Homogeneity test of several high-dimensional covariance matrices for stationary processes under non-normality

Abdullah Qayed and Dong Han

School of Mathematical Sciences, Department of Statistics, Shanghai Jiao Tong University, Shanghai, China

ABSTRACT
We propose a test for testing the equality of several high-dimensional covariance matrices for stationary processes with a general distribution. The asymptotic distribution of the proposed test is proved to be $\chi^2$ distribution. Both the numerical simulation and empirical study illustrate that the proposed test has perfect performance, in particular, its power can approach to 1 on a set of covariance matrices with three known distributions.

ARTICLE HISTORY
Received 17 November 2020
Accepted 21 July 2021

KEYWORDS
Modified box $M$ test; homogeneity of several covariance matrices; non-normal distribution; dependent stationary process; high-dimensional data

1. Introduction
Many statistical methods in multivariate analysis require the assumption of homogeneous covariance matrices; for instance, multivariate analysis of variance, Hotelling $T^2$-test, discriminant function analysis and Mahalanobis’ generalized distance studies, in addition to the statistical models in repeated measures, longitudinal studies and multilevel regression analysis and so on. Moreover, the inference results of many applied sciences are depend on testing the covariance matrices, as in biometrics in the study of homogeneity of covariance matrices in infraspecific paleontologic and neontologic material and, if possible, by comparing with material known to be homochronous. On other hand, with the expansion of exact procedures for measurement theory in many applied sciences such as biology, life and other sciences, data of many different research phenomena may not be distributed normally, as in the exponential distribution for growth, microorganisms, pandemics, cancer cells and compound interest in economic surveys, … and so on. Other example, in genomics, de Torrenté et al. (2019) investigated the prevalence of genes with expression distributions, they found that there are less than 50% of all genes were Normally-distributed, with other distributions including Gamma, Bimodal, Cauchy, and Log-normal were represented. In such a case, when studying the shape of a gene distribution to test for data heterogeneity, the variety of non-normal data distributions is often overlooked when using homogeneity tests that assume that all data groups follow the normal distribution.

CONTACT Abdullah Qayed Albukhaiti qayed@sjtu.edu.cn; sass.center@yahoo.com School of Mathematical Sciences, Shanghai Jiao Tong University, 800 Dongchuan Rd., Minhang Campus, Shanghai 200240, China.
© 2021 Taylor & Francis Group, LLC
In this paper, we consider homogeneity tests of covariance matrices with high dimensional data when \( p \) can be much larger than the sample size \( n_i \). Recently, many homogeneity tests have been developed by many authors to address the deficiencies of classical tests caused by high-dimensional \( p \) and corresponding small sample sizes, and these tests are based on the trace of sample covariance matrices, as in sphericity hypothesis tests proposed by Jiang and Chen, Zhang, and Zhang (2010), Qiu and Chen (2012), Ledoit and Wolf (2002), Srivastava (2005), Cai, Liu, and Xia (2013), Cai and Ma (2013), Peng, Chen, and Zhou (2016), Ishll, Yata, and Aoshima (2019), and Zhang, Hu, and Bai (2020). Despite the classical homogeneity tests (as in Box 1949; Jiang, Jiang, and Yang 2012; Jiang and Yang 2013) are sensitive to departure the normal distribution, almost recently developed tests also assume the normal distribution of data. This raises questions about the accuracy of the various homogeneity tests with relatively small sample sizes of groups or when assumptions of normality and/or large samples are not met.

Moreover, for testing the equality of several covariance matrices under the hypotheses

\[
H_0: \Sigma_1 = \ldots = \Sigma_k = \Sigma \quad \text{vs} \quad H_1: \Sigma_i \neq \Sigma_j, \quad \text{for at least one pair} \ (i, j), \quad i \neq j, \quad (1)
\]

where \( \Sigma_i \) denotes the the \( p \times p \) covariance matrices of the \( i \)th population with \( p \)-dimensional multivariate distribution for \( 1 \leq i \leq k \), where \( k \geq 2 \). Schott (2007) presented the \( J_k \) test based on the trace of sample covariance matrices and proved that the asymptotic distribution of \( J_k \) is normal \( N(0, 1) \) under the null hypothesis \( H_0 \) in (1) with the setting of \( p/n_i \rightarrow b_i \in (0, \infty) \) as \( n_i p \rightarrow \infty \) for \( 1 \leq i \leq k \), where \( n_i \) is the number of samples for \( i \)th population. Srivastava and Yanagihara (2010) proposed two tests \( T_k^2 \) and \( Q_k^2 \) based on the trace of sample covariance matrices for testing \( H_0 \) and showed that the asymptotic null distributions of both tests are \( \chi^2_{k-1} \) in the setting of \( n_i = O(p^\delta), \delta > 1/2 \) as \( n_i p \rightarrow \infty \) for \( 1 \leq i \leq k \). However, all tests \( J_k, T_k^2 \) and \( Q_k^2 \) assume the normal distribution of data. Based on the \( U \)-statistics, Ahmad (2017) and Zhong, Li, and Santo (2019) presented tests \( T_g \) and \( \hat{D}_{nt} \), respectively, for high-dimensional data under nonnormality and high-dimensional longitudinal data, where both tests have asymptotic null distribution and do not need a specific relationship between \( p \) and \( n_i \) as

\[
p = p(n_i), n_i \rightarrow \infty. \quad \text{However, all five tests} \ J_k, T_k^2, Q_k^2, T_g \quad \text{and} \ \hat{D}_{nt} \quad \text{above need to assume that} \quad 0 < a_i = \lim_{p \rightarrow \infty} tr(\Sigma^j) / p < \infty, \quad j = 1, 2, 3, 4, \quad \text{or} \quad 0 < tr(\Sigma^2) / p^2 = \delta_1 < \infty \quad \text{for all} \ p \quad \text{and} \ \inf_p \{ tr(\Sigma^4) / [tr(\Sigma^2)]^2 \} > 0.
\]

The research in this paper is motivated by the conventional likelihood ratio \( M \) test in Box (1949), which established the asymptotic \( \chi^2 \) distribution of the test statistic for fixed-dimensional normally distributed random vectors, when \( n_i > p, 1 \leq i \leq k \). In this paper, a modified \( M \) test, \( L_k \), is proposed to test the null hypothesis in (1) without the normality assumption. We don’t assume the conditions about the relationship between the traces of powers of covariance matrix and \( p \) listed above for testing the null hypothesis \( H_0 \). However, we assume that the relationship between \( p \) and \( n_i \) satisfies \( p \geq c(n_{\text{max}})^{3r/(r-2)} \) for some positive constant \( c \), where \( n_{\text{max}} = \max\{n_1, n_2, \ldots, n_k\} \) and \( r > 2 \). Hence, the test \( L_k \) accommodates the large \( p \), small \( n_{\text{max}} \) situations.

This paper is organized as follows. Section 2 introduces the basic data structure of the G-stationary process and its associated properties, and a modified \( M \) test. In Section 3, we present the main results of this paper, we show that the asymptotic null distribution of \( L_k \),
as \( n_{\min} \to \infty \), is \( \chi^2_{k-1} \) under the null hypothesis \( H_0 \) in (1) followed by the procedures of \( L_k \) test. Reports of numerical simulation are given in Section 4. An empirical study using time-series data to extract clinical information from speech signals is presented in Section 5. The paper concludes in Section 6. The proofs of Theorems 3.1 and 3.2, and the examples results are deferred to the Appendix.

2. The G-stationary process and a modified M test

We first introduce the G-stationary process and its associated properties, then present a modified M test.

