HIROSHIMA MATH. J., **52** (2022), 217–233
doi:10.32917/h2021025

Test for equality of standardized generalized variance with different dimensions under high-dimensional settings

Hiroki WATANABE, Masashi HYODO, Takatoshi SUGIYAMA and Takashi SEO

(Received March 29, 2021)
(Revised December 23, 2021)

ABSTRACT. In this article, the testing equality of the standardized generalized variance (SGV) of k multivariate normal distributions with possibly unequal dimensions is studied. The conventional likelihood-ratio test statistic reveals a serious bias with an increase in dimensions. Therefore, we present a new test statistic that eliminates this bias, and propose an asymptotic approximation-based test. Our proposed test is valid not only in high-dimensional settings but also in large sample settings. Additionally, we obtain the asymptotic non-null distribution of the proposed test and the approximate confidence interval of the SGV under high-dimensional and large sample settings. Finally, we investigate the finite sample and dimension behavior of this test using Monte Carlo simulations.

1. Introduction

For a $p \times p$ covariance matrix $\boldsymbol{\Sigma}$, its determinant, $|\boldsymbol{\Sigma}|$, is known as a generalized variance and is a measure of variability for multivariate data (Wilks [12]). In a multivariate setting, a more suitable measure for comparing the variability in the distributions of different dimensions is one which takes into account the dimensions of the distributions (see, for example, SenGupta [8] and [9]). In this regard, assuming that $\mathbf{x}$ is a p-dimensional random vector with covariance matrix $\boldsymbol{\Sigma}$, SenGupta [8] introduced the concept of the standardized generalized variance (SGV) of $\mathbf{x}$, which is defined as $|\boldsymbol{\Sigma}|^{1/p}$, as a measure for comparing the variability across populations of different dimensions. This measure scales down the values across these populations of different dimensions so as to render them comparable with the univariate case (SenGupta [9]). Thus, it can be stated that while $|\boldsymbol{\Sigma}|$ is used as a measure of multidimensional variability, $|\boldsymbol{\Sigma}|^{1/p}$ can serve the same purpose, while it can also be used for comparing the variability in different dimensions. Pena and Rodriguez [6] revisited the concept of the SGV under the label "effective

---

2020 *Mathematics Subject Classification*. Primary 62H15; Secondary 62E20.
Key words and phrases. generalized variance, asymptotic approximation-based test, high-dimensional data.

variance". Meanwhile, using the SGV concept, Pena and Linde [5] suggested specific dimensionless descriptive measures of multivariate variability and dependence. Comparing the overall variability for populations of different dimensions could be useful for portfolio analysis with different numbers of entries, additional or missing information on components of the same item produced by different factories, the risk of various investment portfolios in financial analysis, the precision in statistical quality control, the homogeneity in cluster analysis, and comparing body size measurements among different species of living organisms (SenGupta [9]).

For $i \in \{1, 2, \dots, k\}$ $(k \geq 2)$, let $\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iN_i}$ be p_i-dimensional random vectors from the i-th multivariate normal population. We can denote the i-th population mean vector by $\boldsymbol{\mu}_i$, the i-th population covariance matrix by $\boldsymbol{\Sigma}_i$, and the i-th multivariate normal population by $\mathcal{N}_{p_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, respectively. Let $\mathbf{x}_{ij}$ for $j \in \{1, 2, \dots, N_i\}$, $i \in \{1, 2, \dots, k\}$ be distributed as $\mathcal{N}_{p_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, and the random vectors $\mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{kN_k}$ be mutually independent. Our primary interest is to test the following hypothesis:

$$\mathscr{H} : |\boldsymbol{\Sigma}_1|^{1/p_1} = |\boldsymbol{\Sigma}_2|^{1/p_2} = \cdots = |\boldsymbol{\Sigma}_k|^{1/p_k} \ \text{vs.}\ \mathscr{A} : \neg\mathscr{H}. \tag{1.1}$$

The most classical approach for the testing problem (1.1) is the log-likelihood ratio test (LRT). The LRT for testing $\mathscr{H}$ versus $\mathscr{A}$ was investigated by SenGupta [8]. Based on Bartlett's test for the homogeneity of variances, SenGupta [9] modified the LRT statistic for testing $\mathscr{H}$ against $\mathscr{A}$ and obtained the exact null distribution of this modified statistic for $p_i = 1, 2$ and mentioned that for $p_i > 2$, the exact null distribution of the this statistic is complicated and computationally expensive (see, SenGupta [9]). According to the general theory of the LRT, the asymptotic null distribution of the log-likelihood ratio statistic is a chi-square distribution with $k - 1$ degrees of freedom; $\min\{N_1, N_2, \dots, N_k\} \to \infty$. However, it is held that this limit distribution does not act as an approximate distribution in high-dimensional settings. Najarzadeh [3] proposed two modifications of the LRT (1.1) based on two well-known distribution approximation methods. While it was confirmed that these methods are good approximations if each dimension is different but almost equal in high-dimensional settings according to numerical simulations, these methods are impractical because they involve complicated moment calculations for moment matching, and there is no theoretical justification when the dimensions are different.

Recently, Sugiyama et al. [11] proposed an effective approximate test for (1.1) when $p_1 = p_2 = \cdots = p_k$ based on the high-dimensional asymptotic theory. In this paper, by improving this approximate test such that it can be applied even if the dimensions of each population are different, we derive a new test for (1.1) that guarantees approximation accuracy not only in large sample settings but also in high-dimensional settings by making appropriate modifications to the LRT. Moreover, our asymptotic approximation is valid in so-called boundary conditions such as $p_i/n_i \to 1$ as $n_i \to \infty$ and $p_i \to \infty$, where $n_i = N_i - 1$.

The remainder of the paper is organized as follows. Section 2 presents the new test statistic, its asymptotic null and non-null distributions, the asymptotic approximation-based test and the approximate confidence interval of the SGV. Section 3 then presents an empirical analysis of the null and non-null distribution of the proposed test statistic before. Section 4 concludes the paper.

2. Main result

2.1. Jensen's inequality-based criteria and its estimation. In this section, we define Jensen's inequality-based criteria for the rational discrimination between the null hypothesis $\mathscr{H}$ and the alternative hypothesis $\mathscr{A}$ and propose an estimator of the criterion that has consistency under both high-dimensional settings and large sample settings.

