Kybernetika

M. Rauf Ahmad

A homogeneity test of large dimensional covariance matrices under non-normality

Kybernetika, Vol. 54 (2018), No. 5, 908–920

Persistent URL: http://dml.cz/dmlcz/147534

KYBERNETIKA — VOLUME 54 (2018), NUMBER 5, PAGES 908–920

A HOMOGENEITY TEST OF LARGE DIMENSIONAL COVARIANCE MATRICES UNDER NON-NORMALITY

M. Rauf Ahmad

A test statistic for homogeneity of two or more covariance matrices is presented when the distributions may be non-normal and the dimension may exceed the sample size. Using the Frobenius norm of the difference of null and alternative hypotheses, the statistic is constructed as a linear combination of consistent, location-invariant, estimators of trace functions that constitute the norm. These estimators are defined as U-statistics and the corresponding theory is exploited to derive the normal limit of the statistic under a few mild assumptions as both sample size and dimension grow large. Simulations are used to assess the accuracy of the statistic.

Keywords: high-dimensional inference, covariance testing, U-statistics, non-normality

Classification: 62H15

1. INTRODUCTION

Let $\mathbf{X}_{ik} = (X_{i1k}, \dots, X_{ipk})'$, $k = 1, \dots, n_i$ be independent random vectors drawn from a non-degenerate multivariate distribution $\mathcal{F}_i$, $i = 1, \dots, g$. Let $\mathrm{E}(\mathbf{X}_{ik}) = \boldsymbol{\mu}_i \in \mathbb{R}^p$, $\mathrm{Cov}(\mathbf{X}_{ik}) = \boldsymbol{\Sigma}_i \in \mathbb{R}^{p\times p}_{>0}$, where $\mathbb{R}^{p\times p}_{>0}$ denotes the space of real, symmetric, positive-definite matrices. Let the vectors be generated by a probability space $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ with its usual meaning, i. e. $\mathcal{X}$ is the sample space, $\mathcal{A}$ an appropriate σ-algebra and $\mathcal{P}$ a probability measure.

Our objective is to construct a statistic to test $H_{0g} : \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}\ \forall\ i$ vs. $H_{1g} : \boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}$ for at least one i, when $p \gg n_i$ and $\mathcal{F}_i$ need not necessarily be multivariate normal. As the classical likelihood-ratio tests collapse for $p > n_i$, they need to be modified. Some recent modifications include [5, 6, 13, 14, 20, 24]; for more details and a review, see [7].

[2] also present a test statistic for H_{0g}, using the component estimators of the test statistic defined as U-statistics of second order symmetric kernels. The kernels are non-invariant bilinear forms of the vectors coming from g independent samples, which makes the test statistic also non-invariant which, in turn, implies that the validity of the test statistic depends on assuming zero centroids for all populations, i. e. $\boldsymbol{\mu}_i = \mathbf{0}\ \forall\ i$. Under H_{0g}, an extension to testing multi-sample sphericity or identity structures of the common covariance matrix is also provided. The present article extends the homogeneity test in [2] to the location-invariant case which is valid whether the mean vectors are assumed

DOI: 10.14736/kyb-2018-5-0908

zero or not; in other words, for the proposed test, the mean vectors can be assumed zero without any loss of generality. A corresponding extension of tests for multi-sample sphericity and identity to the location-invariant setting is given in [1].

The proposed test statistic is a linear combination of unbiased, consistent, location-invariant estimators. Keeping the underlying assumptions minimum and mild, the joint distribution of estimators is used to derive the normal limit of the statistic. The non-parametric nature of estimators, being U-statistics, help us relax normality assumption, and replace it with a more flexible model

$$\mathbf{a}_{ik} = \mathbf{\Lambda}_i \mathbf{b}_{ik}, \tag{1}$$

where $\mathbf{a}_{ik} = \mathbf{X}_{ik} - \boldsymbol{\mu}_i$, $\mathbf{b}_{ik} = (b_{ik1}, \ldots, b_{ikp})'$ are independent vectors with $\mathrm{E}(\mathbf{b}_{ik}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{b}_{ik}) = \mathbf{I}$, and $\mathbf{\Lambda}_i$ is a known $p \times p$ matrix of constants such that $\mathbf{\Lambda}_i'\mathbf{\Lambda}_i = \mathbf{A}_i$ and $\mathbf{\Lambda}_i\mathbf{\Lambda}_i' = \mathbf{\Sigma}_i$, $i = 1, \ldots, g$. The mean-deviated form of $\mathbf{a}_{ik}$ fits into the location-invariant form of the test so that we can assume w.l.o.g. $\boldsymbol{\mu}_i = \mathbf{0}\ \forall\ i$. The test statistic and its properties are studied in the next section and its accuracy is assessed in Sec. 3. Proofs are collected in the Appendix.

2. TEST STATISTICS AND THEIR PROPERTIES

2.1. Two-sample case

We begin with the two sample case ($g = 2$) which will be extended to the general case in the next section. Then, for Model (1), $\mathbf{X}_{ik} \sim \mathcal{F}_i$, $k = 1, \ldots, n_i$, are independent random samples from $\mathcal{F}_i$ with parameters $\boldsymbol{\mu}_i$ and $\mathbf{\Sigma}_i$, $i = 1$, 2, as defined above. To test $H_{02} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, let the Frobenius norm $\tau_{12} = \|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|^2$ measure the distance between the covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$. Note that, if the data set can be assumed to follow a Hilbert space, $\mathcal{H}$, then τ_{12} is also the Hilbert-Schmidt norm. Since $\tau_{12} = \mathrm{tr}(\mathbf{\Sigma}_1^2) + \mathrm{tr}(\mathbf{\Sigma}_2^2) - 2\,\mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)$, a test for any pair of covariance matrices can be constructed by either using efficient estimators of $\mathbf{\Sigma}_i$, $i = 1$, 2, or directly of the traces that compose the norm. We shall pursue the later approach by defining unbiased and consistent estimators of the traces. Let $\widehat{\delta}_i$, $\widehat{\delta}_{12}$ denote these estimators, so that $\widehat{\tau}_{12} = \sum_{i=1}^2 \widehat{\delta}_i - 2\widehat{\delta}_{12}$ is an empirical measure of τ_{12} with $\mathrm{E}(\widehat{\tau}_{12}) = \tau_{12} = 0$ under H_{02}. Then

$$\widehat{\tau}_{12} - \tau_{12} = \sum_{i=1}^{2} \{\widehat{\delta}_i - \mathrm{tr}(\mathbf{\Sigma}_i^2)\} - 2\{\widehat{\delta}_{12} - \mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)\}$$

with $\mathrm{E}(\widehat{\tau}_{12} - \tau_{12}) = 0$ under H_{02} and H_{12}. With a simple re-scaling which will help us allow $p \to \infty$ along with $n \to \infty$, we define

$$\mathrm{T}_2 = (\widehat{\tau}_{12} - \tau_{12})/p^2 = \sum_{i=1}^{2} a_i \widetilde{\delta}_i - 2a_{12}\widetilde{\delta}_{12} \tag{2}$$

with $a_i = \mathrm{tr}(\mathbf{\Sigma}_i^2)/p^2$, $\widetilde{\delta}_i = \widehat{\delta}_i / \mathrm{E}(\widehat{\delta}_i) - 1$, $a_{12} = \mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)/p^2$, $\widetilde{\delta}_{12} = \widehat{\delta}_{12}/\mathrm{E}(\widehat{\delta}_{12}) - 1$. Under H_{02}, $a_i = a_{12} = \mathrm{tr}(\mathbf{\Sigma}^2)/p^2$ with an estimator, say $\widehat{a}$. Note that, T_2 is a linear combination of non-degenerate weighted U-statistics, with coefficients a_i, a_{12} [11]; see also [15, 16] for a detailed treatment of degenerate case. Now, T_2 is a test statistic, also under the null using $\widehat{a}$ for a. With a slight abuse of notation, we henceforth use T_2 as test statistic for both cases, where an efficient estimator of a will be plugged in later after studying the moments of T_2.