A dependent stationary process is called as the G-stationary process if it has the following representation

\[
X_j = G(\ldots, e_{j-1}, e_j, e_{j+1}, \ldots), \quad j \in \mathbb{Z}
\]

and satisfies the conditions of Theorem 2.1 in Berkes, Liu and Wu (2014), where \( e_i, i \in \mathbb{Z} \), are i.i.d. random variables, \( G : \mathbb{R}^\mathbb{Z} \to \mathbb{R} \) is a measurable function, \( \mathbb{Z} \) and \( \mathbb{R} \) denote the set of integers and the set of real numbers, respectively. In fact, under some appropriate conditions, the G-stationary process can include a large class of linear and nonlinear processes, such as the functionals of linear processes, bilinear models, GARCH processes, generalized random coefficient autoregressive models, nonlinear AR models and so on, see Liu and Lin (2009), Berkes, Liu and Wu (2014).

Assume that \( X_j \) has mean zero, \( \mathbb{E}(|X_j|^r) < \infty \) for \( r > 2 \), with covariance function \( \gamma(j) = \mathbb{E}(X_1X_j), j \in \mathbb{Z} \). It is clear that the covariance matrix \( \Sigma_p = (\sigma_{ij})_{p \times p} \) of \( (X_1, X_2, \ldots, X_p) \) satisfies \( \Sigma_p = (\sigma_{ij})_{p \times p} = \Sigma_p = (\gamma(i - j))_{p \times p} \). Let \( \mathbb{S}_p = \sum_{j=1}^p X_j \). The most important result for the G-stationary process is that it satisfies the strong invariance principle, that is, it can arrive the optimal KMT approximation (see Berkes, Liu and Wu 2014), that is, there exists a richer probability space such that

\[
\mathbb{S}_p - \sigma\mathbb{B}(p) = o(p^{1/2}), \quad \text{a.s.}
\]

where \( \mathbb{B}(t) \) is a standard Brownian motion and \( \sigma^2 = \sum_{j \in \mathbb{Z}} \gamma(j) \). From (3) it follows that

\[
\frac{\mathbb{S}_p}{\sigma \sqrt{p}} \Rightarrow N(0, 1) \quad \text{and} \quad \sigma^2 = \lim_{p \to \infty} \frac{\mathbb{E}(\mathbb{S}_p^2)}{p}.
\]

In this paper, our main purpose is to present a test for testing the equality of \( k \) covariance matrices \( \Sigma_1, \Sigma_2, \ldots, \Sigma_k \) which correspond to \( k \) G-stationary processes, \( X_j^{(i)}, 1 \leq j \leq p \) for \( 1 \leq i \leq k \), respectively, where \( \Sigma_i = (\sigma_{ij}^{(i)})_{p \times p} = (\gamma^{(i)}(j - j'))_{p \times p} \) and \( \gamma^{(i)}(j) = \mathbb{E}(X_1^{(i)}X_j^{(i)}) \), \( 1 \leq j \leq p \) for \( 1 \leq i \leq k \). The \( k \) G-stationary processes can be considered as \( k \) \( p \)-dimensional random vectors, \( X^{(i)} = (X_1^{(i)}, X_2^{(i)}, \ldots, X_p^{(i)}) \), \( 1 \leq i \leq k \). For each \( i \), we take \( n_i \) i.i.d. random samples (random vectors)

\[
X_l^{(i)} = (X_{1l}^{(i)}, X_{2l}^{(i)}, \ldots, X_{pl}^{(i)}), \quad 1 \leq l \leq n_i,
\]

to estimate \( \Sigma_i \) for \( 1 \leq i \leq k \). Sometimes, \( p \) is much bigger than \( n_i \). For example, consider whether the fluctuation of hypertension in half a year (180 days) in patients with
hypertension in \( k \) countries or regions is consistent. We may randomly select 100 patients with hypertension in every country and wear a smart health bracelet to every patient with hypertension, which can monitor the blood pressure profile of the patient every minute. Here, \( n_1 = n_2 = \ldots = n_k = 100 \) and \( p = 60 \times 24 \times 180 = 259200 \).

Let \( S_i \) be the sample covariance matrix for \( \Sigma_i, 1 \leq i \leq k \). Note that \( S_i \) can be written as

\[
S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \begin{pmatrix}
X_{il}^{(i)} - \bar{X}_l^{(i)} \\
X_{j,l}^{(i)} - \bar{X}_l^{(i)} \\
\vdots \\
X_{pl}^{(i)} - \bar{X}_p^{(i)}
\end{pmatrix}
\]

and therefore

\[
1S_i I^T = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \sum_{j=l}^{p} (X_{j,l}^{(i)} - \bar{X}_j^{(i)})^2 \geq 0,
\]

where \( \bar{X}_j^{(i)} = \sum_{l=1}^{n_i} X_{j,l}^{(i)}/n_i \) for \( 1 \leq j \leq p \) and \( I = (1, 1, \ldots, 1) \) is \( p \)-dimensional vector with 1 as the component. How to guarantee \( 1S_i I^T > 0 \), a.s.? In fact, if \( S_{pl}^{(i)} := \sum_{j=1}^{p} X_{j,l}^{(i)} \) is a continuous random variable for each \( i, 1 \leq i \leq k \), then

\[
S_{pl}^{(i)} - \frac{1}{n_i} \sum_{l=1}^{n_i} S_{pl}^{(i)} = \left(1 - \frac{1}{n_i}\right) S_{pl}^{(i)} - \frac{1}{n_i} \sum_{l \neq pl} S_{pl}^{(i)} \neq 0, \quad \text{a.s.}
\]

for \( n_i > 1, 1 \leq i \leq k \), since \( S_{pl}^{(i)}, 1 \leq l \leq n_i \), are i.i.d. continuous random variables for each \( i, 1 \leq i \leq k \). This means that

\[
(n_i - 1)1S_i I^T = \sum_{l=1}^{n_i} (S_{pl}^{(i)} - \frac{1}{n_i} \sum_{l=1}^{n_i} S_{pl}^{(i)})^2 > 0, \quad \text{a.s.}
\]

for \( n_i > 1, 1 \leq i \leq k \). For general case, we can prove that \( 1S_i I^T/p \geq 0, \quad \text{a.s. as } p \to \infty \) in Theorem 3.1 in the next section.

The often used test for testing the equality of several covariance matrices is the Box’s \( M \) (1949) test which can be written in the following

\[
M = (n - k)\log|S| - \sum_{i=1}^{k} (n_i - 1) \log |S_i|,
\]

and \( S = \sum_{i=1}^{k} (n_i - 1)S_i/(n - k) \) and \( n - k = \sum_{i=1}^{k} (n_i - 1) \). When \( p \) is fixed and \( \min_{1 \leq i \leq k} n_i \to \infty \), the asymptotic null distribution of \( \varphi M \) test is chi-squared \( \chi^2 \) with \( df = (k - 1)p(p + 1)/2 \) degrees of freedom, where

\[
\varphi = 1 - \frac{2p^2 + 3p - 1}{6(p + 1)(k - 1)} \left( \sum_{i=1}^{k} \frac{1}{n_i - 1} - \frac{1}{n - k} \right).
\]

The Box’s \( M \) test in (7) represents a ratio of the pooled determinant \( |S| \) to the geometric mean of the determinants \( |S_i|, (i = 1, \ldots, k) \), this test is extremely sensitive to departures from normality. On the other hand, the \( M \) test is valid only for \( n_i > p \) for \( 1 \leq i \leq k \).
It can be seen that the modified test statistic $L_k$ shows that the asymptotic distribution of the test statistic $\text{M}^2$ is that it is not easy to calculate the determinant $|S|$ when $p$ is large. In order to overcome these two shortcomings, we propose a modified $M$ test, $L_k$, by replacing $|S|$ in $M$ test (7) with $\text{M}^{-1}S\text{I}^T$, that is,

$$L_k := (n-k)\log \hat{V}_p - \sum_{i=1}^{k}(n_i - 1)\log \hat{S}_i$$

(8)

where

$$\hat{S}_i = \text{M}^{-1}S\text{I}^T, \quad \hat{V}_p = \frac{1}{n-k} \sum_{i=1}^{k}(n_i - 1)\hat{S}_i, \quad i = 1, ..., k.$$  

It can be seen that the modified $M$ test, $L_k$, can not only be defined but also can be easily calculated when $p$ is large.