When $p_i \leq \min\{n_1, n_2, \dots, n_k\}$, SenGupta [8] derived the LRT statistic for the testing problem (1.1) which can be defined as follow:

$$T_L = m\left\{\ln\left(\sum_{i=1}^{k}\frac{N_i p_i}{m}|(n_i/N_i)\mathbf{S}_i|^{1/p_i}\right) - \sum_{i=1}^{k}\frac{N_i p_i}{m}\frac{\ln|(n_i/N_i)\mathbf{S}_i|}{p_i}\right\},$$

where

$$m = \sum_{j=1}^{k} N_j p_j, \qquad \bar{\mathbf{x}}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\mathbf{x}_{ij}, \qquad \mathbf{S}_i = \frac{1}{n_i}\sum_{j=1}^{N_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^{\top}.$$

Here, $n_i = N_i - 1$ for $i \in \{1, 2, \dots, k\}$. According to the general theory of the LRT, an approximate test of size α using T_L is

$$T_L > \chi^2_{k-1}(\alpha) \Leftrightarrow \text{reject } \mathscr{H}, \tag{2.1}$$

where $\chi^2_{k-1}(\alpha)$ is the upper $100 \times \alpha$ percentile of the chi-square distribution with $k-1$ degrees of freedom.

First, we introduced the relationship between a specific criterion related to Jensen's inequality and the LRT statistic and defined Jensen's inequality-based criteria. In the large sample framework, the LRT statistic T_L is a naive estimator of the following criteria:

$$\Lambda_L = m\left[\ln\left\{\sum_{i=1}^{k}\frac{N_i p_i}{m}\exp(a_i)\right\} - \sum_{i=1}^{k}\frac{N_i p_i}{m}\ln\{\exp(a_i)\}\right],$$

where

$$a_i = \frac{\ln|\boldsymbol{\Sigma}_i|}{p_i}$$

for $i \in \{1, 2, \dots, k\}$. It should be noted that Λ_L is a special case of Jensen's inequality-based criteria that is obtained using the following definition:

Definition 2.1. *For any* $\mathbf{c} = (c_1, c_2, \dots, c_k)^\top$, *such that* $c_i > 0$ *for* $i \in \{1, 2, \dots, k\}$ *and* $\mathbf{c}^\top \mathbf{1}_k = 1$, *we can define*

$$\Lambda(\mathbf{c}) = m\left[\ln\left\{\sum_{i=1}^{k} c_i \exp(a_i)\right\} - \sum_{i=1}^{k} c_i \ln\{\exp(a_i)\}\right].$$

It should be noted that if $\mathbf{c} = (N_1 p_1/m, N_2 p_2/m, \dots, N_k p_k/m)^\top$, then $\Lambda(\mathbf{c}) = \Lambda_L$. Furthermore, from Jensen's inequality, $\Lambda(\mathbf{c}) \geq 0$ and $\Lambda(\mathbf{c}) = 0$ holds if and only if $\mathscr{H}$ holds. Therefore, it can be held that $\Lambda(\mathbf{c})$ is reasonable for classifying both the null hypothesis and the alternative hypothesis.

Next, we proposed an estimator of $\Lambda(\mathbf{c})$ that is consistent not only under large sample settings but also under high-dimensional settings. Using the naive estimator a_i to estimate $\ln|\boldsymbol{\Sigma}_i|/p_i$ results in a significant bias in high-dimensional settings. Therefore, we used an exact unbiased estimator of a_i, given by

$$\hat{a}_i = \frac{\ln|\mathbf{S}_i|}{p_i} - \frac{1}{p_i}\sum_{\ell=1}^{p_i}\left\{\psi\left(\frac{n_i - \ell + 1}{2}\right) - \ln\left(\frac{n_i}{2}\right)\right\},$$

where $\psi(x) = \frac{\partial}{\partial z} \ln \Gamma(z)|_{z=x}$ is the digamma function with the gamma function $\Gamma(z)$. Using this estimator, we constructed an estimator of $\Lambda(\mathbf{c})$ as follows:

$$L_H(\mathbf{c}) = m\left[\ln\left\{\sum_{i=1}^{k} c_i \exp(\hat{a}_i)\right\} - \sum_{i=1}^{k} c_i \hat{a}_i\right]$$

and used $L_H(\mathbf{c})$ as a test statistic for the testing problem (1.1).

Finally, we investigated the asymptotic properties of $L_H(\mathbf{c})$ and $\Lambda(\mathbf{c})$ under the following three asymptotic frameworks.

(A0) The dimensions $p_1, p_2, \dots, p_k$ are fixed and each $n_i = n_i(n)$ grows as a function of n, such that n_i also tends to infinity. Furthermore, $n_i(n)/n \to r_i^*$ for some $0 < r_i^* < 1$.

(A1) Each $n_i = n_i(n)$ grows as a function of n, such that n_i also tends to infinity and $\lim_{n\to\infty} n_i(n)/n = r_i^*$ for some $0 < r_i^* < 1$. For $i \in \{1, 2, \dots, k\}$, $p_i = p_i(n)$ grows as a function of n as long as $p_i(n) <$

$n_i(n)$, such that p_i also tends to infinity and $\lim_{n\to\infty} p_i(n)/n_i(n) = q_i^*$ for some $0 \le q_i^* < 1$.

(A1′) Each $n_i = n_i(n)$ grows as a function of n, such that n_i also tends to infinity and $\lim_{n\to\infty} n_i(n)/n = r_i^*$ for some $0 < r_i^* < 1$. For $i \in \{1, 2, \ldots, k\}$, $p_i = n_i(n)$.

The condition (A0) represents a large sample framework, while conditions (A1) and (A1′) represent a high-dimensional framework. When assuming a high-dimensional frameworks (A1) or (A1′), the following condition should also be assumed.

(A2) Each $a_i = \ln|\boldsymbol{\Sigma}_i|/p_i$ grows as a function of p_i, such that $\lim_{p_i\to\infty} a_i = a_i^*$ for some $0 \le a_i^* < \infty$.

We could then obtain the following theorem to obtain the asymptotic property of $L_H(\mathbf{c})$.

Theorem 2.1. *Here, we could assume any one of the conditions* (A0), (A1), *and* (A1′). *When* (A1) *or* (A1′) *is assumed, we can also assume* (A2). *For any constant vector* $\mathbf{c} = (c_1, c_2, \ldots, c_k)^\top$, *such that* $c_i > 0$ *for* $i \in \{1, 2, \ldots, k\}$ *and* $\mathbf{c}^\top \mathbf{1}_k = 1$, $L_H(\mathbf{c})/m$ *is a consistent estimator of* $\Lambda(\mathbf{c})/m$, *that is,*

$$\frac{L_H(\mathbf{c})}{m} = \frac{\Lambda(\mathbf{c})}{m} + o_p(1).$$

Proof. The expected value and the variance of the estimator $\hat{a}_i$ can be given as follows:

$$\mathrm{E}(\hat{a}_i) = a_i, \qquad \mathrm{var}(\hat{a}_i) = \frac{1}{p_i^2} \sum_{\ell=1}^{p_i} \psi'\left(\frac{n_i - \ell + 1}{2}\right).$$

In addition, for each assumption, the variance of the estimator $\hat{a}_i$ was evaluated as follows:

$$\mathrm{var}(\hat{a}_i) = \begin{cases} O(n_i^{-1}) & \text{under (A0)}, \\ O(p_i^{-2}) & \text{under (A1)}, \\ O(p_i^{-2} \ln n_i) & \text{under (A1}'\text{)}, \end{cases} \tag{2.2}$$

(see, e.g., Sugiyama et al. [11]). From (2.2),

$$\hat{a}_i = a_i + o_p(1), \tag{2.3}$$

as $n \to \infty$. We obtain Theorem 2.1 directly from the continuous mapping theorem (see, e.g., Theorem 1.10 (ii) in Shao [10]) and (2.3).