The properties of T_2 obviously follow from those of $\widehat{\delta}_i$, $i = 1, 2$, $\widehat{\delta}_{12}$. Let $\overline{\mathbf{X}}_i = \sum_{k=1}^{n_i} \mathbf{X}_{ki}/n_i$ and $\widehat{\mathbf{\Sigma}}_i = \sum_{k=1}^{n_i} \widetilde{\mathbf{X}}_{ki}\widetilde{\mathbf{X}}'_{ki}/(n_i - 1)$, be unbiased estimators of $\boldsymbol{\mu}_i$, $\mathbf{\Sigma}_i$, respectively, with $\widetilde{\mathbf{X}}_i = \mathbf{X}_{ki} - \overline{\mathbf{X}}_i$. Let $Q_i = \sum_{k=1}^{n_i}(\widetilde{\mathbf{X}}'_{ki}\widetilde{\mathbf{X}}_{ki})^2/(n_i-1)$, $\eta_i = (n_i-1)/[n_i(n_i-2)(n_i-3)]$. Then $\widehat{\delta}_i$ and $\widehat{\delta}_{12}$, which follows by independence, as estimators of $\mathrm{tr}(\mathbf{\Sigma}_i^2)$ and $\mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)$, are defined as

$$\widehat{\delta}_i = \eta_i\{2\,\mathrm{tr}(\widehat{\mathbf{\Sigma}}_i^2) + (n_i^2 - 3n_i + 1)[\mathrm{tr}(\widehat{\mathbf{\Sigma}}_i)]^2 - n_iQ_i\}, \qquad \widehat{\delta}_{12} = \mathrm{tr}(\widehat{\mathbf{\Sigma}}_1\widehat{\mathbf{\Sigma}}_2). \tag{3}$$

Theorem 2.1, proved in Sec. B.2, gives moments of $\widehat{\delta}_i$ and $\widehat{\delta}_{12}$.

Theorem 2.1. Given Model (1), $\widehat{\delta}_i$, $\widehat{\delta}_{12}$, $\widetilde{\delta}_i$, $\widetilde{\delta}_{12}$ as above, M_2, M_3 as in Theorem A.1. Let $n = \sum_{i=1}^2 n_i$, $Q(n_i) = n_i(n_i-1)$, $P(n_i) = Q(n_i)(n_i-2)(n_i-3)$. Then $\mathrm{E}(\widehat{\delta}_i) = \mathrm{tr}(\mathbf{\Sigma}_i^2)$, $\mathrm{E}(\widehat{\delta}_{12}) = \mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)$ and, using $a(n_i) = 2n_i^3 - 9n_i^2 + 9n_i - 16$, $b(n_i) = n_i^2 - 3n_i + 8$,

$$\begin{aligned}
\mathrm{Var}(\widehat{\delta}_i) &= \frac{4}{P(n_i)}\Big[a(n_i)\,\mathrm{tr}(\mathbf{\Sigma}_i^4) + b(n_i)[\mathrm{tr}(\mathbf{\Sigma}_i^2)]^2 + (M_2 + M_3)O(n_i^2)\Big]\\
\mathrm{Var}(\widehat{\delta}_{12}) &= \frac{2}{(n_i-1)(n_j-1)}\Big[(n-1)\,\mathrm{tr}(\mathbf{\Sigma}_i\mathbf{\Sigma}_j)^2 + [\mathrm{tr}(\mathbf{\Sigma}_i\mathbf{\Sigma}_j)]^2 + (M_2+M_3)O(n_i)\Big]\\
\mathrm{Cov}(\widehat{\delta}_1, \widehat{\delta}_{12}) &= \frac{4}{Q(n_1)}\Big[n_1\,\mathrm{tr}(\mathbf{\Sigma}_1^3\mathbf{\Sigma}_2) + M_2O(n_1)\Big].
\end{aligned}$$

Further, $\mathrm{Var}(\widetilde{\delta}_i) = O(1/n_i)$, $\mathrm{Var}(\widetilde{\delta}_{12}) = O(n/n_1n_2)$, $\mathrm{Cov}(\widetilde{\delta}_i, \widetilde{\delta}_{12}) = O(1/n_i)$, $i = 1, 2$.

The last part of Theorem 2.1 refers to moment-ratios, uniformly bounded in p, of scaled estimators, $\widetilde{\delta}$'s. This explains why T_2 is likewise scaled. It will help us determine the non-degenerate limit of T_2 under mild assumptions which will also be stated with same scaling to keep the entire program consistent. We state these assumptions for the general case, $i, j = 1, \ldots g \geq 2$, so that we can also use them in multi-sample extension in the following section without repetition.

Let λ_{is}, $s = 1, \ldots, p$, be the eigenvalues of $\mathbf{\Sigma}_i$, so that $\nu_{is} = \lambda_{is}/p$ are those of $\mathbf{\Sigma}_i/p$. Let $\mathbf{\Sigma}_i^a \odot \mathbf{\Sigma}_j^b$, $\mathbf{\Sigma}_i^a \otimes \mathbf{\Sigma}_j^b$, $a, b \in \mathbb{R}^+$, be Hadamard and Kronecker products of matrices raised to powers indicated. Denote $\mathrm{tr}(\mathbf{\Sigma}_i^a \otimes \mathbf{\Sigma}_j^b) = \sum_s \lambda_{is}^a \sum_s \lambda_{js}^b = \theta_{Kij}^{a,b}$, similarly $\theta_{Hij}^{a,b}$ for Hadamard product. We need the following assumptions, where $i, j = 1, \ldots g$, $i \neq j$, with $g \geq 2$.

Assumption 2.2. $\mathrm{E}(b_{iks}^4) = 3 + \gamma$, $\forall\ s = 1, \ldots, p$, $\gamma \in \mathbb{R}$.

Assumption 2.3. $\lim_{p\to\infty} \sum_{s=1}^p \nu_{is} \to \nu_0 = O(1)$.

Assumption 2.4. $\lim_{p\to\infty} \theta_{12}^H/\theta_{12}^K \to 0$, $a, b = 1, 2, 3$, $a + b \leq 4$.

Assumption 2.5. (i) $p/n_i \to c_i \leq c \in (0, \infty)$, $i = 1, 2$ (ii) $n_i/n_j \to c_0 \in (0, \infty)$.

Assumption 2.2 helps us relax normality and work under Model (1). Assumption 2.3 holds conveniently for most covariance structures. An immediate consequence of Assumption 2.3 is that $\sum_s \nu_s^2 = O(1)$, so that in the sequel, the assumption may also imply its consequence. Assumption 2.5 controls the joint growth of n_i and p. It is more a mathematical requirement as it is shown in Sec. 3 that T_g, $g \geq 2$, performs accurately for moderate n_i and any p. Assumption 2.4 is mild with the numerator only a fraction of the denominator, and it is not required when $\mathcal{F}_i$ is multivariate normal. For the limit of T_2 under the assumptions, we get $\mathrm{E}(\mathrm{T}_2) = 0$ where $\sigma^2_{\mathrm{T}_2} = \mathrm{Var}(\mathrm{T}_2)$ follows as

$$\sigma^2_{\mathrm{T}_2} = \sum_{i=1}^{2} a_i^2 \,\mathrm{Var}(\widetilde{\delta}_i) + 4a_{12}^2 \,\mathrm{Var}(\widetilde{\delta}_{12}) - 4a_{12} \sum_{i=1}^{2} a_i \,\mathrm{Cov}(\widetilde{\delta}_i, \widetilde{\delta}_{12}). \tag{4}$$

Denoting $\kappa = \mathrm{tr}(\mathbf{\Sigma}^4)/[\mathrm{tr}(\mathbf{\Sigma}^2)]^2$, T_2 and $\sigma^2_{\mathrm{T}_2}$ simplify under H_{02} as

$$\mathrm{T}_{20} = a\left[(\widehat{\delta}_1 + \widehat{\delta}_2 - 2\widehat{\delta}_{12})/\,\mathrm{tr}(\mathbf{\Sigma}^2)\right] \tag{5}$$

$$\sigma^2_{\mathrm{T}_{20}} = a^2\left[O\left(1/n_1 + 1/n_2\right)\kappa + 4(1/n_1 + 1/n_2)^2\right]. \tag{6}$$

Ignoring the vanishing terms under the assumptions, $\sigma^2_{\mathrm{T}_{20}}$ can be approximated as

$$\sigma^2_{\mathrm{T}_{20}} = 4\left(1/n_1 + 1/n_2\right)^2 a^2[1 + o(1)]. \tag{7}$$

Since T_2 is a linear combination of $\widetilde{\delta}_i$ and $\widetilde{\delta}_{12}$, with a_i, a_{12} uniformly bounded under Assumption 2.3, and with the samples independent, it conveniently leads to a limit of T_2, where the same limit under H_{02} holds as a special case [8]. For a similar approach in mean testing using U-statistics theory, see also [18].