**3. Main results**

The asymptotic properties of $\text{M}^{-1}S\text{I}^T$ and $L_k$ will be given in this section.

**Theorem 3.1.** For each $i$, $1 \leq i \leq k$, let $\{X_{ij}^{(i)}, 1 \leq j \leq p\}, 1 \leq l \leq n_i$ be $n_i$ i.i.d. $G$-stationary processes with $E(X_{ij}^{(i)}) = 0$ and $E(|X_{ij}^{(i)}|^r) < \infty$ for $r > 2, 1 \leq l \leq n_i, 1 \leq j \leq p$ and $1 \leq i \leq k$. Let $\text{M}^{-1}S\text{I}^T > 0$ and $\sigma_i^2 = \sum_{j \in Z} v^{(i)}(j)$ for $1 \leq i \leq k$. If $p \geq c(n_i)^{r/(r-2)}$ for some positive constant $c$, then

$$\frac{(n_i - 1)\text{M}^{-1}S\text{I}^T}{p\sigma_i^2} \to \chi_i, \quad \chi_i \sim \chi_{n_i-1}^2$$

(10)

as $p \to \infty$ for $1 \leq i \leq k$.

**Remark 3.1.** From (4) it follows that $\text{M}^{-1}S\text{I}^T/(p\sigma_i^2) \to 1$ as $p \to \infty$, that is, $(n_i - 1)\text{M}^{-1}S\text{I}^T/(\Sigma \text{I}^T) \to \chi_i$. Gupta and Nagar (2000, Theorem 3.3.1) has proved the similar result of Theorem 3.1 for $p < n_i$ by using the method of characteristic function assuming the normal distribution of the data. Here, we prove Theorem 3.1. without assuming the normality and permitting $p$ larger than $n_i$ and therefore (10) holds for any data distribution satisfying the conditions of Theorem 3.1.

Let $n_{\min} = \min\{n_1, n_2, ..., n_k\}$ and $n_{\max} = \max\{n_1, n_2, ..., n_k\}$. The following theorem shows that the asymptotic distribution of the test statistic $L_k$ in (8) is $\chi_{k-1}^2$.

**Theorem 3.2.** Let the conditions of Theorem 3.1 hold and $p \geq c(n_{\max})^{3r/(r-2)}$. Under the original hypothesis $H_0$ in (1) with $\Sigma > 0$, we have that the test statistic $L_k$ in (8) converges in distribution to $\chi_{k-1}^2$ as $n_{\min} \to \infty$.

Next we give an example to illustrate how to use Theorem 3.2. to test the null hypothesis (1).

**Example 1.** Let $k = 3, n_1 = n_2 = n_3 = 100$ and $r = 0.05$. Take two positive numbers $\tilde{\chi}_1$ and $\tilde{\chi}_2$ such that $\tilde{\chi}_1 < \tilde{\chi}_2$ and...
\[ P(\chi^2_{k-1} < \tilde{\chi}_1) = 0.025, \quad P(\chi^2_{k-1} > \tilde{\chi}_2) = 0.025. \]

This means that \( P(\tilde{\chi}_1 \leq \chi^2_{k-1} \leq \tilde{\chi}_2) = 1 - 0.05 \), that is, \( R_k = \{0, \tilde{\chi}_1 \cup (\tilde{\chi}_2, \infty)\} \) is a domain of rejecting the null hypothesis \( H_0 \), here, the procedures of \( L_k \) hold for any \( I = (1, \ldots, 1_p), I \) is a \( 1 \times p \) vector.

For high dimensional observation vector \((X^{(i)}_1, \ldots, X^{(i)}_p), 1 \leq i \leq k\), researchers sometimes not only want to infer whether their covariance matrices, \( \Sigma_1, \ldots, \Sigma_k \), are equal, but also want to know whether sub covariance matrices of relative low dimensional observations vectors, \((X^{(i)}_1, \ldots, X^{(i)}_{p_i}), (X^{(i)}_{p_i+1}, \ldots, X^{(i)}_{p_i+p_2}), \ldots, (X^{(i)}_{p_i+\cdots+p_{i-1}+1}, \ldots, X^{(i)}_p)\), are equal, where \( p = \sum_{i=1}^k p_i \), all \( p_i \geq 2 \) and \( l \geq 2 \). Without loss of generality, we assume that \( p \geq n_{\min} = \min\{n_i, 1 \leq i \leq k\} \). Let \( p_j = j(n_{\min} - 1) + p \) for \( 1 \leq j \leq m - 1 \) and take \( m \geq 2 \) satisfying \( p_{m-1} < p_m = p \) and \( m \) is the smallest positive integer number such that \( n_{\min} > p_j - p_{j-1} \) for all \( 1 \leq j \leq m \). By the definition of \( p_j \) for \( 1 \leq j \leq m \), we have \( p_j - p_{j-1} = n_{\min} - 1 \) for \( 1 \leq j \leq m - 1 \) and \( p - p_{m-1} = p_m - p_{m-1} \leq n_{\min} - 1 \). Note that \( m \) depends on \( p \) and \( n_{\min} \) and \( p_j > p_i \) for \( 1 \leq i < j \leq m \) and \( p_j = p_i \) for \( j = i \). For example, when \( p = n_{\min} \), it takes \( p_1 = \min\{(n_{\min} - 1), p\} \) and \( p_2 = n_{\min} - 1 = p \), that is, \( m = 2 \). If \( p = 2n_{\min} \), we have \( p_1 = \min\{(n_{\min} - 1), p\}, p_2 = 2\min\{(n_{\min} - 1), p\} \) and \( p_3 = p_2 + 2 = p \), that is \( m = 3 \).

Now, we divide \( \Sigma_i \) and \( \Sigma_i \) into the following two forms, respectively

\[
\Sigma_i = \begin{pmatrix}
\Sigma_{\hat{p}_1 \times \hat{p}_1} & \Sigma_{\hat{p}_1 \times \hat{p}_2} & \cdots & \cdots \\
\Sigma_{\hat{p}_2 \times \hat{p}_1} & \Sigma_{\hat{p}_2 \times \hat{p}_2} & \cdots & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{\hat{p}_m \times \hat{p}_1} & \Sigma_{\hat{p}_m \times \hat{p}_2} & \cdots & \cdots \\
& \cdots & \cdots & \Sigma_{\hat{p} \times \hat{p}}
\end{pmatrix}
\]

and

\[
S_i = \begin{pmatrix}
S_{\hat{p}_1 \times \hat{p}_1} & S_{\hat{p}_1 \times \hat{p}_2} & \cdots & \cdots \\
S_{\hat{p}_2 \times \hat{p}_1} & S_{\hat{p}_2 \times \hat{p}_2} & \cdots & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
S_{\hat{p}_m \times \hat{p}_1} & S_{\hat{p}_m \times \hat{p}_2} & \cdots & \cdots \\
& \cdots & \cdots & S_{\hat{p} \times \hat{p}}
\end{pmatrix},
\]

for \( 1 \leq i \leq k \), where \( \Sigma_{\hat{p}_j \times \hat{p}_j} \) and \( S_{\hat{p}_j \times \hat{p}_j} \) denote a \( \hat{p}_j \times \hat{p}_j \) sub-matrices for \( 1 \leq i \leq k, 1 \leq j \leq m - 1 \) and \( \hat{p}_j = (p_j - p_{j-1}) \), and \( \Sigma_{\hat{p} \times \hat{p}} \) and \( S_{\hat{p} \times \hat{p}} \) denote a \( \hat{p} \times \hat{p} \) sub-matrices for \( 1 \leq i \leq k \) and \( \hat{p} = (p - p_{m-1}) \).

