the proposed test since the asymptotic lower bound of β_N^S takes its maximum value when $\mathbf{\Sigma}_{\mathrm{C}} = \mathbf{I}_c$ since $\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2) \leq \mathrm{tr}(\mathbf{I}_c^2) = c$.

3.3 Identity test

For the identity hypothesis test (1.2), consider

$$V_N = T_{2N} - 2T_{1N} + r$$

an unbiased estimator of the squared Frobenius norm $\mathrm{tr}\left[(\mathbf{\Sigma}_{\mathrm{R}} - \mathbf{I}_r)^2\right] = \mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}}^2) - 2\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}}) + r$, that equals zero if and only if the null hypothesis holds. Let

$$\begin{aligned}\sigma_{V_N}^2 =& \frac{4}{N(N-1)}\left[\frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}{c^2}\right]^2 \mathrm{tr}^2(\mathbf{\Sigma}_{\mathrm{R}}^2) + \frac{8}{N}\frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}{c^2}\left[\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}}^2 - \mathbf{\Sigma}_{\mathrm{R}})^2\right] \\ &+ \frac{4B}{N}\frac{\mathrm{tr}(\mathbf{\Delta}_{\mathbf{\Sigma}_{\mathrm{C}}}^2)}{c^2}\mathrm{tr}\left[(\mathbf{\Sigma}_{\mathrm{R}}^2 - \mathbf{\Sigma}_{\mathrm{R}}) \circ (\mathbf{\Sigma}_{\mathrm{R}}^2 - \mathbf{\Sigma}_{\mathrm{R}})\right] > 0\,.\end{aligned}$$

Theorem 4 proves that $\sigma_{V_N}^2$ is the asymptotic variance term of V_N and consequently we can derive the general asymptotic distribution of V_N.

Theorem 4. *Under model (2.1) and assumption (2.2), it follows that* $\mathrm{Var}(V_n) = \sigma_{V_N}^2\{1 + o(1)\}$. *Further,*

$$\frac{V_N - \mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}} - \mathbf{I}_r)^2}{\sigma_{V_N}} \xrightarrow{d} \mathrm{N}(0,1)\,.$$

Slutsky's Theorem and Theorems 1 and 4 imply that a test with nominal

α level of significance rejects H_0 in the identity hypothesis test (1.2) when

$$\frac{N-1}{2}\frac{c^2}{T_{5N}}\frac{1}{r}V_N \geq z_{1-\alpha}\,.$$

To investigate the asymptotic power of the proposed test, we need to introduce additional notation. Let

$$\xi_{3N} = \frac{1}{r}\mathrm{tr}\left[(\mathbf{\Sigma}_{\mathrm{R}} - \mathbf{I}_r)^2\right]$$

and

$$\xi_{4N} = \frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}}^2)}{N\mathrm{tr}\left[(\mathbf{\Sigma}_{\mathrm{R}} - \mathbf{I}_r)^2\right]}\,.$$

Since $\mathrm{tr}\left[(\mathbf{\Sigma}_{\mathrm{R}} - \mathbf{I}_r)^2\right] \leq \mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}} - \mathbf{I}_r)\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{R}})$ we obtain that for large N

$$4\left[\frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}{c^2}\right]^2 r^2\xi_{3N}^2\xi_{4N}^2 \leq \sigma_{V_N}^2 \leq 4\frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}{c^2}\mathrm{tr}\left[(\mathbf{\Sigma}_{\mathrm{R}} - \mathbf{I}_r)^2\right]\left[\xi_{4N}^2 + (2+B)\xi_{4N}\right]\,.$$

Theorem 5. *Under model (2.1) and assumption (2.2)*

$$\liminf_N \beta_N^I \geq 1 - \limsup_N \Phi\left(\frac{z_{1-\alpha}}{N\xi_{3N}\xi_{4N}} - \frac{1}{2}\sqrt{\frac{c^2}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{C}}^2)}\frac{1}{\xi_{4N}^2 + (2+B)\xi_{4N}}}\right)$$

where β_N^I is the power function of the proposed identity test.

Theorem 5 suggests that the proposed test is consistent under mild conditions about the row covariance matrix, for example, whenever ξ_{3N} and ξ_{4N} are bounded away from 0. Similar to the proposed sphericity test, the proposed identity test is expected to be more powerful in the presence of weak rather than strong column-wise correlation pattern.