To use T_2 in practice, however, we need to estimate $\mathrm{Var}(\mathrm{T}_2)$, which follows by using $\widehat{\delta}_i$ and $\widehat{\delta}_{ij}$ as plug-in consistent estimators of $\mathrm{tr}(\mathbf{\Sigma}_i^2)$ and $\mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)$ so that $\widehat{\sigma}^2_{\mathrm{T}_2}$ is a consistent estimator where $\widehat{\sigma}_{\mathrm{T}_2}/\sigma_{\mathrm{T}_2} \to 1$. Under H_{02}, we use $\widehat{\delta}_{12}$ as a (pooled) estimator of $\mathrm{tr}(\mathbf{\Sigma}^2)$ in $\sigma^2_{\mathrm{T}_2}$. The following theorem, proved in Sec. B.3, summarizes the result.

Theorem 2.6. For T_2 in Eqn. (2), $(\mathrm{T}_2 - \mathrm{E}(\mathrm{T}_2))/\widehat{\sigma}_{\mathrm{T}_2} \xrightarrow{\mathcal{D}} N(0,1)$ under Assumptions 2.2–2.5, as $n_i, p \to \infty$, where $\widehat{\sigma}^2_{\mathrm{T}_2}$ is a consistent estimator of $\sigma^2_{\mathrm{T}_2}$ as defined above.

2.2. Multi-sample extension

The general case is a straightforward extension of the two-sample case so that we can skip many unnecessary details. Consider again the data set up from Sec. 1 under Model (1), where now $H_{0g} : \mathbf{\Sigma}_i = \mathbf{\Sigma}\ \forall\ i = 1, \ldots, g$ vs $H_{1g} : \mathbf{\Sigma}_i \neq \mathbf{\Sigma}_j$ for at least one pair (i, j), $i \neq j$, $i, j = 1, \ldots, g$. To extend T_2 in Eqn. (2) for g samples, we consider $\tau_g = \sum_{i<j}^{g} \tau_{ij}$, sum of Frobenius norms over all distinct pairs, with $\tau_g = 0$ under H_{0g}. Then

$$\mathrm{T}_g = \sum_{\substack{i=1 \\ i<j}}^{g}\sum_{j=1}^{g} \mathrm{T}_{ij} = (g-1)\sum_{i=1}^{g} a_i\widetilde{\delta}_i - 2\sum_{\substack{i=1 \\ i<j}}^{g}\sum_{j=1}^{g} a_{ij}\widetilde{\delta}_{ij}, \tag{8}$$

where

$$\mathrm{T}_{ij} = (\widehat{\tau}_{ij} - \tau_{ij})/p^2 = a_i\widetilde{\delta}_i + a_j\widetilde{\delta}_j - 2\widetilde{\delta}_{ij}$$

is like T_2 for any pair (i,j) with $\widehat{\tau}_{ij} = \widehat{\delta}_i + \widehat{\delta}_j - 2\widehat{\delta}_{ij}$ as an empirical measure for $\tau_{ij} = \|\mathbf{\Sigma}_i - \mathbf{\Sigma}_j\|^2$. The estimator $\widehat{\delta}_i$ remains as in Eqn. (3), where $\widehat{\delta}_{ij} = \mathrm{tr}(\widehat{\mathbf{\Sigma}}_i\widehat{\mathbf{\Sigma}}_j)$ is defined like $\widehat{\delta}_{12}$ for any pair (i,j), with $\widehat{\mathbf{\Sigma}}_i$ already defined around Eqn. (3). The properties of estimators in Theorem 2.1 extend to the general case by the same token. We, therefore, only focus on the limit of T_g under Assumptions 2.2−2.5, which are already given in general form. For this, we note that

$$\mathrm{E}(\mathrm{T}_g) = (g-1)\sum_{i=1}^{g} a_i\widetilde{\delta}_i - 2\sum_{\substack{i=1 \\ i<j}}^{g}\sum_{j=1}^{g} a_{ij}\widetilde{\delta}_{ij} \tag{9}$$

$$\begin{aligned}\mathrm{Var}(\mathrm{T}_g) = {} & (g-1)^2\sum_{i=1}^{g} a_i^2\,\mathrm{Var}(\widetilde{\delta}_i) + 4\sum_{\substack{i=1 \\ i<j}}^{g}\sum_{j=1}^{g} a_{ij}^2\,\mathrm{Var}(\widetilde{\delta}_{ij}) + 8\sum_{\substack{i=1 \\ i<j<j'}}^{g}\sum_{j=1}^{g}\sum_{j'=1}^{g} a_{ij}a_{ij'}\,\mathrm{Cov}(\widetilde{\delta}_{ij}, \widetilde{\delta}_{ij'}) \\ & + 8\sum_{\substack{i=1 \\ i<i'<j}}^{g}\sum_{i'=1}^{g}\sum_{j=1}^{g} a_{ij}a_{i'j}\,\mathrm{Cov}(\widetilde{\delta}_{ij}, \widetilde{\delta}_{i'j}) - 8(g-1)\sum_{\substack{i=1 \\ i<j}}^{g}\sum_{j=1}^{g} a_i a_{ij}\,\mathrm{Cov}(\widetilde{\delta}_i, \widetilde{\delta}_{ij}). \end{aligned} \tag{10}$$

Consider $\mathrm{Var}(\widetilde{\delta}_i) = 4[\{\mathrm{tr}(\mathbf{\Sigma}_i^4)O(1/n_i) + (M_2 + M_3)O(1/n_i^2)\}/\delta_i^2 + O(1/n_i^2)]$. As $p \to \infty$, first two terms vanish by Assumptions 2.4−2.5, where $\mathrm{tr}(\mathbf{\Sigma}_i^4)/\delta_i^2 \leq 1$, so that $\mathrm{Var}(n_i\widetilde{\delta}_i) = O(1)$. Similarly $\mathrm{Var}(\widetilde{\delta}_{ij})) = 2[\{\mathrm{tr}(\mathbf{\Sigma}_i\mathbf{\Sigma}_j)^2 O(1/n_i + 1/n_j) + (M_2 + M_3)O(1/n_i + 1/n_j)\}/\delta_{ij}^2 + 1/n_in_j]$ so that $\mathrm{Var}(\sqrt{n_in_j}\widetilde{\delta}_{ij}) = O(1)$. Further, all covariances in $\mathrm{Var}(\mathrm{T}_g)$ vanish. Then, as $n_i, p \to \infty$,

$$\left((\widehat{\delta}_i - \mathrm{E}(\widehat{\delta}_i))/\sqrt{\mathrm{Var}(\widehat{\delta}_i)} \quad (\widehat{\delta}_{ij} - \mathrm{E}(\widehat{\delta}_{ij}))/\sqrt{\mathrm{Var}(\widehat{\delta}_{ij})}\right) \xrightarrow{\mathcal{D}} (\mathbf{0},\ \mathbf{I}), \tag{11}$$

with $\mathbf{I}_2$ identity matrix. Under H_{0g}, $\mathrm{T}_g = [a/\delta^2]\{(g-1)\sum_{i=1}^{g}\widehat{\delta}_i - 2\sum_{i<j}^{g}\widehat{\delta}_{ij}\}$ with $\mathrm{E}(\mathrm{T}_g) = 0$ and

$$\mathrm{Var}(\mathrm{T}_g) = 4a^2\left[(g-1)^2\sum_{i=1}^{g}\frac{1}{n_i^2} + \sum_{\substack{i=1 \\ i<j}}^{g}\sum_{j=1}^{g}\frac{2}{n_in_j}\right]. \tag{12}$$

With T_g an extension of T_2 using Frobenius norms of pairwise differences, and due to the independence of samples, the asymptotic distribution for T_g follows on the same lines as for T_2. Further, a consistent estimator of $\mathrm{Var}(\mathrm{T}_g) = \sigma^2_{\mathrm{T}_g}$, follows similarly by plugging the consistent estimators $\widehat{\delta}_i$ and $\widehat{\delta}_{ij}$, where under H_{0g}, the common trace

$a = \text{tr}(\mathbf{\Sigma}^2)$ is estimated by a pooled estimator as $\widehat{a} = \sum_{i<j} P(n_i, n_j)\widehat{\delta}_{ij} / \sum_{i<j} P(n_i, n_j)$, where $P(n_i, n_j) = Q(n_i)Q(n_j)$, $Q(n_i) = n_i(n_i - 1)$. The estimator $\widehat{a}$ reduces to that used for two-sample case for $g = 2$. We state the following theorem on the limiting distribution of T_g, as an extension of Theorem 2.7, without proof.