Let \( y_{p_1} = (1, \ldots, 1_{p_1}), y_{p_2} = (1_{p_1+1}, \ldots, 1_{p_2}), \ldots, y_{p_m} = (1_{p_{m-1}+1}, \ldots, 1_p) \) and

\[
y_1 = (y_{p_1}, 0, \ldots, 0), y_2 = (0, \ldots, 0, y_{p_2}, 0, \ldots, 0), \ldots, y_m = (0, \ldots, 0, y_{p_m}),
\]

be \( m \) \( p \)-dimensional vectors. It follows that
\[ y_j \sum_i y_j^T = y_{p_j} \sum_{i, j} y_j^T, \quad y_j S_i y_j^T = y_{p_j} \sum_{i, j} y_j^T \]

for \( 1 \leq i \leq k, 1 \leq j \leq m \). Note that \( \tilde{p}_j = p_j - p_{j-1} = n_{\text{min}} - 1 \) for \( 1 \leq j \leq m - 1 \) and \( \tilde{p} = p - p_{m-1} \leq n_{\text{min}} - 1 \), that is, \( y_{p_j}, 1 \leq j \leq m - 1 \), are \( \tilde{p}_j \)-dimensional vectors and \( y_{p_{m-1}} \) is a \( \tilde{p} \)-dimensional vector. If \( y_j \neq 0 \), or equivalently, \( y_j \neq 0 \), then \( y_j S_i y_j^T = y_{p_j} \sum_{i, j} y_j^T > 0 \) when \( \sum_{i, j} y_j^T > 0 \) for \( 1 \leq i \leq k, 1 \leq j \leq m \). Now we present \( m \) test statistics as follows:

\[
L_{kj} := (n - k) \log \hat{V}_{p_j} - \sum_{i=1}^{k} (n_i - 1) \log \hat{S}_{ij}
\]

for \( 1 \leq j \leq m \), where

\[
\hat{S}_{ij} = y_j S_i y_j^T = y_{p_j} \sum_{i, j} y_j^T, \quad \hat{V}_{p_j} = \frac{1}{n - k} \sum_{i=1}^{k} (n_i - 1) \hat{S}_{ij}
\]

for \( 1 \leq i \leq k \) and \( 1 \leq j \leq m \).

Hence, by extending Theorem 3.2 we can get the following corollary.

**Corollary 3.1.** Let \( p \geq c(n_{\text{max}})^{3r/2} \). Under the original hypothesis \( H_0 \) in (1) with \( \Sigma > 0 \), for any \( p \)-dimensional vector \( y_j \neq 0 \) defined in (11), we have that for every \( j (1 \leq j \leq m) \), the test statistic \( L_{kj} \) in (12) converges in distribution to \( \chi^2_{k_j} \) as \( n_{\text{min}} \to \infty \).

**Remark 3.2.** Here, we divide the original hypothesis \( H_0 \) and the alternative hypothesis \( H_1 \) into \( m \) original hypotheses and \( m \) alternative hypotheses in the following

\[
H_{0j} : \Sigma_{p_j}^{(i)} = \Sigma_{p_j}^{(l)} \quad \text{vs} \quad H_{lj} : \Sigma_{p_j}^{(i)} \neq \Sigma_{p_j}^{(l)} \quad \text{at least one pair } (i, l), i \neq l
\]

for \( 1 \leq j \leq m \). Thus, we can use the test statistic \( L_{kj} \) to do the original hypothesis \( H_{0j} \) and the alternative hypothesis \( H_{lj} \) for \( 1 \leq j \leq m \). Moreover, though the above division may lose some information, for example, \( \Sigma_{p_j}^{(i)}, j \neq l \), one of its advantages is that it can help us to test which part \( (H_{0j}) \) of the original hypothesis \( H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_k = \Sigma \) is inconsistent.

**Remark 3.3.** If necessary, one can divide the original hypothesis \( H_0 \) and the alternative hypothesis \( H_1 \) into more or less sub hypotheses, \( H_{0j} \) and \( H_{lj}, 1 \leq j \leq l \).

The following example will discuss on how to use Corollary 3.1. for testing the null hypothesis.

**Example 2.** Let \( k = 3, n_1 = n_2 = n_3 = 101 \) and \( p = 350 \). Let \( 1_a = (1, 1, \ldots, 1) \) and \( 0_a = (0, 0, \ldots, 0) \) be two 100-dimensional vectors and \( 1_b = (1, 1, \ldots, 1) \) and \( 0_b = (0, 0, \ldots, 0) \) denote two 50-dimensional vectors. Take four 350-dimensional vectors \( y_j, 1 \leq j \leq 4 \), defined in the following

\[
y_1 = (1_a, 0_a, 0_a, 0_a), \quad y_2 = (0_a, 1_a, 0_a, 0_b)
\]
\[ y_3 = (0_a, 0_a, 1_a, 0_b), \quad y_4 = (0_a, 0_a, 0_a, 1_b). \]

Divide \( S_i \) into four sub sample covariance matrices: \( S^{(i)}_{pj \times \tilde{p}_j}, j = 1, 2, 3, 4 \) for \( i = 1, 2, 3, \) where \( p_1 = 100, p_2 = 200, p_3 = 300 \) and \( p_4 = 350 \). By (8) we can get four test statistics \( L_{kj}, 1 \leq j \leq 4 \). Thus, we will reject the null hypothesis \( H_0 \) if at least one of \( L_{kj}, 1 \leq j \leq 4 \) belong to the rejecting region \( R_k \). If \( L_{k4} \in R_k \) and \( L_{kj} \notin R_k \) for \( 1 \leq j \leq 3 \), this means that the three covariances, \( \Sigma^{(1)}_{p_1 \times \tilde{p}_1}, \Sigma^{(2)}_{p_2 \times \tilde{p}_2} \) and \( \Sigma^{(3)}_{\tilde{p}_3 \times p_3} \) corresponding to \( (X_{301}^{(1)}, \ldots, X_{350}^{(1)}), (X_{301}^{(2)}, \ldots, X_{350}^{(2)}) \) and \( (X_{301}^{(1)}, \ldots, X_{350}^{(1)}) \), are not equal.

**Remark 3.4.** In practice, \( \rho L_k \) converges to \( \chi^2_{k-1} \), where \( \rho = \rho_n \) is the scale factor given in Bartlett (1937) as follows:

\[
\rho = \rho_n = \frac{1}{C}, \quad \text{where} \quad C = 1 + \frac{1}{3(k-1)} \left( \sum_{i=1}^{k} \frac{1}{n_i-1} - \frac{1}{n-k} \right) \tag{13}
\]

and \( \rho \to 1 \) for fixed \( k \) as \( n_{\min} \to \infty \). Here, the use of \( \rho \) tends to over-correct the exaggerated significance levels in small samples. The scale factor \( \rho \) also greatly increases the approximation of significance level and uses as a gauge of its convergence. The empirical sizes of \( \rho L_k \) in Table 1 are corresponding to the correct significance levels of \( \chi^2_{k-1} \) for several cases of \( n, p \) and \( k \) with the nominal level 0.05.