3.4 Diagonality test

A test statistic for assessing the diagonality hypothesis test (1.3) or (1.4) can be constructed by following a similar strategy. In particular, consider

$$W_N = T_{2N} - T_{3N}\,,$$

an unbiased estimator of the squared Frobenius norm

$$\mathrm{tr}\left[(\boldsymbol{\Sigma}_{\mathrm{R}} - \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}})^2\right] = \mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}}^2) - 2\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}}\boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}}) + \mathrm{tr}(\boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}}^2) = \mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{R}}^2) - \mathrm{tr}(\boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}}^2)$$

that equals zero if and only if the null hypothesis in the diagonality hypothesis test (1.3) holds. The asymptotic variance of W_N is

$$\begin{aligned}\sigma_{W_N}^2 =& \frac{4}{N^2}\left[\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{C}}^2)}{c^2}\right]^2 \mathrm{tr}^2(\boldsymbol{\Sigma}_{\mathrm{R}}^2) + \frac{8}{N}\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{C}}^2)}{c^2}\mathrm{tr}\left[\boldsymbol{\Sigma}_{\mathrm{R}}(\boldsymbol{\Sigma}_{\mathrm{R}} - \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}})\boldsymbol{\Sigma}_{\mathrm{R}}(\boldsymbol{\Sigma}_{\mathrm{R}} - \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}})\right] \\ &+ \frac{4B}{N}\frac{\mathrm{tr}(\boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{C}}}^2)}{c^2}\mathrm{tr}\left\{\left[\boldsymbol{\Sigma}_{\mathrm{R}}^{1/2}(\boldsymbol{\Sigma}_{\mathrm{R}} - \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}})\boldsymbol{\Sigma}_{\mathrm{R}}^{1/2}\right] \circ \left[\boldsymbol{\Sigma}_{\mathrm{R}}^{1/2}(\boldsymbol{\Sigma}_{\mathrm{R}} - \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}})\boldsymbol{\Sigma}_{\mathrm{R}}^{1/2}\right]\right\}.\end{aligned}$$

Theorem 6. *Under model (2.1) and assumption (2.2), it follows that* $\mathrm{Var}(W_N) = \sigma_{W_N}^2\{1 + o(1)\}$. *Further,*

$$\frac{W_N - \mathrm{tr}\left[(\boldsymbol{\Sigma}_{\mathrm{R}} - \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}})^2\right]}{\sigma_{W_N}} \xrightarrow{d} \mathrm{N}(0,1)\,.$$

As before, the general asymptotic distribution of W_N in Theorem 6 will be used to find a rejection area. Slutsky's Theorem and Theorems 1 and 6 imply that a test with nominal α level of significance rejects H_0 in the diagonality hypothesis test (1.3) when

$$\frac{N-1}{2}\frac{c^2}{T_{5N}}\frac{1}{T_{3N}}W_N \geq z_{1-\alpha}\,.$$

To investigate the asymptotic power of the proposed test, let

$$0 \leq \xi_{5N} = \frac{\operatorname{tr}\left[(\boldsymbol{\Sigma}_{\mathrm{R}} - \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}})^2\right]}{\operatorname{tr}(\boldsymbol{\Sigma}_{\mathrm{R}}^2)} = 1 - \frac{\operatorname{tr}(\boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}}^2)}{\operatorname{tr}(\boldsymbol{\Sigma}_{\mathrm{R}}^2)} < 1$$

and note that for large N

$$\frac{4}{N^2}\left[\frac{\operatorname{tr}(\boldsymbol{\Sigma}_{\mathrm{C}}^2)}{c^2}\right]^2 \operatorname{tr}^2(\boldsymbol{\Sigma}_{\mathrm{R}}^2) \leq \sigma_{W_N}^2 \leq \frac{\operatorname{tr}(\boldsymbol{\Sigma}_{\mathrm{C}}^2)}{c^2}\operatorname{tr}^2(\boldsymbol{\Sigma}_{\mathrm{R}}^2)\left[\frac{4}{N^2} + \frac{4(8+B)}{N}\right].$$

Theorem 7. *Under model (2.1) and assumption (2.2)*

$$\liminf_N \beta_N^D \geq 1 - \Phi\left(z_{1-a} - \frac{1}{2}\liminf_N \sqrt{\frac{c^2}{\operatorname{tr}(\boldsymbol{\Sigma}_{\mathrm{C}}^2)}\frac{N^2\xi_{5N}}{1+(8+B)N}}\right)$$

where β_N^D is the power of the proposed diagonality test.