Theorem 2.7. For T_g in Eqn. (8), $(\text{T}_g - \text{E}(\text{T}_g))/\widehat{\sigma}_{\text{T}_g} \xrightarrow{\mathcal{D}} N(0,1)$ under Assumptions 2.2−2.5, as $n_i, p \to \infty$ with $\widehat{\sigma}^2_{\text{T}_g}$ a consistent estimator of $\sigma^2_{\text{T}_g} = \text{Var}(\text{T}_g)$ given above.

For power of T_g, let z_α denote 100α%th quantile of $N(0,1)$, $\beta(\boldsymbol{\theta})$ be the power function of T_g with $\boldsymbol{\theta} = \{\mathbf{\Sigma}_i,\ 1 \le i \le g\}$ and $\mathbf{\Theta}_0$, $\mathbf{\Theta}_1$ be the parameter spaces under H_{0g} and H_{1g}, where $\mathbf{\Theta}_0 = \{\mathbf{\Sigma}\}$. By Theorem 2.7, $\beta(\boldsymbol{\theta}) = \text{P}((\text{T}_g - \tau_g)/\sigma_{\text{T}_{g0}} \ge z_\alpha)$ with $\beta(\boldsymbol{\theta}|\mathbf{\Theta}_0) = \alpha$, $\beta(\boldsymbol{\theta}|\mathbf{\Theta}_1) = 1 - \beta$, where $\tau_g = \sum_{i<j}^{g} \|\mathbf{\Sigma}_i - \mathbf{\Sigma}_j\|^2$ so that $\tau_g = 0$ under H_{0g}, and $\sigma^2_{\text{T}_{g0}}$ denotes $\text{Var}(\text{T}_g)$ under H_{0g}. Denote $\gamma = \sigma_{\text{T}_{g0}}/\sigma_{\text{T}_g}$ and $\delta = \tau_g/\sigma_{\text{T}_g}$. The power of T_g follows as $\beta(\boldsymbol{\theta}|\mathbf{\Theta}_1) = P(\text{T}_g/\sigma_{\text{T}_g} \ge \gamma z_\alpha + \delta) = 1 - \beta$. From Theorem 2.1 and under the assumptions, $\gamma = O(1)$ and $\delta = O(p)$ after some simplification, so that the power increases with increasing p. See also Sec. 3 for a proof of this through simulations.

3. SIMULATIONS

To evaluate T_g for its accuracy in size and power, we consider $g = 3$ and generate data from Exp(1) and U[0, 1] distributions. We use $n_1 \in \{20, 30, 50\}$, $n_2 = n_1 + 10$, $n_3 = n_1 + 20$ for size and $n_1 \in \{10, 20, 30\}$, $n_2 = n_1 + 5$, $n_3 = n_1 + 10$ for power, where each sample size triplet is combined with dimension $p = \{50, 100, 300, 500, 1000\}$. For $\mathbf{\Sigma}_i$, we use three structures, Compound Symmetry (CS), Autoregressive of order 1, AR(1), and unstructured (UN). The CS and AR(1) are defined as $\kappa\mathbf{I} + \rho\mathbf{J}$ and $\text{Cov}(X_k, X_l) = \kappa\rho^{|k-l|}$, $\forall\ k, l$, respectively, where UN is defined as $\mathbf{\Sigma} = (\sigma_{ij})_{i,j=1}^{d}$ with $\sigma_{ij} = 1(1)d\ (i = j)$, $\rho_{ij} = (i-1)/d\ (i > j)$, with $\mathbf{I}$ as identity matrix, $\mathbf{J}$ as matrix of 1s. We take $\rho = 0.5$, $\kappa = 1$. Under H_{0g}, each covariance structure is used as common $\mathbf{\Sigma}$. Under H_{1g}, two cases, once with one population having a different structure than the other two and once with all three populations having different structures, are considered.

For size, we use $\alpha \in \{0.01, 0.05, 0.10\}$ and estimate it by averaging $P(z_g \ge z_c|H_{0g})$ over $m = 1000$ simulation runs, where $z_g = (\text{T}_g - \text{E}(\text{T}_g)(/\sigma_{\text{T}_g}$ and z_c is the critical value. For power, $1 - \beta$, we use $\alpha = 0.05$ and estimate it by averaging $P(z_g \ge z_c|H_{1g})$ over m runs. Tables 1−2 report estimated sizes for two distributions and Table 3 reports estimated power for both distributions.

Generally, the test performs accurately, both for size and power, for all parameter settings. The accuracy increases with sample size and is not disturbed with increasing dimension. Whereas the Exponential distribution requires slightly larger sample size for better accuracy, the results improve with increasing n_i. In particular, we observe a high accuracy of the test for the case when all three populations have different covariance structures. This is the case of complete heteroscedasticity and the test seems to perform accurately even for small sample sizes.

Note that, compared to AR and UN, CS belongs to the class of spiked covariance matrices, where a few, say k, eigenvalues dominate the rest $p - k$; for CS, $k = 1$. The test seems to perform accurately even for this extreme case which enhances its applicability to a wide variety of situations.

n_1, n_2, n_3	p	CS			AR			UN		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
20,30, 40	50	0.010	0.053	0.133	0.018	0.064	0.127	0.014	0.069	0.122
	100	0.007	0.046	0.125	0.012	0.061	0.112	0.015	0.064	0.121
	300	0.008	0.052	0.121	0.016	0.057	0.109	0.012	0.059	0.112
	500	0.006	0.050	0.118	0.018	0.055	0.113	0.013	0.057	0.114
	1000	0.009	0.049	0.110	0.013	0.053	0.112	0.010	0.054	0.114
30,40,50	50	0.004	0.039	0.124	0.014	0.062	0.120	0.016	0.058	0.119
	100	0.009	0.043	0.115	0.024	0.075	0.121	0.019	0.060	0.111
	300	0.004	0.045	0.127	0.008	0.055	0.114	0.010	0.045	0.096
	500	0.006	0.043	0.125	0.014	0.061	0.112	0.022	0.056	0.118
	1000	0.011	0.049	0.114	0.008	0.051	0.102	0.016	0.055	0.112
50,60,80	50	0.012	0.041	0.112	0.012	0.054	0.107	0.012	0.045	0.093
	100	0.010	0.049	0.100	0.013	0.050	0.111	0.017	0.046	0.112
	300	0.010	0.054	0.115	0.008	0.050	0.116	0.015	0.049	0.103
	500	0.008	0.056	0.109	0.012	0.052	0.093	0.017	0.055	0.108
	1000	0.007	0.051	0.104	0.011	0.055	0.105	0.014	0.048	0.110

Tab. 1. Estimated test size of T_3: Uniform distribution.

n_1, n_2	p	CS			AR			UN		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
20, 30, 40	50	0.012	0.040	0.123	0.016	0.069	0.118	0.013	0.057	0.107
	100	0.012	0.044	0.134	0.017	0.066	0.123	0.014	0.066	0.120
	300	0.016	0.054	0.133	0.014	0.062	0.122	0.012	0.061	0.114
	500	0.009	0.057	0.132	0.013	0.056	0.113	0.015	0.059	0.116
	1000	0.007	0.055	0.127	0.012	0.047	0.114	0.014	0.057	0.115
30, 40, 50	50	0.004	0.038	0.118	0.016	0.056	0.111	0.012	0.060	0.100
	100	0.005	0.039	0.116	0.015	0.059	0.108	0.020	0.063	0.117
	300	0.004	0.037	0.113	0.016	0.058	0.108	0.008	0.052	0.098
	500	0.007	0.047	0.113	0.012	0.059	0.108	0.019	0.063	0.119
	1000	0.013	0.052	0.115	0.014	0.055	0.109	0.012	0.053	0.114
50, 60, 80	50	0.004	0.034	0.107	0.015	0.072	0.129	0.008	0.055	0.098
	100	0.006	0.035	0.108	0.013	0.057	0.102	0.008	0.057	0.114
	300	0.004	0.039	0.110	0.018	0.061	0.111	0.012	0.056	0.124
	500	0.008	0.033	0.114	0.015	0.052	0.122	0.023	0.061	0.114
	1000	0.007	0.045	0.110	0.011	0.056	0.119	0.009	0.052	0.111

Tab. 2. Estimated test size of T_3: Exponential distribution.