### 4. Numerical simulation

We report results of simulation which were designed to evaluate the performance of the proposed tests \( \rho L_k \) for testing the null hypothesis in (1). \( \rho \) was chosen as in (13) so that the convergence of the significance level by the empirical distribution of \( \rho L_k \) test are close to \( \chi^2_{k-1} \). We also experimented the high dimensional test proposed by Ahmad (2017) \( (T_g \) hereafter). To mimic the large \( p \), small \( n_i \) scenario, we set \( n_i = 10, 20, 50 \) and 100, and for each \( n_i \) we let \( p = 20, 50, 100, 200 \) and 300. We experimented the equality tests of three groups \( k = 3 \) where \( n_1 = n_2 = n_3 \). All simulation results in this paper are obtained using \( 10^3 \) repetitions with the significance level \( \alpha = 0.05 \). We considered two scenarios with respect to the data.

**Scenario I:** in this Scenario the data was generated from a Gaussian AR(1) model. The covariance structures were designed as in Srivastava and Yanagihara (2010) in the following setup:

\[
H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3 = \Omega \Delta_0 \Omega \\
H_1 : \Sigma_i = \Omega \Delta_{i-1} \Omega, i = 1, 2, 3.
\]

where \( \Omega = \text{diag}(\omega_1, \ldots, \omega_p) \) with \( \omega_1, \ldots, \omega_p \sim \text{Unif} \ (1, 5) \) and \( \Delta_j \) is \( p \times p \) matrix with \((i,j)\) th element being \((-1)^{(i+j)}(0.2 \times (J + 2))^{(i-j)/10}, \) \( J = 0, 1, 2. \)

**Scenario II:** The data structures were generated in the exponential and uniform distributions. The null hypothesis was \( H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3 = \Delta_0, \) where \( \Delta_0 \) is defined above. However, the alternatives settings were designed in three different covariance structures
The three \( p \times p \) covariance structures are taken as compound symmetry (CS), autoregressive of order 1, AR(1), and completely unstructured (UN) where the first (CS) is defined as
\[
R_1 = I + \phi J
\]
However, the second and third, \( R_2 \) and \( R_3 \), are defined with the \( ij \)th element being
\[
\phi^{\lvert i-j \rvert} \quad \text{for AR(1)} \quad \text{and} \quad \sigma_{ij} = 1, \quad i = j \quad \text{as diagonal elements}, \quad \sigma_{ij} = \frac{1}{p}, \quad i \neq j \quad \text{as off diagonal elements} \quad \text{for UN structure} \quad \forall i, j, \quad \text{with} \quad I \quad \text{as identity matrix}, \quad J \quad \text{as the matrix of 1's} \quad \text{and} \quad \phi = 0.70 \quad \text{as constant.}
\]
The empirical sizes and power results for the two tests \( \rho L_k \) and \( T_g \) of scenarios I and II were recorded in Tables 1 and 2, respectively.

We evaluated the performance of the \( L_{kj} \), \( 1 \leq j \leq m \) tests \( (\rho L_{kj} \) hereafter) in (12) using the settings of scenarios I and II. The empirical size and power results were simulated by applying the \( \rho L_{kj} \) tests in Example 2, \( m = 4 \), the results were recorded in Tables 3.

### 4.1. Results and discussion

According to scenarios I and II, Tables 1 and 2 show the attained empirical size under \( H_0 \) and the power under \( H_1 \) for the \( \rho L_k \) and \( T_g \) tests. We observe from the results in Tables 1 and 2 that the empirical sizes of the two tests had substantial differences. In the normal distribution, we observe from Table 1 that the sizes of the \( \rho L_k \) test are close to the significance level \( \alpha = 0.05 \) regardless of the relationship between the dimension \( p \) and \( n_i \). However, sizes of \( T_g \) test in Table 2 appear to be smaller than those of \( \rho L_k \), and the smaller the \( T_g \) sample size, the greater the deviation of the empirical size from the significance level 0.05.

For both exponential and uniform distributions in Tables 1 and 2, the \( T_g \) test encountered serious size distortion while the \( \rho L_k \) test still had perfect sizes. This distortion

### Table 1.
The empirical size and power of \( \rho L_k \) test in the normal, exponential, and uniform distributions.

| \( p \) | \( n_1 = n_2 = n_3 \) | Normal | Exp | Uniform | Normal | Exp | Uniform |
|---|---|---|---|---|---|---|---|
| 20 | 10 | 0.049 | 0.054 | 0.057 | 0.981 | 0.988 | 0.984 |
| 20 | 20 | 0.049 | 0.056 | 0.056 | 0.989 | 0.999 | 0.993 |
| 50 | 50 | 0.051 | 0.047 | 0.061 | 1.000 | 1.000 | 1.000 |
| 100 | 50 | 0.050 | 0.054 | 0.067 | 1.000 | 1.000 | 1.000 |
| 100 | 10 | 0.054 | 0.052 | 0.052 | 0.992 | 0.999 | 0.990 |
| 100 | 20 | 0.046 | 0.050 | 0.049 | 0.991 | 1.000 | 1.000 |
| 100 | 50 | 0.045 | 0.047 | 0.056 | 1.000 | 1.000 | 1.000 |
| 100 | 100 | 0.053 | 0.050 | 0.051 | 1.000 | 1.000 | 1.000 |
| 50 | 10 | 0.050 | 0.050 | 0.051 | 0.989 | 1.000 | 1.000 |
| 50 | 20 | 0.048 | 0.042 | 0.052 | 1.000 | 1.000 | 1.000 |
| 50 | 50 | 0.053 | 0.054 | 0.046 | 1.000 | 1.000 | 1.000 |
| 50 | 100 | 0.052 | 0.050 | 0.052 | 1.000 | 1.000 | 1.000 |
| 100 | 10 | 0.047 | 0.046 | 0.041 | 0.987 | 1.000 | 1.000 |
| 100 | 20 | 0.046 | 0.050 | 0.055 | 1.000 | 1.000 | 1.000 |
| 100 | 50 | 0.050 | 0.047 | 0.053 | 1.000 | 1.000 | 1.000 |
| 100 | 100 | 0.051 | 0.052 | 0.050 | 1.000 | 1.000 | 1.000 |
| 200 | 10 | 0.053 | 0.051 | 0.054 | 0.995 | 1.000 | 1.000 |
| 200 | 20 | 0.046 | 0.050 | 0.048 | 1.000 | 1.000 | 1.000 |
| 200 | 50 | 0.050 | 0.047 | 0.051 | 1.000 | 1.000 | 1.000 |
| 200 | 100 | 0.051 | 0.051 | 0.053 | 1.000 | 1.000 | 1.000 |
| 300 | 10 | 0.053 | 0.051 | 0.054 | 0.995 | 1.000 | 1.000 |
| 300 | 20 | 0.046 | 0.050 | 0.048 | 1.000 | 1.000 | 1.000 |
| 300 | 50 | 0.050 | 0.047 | 0.051 | 1.000 | 1.000 | 1.000 |
| 300 | 100 | 0.051 | 0.051 | 0.053 | 1.000 | 1.000 | 1.000 |
may be attributed to the fact that the $T_g$ test was constructed as an extension of Ahmad’s (2014) test statistic ($T_2$) to test the equality of two covariance matrices. Say $T_g$ as an estimator of the sum of Frobenius norms over all possible distinct pairs $\sum_{i < j} \| \Sigma_i - \Sigma_j \|$, $i, j = 1, ..., k$, however, using simple trace properties and the moments of bilinear forms may cause large type 1 errors. It was for this reason we do not report the power of the $T_g$ test in Tables 1 and 2. Under the settings of both scenarios I and II, we observe that the power and size of the $qL_k$ test is better than those of the $T_g$ test. The reason why the performance of the proposed test is better should be that the test statistic $\rho L_k$ is based on the $k$ estimators $1S_i I^T$, $1 \leq i \leq k$.

