A Modified Bartlett Test
for Heteroscedastic Two-way MANOVA

Jin-Ting Zhang[1*], Bu Zhou[1], Jia Guo[1] and Xuefeng Liu[2]

[1]Department of Statistics and Applied Probability, National University of Singapore, Singapore
[2]IBM, Xi-An, China
Email: stazjt@nus.edu.sg

Abstract A two-way multivariate analysis of variance (MANOVA) aims to compare the effects of several levels of two factors in a factorial experiment with two-way layout. It is widely used in experimental sciences, e.g., biology, psychology, physics, among others. When the cell covariance matrices are the same, it can be solved using the well-known Wilks likelihood ratio, Lawley-Hotelling trace, Bartlett-Nanda-Pillai trace and Roy's largest root tests ([1]). However, when the homogeneous assumption is violated, these tests may become seriously biased. To overcome this problem, several authors have proposed and studied different approximation solutions. In this paper, we propose and study a Modified Bartlett (MB) test using a Wald-type statistic and the modified Bartlett correction ([2]) for heteroscedastic two-way MANOVA problems. The MB test can be easily implemented using the usual χ^2-distribution with known degrees of freedom. We show that it admits several invariant properties. Simulation studies show that the MB test generally outperforms the classical Lawley-Hotelling trace (LHT) test and a modified LHT test of [3] under various parameter configurations in terms of size controlling and power. A real data example illustrates our method and the effect of heteroscedasticity.

Keywords: Heteroscedastic two-way MANOVA, Tests of linear hypotheses, Modified Bartlett correction, Wald-type statistic.

1 Introduction

A two-way multivariate analysis of variance (MANOVA) aims to compare the effects of several levels of two factors in a factorial experiment with two-way layout. It is a multivariate version of two-way ANOVA model and is widely used in experimental sciences, e.g., biology, psychology, physics, among others; examples may be found in [4], [5], and [6], among others. When the cell covariance matrices are known to be the same, this problem can be solved using the Wilks likelihood ratio (WLR), Lawley-Hotelling trace (LHT), Bartlett-Nanda-Pillai (BNP) and Roy's largest root tests as discussed in [1]. However, when the homogeneous assumption is violated, these tests may become seriously biased, which means their sizes may be severely inflated or deflated. For example, in our simulations which are presented in Section 3, we set the nominal size $\alpha = 5\%$, the empirical size of the LHT test for testing interaction effect could be as large as 75% or as small as 0%. This is a serious problem. In real data analysis, Box's M test ([7]) is usually used to check whether the cell covariance matrices are equal and when the null hypothesis is rejected, those tests mentioned above are not suitable for the main effect testing or interaction effect testing. In this case, a test for heteroscedastic two-way MANOVA is needed.

 To our knowledge, this problem for heteroscedastic two-way MANOVA has not been well addressed in the literature. Recently, [3] tried to solve this problem via modifying the WLR, LHT and BNP tests. Their main ideas focus on modifying the degrees of freedom of the random matrices involved in the test statistics so that the heteroscedasticity of the cell covariance matrices is taken into account and the WLR, LHT and BNP tests can still be used but with the degrees of freedom estimated from the data via matching the first two moments. Although their approaches are simple to understand, these approaches admit the following three main drawbacks: (1) one needs to estimate the degrees of freedom of both the random matrices involved in the test statistics; (2) the estimated degrees of freedom, as given in Section 3 of [3], are complicated, case-sensitive, and not affine invariant; and (3) the null distributions of the WLR, LHT and BNP tests with known degrees of freedom are not immediately available; further approximations

based on χ^2 or normal asymptotic expansions are needed, as shown in Sections 3.1 and 3.2 of [3]. To overcome these drawbacks, using Wishart-approximation, [8] proposed an approximate Hotelling T^2 test, and [9] improved [3]'s modified MANOVA tests. [10,11] studied an approximate degrees of freedom F-test for heteroscedastic two-way ANOVA and one-way MANOVA respectively.