4. DISCUSSION

A test statistic for homogeneity of several large-dimensional covariance matrices is proposed which can be used for any distribution having first four finite moments across the p dimensions of the independent vectors. The applicability of the test thus holds for a wide variety of models, including multivariate normal. The test is composed of location-invariant, consistent, computationally efficient, estimators, which can be equiv-

n_1, n_2, n_3	p	U[0, 1]		Exp(1)	
		CS-CS-AR	CS-AR-UN	CS-CS-AR	CS-AR-UN
10, 15, 20	50	0.582	0.443	0.498	0.401
	100	0.698	0.592	0.556	0.473
	300	0.772	0.686	0.615	0.544
	500	0.825	0.751	0.709	0.628
	1000	0.901	0.859	0.795	0.702
20, 25, 30	50	0.771	0.643	0.725	0.705
	100	0.825	0.766	0.812	0.759
	300	0.908	0.885	0.898	0.815
	500	0.993	0.925	0.966	0.905
	1000	1.000	0.985	1.000	0.963
30, 35, 40	50	0.899	0.875	0.854	0.832
	100	0.965	0.928	0.916	0.903
	300	1.000	0.997	0.983	0.971
	500	1.000	1.000	1.000	1.0000
	1000	1.000	1.000	1.000	1.000

Tab. 3. Estimated power of T_3: Both distributions.

alently defined as U-statistics of fourth order symmetric kernels. Theoretical properties of the test are studied under a few mild conditions, and simulations are used to show its accuracy across different parameter settings.

A. SOME BASIC RESULTS

A.1. Moments of quadratic & bilinear forms

Given $\mathbf{a}_{ik}$ in Eqn. (1), let $A_{ik} = \mathbf{X}'_{ik}\mathbf{\Sigma}_i\mathbf{X}_{ik} = \mathbf{Y}'_{ik}\mathbf{A}^2\mathbf{Y}_{ik}$, $A_{ijkl} = \mathbf{X}'_{ik}\mathbf{X}_{jl} = \mathbf{Y}'_{ik}\mathbf{\Lambda}_i\mathbf{\Lambda}_j\mathbf{Y}_{jl}$, $k \neq l$, be quadratic and bilinear forms, with $A_{ik} = \mathbf{Y}_{ik}\mathbf{A}_i\mathbf{Y}_{ik} = Q_{ik}$ for $\mathbf{\Sigma}_i = \mathbf{I}$. Theorem A.1 gives basic moments of A_{ik}, A_{ijkl} which are used to derive extended moments in Lemma A.2. Proofs of these results are tedious but not directly related here, hence skipped; see e. g. [1, 3].

Theorem A.1. For A_{ik} and A_{ijkl}, as defined above, we have

$$\mathrm{E}\left(Q_{ik}^2\right)^2 = 2\,\mathrm{tr}(\mathbf{\Sigma}_i^2) + [\mathrm{tr}(\mathbf{\Sigma}_i)]^2 + M_1 \tag{13}$$

$$\mathrm{E}\left(A_{ik}^2\right)^2 = 2\,\mathrm{tr}(\mathbf{\Sigma}_i^4) + [\mathrm{tr}(\mathbf{\Sigma}_i^2)]^2 + M_2 \tag{14}$$

$$\mathrm{E}\left(A_{ik}A_{jk}\right) = 2\,\mathrm{tr}(\mathbf{\Sigma}_i^3\mathbf{\Sigma}_j) + \mathrm{tr}(\mathbf{\Sigma}_i^2)\,\mathrm{tr}(\mathbf{\Sigma}_i\mathbf{\Sigma}_j) + M_2 \tag{15}$$

$$\mathrm{E}\left(A_{ijkl}^4\right) = 6\,\mathrm{tr}(\mathbf{\Sigma}_i\mathbf{\Sigma}_j)^2 + 3\,[\mathrm{tr}(\mathbf{\Sigma}_i\mathbf{\Sigma}_j)]^2 + M_3 \tag{16}$$

$$\mathrm{E}\left(Q_{ik}Q_{jk}A_{ijkl}^2\right) = 4\,\mathrm{tr}(\mathbf{\Sigma}_i\mathbf{\Sigma}_j)^2 + 4\,\mathrm{tr}(\mathbf{\Sigma}_i^3)\,\mathrm{tr}(\mathbf{\Sigma}_j) + [\mathrm{tr}(\mathbf{\Sigma}_i)]^2\,\mathrm{tr}(\mathbf{\Sigma}_j^2) + M_4 \tag{17}$$

where $\mathbf{D} = diag(\mathbf{A})$, $M_1 = \gamma\,\mathrm{tr}(\mathbf{A} \odot \mathbf{A})$, $M_2 = \gamma\,\mathrm{tr}(\mathbf{A}^2 \odot \mathbf{A}^2)$, $M_3 = 6\gamma\,\mathrm{tr}(\mathbf{A}^2 \odot \mathbf{A}^2) + \gamma^2 \sum_{s=1}^p \sum_{t=1}^p A_{st}^4$ and $M_4 = 2\gamma\,\mathrm{tr}(\mathbf{\Sigma}_i)\,\mathrm{tr}(\mathbf{A}^2 \odot \mathbf{A}) + 4\gamma\,\mathrm{tr}(\mathbf{A}^3 \odot \mathbf{A}) + \gamma\,\mathrm{tr}(\mathbf{A} \odot \mathbf{ADA})$.

Lemma A.2. For $\mathbf{a}_{ik}$ in Eqn. (1), we have the following.

$$\begin{aligned}
\mathrm{E}[\mathbf{a}'_{it}\mathbf{a}_{iu}\mathbf{a}'_{it}\mathbf{a}_{iv}\mathbf{a}'_{iu}\boldsymbol{\Sigma}_i\mathbf{a}_{iv}] &= \mathrm{tr}(\boldsymbol{\Sigma}_i^4) && (18)\\
\mathrm{E}[\mathbf{a}'_{it}\mathbf{a}_{iu}\mathbf{a}'_{iw}\mathbf{a}_{iu}\mathbf{a}'_{it}\mathbf{a}_{iv}\mathbf{a}'_{iw}\mathbf{a}_{iv}] &= \mathrm{tr}(\boldsymbol{\Sigma}_i^4) && (19)\\
\mathrm{E}(\mathbf{a}'_{jt}\mathbf{a}_{iu}\mathbf{a}'_{jt}\mathbf{a}_{iv}\mathbf{a}'_{iu}\boldsymbol{\Sigma}_j\mathbf{a}_{iv}) &= \mathrm{tr}\{(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)^2\} && (20)\\
\mathrm{Cov}(\mathbf{a}'_{it}\boldsymbol{\Sigma}_i\mathbf{a}_{iu}, \mathbf{a}'_{it}\boldsymbol{\Sigma}_j\mathbf{a}_{iu}) &= \mathrm{tr}\{(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)^2\} && (21)\\
\mathrm{E}[(\mathbf{a}'_{iu}\mathbf{a}_{iv})^2\mathbf{a}'_{it}\boldsymbol{\Sigma}_j\mathbf{a}_{it}] &= \mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)\,\mathrm{tr}(\boldsymbol{\Sigma}_i)^2 && (22)\\
\mathrm{E}[(\mathbf{a}'_{it}\mathbf{a}_{iu})^2\mathbf{a}'_{it}\boldsymbol{\Sigma}_i\mathbf{a}_{it}] &= 2\,\mathrm{tr}(\boldsymbol{\Sigma}_i^4) + [\mathrm{tr}(\boldsymbol{\Sigma}_i^2)]^2 + M_2 && (23)\\
\mathrm{Var}(\mathbf{a}'_{it}\mathbf{a}_{iu}\mathbf{a}'_{iv}\mathbf{a}_{iu}) &= 2\,\mathrm{tr}(\boldsymbol{\Sigma}_i^4) + [\mathrm{tr}(\boldsymbol{\Sigma}_i^2)]^2 + M_2 && (24)\\
\mathrm{Cov}[(\mathbf{a}'_{it}\mathbf{a}_{iu})^2, (\mathbf{a}'_{it}\mathbf{a}_{iv})^2] &= 2\,\mathrm{tr}(\boldsymbol{\Sigma}_i^4) + M_2 && (25)\\
\mathrm{E}[(\mathbf{a}'_{it}\mathbf{a}_{ju})^2\mathbf{a}'_{it}\boldsymbol{\Sigma}_j\mathbf{a}_{it}] &= 2\,\mathrm{tr}\{(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)^2\} + \left[\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)\right]^2 + M_2 && (26)\\
\mathrm{Var}(\mathbf{a}'_{it}\mathbf{a}_{ju}\mathbf{a}'_{iv}\mathbf{a}_{ju}) &= 2\,\mathrm{tr}\{(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)^2\} + \left[\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)\right]^2 + M_2 && (27)\\
\mathrm{Cov}[(\mathbf{a}'_{jt}\mathbf{a}_{iu})^2, (\mathbf{a}'_{jt}\mathbf{a}_{iv})^2] &= 2\,\mathrm{tr}\{(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)^2\} + M_2 && (28)\\
\mathrm{Cov}[(\mathbf{a}'_{it}\mathbf{a}_{iu})^2, \mathbf{a}'_{it}\boldsymbol{\Sigma}_j\mathbf{a}_{it}] &= 2\,\mathrm{tr}(\boldsymbol{\Sigma}_i^3\boldsymbol{\Sigma}_j) + M_2 && (29)
\end{aligned}$$