Note that ξ_{5N} converges to 0 if all elements of $\boldsymbol{\Sigma}_{\mathrm{R}} - \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}}$ converge to zero. In this case, $\operatorname{tr}\left[(\boldsymbol{\Sigma}_{\mathrm{R}} - \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}})^2\right] \to 0$ and hence the proposed test is expected to suffer power loss. On the other hand, the test will be asymptotically consistent provided that $\boldsymbol{\Sigma}_{\mathrm{R}}$ and $\boldsymbol{\Delta}_{\boldsymbol{\Sigma}_{\mathrm{R}}}$ differ in at least one element as $N \to \infty$ and $r \to \infty$ as long as this difference is bounded away from zero and regardless of its magnitude.

3.5 Special Cases

When the subject-specific data are vector-valued instead of matrix-valued ($c = 1$), it can be the shown that the proposed sphericity and identity tests reduce to the corresponding sphericity and identity tests proposed by Srivastava, Yanagihara and Kubokawa (2014), which Mao (2016) showed

are the same except for a scale factor to those proposed by Chen, Zhang and Zhong (2010). Further, the proposed diagonality test is asymptotically equivalent, but not identical, to the bandness test with fixed bandwidth equal to 1 proposed by Qiu and Chen (2012).

When the column features are independent, in which case $\mathbf{\Sigma}_{\mathrm{C}} = \mathbf{I}_c$, and $\mathbf{M} = \boldsymbol{\mu}\mathbf{1}_c$ for an r-dimensional mean vector $\boldsymbol{\mu}$, then the proposed tests are asymptotically equivalent to the corresponding test statistics of Srivastava, Yanagihara and Kubokawa (2014), Chen, Zhang and Zhong (2010) and Qiu and Chen (2012) when treating the columns as independent. However, if $\mathbf{M} \neq \boldsymbol{\mu}\mathbf{1}_c$ the asymptotic equivalence between the proposed tests and the existing vector-based tests no longer holds.

3.6 Software availability

The function `covmat.ts()` of the R package HDTD (Touloumis, Marioni and Tavaré, 2016) implements the proposed sphericity, identity and diagonality tests. These can be applied to either the row or column covariance matrix by specifying the `voi` argument. The software is available from the Bioconductor repository at http://bioconductor.org/packages/HDTD/.

4. Simulations

We investigated the performance of the proposed procedures for testing hypotheses (1.1), (1.2) and (1.3) via numerical studies. Due to the location

invariance property of the proposed test statistics, we generated $r \times c$ matrix-variate random variables $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ according to model (2.1) with $\mathbf{M} = \mathbf{0}$. To assess the nonparametric nature, we simulated $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N$ under a standard matrix-variate normal scenario, where $Z_{ir_1c_1} \overset{i.i.d}{\sim} \mathrm{N}(0,1)$ such that $B = 0$, and under three standardized Gamma scenarios, where $Z_{ir_1c_1} = \left(Z^*_{ir_1c_1} - \alpha/\beta\right)/\sqrt{\beta}$ with $Z^*_{ir_1c_1} \overset{i.i.d}{\sim} \mathrm{Gamma}(\alpha, \beta)$: (i) $\mathrm{Gamma}(1, 0.5)$ such that $B = 6$, (ii) $\mathrm{Gamma}(0.6, 1)$ such that $B = 10$, and (iii) $\mathrm{Gamma}(0.3, 1)$ such that $B = 20$. To reflect high-dimensional settings, we considered $N = 20, 40, 60, 100, 200$, $r = 10, 50, 100, 300, 600$ and $c = 10, 100, 600$ so that the number of subject-specific observations $(r \times c)$ was larger than the sample size (N) in all instances except when $N = 200$ and $r = c = 10$, without specifying the relationship among N, r and c. For the "nuisance" covariance matrix $\mathbf{\Sigma}_{\mathrm{C}}$, we employed a first order autoregressive correlation matrix with elements $\left(\mathbf{\Sigma}_{\mathrm{C}}\right)_{c_1c_2} = 0.85^{|c_1 - c_2|}$. This configuration generated complex pairwise correlation patterns in which the strength of the pairwise correlation among the column variables varied from moderate to strong $(c = 10)$ and from weak to strong $(c = 100, 600)$.