The most related topic to the heteroscedastic two-way MANOVA is the heteroscedastic one-way MANOVA which tests the effect of a factor, having k levels, in an experimental design. When $k = 2$, this problem is often referred to as the multivariate Behrens-Fisher (BF) problem and it has been well addressed in the literature by various authors including [12], [13], [14], [15], [16], [17], [18], and [19], among others. [18]'s test is based on a Wald-type statistic's asymptotic distribution and the modified Bartlett correction of [2]. [2] proposed several monotone transformations that can be applied to a wide class of approximately chi-squared distributed statistics, aiming to improve the chi-squared approximation accuracy. According to [18]'s simulation studies, the MB test based on [2]'s method works well for the two-sample multivariate Behrens-Fisher problem. We feel this idea also works for heteroscedastic two-way MANOVA models, so we propose and study the MB test in this paper. Following [18], we use a Wald-type statistic (see for example, [20]) and adopt the modified Bartlett correction. Our MB test admits several nice properties: (1) it has a simple form and its P-value is easy to compute using a chi-square distribution with known degrees of freedom; (2) it is shown to be affine-invariant; (3) all the related tests under the two-way MANOVA, such as the main effect, interaction effect, post hoc, and contrast tests among others, can be unified under a common framework — a general linear hypothesis test. The MB test is shown to be invariant under different choices of the matrices used to define the same hypothesis; and (4) it works well. Simulation results reported in Section 3 show that the MB test generally outperforms the LHT test in terms of size controlling and [3]'s LHT_m test in terms of size and power under heteroscedastic cases. The simulations also show that the MB test does not lose much power in homogenous cases compared to the LHT test, indicating the MB test also works well when the cell covariance matrices are the same. We would also like to mention that it is straight forward to extend the ideas and methodologies for two-way heteroscedastic MANOVA to three and higher-way heteroscedastic MANOVA where more factors are considered as done in [8].

It is worthwhile to mention that heteroscedastic ANOVA and MANOVA problems have been paid much attention in the past decades due to their interesting and challenging nature. There are two major kinds of methods: simulation-based approaches and approximation-based approaches. The references mentioned earlier among others are generally approximation-based approaches. In the literature, simulation-based approaches are also popular due to their simplicity and recent advance of computation powers. For example, to deal with heteroscedasticity, [21] proposed a simulation-based approximate test. [22] proposed another simulation-based test using the concept of generalized P-values, resulting in a so-called generalized F-test. The generalized P-values method is further studied by [23] and [24]. [25] compared the performance of a parametric bootstrap test with the generalized F-test of [22]. [26] proposed a bootstrap method and [27] studied a permutation test for general ANOVA problems. A drawback of simulation-based approaches is that they are often time-consuming especially when the sample sizes are large.

The rest of the paper is organized as follows. The methodologies for the MB test are presented in Section 2. Simulation results are presented in Section 3. An example using a data set from a smoking cessation trial is presented in Section 4. Finally, some technical proofs of the main results are given in the Appendix.

2 Methodologies

2.1 Main and Interaction Effects in Two-way MANOVA Models

Consider a two-way experiment with two factors A and B having a and b levels, respectively, with a total of ab factorial combinations or cells. Suppose at the (i, j)-th cell, we have a p-dimensional random sample: $\boldsymbol{y}_{ijk}, k = 1, 2, \cdots, n_{ij}$, satisfying the following model:

$$\boldsymbol{y}_{ijk} = \boldsymbol{\mu}_{ij} + \boldsymbol{\epsilon}_{ijk}, \quad \boldsymbol{\epsilon}_{ijk} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{ij}), \ k = 1, \cdots, n_{ij}, \tag{1}$$

where $\boldsymbol{\mu}_{ij} : p \times 1$ and $\boldsymbol{\Sigma}_{ij} : p \times p$ are the cell mean vector and cell covariance matrix of the random sample at the (i, j)-th cell. All these ab samples are assumed to be independent with each other. In this

subsection, we aim to represent the main and interaction effects as linear combinations of the cell means which are estimable in the two-way MANOVA model (2) as described below.