where terms like $\mathrm{E}[(\mathbf{a}'_{it}\mathbf{a}_{iu})^2\mathbf{a}'_{it}\mathbf{a}_{iu}\mathbf{a}'_{it}\mathbf{a}_{iv}]$, $\mathrm{E}[\mathbf{a}'_{it}\mathbf{a}_{iu}\mathbf{a}'_{it}\mathbf{a}_{iv}\mathbf{a}'_{it}\boldsymbol{\Sigma}_i\mathbf{a}_{it}]$, $\mathrm{E}[\mathbf{a}'_{it}\mathbf{a}_{iu}\mathbf{a}'_{it}\mathbf{a}_{iv}\mathbf{a}'_{it}\boldsymbol{\Sigma}_i\mathbf{a}_{iu}]$, $\mathrm{E}[(\mathbf{a}'_{it}\mathbf{a}_{iu})^2\mathbf{a}'_{it}\boldsymbol{\Sigma}_i\mathbf{a}_{iu}]$, $\mathrm{E}[(\mathbf{a}'_{it}\mathbf{a}_{iu})^2\mathbf{a}'_{it}\boldsymbol{\Sigma}_j\mathbf{a}_{iu}]$, $\mathrm{E}[(\mathbf{a}'_{it}\mathbf{a}_{iu})^2\mathbf{a}'_{it}\mathbf{a}_{iv}\mathbf{a}'_{iu}\mathbf{a}_{iv}]$ all vanish.

The results in Lemma A.2 without M_2, M_3 hold even if $\mathbf{a}_{ik}$ are not normally distributed. If they do, then $M_2 = 0 = M_3$ and all results coincide with those under normality.

B. MAIN PROOFS

B.1. Some basic results

First, we need to set some notations. For details, see e. g. [10, 12, 23, 25]. For iid X_i, let $h(X_1, \ldots, X_m) : \mathbb{R}^m \to \mathbb{R}$ denote the kernel of an mth order U-statistic, U_n, with $\mathrm{E}(U_n) = \theta = \mathrm{E}[h(\cdot)]$ with its projection $h_c(x_1, \ldots, x_c) = \mathrm{E}[h(\cdot)|x_1, \ldots, x_c]$, $h_m(\cdot) = h(\cdot)$ and $\xi_c = \mathrm{Var}[h_c(\cdot)$, $c = 1, \ldots, m$, so that $\mathrm{Var}(U_n) = \sum_{c=1}^m \binom{m}{c}\binom{n-m}{m-c}\xi_c/\binom{n}{m}$. If $0 < \xi_c < \infty\ \forall\ c$, then $(U_n - \mathrm{E}(U_n))/\sqrt{\mathrm{Var}(U_n)} \xrightarrow{\mathcal{D}} N(0,1)$. For two U-statistics, U_{ni}, of order m_i, kernels $h_i(\cdot)$, projections $h_{ic}(\cdot)$, $i = 1, 2$, let $\xi_{cc} = \mathrm{Cov}[h_{1c}(\cdot), h_{2c}(\cdot)]$, $c = 1, \ldots, m_1 \le m_2$. Then $\mathrm{Cov}(U_{n1}, U_{n2}) = \sum_{c=1}^{m_1}\binom{m_2}{c}\binom{n-m_2}{m_1-c}\xi_{cc}/\binom{n}{m_1}$. Let $U_{n_1n_2}$ be a U-statistic of two independent samples, with kernel $h(X_{11}, \ldots, X_{1m_1}, X_{21}, \ldots, X_{2m_2})$, symmetric in each sample, projection $h_{c_1c_2} = \mathrm{E}[h(\cdot)|X_{11}, \ldots, X_{1c_1}; X_{21}, \ldots, X_{2c_2}]$, $\xi_{c_1c_2} = \mathrm{Cov}[h(\cdot), h_{c_1c_2}(\cdot)]$, $\xi_{00} = 0$, $c_i = 0, 1, \ldots, m_i$. If $0 \le n_i/n \le 1$, $n = n_1 + n_2$, $0 < \xi_{c_1c_2} < \infty\ \forall\ c_i$, then $(U_{n_1n_2} - \mathrm{E}(U_{n_1n_2}))/\sqrt{\mathrm{Var}(U_{n_1n_2})} \xrightarrow{\mathcal{D}} N(0,1)$ where $\mathrm{Var}(U_{n_1n_2}) = \sum_{c_1=0}^{m_1}\sum_{c_2=0}^{m_2}\binom{m_1}{c_1}\binom{n_1-m_1}{m_1-c_1}\binom{m_2}{c_2}\binom{n_2-m_2}{m_2-c_2}\xi_{c_1c_2}/\binom{n_1}{m_1}\binom{n_2}{m_2}$.

B.2. Proof of Theorem 2.1

Consider $\widehat{\delta}_1$, a one-sample U-statistic. Using notations in Sec. B.1 and moments in Theorem A.1 and Lemma A.2, the projections $h_c(\cdot)$, $c = 1, \ldots, 4$, are computed as

$$\begin{aligned}
h_1 &= 6\mathbf{a}'_{1k}\boldsymbol{\Sigma}_1\mathbf{a}_{1k} + 6\operatorname{tr}(\boldsymbol{\Sigma}_1^2)\\
h_2 &= 4\mathbf{a}'_{1k}\boldsymbol{\Sigma}_1\mathbf{a}_{1k} + 4\mathbf{a}'_{1r}\boldsymbol{\Sigma}_1\mathbf{a}_{1r} + 2(\mathbf{a}'_{1k}\mathbf{a}_{1r})^2 - 4\mathbf{a}'_{1k}\boldsymbol{\Sigma}_1\mathbf{a}_{1r} + 2\operatorname{tr}(\boldsymbol{\Sigma}_1^2)\\
h_3 &= 2[(\mathbf{a}'_{1k}\mathbf{a}_{1k'})^2 + \mathbf{a}'_{1k}\boldsymbol{\Sigma}_1\mathbf{a}_{1k} + (\mathbf{a}'_{1r}\mathbf{a}_{1k'})^2 + \mathbf{a}'_{1r}\boldsymbol{\Sigma}_1\mathbf{a}_{1r} - \mathbf{a}'_{1k}\mathbf{a}_{1k'}\mathbf{a}'_{1r}\mathbf{a}_{1k'} - \mathbf{a}'_{1k}\mathbf{a}_{1r}\mathbf{a}'_{1k'}\mathbf{a}_{1r},
\end{aligned}$$