We employed identity, heteroscedastic (2-3) and tridiagonal (4-5) structures for the row covariance matrix $\mathbf{\Sigma}_{\mathrm{R}}$:

1. The identity matrix $\mathbf{\Sigma}_{\mathrm{R}} = \mathbf{I}_r$.

2. Diagonal $\mathbf{\Sigma}_{\mathrm{R}}$ with $(\mathbf{\Sigma}_{\mathrm{R}})_{r_1r_1} \overset{i.i.d}{\sim} \mathrm{U}(0.5, 1.5)$, where $\mathrm{U}(a, b)$ denotes the uniform distribution with parameters a and b.

3. Diagonal $\mathbf{\Sigma}_{\mathrm{R}}$ with $(\mathbf{\Sigma}_{\mathrm{R}})_{r_1r_1} = 1 + I(r_1 \leq 0.9r)$, where $I(A)$ is the indicator function of the event A.

4. Tridiagonal $\mathbf{\Sigma}_{\mathrm{R}}$ with elements $(\mathbf{\Sigma}_{\mathrm{R}})_{r_1r_2} = 0.10^{|r_1-r_2|}I(|r_1 - r_2| \leq 1)$.

5. Tridiagonal $\mathbf{\Sigma}_{\mathrm{R}}$ with elements $(\mathbf{\Sigma}_{\mathrm{R}})_{r_1r_2} = 0.15^{|r_1-r_2|}I(|r_1 - r_2| \leq 1)$.

In each simulation scheme, we used 1000 replicates and we calculated the proportions of rejections at a 5% nominal significance level based on the proposed test statistics for the sphericity, identity and diagonality hypotheses. The empirical level of the proposed sphericity and identity test was calculated when $\mathbf{\Sigma}_{\mathrm{R}} = \mathbf{I}_r$ while their empirical powers were recorded whenever any of the other four structures for $\mathbf{\Sigma}_{\mathrm{R}}$ were used. For the proposed diagonality test, the empirical level was calculated with the identity and heteroscedastic structures and its empirical power was calculated under the tridiagonal structures. Tables 1-10 in Section S9 in the Supplementary Materials contains all simulation results for the sphericity and diagonality tests. Results for the proposed identity test are not discussed or presented as they were similar to those of the sphericity test in all sampling schema.

Table 1 in the Supplementary Materials suggests that the nominal size

of the proposed sphericity test was well approximated with Normal instances. For Gamma instances, its empirical sizes were slightly inflated when $r = 10$ or $r = 50$ but they were getting closer to the nominal size once $r \geq 100$. The empirical sizes of the proposed diagonality test were close to the nominal size regardless of the distributional scenario and the number of row variables as shown in Table 6 in the Supplementary Materials. The difference in the behavior of the two tests with skewed data and small r could be attributed to the fact that the variance of W_N is approximated more accurately by $\sigma^2_{W_N}$ than that of U_N by $\sigma^2_{U_N}$. From Tables 7 and 8 in the Supplementary Materials, it can also be checked that the proposed diagonality test preserved its size under both heteroscedastic structures as desired.

As expected from Theorem 3, the empirical power of the proposed sphericity test under the heteroscedastic and tridiagonal structures approached 1.0 for a large number of column variables ($c = 100, 600$) , as shown in Tables 2-5 in the Supplementary Materials. Therefore, we restrict our attention in sampling schema with $c = 10$. Conditional on $\mathbf{\Sigma}_{\mathrm{R}}$ and r, the empirical power was not severely affected by the distributional scenario and this can be viewed as a confirmation of the nonparametric nature of the proposed test. For fixed r, the empirical powers approached 1.0 as N

increased to 200 under both heteroscedastic and tridagonal structures for $\mathbf{\Sigma}_{\mathrm{R}}$ but the exact gains depended on the implied value of ξ_{1N}. For the two structures that lead to smaller values of ξ_{1N}, that is the heteroscedastic structure with $(\mathbf{\Sigma}_{\mathrm{R}})_{r_1 r_1} = 1 + I(r_1 \leq 0.9r)$ and the tridiagonal structure with non-zero correlation parameter equal to 0.10, the empirical powers were low even for $N = 60$. For the other two structures, the larger values of ξ_{1N} were obtained and this was reflected in their empirical powers for $N = 40$ and $N = 60$. Therefore, we conclude that for a small number of strongly dependent column variables, the consistency of the proposed sphericity test appears to depend on the magnitude of ξ_{1N}. The results for the power of the proposed diagonality test were almost identical to those above and can be found in Tables 9 and 10 in the Supplementary Materials.