Table 3 shows the simulation results of size and power of the $\rho L_{kY_j}$ ($L_{kj}$) tests in (12) for the $y_j$ vectors in Example 2, $1 \leq j \leq 4$. We observe from Table 3 that the four tests $\rho L_{kY_1}, \rho L_{kY_2}, \rho L_{kY_3}$ and $\rho L_{kY_4}$ have good performance for normal, uniform and exponential distributions, we also observe that the results of size and power in Table 3 are consistent with those of the $\rho L_k$ in Tables 1 and 2.

### 5. Experimental study

To demonstrate the performance of the proposed $\rho L_k$ test we use LSVT Voice rehabilitation data set, provided by https://archive.ics.uci.edu. This data set represents nonlinear time series data, it consists of a range of biomedical speech signal processing algorithms from 14 people who have been diagnosed with Parkinson’s disease undergoing LSVT (a program assisting voice rehabilitation), see Tsanas, Little, and Fox (2014). This data set aims to assess whether voice rehabilitation treatment lead to phonations (speech signals) considered” acceptable” (a clinician would allow persisting during in-person rehabilitation treatment) or “unacceptable” (a clinician would not allow persisting during in-person rehabilitation treatment). The original LSTV scale was used to characterize each of
the 126 phonations (as instances) with 309 dysphonia measures (as variables). For this purpose, they used the correlation analysis to find the variability between 309 dysphonia algorithms (between variables). The results showed that all dysphonia measures (309) were statistically significant ($P_{value} < 0.01$). These results confirm the validity (internal consistency) of the LSVT voice rehabilitation scale. In this paper, we aim to evaluate the validity (internal consistency) of the LSVT voice rehabilitation scale by testing the heterogeneity of multi-samples (groups) of phonations.

5.1. Results and discussion

To achieve the above aim, we first distributed the phonations (according to their arrangement in dysphonia scale) into three samples (groups), with $n_i = 42$ phonations per sample. Moreover, to investigate the heterogeneity of small and moderate samples sizes of phonations, two phonations samples were randomly selected with sizes $n_i = 10$ and $n_i = 20$, $1 \leq i \leq 3$. We test the null hypothesis $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3$ versus the alternative $H_0 : \Sigma_1 \neq \Sigma_2 \neq \Sigma_3$, where $\Sigma_i$ is the covariance matrix of the $i$th sample (group) of phonations (as instances) for all 309 dysphonia algorithms (as a variables). Given the fact that each instance of phonations measures one characteristic (of dysphonia algorithms) that differs from those other characteristics in voice rehabilitation LSVT data. That is, the (309) dysphonia algorithms in the LSVT voice rehabilitation scale are heterogeneous between phonations samples. Therefore, rejecting $H_0$ demonstrates the validity (internal consistency) of the LSVT scale.

Table 4 show the statistic measure and $P_{value}$ of $T_g$ and $\rho L_k$ tests. We observe from Table 4 that the $\rho L_k$ test rejects the null hypothesis $H_0$, the results were statistically significant ($P_{value} < 0.01$) for all samples. The results of the $\rho L_k$ test are consistent with those of the original study of LSVT voice rehabilitation (see Tsanas, Little, and Fox 2014), and the results of $\rho L_k$ may consistent with those of $T_g$, however, for $n_i = 10$ the $P_{value}$ of $\rho L_k$ test equals 0.0039, that is, $T_g$ appears more conservative than the $\rho L_k$ to reject $H_0$ when $H_0$ is correct.

### Table 3. The empirical size and power of $\rho L_k y_j$ tests in Example 2, $1 \leq j \leq 4$.

|          | Normal | Uniform | Exponential |
|----------|--------|---------|-------------|
|          | Size   | Power   | Size        | Power   | Size        | Power   |
| $\rho L_k y_1$ | 0.45   | 1.000   | 0.053       | 0.965   | 0.052       | 0.956   |
| $\rho L_k y_2$ | 0.047  | 0.986   | 0.047       | 1.000   | 0.046       | 0.979   |
| $\rho L_k y_3$ | 0.051  | 1.000   | 0.052       | 0.987   | 0.051       | 1.000   |
| $\rho L_k y_4$ | 0.050  | 1.000   | 0.048       | 1.000   | 0.049       | 1.000   |

### Table 4. The statistic measure and $P_{value}$ of $T_g$ and $\rho L_k$ tests, of LSVT voice rehabilitation dataset.

| $n_1 = n_2 = n_3$ | $T_g$ | $P_{value}$ | $(\chi^2)_{\rho L_k}$ | $P_{value}$ |
|-------------------|-------|------------|------------------------|------------|
| 10                | 4.2005| 0.0000     | 3.0833                 | 0.0039     |
| 20                | 9.4051| 0.0000     | 15.1066                | 0.0005     |
| 42                | 13.0521| 0.0000   | 39.7810                | 0.0000     |
Table 5. The statistic measure and $P_{value}$ of $\rho L_k$ test for testing a stationary processes AR(1), in Examples 1 and 2.

| Example 1 | Example 2 |
|-----------|-----------|
| $\chi^2$ | $P_{value}$ | $\chi^2$ | $P_{value}$ | $\chi^2$ | $P_{value}$ |
| $\rho_{L_k y_1}$ | - | - | - | - | - | - |
| $\rho_{L_k y_2}$ | - | - | - | - | - | - |
| $\rho_{L_k y_3}$ | - | - | - | - | - | - |
| $\rho_{L_k (1)}$ | 1.0593 | 0.6528 | 1.6076 | 0.7230 | - | - |

6. Concluding remarks

This article presents a modified test, $L_k$, to test the null hypothesis in (1) for high-dimensional settings with ignoring the assumption of normality. We prove that the $L_k$ converges in distribution to $\chi^2_{k-1}$. Here, we don’t assume the usual conditions about the relationship between the traces of powers of covariance matrix and $p$, but we assume that the relationship between $p$ and $n_{max}$ satisfies $p \geq c(n_{max})^{3/(r-2)}$ for a positive constant $c$. Hence, the $L_k$ test accommodates the large $p$, small $n_{max}$ situations. The most significant two advantages of the modified test, $L_k$, are that (1) it is easy to calculate; (2) it can effectively test the equality of several high-dimensional covariance matrices when $p$ is much bigger than $n_{max}$. The simulation results in Tables 1 and 2 demonstrate that the $\rho L_k$ test has perfect performance and its power tends to 1.

Funding

This work was supported by National Basic Research Program of China (973 Program, 2015CB856004) and the National Natural Science Foundation of China (11531001).

References

Ahmad, M. R. 2017. Testing homogeneity of several covariance matrices and multi-sample sphericity for high-dimensional data under non-normality. Communications in Statistics - Theory and Methods 46 (8):3738–53. doi:10.1080/03610926.2015.1073310.