2.2. Approximate test. In this section, we determine an appropriate $\mathbf{c}$ in $L_H(\mathbf{c})$ and propose an approximate size α test that is valid in both high-dimensional and large sample settings. We also explain how the proposed test statistic is asymptotically equivalent to the LRT statistic under large sample settings. In addition, we obtain the asymptotic distribution of the non-null distribution of the proposed test statistic.

First, we prepared the following lemma, given by Cai et al. [1], to derive the asymptotic distribution of $L_H(\mathbf{c})$.

Lemma 2.1 (Cai et al. [1]). *Let* $\mathbf{W}_i \sim \mathscr{W}_{p_i}(n_i, \boldsymbol{\Sigma}_i)$, $\mathbf{S}_i = \mathbf{W}_i/n_i$, $a_i = \ln|\boldsymbol{\Sigma}_i|/p_i$, *and*

$$\hat{a}_i = \frac{\ln|\mathbf{S}_i|}{p_i} - \frac{1}{p_i}\sum_{\ell=1}^{p_i}\left\{\psi\left(\frac{n_i - \ell + 1}{2}\right) - \ln\left(\frac{n_i}{2}\right)\right\}.$$

Here, we could assume any one of the conditions (A0), (A1), *and* (A1′). *Then,*

$$\frac{p_i(\hat{a}_i - a_i)}{\varsigma_i} \rightsquigarrow \mathscr{N}(0, 1),$$

as $n \to \infty$. *Here, "$\rightsquigarrow$" denotes convergence in distribution, and*

$$\varsigma_i = \sqrt{\sum_{\ell=1}^{p_i}\psi'\left(\frac{n_i - \ell + 1}{2}\right)}.$$

Proof. See, Cai et al. [1].

We can now discuss the selection of $\mathbf{c}$ and propose a new test statistic. From Lemma 2.1, we could choose $\mathbf{c}$ such that the null asymptotic distribution of the statistic $L_H(\mathbf{c})$ multiplied by an appropriate constant is a chi-square distribution with $k - 1$ degrees of freedom. This result can be summarized according to the following theorem:

Theorem 2.2. *Here, we could assume any one of the conditions* (A0), (A1), *and* (A1′). *Let us now define*

$$T_H = \frac{2\sum_{j=1}^{k} p_j^2/\varsigma_j^2}{m} L_H(\mathbf{c}_H),$$

where

$$\mathbf{c}_H = \frac{1}{\sum_{i=1}^{k} p_i^2/\varsigma_i^2}\left(\frac{p_1^2}{\varsigma_1^2}, \frac{p_2^2}{\varsigma_2^2}, \ldots, \frac{p_k^2}{\varsigma_k^2}\right)^{\top}.$$

Under $\mathscr{H}$, $T_H \rightsquigarrow \chi^2_{k-1}$ *as* $n \to \infty$.

Proof. Let $\tilde{z}_i = p_i(\hat{a}_i - a_i)/\varsigma_i$ and $d(\mathbf{c}) = \sum_{i=1}^{k} c_i\varsigma_i^2/(2p_i^2)$. Under $\mathscr{H}$, $L_H(\mathbf{c})$ is expanded as follows:

$$\frac{L_H(\mathbf{c})}{d(\mathbf{c})m} = \sum_{i=1}^{k} \frac{c_i\varsigma_i^2}{2p_i^2 d(\mathbf{c})}\tilde{z}_i^2 - \left(\sum_{i=1}^{k} \frac{c_i\varsigma_i}{\sqrt{2d(\mathbf{c})}p_i}\tilde{z}_i\right)^2 + o_p(1)$$

as $n \to \infty$. If we choose $\mathbf{c} = \mathbf{c}_H$,

$$\frac{L_H(\mathbf{c})}{d(\mathbf{c})m} = T_H = \tilde{\mathbf{z}}^\top(\mathbf{I}_k - \mathbf{b}\mathbf{b}^\top)\tilde{\mathbf{z}} + o_p(1) \tag{2.4}$$

as $n \to \infty$, where

$$\mathbf{b} = \frac{1}{\sqrt{\sum_{i=1}^{k} p_i^2/\varsigma_i^2}}\left(\frac{p_1}{\varsigma_1}, \frac{p_2}{\varsigma_2}, \dots, \frac{p_k}{\varsigma_k}\right)^\top.$$

From the mutual independence of $\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_k$ and Lemma 2.1, if we assume any one of the conditions (A0), (A1), and (A1′), then,

$$\tilde{\mathbf{z}} = (\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_k)^\top \rightsquigarrow \mathscr{N}_k(\mathbf{0}, \mathbf{I}_k)$$

as $n \to \infty$. Since k-dimensional random vectors $\tilde{\mathbf{z}}$ converge in distribution to a multivariate normal random variable $\mathbf{z}$, according to the continuous mapping theorem (see, e.g., Theorem 1.10 (iii) in Shao [10]), the sequence of random variables $g(\mathbf{z})$ converges in distribution to $g(\tilde{\mathbf{z}})$ if $g : \mathbb{R}^k \to \mathbb{R}$ is continuous. It should be noted that $g(\mathbf{x}) = \mathbf{x}^\top(\mathbf{I}_k - \mathbf{b}\mathbf{b}^\top)\mathbf{x}$ is a continuous function of $\mathbf{x} \in \mathbb{R}^k$. Thus, $\tilde{\mathbf{z}}^\top(\mathbf{I}_k - \mathbf{b}\mathbf{b}^\top)\tilde{\mathbf{z}} \rightsquigarrow \mathbf{z}^\top(\mathbf{I}_k - \mathbf{b}\mathbf{b}^\top)\mathbf{z}$. Meanwhile, $\mathbf{I}_k - \mathbf{b}\mathbf{b}^\top$ is an idempotent matrix and $\mathrm{rank}(\mathbf{I}_k - \mathbf{b}\mathbf{b}^\top) = k - 1$. From Cochran's theorem, $\mathbf{z}^\top(\mathbf{I}_k - \mathbf{b}\mathbf{b}^\top)\mathbf{z} \sim \chi_{k-1}^2$. Hence,

$$\tilde{\mathbf{z}}^\top(\mathbf{I}_k - \mathbf{b}\mathbf{b}^\top)\tilde{\mathbf{z}} \rightsquigarrow \chi_{k-1}^2. \tag{2.5}$$

From (2.4) and (2.5), under $\mathscr{H}$, $T_H \rightsquigarrow \chi_{k-1}^2$.