where $h_4(\cdot) = h(\cdot)$. The variances of these projections follow as

$$\begin{aligned}
\xi_1 &= 72\operatorname{tr}(\boldsymbol{\Sigma}_1^4) + M_2O(1)\\
\xi_2 &= 168\operatorname{tr}(\boldsymbol{\Sigma}_1^4) + 8[\operatorname{tr}(\boldsymbol{\Sigma}_1^2)]^2 + (M_2 + M_3)O(1)\\
\xi_3 &= 300\operatorname{tr}(\boldsymbol{\Sigma}_1^4) + 36[\operatorname{tr}(\boldsymbol{\Sigma}_1^2)]^2 + (M_2 + M_3)O(1)\\
\xi_4 &= 96\{9\operatorname{tr}(\boldsymbol{\Sigma}_1^4) + 3[\operatorname{tr}(\boldsymbol{\Sigma}_1^2)]^2\} + M_2O(1),
\end{aligned}$$

where the terms involving M_2, M_3 are merged into $O(1)$ for simplicity since all such terms eventually vanish under Assumption 2.4. Substituting ξ_c in $\operatorname{Var}(U_n)$ in Sec. B.1 gives the required variance. Now, consider $\widehat{\delta}_{12}$ which is a two-sample U-statistic. From Sec. B.1 again, we get

$$\begin{aligned}
h_{10} &= 2\mathbf{a}'_{1k}\boldsymbol{\Sigma}_2\mathbf{a}_{1k} + 2\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)\\
h_{20} &= 2(\mathbf{a}_{1k} - \mathbf{a}_{1r})'\boldsymbol{\Sigma}_2(\mathbf{a}'_{1k} - \mathbf{a}_{1r})\\
h_{11} &= (\mathbf{a}'_{1k}\mathbf{a}_{2l})^2 + \mathbf{a}'_{1k}\boldsymbol{\Sigma}_2\mathbf{a}_{1k} + \mathbf{a}'_{2l}\boldsymbol{\Sigma}_1\mathbf{a}_{2l} + \operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)\\
h_{21} &= (\mathbf{a}_{1k} - \mathbf{a}_{1r})'\boldsymbol{\Sigma}_2(\mathbf{a}_{1k} - \mathbf{a}_{1r}) + [(\mathbf{a}_{1k} - \mathbf{a}_{1r})'\mathbf{a}_{2l}]^2,
\end{aligned}$$

where $h_{22}(\cdot) = h(\cdot)$, with their corresponding variances

$$\begin{aligned}
\xi_{10} &= 8\operatorname{tr}\{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2\} + M_2O(1)\\
\xi_{20} &= 32\{\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2\} + M_2O(1)\\
\xi_{11} &= 18\operatorname{tr}\{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2\} + 2[\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)]^2 + (M_2 + M_3)O(1)\\
\xi_{21} &= 48\operatorname{tr}\{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2\} + 8[\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)]^2 + (M_2 + M_3)O(1)\\
\xi_{22} &= 96\operatorname{tr}\{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2\} + 32[\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)]^2 + M_3O(1).
\end{aligned}$$

Substituting in $\operatorname{Var}(U_{n_1n_2})$ gives $\operatorname{Var}(\widehat{\delta}_{12})$. Similarly, $\operatorname{Cov}(\widehat{\delta}_1, \widehat{\delta}_{12})$ follows from $\operatorname{Cov}(U_{n_1}, U_{n_2})$ in Sec. B.1 by noting that

$$\begin{aligned}
h_{11} &= \mathbf{a}'_{1k}\boldsymbol{\Sigma}_2\mathbf{a}_{1k} + \operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2),\ h_{12} = \mathbf{D}'_{1kr}\boldsymbol{\Sigma}_2\mathbf{D}_{1kr}, h_{21} = 6\mathbf{a}'_{1k}\boldsymbol{\Sigma}_1\mathbf{a}_{1k} + 6\operatorname{tr}(\boldsymbol{\Sigma}_1^2)\\
h_{22} &= 4\mathbf{a}'_{1k}\boldsymbol{\Sigma}_1\mathbf{a}_{1k} + 4\mathbf{a}'_{1r}\boldsymbol{\Sigma}_1\mathbf{a}_{1r} + 2(\mathbf{a}'_{1k}\mathbf{a}_{1r})^2 - 4\mathbf{a}'_{14}\boldsymbol{\Sigma}_1\mathbf{a}_{1r} + 2\operatorname{tr}(\boldsymbol{\Sigma}_1^2)
\end{aligned}$$

with $\xi_{11} = 12\operatorname{tr}(\boldsymbol{\Sigma}_1^3\boldsymbol{\Sigma}_2) + M_2O(1)$, $\xi_{22} = \operatorname{Cov}(h_{12}, h_{22}) = 32\operatorname{tr}(\boldsymbol{\Sigma}_1^3\boldsymbol{\Sigma}_2) + M_2O(1)$. The bounds follow by a simple application of the Cauchy–Schwarz inequality. □

B.3. Proof of Theorem 2.7

The proof essentially follows by the argument after Eqn. (7), given that distributions of individual components and their covariance terms are taken care of. Note that, the kernels of these component U-statistics vary with n_i (and p through n_i). Whereas the theory of U-statistics with kernel varying with n has been extensively explored [9, 10, 16], it has also recently been applied to high-dimensional inference. Of particular mention are [18], who use it for U-statistics similar to those in the present case, and [26] who also provide a general discussion on determining the limit of U-statistics for high-dimensional data.

It follows from [26] that the limit of such a U-statistic rests on the behavior of projection variances ξ_c (see Sec. B.1). In the present case, it follows from the proof in Sec. B.2 that the projection variances for both $\widehat{\delta}_i$ and $\widehat{\delta}_{ij}$ are uniformly bounded under the assumptions, as shown below. This, by Eqn. (7) or, generally Eqn. (10), implies that both components may have a non-degenerate limit. We proceed as following.

Consider $\widetilde{\delta}_i = \widehat{\delta}_i/\operatorname{tr}(\boldsymbol{\Sigma}_i^2) - 1 = \widetilde{U}_{n_i}$ with $\widetilde{h} = h(\cdot)/\operatorname{tr}(\boldsymbol{\Sigma}_i^2) - 1$ (see Sec. B.1) and its first-order (Hájek) projection $\widetilde{U}_n$ [23, 25] $\widehat{U}_{n_i} = m\sum_{k=1}^{n_i}\widetilde{h}_{i1}(\mathbf{x}_{1k})/n_i$, where

$$\widetilde{h}_{i1}(\mathbf{x}_{1k}) = \left[\mathbf{a}_{ik}'\boldsymbol{\Sigma}_i\mathbf{a}_{ik} + \operatorname{tr}(\boldsymbol{\Sigma}_i^2)\right]/2\operatorname{tr}(\boldsymbol{\Sigma}_i^2) - 1 \Rightarrow \operatorname{Var}[\widetilde{h}_{i1}(\mathbf{x}_{1k})] = [2\operatorname{tr}(\boldsymbol{\Sigma}_i^4) + M_2]/4[\operatorname{tr}(\boldsymbol{\Sigma}_i^2)]^2.$$

The term with M_2 vanishes by Assumption 2.4, and $\operatorname{E}(\widehat{U}_{n_i}) = 0$, $\operatorname{Var}(\widehat{U}_{n_i}) = 4[2\operatorname{tr}(\boldsymbol{\Sigma}_i^4) + M_2]/n_i[\operatorname{tr}(\boldsymbol{\Sigma}_i^2)]^2 \le 8/n_i = O(1/n_i)$ which is independent of p. From Theorem A.1, $\operatorname{Cov}(\widetilde{U}_{n_i}, \widehat{U}_{n_i}) = \operatorname{Var}(\widehat{U}_{n_i})$ as required, so that $\widetilde{U}_{n_i} = \widehat{U}_{n_i} + o_P(1)$ with $\operatorname{Var}(\widetilde{U}_{n_i})/\operatorname{Var}(\widehat{U}_{n_i}) \to 1$. By Slutsky's lemma, $\widehat{U}_{n_i}$ can replace $\widetilde{U}_{n_i}$. The second-order projection is

$$\widetilde{h}_{i2}(\mathbf{x}_{ik}, \mathbf{x}_{ir}) = [-2\mathbf{a}_{1k}'\boldsymbol{\Sigma}_i\mathbf{a}_{1k} - 2\mathbf{a}_{1r}'\boldsymbol{\Sigma}_i\mathbf{a}_{1r} + 2(\mathbf{a}_{1k}'\mathbf{a}_{1r})^2 - 4\mathbf{a}_{1k}'\boldsymbol{\Sigma}_i\mathbf{a}_{1r} + 2\operatorname{tr}(\boldsymbol{\Sigma}_i^2)]/12[\operatorname{tr}(\boldsymbol{\Sigma}_i^2)]^2$$