Ahmad, M. R. 2014. A U-statistic approach for a high-dimensional two-sample mean testing problem under non-normality and Behrens-Fisher setting. Annals of the Institute of Statistical Mathematics 66 (1):33–61. doi:10.1007/s10463-013-0404-2.

Bartlett, M. S. 1937. Properties of sufficiency and statistical tests. Proc. R. Soc. Lond. A 160: 268–82.

Berkes, I., W. Liu, and W. B. Wu. 2014. Komlós,- Major- Tusnády approximation under dependence. The Annals of Probability 42 (2):794–817. doi:10.1214/13-AOP850.

Box, G. E. P. 1949. A general distribution theory for a class of likelihood criteria. Biometrika 36 (3-4):317–46. doi:10.2307/2332671.

Cai, T., and Z. M. Ma. 2013. Optimal hypothesis testing for high dimensional covariance matrices. Bernoulli 19 (5B):2359–88. doi:10.3150/12-BEJ455.

Cai, T., W. Liu, and Y. Xia. 2013. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. Journal of the American Statistical Association 108: 265–77.
Chao, M., and R. Glaser. 1978. The exact distribution of Bartlett’s test statistic for homogeneity of variances with unequal sample sizes. *Journal of the American Statistical Association* 73 (362): 422–6. doi:10.1080/01621459.1978.10481594.

Chen, S. X., L. X. Zhang, and P. S. Zhang. 2010. Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association* 105 (490):810–9. doi:10.1198/jasa.2010.tm09560.

Dykstra, R. L. 1970. Establishing the positive definiteness of the sample covariance matrix. *The Annals of Mathematical Statistics* 41 (6):2153–4. doi:10.1214/aoms/1177696719.

Gupta, A. K., and D. K. Nagar. 2000. *Matrix variate distribution*. New York: Chapman and Hall/CRC.

Ishll, A., K. Yata, and M. Aoshima. 2019. Equality tests of high-dimensional covariance matrices under the strongly spiked eigenvalue model. *Journal of Statistical Planning and Inference* 202: 99–111.

Jiang, D., T. Jiang, and F. Yang. 2012. Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *Journal of Statistical Planning and Inference* 142 (8): 2241–56. doi:10.1016/j.jspi.2012.02.057.

Jiang, T., and F. Yang. 2013. Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *The Annals of Statistics* 41 (4):2029–74. doi:10.1214/13-AOS1134.

Ledoit, O., and M. Wolf. 2002. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics* 30 (4):1081–102. doi:10.1214/aos/1031689018.

Liu, W., and Z. Lin. 2009. Strong approximation for a class of stationary processes. *Stochastic Processes and Their Applications* (1)119:249–80. doi:10.1016/j.spa.2008.01.012.

Peng, L. H., S. X. Chen, and W. Zhou. 2016. More powerful for tests sparse high-dimensional covariance matrices. *Journal of Multivariate Analysis* 149:124–43. doi:10.1016/j.jmva.2016.03.008.

Qiu, Y., and S. X. Chen. 2012. Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *The Annals of Statistics* (3)40:1285–314. doi:10.1214/12-AOS1002.

Schott, J. R. 2007. A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Computational Statistics & Data Analysis* (12)51:6535–42. doi:10.1016/j.csda.2007.03.004.

Srivastava, M. S., and H. Yanagihara. 2010. Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis* (6)101:1319–29. doi:10.1016/j.jmva.2009.12.010.

Srivastava, M. S. 2005. Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society* (2)35:251–72. doi:10.14490/jjss.35.251.

de TorrenTE, L., S. Zimmerman, M. Suzuki, M. Christopeit, J. M. Greally, and J. C. Mar. 2020. The shape of gene expression distributions matter: How incorporating distribution shape improves the interpretation of cancer transcriptomic data. *BMC Bioinformatics* (Suppl 21)21: 562. doi:10.1186/s12859-020-03892-w.

Tsanas, M. A., C. Little, and L. O. Fox. 2014. Objective automatic assessment of rehabilitative speech treatment in Parkinson’s disease. *IEEE Transaction on Neural Systems and Rehabilitation Engineering* 22: 181–90.

Zhang, Q. Y., J. Hu, and Z. D. Bai. 2020. Modified Pillai’s trace statistics for two high-dimensional sample covariance matrices. *Journal of Statistical Planning and Inference* 207:255–75. doi:10.1016/j.jspi.2020.01.002.

Zhong, P.-S., R. Li, and S. Santo. Homogeneity tests of covariance matrices with high-dimensional longitudinal data. *Biometrika* (3)106 2019: 619–34. doi:10.1093/biomet/asz011.
Appendix

A.1. Proof of Theorem 3.1

Let \( S_l^{(i)} = \sum_{j=1}^{p} X_l^{(i,j)} \) and \( \overline{S}_p^{(i)} = \sum_{l=1}^{n} S_l^{(i)}/n_l \). By the strong invariance principle in (3), \( S_p^{(i)} \) can be written as

\[
S_p^{(i)} = \sigma_i \overline{B}_l^{(i)}(p) + \epsilon_p^{(i)}, \quad \text{a.s.,}
\]

for \( 1 \leq l \leq n_i, 1 \leq i \leq k \), and therefore,

\[
S_p^{(i)} - \overline{S}_p^{(i)} = \sigma_i (B_l^{(i)}(p) - \overline{B}_l^{(i)}(p)) + \epsilon_p^{(i)} - \epsilon_p^{(i)}, \quad \text{a.s.}
\]

where \( \overline{B}_l^{(i)}(p) = \sum_{i=1}^{n_i} B_l^{(i)}(p)/n_i \), for each \( i \), \( B_l^{(i)}(t), 1 \leq l \leq n_i \), are mutually independent \( n_i \) standard Brownian motions, and the random variables \( \epsilon_p^{(i)} \) satisfies that \( \epsilon_p^{(i)}/p^{1/r} \rightarrow 0 \), a.s. as \( p \rightarrow \infty \) for all \( 1 \leq l \leq n_i, 1 \leq i \leq k \). Note that

\[
\chi_i := \frac{\sum_{l=1}^{n_i} (B_l^{(i)}(p) - \overline{B}_l^{(i)}(p))^2}{p} \sim \chi^2_{n-1}.
\]

It means that \( \chi_i \) does not depend on \( p \) for all \( 1 \leq l \leq n_i, n_i > 1 \) and \( 1 \leq i \leq k \). From (6), (A.1) and (A.2), it follows that

\[
\frac{(n_i - 1) \mathbf{1}_i S_i \mathbf{1}_T}{p \sigma_i^2} = \frac{1}{p \sigma_i^2} \sum_{l=1}^{n_i} (S_l^{(i)} - \overline{S}_p^{(i)})^2 \sim \chi_i + \zeta_i(p)
\]

for large \( p, 1 \leq l \leq n_i, n_i > 1 \) and \( 1 \leq i \leq k \), where

\[
\zeta_i(p) = 2 \frac{\sum_{l=1}^{n_i} (B_l^{(i)}(p) - \overline{B}_l^{(i)}(p)) (\epsilon_p^{(i)} - \overline{\epsilon}_p^{(i)})}{\sqrt{p}} + \frac{1}{p \sigma_i^2} \sum_{l=1}^{n_i} (\epsilon_p^{(i)} - \overline{\epsilon}_p^{(i)})^2
\]

\[
= \frac{\sqrt{n_i} p^{1/r}}{\sqrt{p}} \left( \frac{\sum_{l=1}^{n_i} (B_l^{(i)}(p) - \overline{B}_l^{(i)}(p)) (\epsilon_p^{(i)} - \overline{\epsilon}_p^{(i)})}{\sqrt{n_i} p} + \frac{1}{p} \sum_{l=1}^{n_i} (\epsilon_p^{(i)} - \overline{\epsilon}_p^{(i)})^2 \right).
\]