5 Examples

5.1 Mouse aging project

In a project to study aging in mice, Zahn et al. (2007) measured gene expression levels in up to 16 tissues per mouse ($N = 40$). Herein we focused on inferring the dependence structure among nine tissues ($r = 9$), namely the adrenal glands, cerebrum, hippocampus, kidney, lung, muscle, spinal cord, spleen and thymus, based on the expression levels from 46 genes ($c = 46$) that play a role in the mouse endothelial growth factor (VEGF) signalling

pathway. Since Ning and Liu (2013) argued against a normality assumption, we applied the non-parametric bootstrap test of Aston, Pigoli and Tavakoli (2017) to assess the plausibility of a Kronecker product dependence decomposition for the covariance structure ($p-$value $= 0.616$). This finding partially supports the Kronecker product covariance decomposition modelling approach adopted in previous analysis (Yin and Li, 2012; Ning and Liu, 2013) for the construction of gene and tissue networks and justifies utilization of our proposed testing methods.

The tissue correlation matrix implied by the tissue-wise shrinkage covariance matrix estimate (Touloumis, Marioni and Tavaré, 2016) revealed a rather weak correlation pattern; all pairwise tissue correlations were estimated to be smaller than 0.1 in absolute value except that between the lung and spinal tissues which was equal to 0.2754. At a 5% significance level, we tested and failed to reject the null hypothesis in the diagonality hypothesis test for the tissue covariance matrix ($p-$value $= 0.0686$). Combining these results, it appeared that both Yin and Li (2012) and Ning and Liu (2013) might have overestimated the strength of the tissue dependencies. The tissue networks presented therein might be influenced by networks of genes that co-vary consistently between tissues. Controlling for this, the apparent "relatedness" between tissues is less than previously reported. We further

concluded that the tissues cannot be assumed to be equi-variant since we rejected the sphericity hypothesis ($p-$value < 0.0001). Therefore, it seems sensible to treat the nine tissues as uncorrelated but with differing variances. Using the sample tissue variances, the hippocampus tissue appeared to be the least variable followed by the muscle, kidney, adrenal, spleen, spinal, thymus, cerebrum and lung tissues in ascending order.

5.2 EEG Data

The EEG dataset Zahn et al. (2007) et al., available at `http://kdd.ics.uci.edu/databases/eeg/eeg.data.html`, describes a study that explores whether EEG data suggest a correlation between alcoholism and genetic predisposition. The 122 subjects who participated in this study were classified into either an alcoholic group (77 subjects) or a control group (45 subjects). For each subject, voltage fluctuations were recorded from 64 electrodes placed on the subject's scalp. Each subject was shown either one stimulus or two (matched or unmatched) stimuli and the voltage measures were recorded at 256 consecutive time points. This procedure was then repeated for up to 120 trials. For each of the 122 subjects, we created a two-dimensional data matrix such that the rows correspond to the 64 electrodes, the columns to the 256 time points and the values represent the average of the corresponding voltage measures across the available number

of trials.

Xia and Li (2017) analyzed this dataset assuming a matrix-variate normal distribution, an assumption that will follow in our analysis as well. Their goal was to construct a brain connectivity network for each of the two groups. The key to the construction of the networks is to decorrelate the 256 time points and in effect increase the sample size from 77 to $19712 = 77 \times 256$ in the alcoholic group and from 64 to $11520 = 64 \times 256$ in the control group. Application of the proposed diagonality test to the temporal covariance matrix in each group indicates that at least some of the time points were correlated (the p-values are close to 0 in each group). To decorrelate the columns, Xia and Li (2017) employed and estimated a banded structure (with bandwidth equal to 3) for the temporal covariance matrix at both groups. If this is the case, then the time points in each of the following three sets are expected to be uncorrelated: (i) $\{1, 5, \ldots, 253\}$, (ii) $\{2,6,\ldots,254\}$ and (iii) $\{3, 7, \ldots, 255\}$. To assess this hypothesis, we applied the sphericity test to each set for both groups. The corresponding p-values were again close to zero, suggesting that the time points in each set were correlated regardless of the group. Our finding suggests that Xia and Li (2017) might not have completely decorrelated the rows, and the construction of their two brain connectivity networks might have been affected by

the presence of significant temporal correlations.