Finally, we can propose a new approximation test for the testing problem (1.1). From Theorem 2.2, the asymptotic null distribution of T_H is invariant even under large sample settings (A0); that is, the proposed method is also valid under such settings. Thus, we propose an approximate test of size α by using T_H as follows:

$$T_H > \chi_{k-1}^2(\alpha) \Leftrightarrow \text{reject } \mathscr{H}. \tag{2.6}$$

The following remark summarizes the relationship between the proposed test statistic T_H which is given in Theorem 2.2, and the log-likelihood-ratio test statistic under large sample settings.

REMARK 2.1. *Under* (A0),

$$\frac{T_H}{m} = \frac{T_L}{m} + o_p(1),$$

that is, T_H/m is asymptotically equivalent to the LRT statistic divided by m under the large sample framework.

Finally, we could obtain the asymptotic non-null distribution of test statistic T_H by applying Lemma 1.

THEOREM 2.3. *Here, we could assume any one of the conditions* (A0), (A1), *and* (A1′)*, while also assuming the following condition:*

$$\Delta_i = \frac{2p_i}{\varsigma_i}\left(\frac{\exp(a_i^*)\sum_{j=1}^k p_j^2/\varsigma_j^2}{\sum_{j=1}^k p_j^2/\varsigma_j^2 \exp(a_j^*)} - 1\right) = \Delta_i^* + o(1). \tag{2.7}$$

Let $\mathbf{z}$ be a random vector according to the multivariate normal distribution $\mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$, and let

$$\mathbf{D}^* = \mathrm{diag}(\sqrt{\exp(a_1^*)}, \sqrt{\exp(a_2^*)}, \dots, \sqrt{\exp(a_k^*)}), \qquad (\boldsymbol{\delta}^*)_i = \Delta_i^*$$

$$(\mathbf{b}^*)_i = \lim_{n\to\infty}\sqrt{\frac{\exp(a_i^*)}{\sum_{j=1}^k p_j^2\varsigma_i^2/(p_i^2\varsigma_j^2)\exp(a_j^*)}}$$

$$= \begin{cases} \sqrt{\frac{\exp(a_i^*)}{\sum_{j=1}^k p_j r_j^*/(p_i r_i^*)\exp(a_j^*)}} & \textit{under } (\mathrm{A0}), \\ \sqrt{\frac{\exp(a_i^*)}{\sum_{j=1}^k q_j^{*2} r_j^{*2}/(q_i^{*2} r_i^{*2})\ \ln(1-q_i^*)/\ln(1-q_j^*)\exp(a_j^*)}} & \textit{under } (\mathrm{A1}), \\ \sqrt{\frac{\exp(a_i^*)}{\sum_{j=1}^k q_j^{*2} r_j^{*2}/(q_i^{*2} r_i^{*2})\exp(a_j^*)}} & \textit{under } (\mathrm{A1}'), \end{cases}$$

$$c^* = \lim_{n\to\infty}\frac{\sum_{j=1}^k p_j^2/\varsigma_j^2}{\sum_{j=1}^k p_j^2/\varsigma_j^2\exp(a_i)} = \begin{cases} \frac{\sum_{j=1} p_j r_j^*}{\sum_{j=1} p_j r_j^* \exp(a_i)} & \textit{under } (\mathrm{A0}), \\ \frac{\sum_{j=1} q_j^{*2} r_j^{*2}/\ln(1-q_j^*)}{\sum_{j=1} q_j^{*2} r_j^{*2}/\ln(1-q_j^*)\exp(a_i)} & \textit{under } (\mathrm{A1}), \\ \frac{\sum_{j=1} q_j^{*2} r_j^{*2}}{\sum_{j=1} q_j^{*2} r_j^{*2}\exp(a_i)} & \textit{under } (\mathrm{A1}'). \end{cases}$$

Then,

$$T_H - \frac{2\sum_{j=1}^k p_j^2/\varsigma_j^2}{m}\Lambda_H(\mathbf{c}_H) \rightsquigarrow \boldsymbol{\delta}^{*\top}\mathbf{z} + c^*\mathbf{z}^\top\mathbf{D}^*(\mathbf{I}_k - \mathbf{b}^*\mathbf{b}^{*\top})\mathbf{D}^*\mathbf{z}$$

as $n \to \infty$.

PROOF. Let $\tilde{z}_i = p_i(\hat{a}_i - a_i)/\varsigma_i$. By using $\mathbf{z}$, under (A1) and (A2), $L_H(\mathbf{c}_H^*)$ is expanded as follows:

$$\begin{aligned}
T_H = {} & \frac{2\sum_{j=1}^k p_j^2/\varsigma_j^2}{m} \Lambda_H(\mathbf{c}_H^*) \\
& + \sum_{i=1}^k \frac{2p_i}{\varsigma_i}\left(\frac{\exp(a_i^*)\sum_{j=1}^k p_j^2/p_i^2\varsigma_j^2/\varsigma_i^2}{\sum_{j=1}^k p_j^2/p_i^2\varsigma_j^2/\varsigma_i^2 \exp(a_j^*)} - 1\right)\tilde{z}_i \\
& + \frac{\sum_{j=1}^k p_j^2/\varsigma_j^2}{\sum_{j=1}^k p_j^2/\varsigma_j^2 \exp(a_j)} \sum_{i=1}^k \exp(a_i)\tilde{z}_i^2 \\
& - \left(\frac{\sqrt{\sum_{j=1}^k p_j^2/\varsigma_j^2}}{\sum_{j=1}^k p_j^2/\varsigma_j^2 \exp(a_j)} \sum_{i=1}^k \frac{p_i \exp(a_i)}{\varsigma_i}\tilde{z}_i\right)^2 + o_p(1).
\end{aligned}$$

Thus, if we assume any one of the conditions (A0), (A1), and (A1′), then,

$$T_H = \boldsymbol{\delta}^{*\top}\tilde{\mathbf{z}} + c^*\tilde{\mathbf{z}}^\top\mathbf{D}^*(\mathbf{I}_k - \mathbf{b}^*\mathbf{b}^{*\top})\mathbf{D}^*\tilde{\mathbf{z}} + o_p(1). \tag{2.8}$$