Since \( (B_l^{(i)}(p) - \overline{B}_l^{(i)}(p))/\sqrt{n_i p} \sim N(0, (n_i - 1)/n_i^2), (\epsilon_p^{(i)} - \overline{\epsilon}_p^{(i)})/p^{1/r} \rightarrow 0 \), a.s. as \( p \rightarrow \infty \) and \( \sqrt{n_i p^{1/r}}/\sqrt{p} \) is bounded, or \( p \geq c(n_i)^{r/(r-2)} \) for some positive constant \( c \), it follows that \( \zeta_i(p) \rightarrow 0 \), as \( p \rightarrow \infty \) for \( 1 \leq i \leq k \). Thus, by (A.4) and (A.5) we have

\[
\frac{(n_i - 1) \mathbf{1}_i S_i \mathbf{1}_T}{p \sigma_i^2} \rightarrow \chi_i, \quad \chi_i \sim \chi^2_{n-1}
\]

as \( p \rightarrow \infty \) for \( 1 \leq i \leq k \).

A.2. Proof of Theorem 3.2

Under the original hypothesis \( H_0 \) in (1) with \( \Sigma > 0 \), we have \( \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2 \). Let \( a_i = (n_i - 1)/(n - k) \) and \( N = (n - k)/2 \). By (A.4) \( I_k \) can be written as \( I_k = -2N \log J_{n,k} = -2 \log J_{n,k}^N \) for \( 1 \leq i \leq k \), where
It follows that

\[ C_{n,k}^{-1} \prod_{i=1}^{k} (Z_i)^{a_i} \]

where \( \hat{\Sigma}_i := \hat{\Sigma}_i / \rho \sigma^2 \).

On the other hand, if \( Y_i > 1 \), then

\[ (Z_i Y_i)^{a_i} = (Z_i Y_i - 1 + Z_i)^{a_i} \]

\[ = (Z_i)^{a_i} + (n_i - 1)Z_i(Y_i - 1)\sum_{j=1}^{n_i-1} C_{n_i-2-j}^{a_i-1} [Z_i(Y_i - 1)]^{j-1}(Z_i)^{n_i-2-(j-1)} \]

\[ \leq (Z_i)^{a_i} + (n_i - 1)Z_i(Y_i - 1)Z_i(Y_i - 1) + Z_i)^{a_i} \]

\[ = (Z_i)^{a_i} + (n_i - 1)Z_i(Y_i - 1)(Z_i Y_i)^{a_i-2}, \quad n_i \geq 2 \]

\[ (Z_i Y_i)^{-\alpha} \leq (Z_i)^{-\alpha}, \quad \alpha > 0 \]

for \( 1 \leq i \leq k \). Similarly, if \( Y_i \leq 1 \), then
$$ (Z_i Y_i)^{n-1} = |Z_i - Z_i(1 - Y_i)|^{n-1} $$

$$ = (Z_i)^{n-1} - (n_i - 1)Z_i(1 - Y_i) \sum_{j=1}^{n_i-1} C_{n_i-2}^{j-1} \frac{1}{j} [-Z_i(1 - Y_i)]^j (Z_i)^{n_i-2-j} $$

$$ \geq (Z_i)^{n-1} - (n_i - 1)Z_i(1 - Y_i) |Z_i - Z_i(1 - Y_i)|^{n-2} $$

$$ \geq (Z_i)^{n-1} - (n_i - 1)Z_i(1 - Y_i) (Z_i)^{n_i-2}, \quad n_i \geq 2 $$

$$ (Z_i Y_i)^{-\alpha} \geq (Z_i)^{-\alpha}, \quad \alpha > 0 $$

for $1 \leq i \leq k$.

Note that $0 \leq Z_i Y_i \leq 1, 0 < Z_i < 1$, and therefore, $|Z_i(1 - Y_i)| = |Z_i(1 - Y_i)| \leq 1$. By (A.5) we know that

$$ n_i |Y_i - 1| = \frac{n_i |\hat{z}_i(p)/Z_i - \sum_{j=1}^{k} \hat{z}_i(p)/\sum_{j=1}^{k} Z_i|}{1 + \sum_{j=1}^{k} \hat{z}_i(p)/\sum_{j=1}^{k} Z_i} \to 0, \quad a.s. $$

as $n_{\text{min}} \to \infty$ for $1 \leq i \leq k$ when $n_{\text{max}} \sqrt{n_{\text{max}}} \Gamma(r)/\sqrt{p}$ is bounded, or $p \geq c(n_{\text{max}})^{3r/(r-2)}$ for some positive constant $c$. Thus, by (A.7), (A.8), (A.9) and the control convergence theorem, we can get that

$$ \lim_{n_{\text{min}} \to \infty} \mathbb{E}(J_{n,k}^N) = \lim_{n_{\text{min}} \to \infty} \mathbb{E}\left( \left[ C_{n_i}^{k} \prod_{i=1}^{k} (Z_i Y_i)^{a_i} \right]^{N_i} \right) \leq (1 + \gamma)^{-\frac{(k-1)}{2}} $$

$$ \lim_{n_{\text{min}} \to \infty} \mathbb{E}(J_{n,k}^N) = \lim_{n_{\text{min}} \to \infty} \mathbb{E}\left( \left[ C_{n_i}^{k} \prod_{i=1}^{k} (Z_i Y_i)^{a_i} \right]^{N_i} \right) \geq (1 + \gamma)^{-\frac{(k-1)}{2}}. $$

for $p \geq c(n_{\text{max}})^{3r/(r-2)}$. This means that

$$ \mathbb{E}(J_{n,k}^N) \to (1 + \gamma)^{-\frac{(k-1)}{2}} \quad (A.10) $$

as $n_{\text{min}} \to \infty$ for $p \geq c(n_{\text{max}})^{3r/(r-2)}$. Note that the random variable $-2 \log X$ is subject to $\chi^2_{k-1}$ distribution if and only if the $\gamma$-th order moment of the positive random variable $X$ is equal to $(1 + \gamma)^{-\frac{(k-1)}{2}}$. Thus, let $X = J_{n,k}^N$, by (A.10) we have

$$ L_k = -2N \log J_{n,k} = -2 \log J_{n,k}^N \to \chi^2_{k-1}, \quad \text{as} \quad n_{\text{min}} \to \infty. $$

It is clear that for positive number $\rho_n \to 1$ we have $\rho_n L_k \to \chi^2_{k-1}$ as $n_{\text{min}} \to \infty$ for fixed $k \geq 2$.

### A.3. Results of Examples 1 and 2

In both Examples 1 and 2, we considered the null hypothesis test for large dimension $p = 350$, however, the sample sizes were $n_i = 100$ and $n_i = 101$, respectively. We test the null hypothesis $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma$, where $\Sigma$ is a $p \times p$ matrix whose $(a,b)$th element are defined by $(-1)^{(a-b)}(0.40)^{|a-b|}$. We considered the null hypothesis test for two cases, the Gaussian and exponential AR(1) models, respectively, with the dimension $p = 350$. Table 5 shows the statistic measure and $P_{\text{value}}$ of the $\rho L_k$ test, we observe from Table 5 that the $\rho L_k$ test does not reject the null hypothesis $H_0$, the results are statistically significant, that is, $P_{\text{values}}$ of the $\rho L_k$ tests does not fall into the rejection region $R_k$. The results of $\rho L_k$ test in Table 5 are completely consistent with the results in Tables 1–3.