6. Discussion

We considered test statistics for assessing the sphericity, identity and diagonality hypothesis tests for the row or column covariance matrix in high-dimensional transposable data, conditional upon the N i.i.d. random matrices having a Kronecker product dependence structure, a reasonable theoretical and practical assumption with high-dimensional transposable data. From a computational perspective, all three tests proposed are parsimonious in construction as estimation of just five parameters is required and there is no need to estimate the full column covariance matrix. Based on the results of the simulation study, it appears that the proposed diagonality test preserves the nominal size regardless of the distributional scheme, the sample size and the number of row and column variables. The proposed sphericity and identity tests also appeared to maintain the nominal size under normality but they might be slightly liberal when there are few column variables, say 10 or less, under non-normality. All three proposed tests seemed to be extremely powerful when there is a large number of 'nuisance' (column) variables but they suffered some power loss in the presence of strongly correlated column variables unless the sample size is greater than 100. We have also created the R package HDTD that implements the pro-

posed testing methods. The implementation of the proposed tests in HDTD takes advantage of the computationally inexpensive formulae presented in the Supplementary Materials, making the proposed methodologies suitable for use with high-dimensional transposable data even for very large numbers of row and/or column variables.

In future works, we aim to investigate the implications of the proposed tests when the true covariance structure does not satisfy a Kronecker product assumption, extend our methodology to account for covariance matrices that do not satisfy assumption (2.2), such as a covariance matrix with bounded variances that implies a compound symmetry correlation pattern, and consider extensions of these methods to array-variate random variables.

Supplementary Materials

The Supplementary Materials contain technical details, alternative formulae for the proposed test statistics, additional simulation results and the R code for reproducing the results in Section 5.

Acknowledgements

We are grateful to the associate editor and two anonymous reviewers for their valuable comments and suggestions that greatly improved the presentation of the article. The vast majority of the simulations for this

paper were performed on a dedicated workstation provided to A.T. by the University of Brighton via the Rising Stars grant.

References

Ahmad, M. R. and von Rosen, D. (2015). Tests of covariance matrices for high dimensional multivariate data under non normality. *Communications in Statistics-Theory and Methods* **44**, 1387–1398.

Allen, G. I. and Tibshirani, R. (2010). Transposable regularized covariance models with an application to missing data imputation. *The Annals of Applied Statistics* **4**, 764–790.

Allen, G. I. and Tibshirani, R. (2012). Inference with transposable data: Modelling the effects of row and column correlations. *Journal of the Royal Statistical Society B* **74**, 721–743.

Aston, J. A., Pigoli, D. and Tavakoli, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics* **74**, 1431–1461.

Bai, Z. D. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica* **6**, 311–330.

Carvalho, C. M. and West, M. (2007). Dynamic matrix-variate graphical models. *Bayesian Analysis* **2**, 69–97.

Chen, S. X. and Qin, Y. L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics* **38**, 808–835.

Chen, S. X., Zhang, L. X. and Zhong, P. S. (2010). Tests for high-dimensional covariance

matrices. *Journal of the American Statistical Association* **105**, 810–819.

Dawid, P. A. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika* **68**, 265–274.

Efron, B. (2009). Are a set of microarrays independent of each other? *The Annals of Applied Statistics* **3**, 922–942.

Galecki, A. T. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics-Theory and Methods* **23**, 3105–3119.

Genton, M. G. (2007). Spatial-temporal analysis of multivariate environmental monitoring data. *Environmetrics* **18**, 681–695.

Gupta, A. K. and Nagar, D. K. (2000). *Matrix Variate Distributions*. Chapman & Hall/CRC.

Hafner, C.M., Linton, O. and Tang, H. (2016). Estimation of a multiplicative covariance structure in the large dimensional case. *arXiv preprint arXiv:1605.04838* https://arxiv.org/abs/1605.04838.