From the mutual independence of $\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_k$ and Lemma 1, if we assume any one of the conditions (A0), (A1), and (A1′), then,

$$\tilde{\mathbf{z}} = (\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_k)^\top \rightsquigarrow \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$$

as $n \to \infty$. Since the k-dimensional random vector $\tilde{\mathbf{z}}$ converges in distribution to a multivariate normal random variable $\mathbf{z}$, according to the continuous mapping theorem (see, e.g., Theorem 1.10 (iii) in Shao [10]), the sequence of random variables $g(\tilde{\mathbf{z}})$ converges in distribution to $g(\mathbf{z})$ if $g : \mathbb{R}^k \to \mathbb{R}$ is continuous. It should be noted that $g(\mathbf{x}) = \boldsymbol{\delta}^{*\top}\mathbf{x} + c^*\mathbf{x}^\top\mathbf{D}^*(\mathbf{I}_k - \mathbf{b}^*\mathbf{b}^{*\top})\mathbf{D}^*\mathbf{x}$ is a continuous function of $\mathbf{x} \in \mathbb{R}^k$. Thus, from the continuous mapping theorem,

$$\boldsymbol{\delta}^{*\top}\tilde{\mathbf{z}} + c^*\tilde{\mathbf{z}}^\top\mathbf{D}^*(\mathbf{I}_k - \mathbf{b}^*\mathbf{b}^{*\top})\mathbf{D}^*\tilde{\mathbf{z}} \rightsquigarrow \boldsymbol{\delta}^{*\top}\mathbf{z} + c^*\mathbf{z}^\top\mathbf{D}^*(\mathbf{I}_k - \mathbf{b}^*\mathbf{b}^{*\top})\mathbf{D}^*\mathbf{z}. \tag{2.9}$$

In combining (2.8) and (2.9), we can prove Theorem 2.3.

2.3. Interval estimation of $|\boldsymbol{\Sigma}_i|^{1/p_i}$. In this section, we apply Lemma 2.1 to propose an approximate confidence interval for the SGV $|\boldsymbol{\Sigma}_i|^{1/p_i}$.

THEOREM 2.4. *Here, we could assume any one of the conditions* (A0), (A1), *and* (A1′). *Then,*

$$\frac{p_i\{\exp(\hat{a}_i) - |\boldsymbol{\Sigma}_i|^{1/p_i}\}}{\exp(\hat{a}_i)\varsigma_i} \rightsquigarrow \mathcal{N}(0, 1),$$

as $n \to \infty$.

Proof. From Lemma 2.1 and the delta method, we could obtain the following:

$$\frac{p_i\{\exp(\hat{a}_i) - |\boldsymbol{\Sigma}_i|^{1/p_i}\}}{\exp(a_i)\varsigma_i} \rightsquigarrow \mathcal{N}(0,1).$$

In combining this result with (2.3) and Slutsky's theorem, we could prove Theorem 2.4.

By applying Theorem 2.4, we could propose an approximate $1-\alpha$ confidence interval of $|\boldsymbol{\Sigma}_i|^{1/p_i}$ as

$$\left(\exp(\hat{a}_i) - \frac{\exp(\hat{a}_i)\varsigma_i}{p_i} z\left(\frac{\alpha}{2}\right), \exp(\hat{a}_i) + \frac{\exp(\hat{a}_i)\varsigma_i}{p_i} z\left(\frac{\alpha}{2}\right)\right), \tag{2.10}$$

where $z(\alpha/2)$ is the upper $100 \times (\alpha/2)$ percentile of the standard normal distribution.

3. Simulation studies and applications

3.1. Simulation studies. First, we compared the proposed test (2.6) with the approximate LRT test (2.1). In this simulation study, we make the following assumptions: all the mean vectors $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_k$ are all equal to $\mathbf{0}$, the covariance matrix $\boldsymbol{\Sigma}_i$ is $\boldsymbol{\Sigma}_i = i^\delta \mathbf{I}_{p_i}$ where $\delta \in \{0.0, 0.1, 0.2, 0.3\}$, the nominal significance level α is $\alpha = 0.05$, and the number of populations k is $k \in \{4, 6\}$. The settings for dimension $\boldsymbol{p} = (p_1, p_2, \dots, p_k)$ and sample size $\boldsymbol{N} = (N_1, N_2, \dots, N_k)$ are summarized in Table 1.

For any combination of $(\boldsymbol{p}_i, \boldsymbol{N}_{ij})$ and $\delta \in \{0.0, 0.1, 0.2, 0.3\}$, we can use the following steps to calculate the empirical type-I error or power of each test method.

1 We set $\boldsymbol{\Sigma}_i = i^\delta \mathbf{I}_{p_i}$ where $\delta \in \{0.0, 0.1, 0.2, 0.3\}$.

2 We generated an independent sample $\boldsymbol{x}_{i1}^{(b)}, \boldsymbol{x}_{i2}^{(b)}, \dots, \boldsymbol{x}_{iN_i}^{(b)}$ drawn from the multivariate normal distribution $\mathcal{N}_{p_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ for $i \in \{1, 2, 3, 4\}$.

3 For the sample $\boldsymbol{x}_{11}^{(b)}, \boldsymbol{x}_{12}^{(b)}, \dots, \boldsymbol{x}_{kN_k}^{(b)}$, we calculated the realized value of T_H, which is denoted as $t_H^{(b)}$ and the realized value of T_L, denoted as $t_L^{(b)}$.

4 We computed the estimated probability of method $M \in \{H, L\}$ as

$$P_M(\delta) = \frac{1}{100{,}000} \sum_{b=1}^{100{,}000} I(t_H^{(b)} > \chi^2_{0.05}),$$

respectively, where $I(\cdot)$ is the indicator function. Meanwhile, $P_M(0)$ represents an empirical type-I error and $P(\delta)$ for $\delta \in \{0.1, 0.2, 0.3\}$ represents an empirical power.

Table 1. The dimension $\boldsymbol{p}$ and sample size $\boldsymbol{N}$ settings

$\boldsymbol{p} = (p_1, p_2, \dots, p_k)$	$\boldsymbol{N} = (N_1, N_2, \dots, N_k)$
$\boldsymbol{p}_1 = (5, 6, 7, 8)$	$\boldsymbol{N}_{11} = (25, 26, 27, 28)$
$\boldsymbol{p}_1 = (5, 6, 7, 8)$	$\boldsymbol{N}_{12} = (25, 36, 47, 58)$
$\boldsymbol{p}_2 = (10, 15, 20, 25)$	$\boldsymbol{N}_{21} = (30, 35, 40, 45)$
$\boldsymbol{p}_2 = (10, 15, 20, 25)$	$\boldsymbol{N}_{22} = (30, 45, 60, 75)$
$\boldsymbol{p}_3 = (50, 60, 70, 80)$	$\boldsymbol{N}_{31} = (70, 80, 90, 100)$
$\boldsymbol{p}_3 = (50, 60, 70, 80)$	$\boldsymbol{N}_{32} = (70, 90, 110, 130)$
$\boldsymbol{p}_4 = (100, 120, 140, 160)$	$\boldsymbol{N}_{41} = (120, 140, 160, 180)$
$\boldsymbol{p}_4 = (100, 120, 140, 160)$	$\boldsymbol{N}_{42} = (120, 150, 180, 210)$
$\boldsymbol{p}_5 = (3, 4, 5, 6, 7, 8)$	$\boldsymbol{N}_{51} = (23, 24, 25, 26, 27, 28)$
$\boldsymbol{p}_5 = (3, 4, 5, 6, 7, 8)$	$\boldsymbol{N}_{52} = (23, 34, 45, 56, 67, 78)$
$\boldsymbol{p}_6 = (5, 10, 15, 20, 25, 30)$	$\boldsymbol{N}_{61} = (25, 30, 35, 40, 45, 50)$
$\boldsymbol{p}_6 = (5, 10, 15, 20, 25, 30)$	$\boldsymbol{N}_{62} = (25, 40, 55, 70, 85, 100)$
$\boldsymbol{p}_7 = (30, 40, 50, 60, 70, 80)$	$\boldsymbol{N}_{71} = (50, 60, 70, 80, 90, 100)$
$\boldsymbol{p}_7 = (30, 40, 50, 60, 70, 80)$	$\boldsymbol{N}_{72} = (50, 70, 90, 110, 130, 150)$
$\boldsymbol{p}_8 = (100, 120, 140, 160, 180, 200)$	$\boldsymbol{N}_{81} = (120, 140, 160, 180, 200, 220)$
$\boldsymbol{p}_8 = (100, 120, 140, 160, 180, 200)$	$\boldsymbol{N}_{82} = (120, 150, 180, 210, 240, 270)$