with $\operatorname{E}[\widetilde{h}_{i2}(\cdot)] = 0$ and

$$\operatorname{Var}[h_{i2}(\cdot)] = \left[3\operatorname{tr}(\boldsymbol{\Sigma}_i^4) + [\operatorname{tr}(\boldsymbol{\Sigma}_i^2)]^2 + M_3\right]/18[\operatorname{tr}(\boldsymbol{\Sigma}_i^2)]^2$$

so that $\widehat{U}_{n_i} = 12\sum_{k\ne r}^{n_i}\widetilde{h}_{i2}(\cdot)/Q(n_i)$ with $\operatorname{E}(\widehat{U}_{n_i}) = 0$ and

$$\operatorname{Var}(\widehat{U}_{n_i}) = 8[3\operatorname{tr}(\boldsymbol{\Sigma}_i^4) + [\operatorname{tr}(\boldsymbol{\Sigma}_i^2)]^2 + 20M_3]/Q(n_i)[\operatorname{tr}(\boldsymbol{\Sigma}_i^2)]^2 = O(1/n_i^2),$$

independent of p. Further, $\operatorname{Cov}(\widetilde{U}_{n_i}, \widehat{U}_{n_i}) = \operatorname{Var}(\widehat{U}_{n_i})$, as required. This gives the normal limit of $\widehat{\delta}_i$. Now, for $\widetilde{\delta}_{12} = \widehat{\delta}_{12}/\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2) - 1$, let $n_0 = n/n_1n_2$ and write

$$\widehat{U}_{n_1n_2} = [\sum_{k=1}^{n_1}\mathbf{a}_{ik}'\boldsymbol{\Sigma}_2\mathbf{a}_{jk}/n_1 + \sum_{k=1}^{n_2}\mathbf{a}_{jk}'\boldsymbol{\Sigma}_1\mathbf{a}_{jk}/n_2]/\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2) - n_0$$

with $\operatorname{E}(\widehat{U}_{n_1n_2}) = 0$ and

$$\operatorname{Var}(\widehat{U}_{n_1n_2}) = n_0[2\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2 + M_2]/[\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)]^2.$$

The normal limit follows, under assumptions, as for one-sample case. As $\operatorname{Cov}(\widetilde{\delta}_1, \widetilde{\delta}_{12})$ converges to a fixed limit, 0, and samples are independent, it gives the joint limit in (11) and also that of T_2. Writing $(\mathrm{T}_2 - \operatorname{E}(\mathrm{T}_2))/\sigma_{\mathrm{T}_2} = \{(\mathrm{T}_2 - \operatorname{E}(\mathrm{T}_2))/\widehat{\sigma}_{\mathrm{T}_2}\}\{\widehat{\sigma}_{\mathrm{T}_2}/\sigma_{\mathrm{T}_2}\}$, and using the consistency of $\widehat{\sigma}_{\mathrm{T}_2}$, the limit remains valid by replacing σ_{T_2} with $\widehat{\sigma}_{\mathrm{T}_2}$. This completes the proof. □

ACKNOWLEDGEMENT

The author is thankful to the editor, the associate editor and an anonymous referee for their constructive comments which helped improve the original version of the manuscript.

(Received November 13, 2017)

REFERENCES

[1] M. R. Ahmad: Location-invariant multi-sample U-tests for covariance matrices with large dimension. Scand. J. Stat. *44* (2017b), 500–523. DOI:10.1111/sjos.12262

[2] M. R. Ahmad: Testing homogeneity of several covariance matrices and multi-sample sphericity for high-dimensional data under non-normality. Comm. Stat. Theory Methods *46* (2017a), 3738–3753. DOI:10.1080/03610926.2015.1073310

[3] M. R. Ahmad: On testing sphericity and identity of a covariance matrix with large dimensions. Math. Meth. Stat. *25* (2016), 121–132. DOI:10.3103/s1066530716020034

[4] T.W. Anderson: An Introduction to Multivariate Statistical Analysis. Third edition. Wiley, NY 2003.

[5] M. Aoshima and K. Yata: Two-stage procedures for high-dimensional data. Seq. An. *30* (2011), 356–399. DOI:10.1080/07474946.2011.619088

[6] T. Cai, W. Liu, and Y. Xia: Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. J. Amer. Statist. Assoc. *108* (2013), 265–277. DOI:10.1080/01621459.2012.758041

[7] Y. Fujikoshi, V. V. Ulyanov, and R. Shimizu: Multivariate statistics: High-dimensional and large-sample approximations. Wiley, NY 2010.

[8] J. Hájek, Z. Šidák, and P. K. Sen: Theory of Rank Tests. Academic Press, SD 1999.

[9] T. Y. Kim, Z-M. Luo, and C. Kim: The central limit theorem for degenerate variable U-statistics under dependence. J. Nonparam. Stat. *23* (2011), 683–699. DOI:10.1080/10485252.2011.556193

[10] V. S. Koroljuk and Y. V. Borovskich: Theory of U-statistics. Kluwer Academic Press, Dordrecht 1994. DOI:10.1007/978-94-017-3515-5

[11] A. J. Lee: U-statistics: Theory and Practice. CRC Press, Boca Raton 1990.

[12] E. L. Lehmann: Elements of Large-sample Theory. Springer, NY 1999. DOI:10.1007/b98855

[13] J. Li and S. X. Chen: Two sample tests for high-dimensional covariance matrices. Ann. Stat. *40* (2012), 908–940. DOI:10.1214/12-aos993

[14] B. Liu, L. Xu, S. Zheng, and G-L. Tian: A new test for the proportionality of two large-dimensional covariance matrices. J. Multiv. An. *131* (2014), 293–308. DOI:10.1016/j.jmva.2014.06.008

[15] T. Mikosch: Weak invariance principles for weighted U-statistics. J. Theoret. Prob. *7* (1991), 147–173. DOI:10.1007/bf02213365

[16] T. Mikosch: A weak invariance principle for weighted U-statistics with varying kernels. J. Multiv. An. *47* (1993), 82–102. DOI:10.1006/jmva.1993.1072

[17] R. J. Muirhead: Aspects of Multivariate Statistical Theory. Wiley, NY 2005 DOI:10.1002/9780470316559

[18] A. Pinheiro, K. Sen, and H. P. Pinheiro: Decomposibility of high-dimensional diversity measures: Quasi-U-statistics, martigales, and nonstandard asymptotics. J. Multiv. An. *100* (2009), 1645–1656. DOI:10.1016/j.jmva.2009.01.007

[19] Y. Qiu and S. X. Chen: Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. Ann. Stat. *40* (2012) 1285–1314. DOI:10.1214/12-aos1002

[20] J. R. Schott: A test for the equality of covariance matrices when the dimension is large relative to the sample size. Computat. Statist. Data Analysis *51* (2007), 6535–6542. DOI:10.1016/j.csda.2007.03.004

[21] G. A. F. Seber: Multivariate Observations. Wiley, NY 2004. DOI:10.1002/9780470316641

[22] P. K. Sen: Robust statistical inference for high-dimensional data models with applications in genomics. Aust. J. Stat. *35* (2006), 197–214.

[23] R. J. Serfling: Approximation Theorems of Mathematical Statistics. Wiley, Weinheim 1980. DOI:10.1002/9780470316481

[24] M. S. Srivastava and H. Yanagihara: Testing the equality of several covariance matrices with fewer observations than the dimension. J. Multiv. An. *101*, 1319–1329. DOI:10.1016/j.jmva.2009.12.010

[25] A. W. van der Vaart: Asymptotic Statistics. Cambridge University Press, 1998. DOI:10.1017/cbo9780511802256

[26] P-S. Zhong and S. X. Chen: Tests for high-dimensional regression coefficients with factorial designs. J. Amer. Statist. Assoc. *106* (2011), 260–274. DOI:10.1198/jasa.2011.tm10284

M. Rauf Ahmad, Department of Statistics, Uppsala University, Ekonomikum, Kyrkogårdsgatan 10, 75120, Uppsala. Sweden.
e-mail: rauf.ahmad@statistik.uu.se