Lee, K., Daniels, M. J. and Joo, Y. (2013). Flexible marginalized models for bivariate longitudinal ordinal data. *Biostatistics* **14** 462–476.

Leng, C. and Tang, C. Y. (2012). Sparse matrix graphical models. *Journal of the American Statistical Association* **107**, 1187–1200.

Mao, G. (2016). A note on tests for high-dimensional covariance matrices. Statistics & Proba-

bility Letters **117**, 89–92.

Mardia, K. V. and Goodall, C. (1993). Spatial-temporal analysis of multivariate environmental monitoring data. *Multivariate Enviromental Statistics* **6** 347–386.

Naik, D. N. and Rao, S. S. (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *Journal of Applied Statistics* **28**, 91–105.

Ning, Y. and Liu, H. (2013). High-dimensional semiparametric bigraphical models. *Biometrika* **100**, 655–670.

Oja, H. (2010). *Multivariate Nonparametric Methods with R*. Springer.

Qiu, Y. and Chen, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *The Annals of Statistics* **40**, 1285–1314.

Srivastava, M. S., von Rosen, T. and von Rosen, D. (2008). Models with a Kronecker product covariance structure: estimation and testing. *Mathematical Methods of Statistics* **17**, 357–370.

Srivastava, M. S., Yanagihara, H. and Kubokawa, T. (2014). Tests for covariance matrices in high dimension with less sample size. *Journal of Multivariate Analysis* **130**, 289–309.

Teng, S. L. and Huang, H. (2009). A statistical framework to infer functional gene relationships from biologically interrelated microarray experiments. *Journal of the American Statistical Association* **104**, 465–473.

Theobald, D. L. and Wuttke, D. S. (2006). Empirical Bayes hierarchical models for regularizing

maximum likelihood estimation in the matrix Gaussian Procrustes problem. *Proceedings of the National Academy of Sciences* **103**, 18521–5360.

Tsiligkaridis, T. and Hero, A. O. (2013). Covariance estimation in high dimensions via kronecker product expansions. *IEEE Transactions on Signal Processing* **61**, 5347–18527.

Touloumis, A., Marioni, J. C. and Tavaré, S. (2016). HDTD: analyzing multi-tissue gene expression data. *Bioinformatics* **32**, 2193–2195.

Touloumis, A., Tavaré, S. and Marioni, J. C. (2015). Testing the mean matrix in high-dimensional transposable data. *Biometrics* **71**, 157–166.

Xia, Y. and Li, L. (2017). Hypothesis testing of matrix graph model with application to brain connectivity analysis. *Biometrics* **73**, 780–791.

Yin, J. and Li, H. (2012). Model selection and estimation in the matrix normal graphical model. *Journal of Multivariate Analysis* **107**, 119–140.

Zahn, J. M., Poosala, S., Owen, A. B., Ingram, D. K., Lustig, A., Carter, A., Weeraratna, A. T., Taub, D. D., Gorospe, M., Mazan-Mamczarz, K., Lakatta, E. G., Boheler, K. R., Xu, X., Mattson, M. P., Falco, G., Ko, S. H. M., Schlessinger, D., Firman, J., Kummerfeld, S. K.,Wood III, W. H., Zonderman, A. B., Kim, S. K. and Becker, K. G. (2007). AGEMAP: A gene expression database for aging in mice. *PLoS Genetics* **3**, 2326–2337.

Zhang, X. L., Begleiter, H., Porjesz, B., Wang, W. and Litke, A. (1995). Event related potentials during object recognition tasks. *Brain Research Bulletin* **38**, 531–538.

Zhou, S. (2014). Gemini: Graph estimation with matrix variate normal instances. *The Annals of Statistics* **42**, 532–562.

Zhu, Y. and Li, L. (2018). Multiple matrix Gaussian graphs estimation. *Journal of the Royal Statistical Society: Series B* **80**, 927–950.

Anestis Touloumis, Centre for Secure, Intelligent and Usable Systems, University of Brighton

E-mail: A.Touloumis@brighton.ac.uk

John C. Marioni, University of Cambridge, EMBL-EBI

E-mail: John.Marioni@cruk.cam.ac.uk

Simon Tavaré, Columbia University

E-mail: st3193@columbia.edu