The empirical sizes and the empirical powers calculated using 10^5 replications are listed in Table 2 and Table 3 for $k = 4$ and $k = 6$, respectively. Here, it was confirmed that the power tended to increase as δ increased, while the power tended to increase as the dimension p and number of groups k increased. These trends can be regarded as natural.

Next, we investigated the accuracy of the approximation of the non-null distributions. Here, we could generate Monte Carlo samples $t_H^{(1)}, t_H^{(2)}, \dots, t_H^{(B)}$ of test statistic T_H by repeating the following procedure B times.

1 We generated an independent sample $\boldsymbol{x}_{i1}^{(b)}, \boldsymbol{x}_{i2}^{(b)}, \dots, \boldsymbol{x}_{iN_i}^{(b)}$ drawn from the multivariate normal distribution $\mathcal{N}_{p_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ for $i \in \{1, 2, 3, 4\}$.

2 For the sample $\boldsymbol{x}_{11}^{(b)}, \boldsymbol{x}_{12}^{(b)}, \dots, \boldsymbol{x}_{kN_k}^{(b)}$, we calculated the realized value of T_H, which is denoted as $t_H^{(b)}$.

Using the probability expression given in Theorem 2.3, the distribution of T_H can be approximated according to the distribution of $\tilde{T}_H$, given by

$$\tilde{T}_H = \frac{2\sum_{j=1}^{k} p_j^2/\varsigma_j^2}{m} \Lambda_H(\mathbf{c}_H) + \boldsymbol{\delta}^\top \mathbf{z} + c\mathbf{z}^\top \mathbf{D}(\mathbf{I}_k - \mathbf{b}\mathbf{b}^\top)\mathbf{D}\mathbf{z},$$

where $\mathbf{z}$ is a random vector distributed as $\mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$ and

Table 2. When $k = 4$, the empirical probabilities $P_{T_H}(\delta)$ and $P_{T_L}(\delta)$ can be calculated for any combination of $(\boldsymbol{p}_i, \boldsymbol{N}_{ij})$ and $\delta \in \{0.0, 0.1, 0.2, 0.3\}$. Here, $P_M(0)$ represents an empirical type-I error and $P(\delta)$ for $M \in \{T_H, T_L\}$, $\delta \in \{0.1, 0.2, 0.3\}$ represents an empirical power.

$\boldsymbol{p}$	$\boldsymbol{N}_{ij}$	δ	$P_{T_H}(\delta)$	$P_{T_L}(\delta)$	$\boldsymbol{p}$	$\boldsymbol{N}_{ij}$	δ	$P_{T_H}(\delta)$	$P_{T_L}(\delta)$
$\boldsymbol{p}_1$	$\boldsymbol{N}_{11}$	0.0	0.047	0.099	$\boldsymbol{p}_3$	$\boldsymbol{N}_{31}$	0.0	0.050	0.822
		0.1	0.089	0.116			0.1	0.877	0.463
		0.2	0.240	0.261			0.2	1.000	0.999
		0.3	0.511	0.526			0.3	1.000	1.000
	$\boldsymbol{N}_{12}$	0.0	0.049	0.095		$\boldsymbol{N}_{32}$	0.0	0.051	0.946
		0.1	0.101	0.256			0.1	0.935	1.000
		0.2	0.309	0.574			0.2	1.000	1.000
		0.3	0.648	0.863			0.3	1.000	1.000
$\boldsymbol{p}_2$	$\boldsymbol{N}_{21}$	0.0	0.050	0.400	$\boldsymbol{p}_4$	$\boldsymbol{N}_{41}$	0.0	0.050	1.000
		0.1	0.174	0.145			0.1	1.000	1.000
		0.2	0.610	0.287			0.2	1.000	1.000
		0.3	0.945	0.726			0.3	1.000	1.000
	$\boldsymbol{N}_{22}$	0.0	0.049	0.117		$\boldsymbol{N}_{42}$	0.0	0.050	1.000
		0.1	0.216	0.468			0.1	1.000	1.000
		0.2	0.729	0.913			0.2	1.000	1.000
		0.3	0.983	0.998			0.3	1.000	1.000

$$\mathbf{D} = \mathrm{diag}(\sqrt{\exp(a_1)}, \sqrt{\exp(a_2)}, \dots, \sqrt{\exp(a_k)}),$$

$$(\boldsymbol{\delta})_i = \frac{2p_i}{\varsigma_i}\left(\frac{\exp(a_i)\sum_{j=1}^k p_j^2/\varsigma_j^2}{\sum_{j=1}^k p_j^2/\varsigma_j^2 \exp(a_j)} - 1\right),$$

$$(\mathbf{b})_i = \sqrt{\frac{\exp(a_i)}{\sum_{j=1}^k p_j^2\varsigma_i^2/(p_i^2\varsigma_j^2)\exp(a_j)}}, \qquad c = \frac{\sum_{j=1}^k p_j^2/\varsigma_j^2}{\sum_{j=1}^k p_j^2/\varsigma_j^2 \exp(a_i)}.$$

We could then generate Monte Carlo samples of $\tilde{T}_H$ as $\tilde{t}_H^{(1)}, \tilde{t}_H^{(2)}, \dots, \tilde{t}_H^{(B)}$ by repeating the following procedure B times.

1 We generated an independent sample $\boldsymbol{z}^{(b)}$ drawn from $\mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$.

2 For the sample $\boldsymbol{z}_b$, we calculated

$$\tilde{t}_H^{(b)} = \frac{2\sum_{j=1}^k p_j^2/\varsigma_j^2}{m}\Lambda_H(\mathbf{c}_H) + \boldsymbol{\delta}^\top \boldsymbol{z} + c\boldsymbol{z}^\top \mathbf{D}(\mathbf{I}_k - \mathbf{b}\mathbf{b}^\top)\mathbf{D}\boldsymbol{z}.$$

In all simulations, we set $B = 10^5$, while we implemented the procedure described above using specific parameter settings. Meanwhile, in all of the simulations, without any loss of generality, we supposed that $\boldsymbol{\mu}_i = \mathbf{0}$ for $i \in$

Table 3. When $k = 6$, the empirical probabilities $P_{T_H}(\delta)$ and $P_{T_L}(\delta)$ can be calculated for any combination of $(\boldsymbol{p}_i, \boldsymbol{N}_{ij})$ and $\delta \in \{0.0, 0.1, 0.2, 0.3\}$. Here, $P_M(0)$ represents an empirical type-I error and $P(\delta)$ for $M \in \{T_H, T_L\}$, $\delta \in \{0.1, 0.2, 0.3\}$ represents an empirical power.

$\boldsymbol{p}$	$\boldsymbol{N}_{ij}$	δ	$P_{T_H}(\delta)$	$P_{T_L}(\delta)$	$\boldsymbol{p}$	$\boldsymbol{N}_{ij}$	δ	$P_{T_H}(\delta)$	$P_{T_L}(\delta)$
$\boldsymbol{p}_5$	$\boldsymbol{N}_{51}$	0.0	0.048	0.119	$\boldsymbol{p}_7$	$\boldsymbol{N}_{71}$	0.0	0.050	0.997
		0.1	0.088	0.113			0.1	0.903	0.372
		0.2	0.251	0.243			0.2	1.000	0.961
		0.3	0.566	0.527			0.3	1.000	1.000
	$\boldsymbol{N}_{52}$	0.0	0.049	0.079		$\boldsymbol{N}_{72}$	0.0	0.050	0.636
		0.1	0.110	0.233			0.1	0.972	1.000
		0.2	0.389	0.608			0.2	1.000	1.000
		0.3	0.787	0.919			0.3	1.000	1.000
$\boldsymbol{p}_6$	$\boldsymbol{N}_{61}$	0.0	0.050	0.711	$\boldsymbol{p}_8$	$\boldsymbol{N}_{81}$	0.0	0.050	0.721
		0.1	0.181	0.258			0.1	1.000	0.979
		0.2	0.672	0.197			0.2	1.000	1.000
		0.3	0.975	0.571			0.3	1.000	1.000
	$\boldsymbol{N}_{62}$	0.0	0.050	0.103		$\boldsymbol{N}_{82}$	0.0	0.050	1.000
		0.1	0.266	0.362			0.1	1.000	1.000
		0.2	0.877	0.921			0.2	1.000	1.000
		0.3	0.999	0.999			0.3	1.000	1.000

$\{1, 2, \ldots, k\}$. We set the covariance structure as $\boldsymbol{\Sigma}_i = i^{\delta}\mathbf{I}_{p_i}$ where $\delta \in \{0.1, 0.2, 0.3\}$, each sample size as $N_i = p_i + 20$ for $i \in \{1, 2, 3, 4\}$, and the dimensions as $\boldsymbol{p}_1 = (5, 6, 7, 8)^\top$, $\boldsymbol{p}_2 = (20, 30, 40, 50)^\top$ and $\boldsymbol{p}_3 = (100, 120, 140, 160)^\top$.

As Figure 1 shows, for any combination of $\boldsymbol{p}$ and $\delta \in \{0.1, 0.2, 0.3\}$, we compared the smoothed histogram of $t_H^{(1)}, t_H^{(2)}, \ldots, t_H^{(B)}$ with one of $\tilde{t}_H^{(1)}, \tilde{t}_H^{(2)}, \ldots, \tilde{t}_H^{(B)}$. In all of the figures, the $\tilde{t}_H^{(b)}$ histogram is very close to the $t_H^{(b)}$ histogram. These behaviors are consistent with Theorem 2.3. In Addition, the $\tilde{t}_H^{(b)}$ histogram and the $t_H^{(b)}$ histogram deviated from the chi-square distribution with $k - 1$ degrees of freedom as δ increased, that is, the power of the proposed test increased as δ became larger. The $\tilde{t}_H^{(b)}$ histogram and the $t_H^{(b)}$ histogram also deviated from the chi-square distribution with $k - 1$ degrees of freedom as the number of populations k and dimension p increased. From these results, the natural behavior of the power of the proposed test could be confirmed.

3.2. Applications. The data related to the Leptograpsus crab (Campbell and Mahon [2]) are available in the MASS package (Ripley et al. [7]) for R. This data set consists of 200 subjects: 100 of species orange (50 male and 50 female) and 100 of species blue (50 male and 50 female). For each subject,

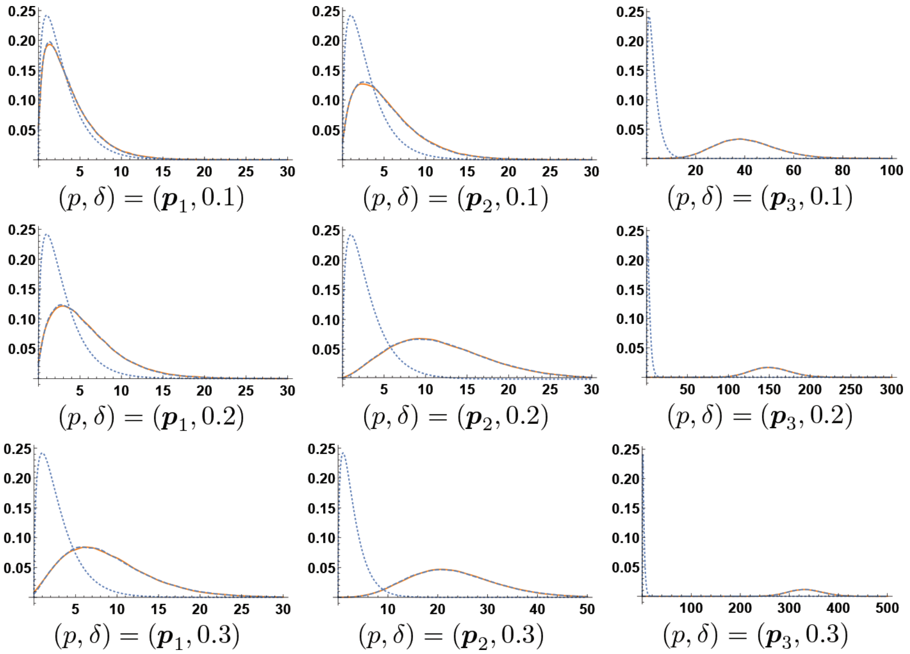

Fig. 1. For any combination of $\boldsymbol{p} \in \{\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3\}$ and $\delta \in \{0.1, 0.2, 0.3\}$, the smoothed histogram of $t_H^{(1)}, t_H^{(2)}, \dots, t_H^{(B)}$ is represented by the dashed lines, the smoothed histogram of $\tilde{t}_H^{(1)}, \tilde{t}_H^{(2)}, \dots, \tilde{t}_H^{(B)}$ by the solid lines and the chi-squared density function by the dotted lines.

five measurements, $x_1 =$ frontal lobe size (mm), $x_2 =$ rear width (mm), $x_3 =$ carapace length (mm), $x_4 =$ carapace width (mm), and $x_5 =$ body depth (mm) were recorded. In the following sections, the blue male, blue female, orange male, and orange female crabs are denoted by 1, 2, 3, and 4, respectively. We could then consider the power set of $\{x_1, x_2, x_3, x_4\}$ which is denoted by $\mathscr{P}$. For each set of variables belonging to $\mathscr{P}$, we verified whether or not the SGVs in the four groups were all equal with the results summarized in Table 4.

4. Conclusion

In this study, we proposed an asymptotic approximation-based test for the equality of the SGVs of k multivariate normal populations in both high-dimensional and large sample settings. SenGupta [8] and [9] investigated a LRT statistic for this testing problem and proposed a reasonable approximation test for a large sample setting. Our test presents an improvement to the LRT statistic that is valid in high-dimensional settings. Using the asymptotic

Table 4. For the results presented in this table, the SGV equivalence test was performed on the corresponding set of variables. The p-value of proposed statistic T_H is denoted as P_{T_H}, while the p-value of proposed statistic T_L is denoted as P_{T_L}. P-values of less than 0.05 are presented in boldface.

p	variables	P_{T_H}	P_{T_L}	p	variables	P_{T_H}	P_{T_L}
1	x_1	0.233	0.217	3	$x_1x_2x_3$	**0.039**	**0.030**
	x_2	0.750	0.738		$x_1x_2x_4$	**0.028**	**0.021**
	x_3	0.138	0.125		$x_1x_2x_5$	**0.019**	**0.014**
	x_4	0.175	0.160		$x_1x_3x_4$	**0.022**	**0.016**
	x_5	0.227	0.211		$x_1x_3x_5$	0.239	0.213
2	x_1x_2	0.184	0.165		$x_1x_4x_5$	0.078	0.065
	x_1x_3	0.065	0.054		$x_2x_3x_4$	0.137	0.117
	x_1x_4	**0.021**	**0.017**		$x_2x_3x_5$	0.111	0.094
	x_1x_5	0.510	0.487		$x_2x_4x_5$	0.078	0.064
	x_2x_3	0.267	0.245		$x_3x_4x_5$	0.506	0.478
	x_2x_4	0.265	0.243	4	$x_1x_2x_3x_4$	**0.007**	**0.004**
	x_2x_5	0.267	0.245		$x_1x_2x_3x_5$	**0.028**	**0.021**
	x_3x_4	0.575	0.554		$x_1x_2x_4x_5$	**0.009**	**0.006**
	x_3x_5	0.735	0.719		$x_1x_3x_4x_5$	**0.035**	**0.026**
	x_4x_5	0.759	0.744		$x_2x_3x_4x_5$	0.053	**0.041**
				5	$x_1x_2x_3x_4x_5$	**0.004**	**0.002**

results reported by Cai et al. [1], we obtained the null and non-null asymptotic distributions of the proposed test statistic. The features of the proposed test statistic are valid not only for high-dimensional and large sample settings but also for large sample settings. Furthermore, using various parameter settings, we investigated the finite sample and dimension behavior of our test statistic using Monte Carlo simulations. The simulation results confirmed that our asymptotic results work well as approximations in term of a finite sample and finite dimensions. Overall, we demonstrated that the proposed test is unique in that it has the capacity to work within a wider range than the conventional method.

Acknowledgement

The first author was supported by JSPS KAKENHI Grant Number 21K17716. The second author was supported by JSPS KAKENHI Grant Number 20K11712. The fourth author was supported by JSPS KAKENHI Grant Number 21K11795. We would also like to express our gratitude to Professor Yasunori Fujikoshi for many valuable comments and discussions.

References

[1] T. T. Cai, T. Liang and H. H. Zhou, Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions. J. Multivariate Anal. **137** (2015), 161–172.

[2] N. Campbell and R. Mahon, A multivariate study of variation in two species of rock crab of the genus leptograpsus. Aust. J. Zool. **22** (1974), 417–425.

[3] D. Najarzadeh, Testing equality of standardized generalized variances of k multivariate normal populations with arbitrary dimensions. Stat. Methods Appl. **28** (2019), 593–623.

[4] D. Najarzadeh, Testing equality of generalized variances of k multivariate normal populations. Comm. Statist. Simulation Comput. (2017), 1–10.

[5] D. Pena and A. Linde, Dimensionless measures of variability and dependence for multivariate continuous distributions. Comm. Statist. Theory Methods **36** (2007), 1845–1854.

[6] D. Pena and J. Rodriguez, Descriptive measures of multivariate scatter and linear dependence. J. Multivariate Anal. **85** (2003), 361–374.

[7] B. Ripley, B. Venables, D. M. Bates, K. Hornik and A. Gebhardt, Firth D, Ripley MB. (2013) Package mass. Cran R.

[8] A. SenGupta, Generalizations of Barlett's and Hartley's tests of homogeneity using overall variability. Comm. Statist. Theory Methods **16** (1987), 987–996.

[9] A. SenGupta, Tests for standardized generalized variances of multivariate normal populations of possibly different dimensions. J. Multivariate Anal. **23** (1987), 209–219.

[10] J. Shao, Mathematical statistics. Springer (2007).

[11] T. Sugiyama, M. Hyodo, H. Watanabe, S. Tsukada and T. Seo, Test for equality of generalized variance in high-dimensional and large sample settings. SUT. J. Math. **55** (2019), 139–154.

[12] S. S. Wilks, Certain generalizations in the analysis of variance. Biometrika **24** (1932), 471–494.

Hiroki Watanabe
Department of Health Sciences
Oita University of Nursing and Health Sciences
Oita 870-1201 *Japan*
E-mail: watahiro919@gmail.com

Masashi Hyodo
Department of Economics
Kanagawa University
Kanagawa 221-8686 *Japan*
E-mail: ft102119jw@kanagawa-u.ac.jp

Takatoshi Sugiyama
Department of Applied Mathematics
Tokyo University of Science
Tokyo 162-8601 *Japan*
E-mail: sugiyamatakatoshi@gmail.com

Takashi Seo
Department of Applied Mathematics
Tokyo University of Science
Tokyo 162-8601 *Japan*
E-mail: seo@rs.tus.ac.jp