In two-way MANOVA, the cell mean vectors $\boldsymbol{\mu}_{ij}$ are usually decomposed into the form $\boldsymbol{\mu}_{ij} = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij}$, $i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$, where $\boldsymbol{\mu}_0$ is the grand mean vector, $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ are the i-th and j-th main effects of factors A and B, respectively, and $\boldsymbol{\gamma}_{ij}$ is the (i, j)-th interaction effect between factors A and B so that (1) can be further written as the following well-known two-way MANOVA model:

$$\boldsymbol{y}_{ijk} = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij} + \boldsymbol{\epsilon}_{ijk}, \ \boldsymbol{\epsilon}_{ijk} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{ij}),$$
$$k = 1, 2, \cdots, n_{ij}; \ i = 1, 2, \cdots, a; j = 1, 2, \cdots, b. \tag{2}$$

For this model, we are interested in the following three null hypotheses:

$$\begin{aligned} H_{0A} \ &: \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \cdots = \boldsymbol{\alpha}_a = \boldsymbol{0}, \\ H_{0B} \ &: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_b = \boldsymbol{0}, \\ H_{0AB} &: \boldsymbol{\gamma}_{11} = \cdots = \boldsymbol{\gamma}_{1b} = \cdots = \boldsymbol{\gamma}_{a1} = \cdots = \boldsymbol{\gamma}_{ab} = \boldsymbol{0}. \end{aligned} \tag{3}$$

The first two null hypotheses aim to test if the main effects of the two factors are statistically significant while the last one aims to test if the interaction effect between the two factors is statistically significant. The model (2) is not identifiable since the parameters $\boldsymbol{\mu}_0, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}_{ij}$ are not uniquely defined unless some constraints are imposed. Given a sequence of positive weights $w_{ij}, \ i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$, we impose the following constraints

$$\sum_{i=1}^{a} w_{i\cdot} \boldsymbol{\alpha}_i = 0, \ \ \sum_{j=1}^{b} w_{\cdot j} \boldsymbol{\beta}_j = 0, \tag{4}$$

$$\sum_{j=1}^{b} w_{ij} \boldsymbol{\gamma}_{ij} = 0, \ \ i = 1, 2, \cdots, a-1, \tag{5}$$

$$\sum_{i=1}^{a} w_{ij} \boldsymbol{\gamma}_{ij} = 0, \ \ j = 1, 2, \cdots, b-1, \tag{6}$$

$$\sum_{i=1}^{a} \sum_{j=1}^{b} w_{ij} \boldsymbol{\gamma}_{ij} = 0, \tag{7}$$

where $w_{i\cdot} = \sum_{j=1}^{b} w_{ij}$ and $w_{\cdot j} = \sum_{i=1}^{a} w_{ij}$. Notice that we here use only $a + b + 1$ constraints which imply the $a + b + 2$ constraints suggested by [28] and adopted by [22]. This is because the constraint (5) [resp. (6)], jointly with the constraint (7), implies that the constraint (6) [resp. (5)] holds for $j = b$ [resp. for $i = a$]. Set

$$\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T, \cdots, \boldsymbol{\alpha}_a^T]^T, \boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \cdots, \boldsymbol{\beta}_b^T]^T, \boldsymbol{\gamma} = [\boldsymbol{\gamma}_{11}^T, \cdots, \boldsymbol{\gamma}_{1b}^T, \cdots, \boldsymbol{\gamma}_{a1}^T, \cdots, \boldsymbol{\gamma}_{ab}^T]^T.$$

Then under the constraints (4)∼(7), simple algebra shows that the three null hypotheses (3) can be equivalently written as

$$\begin{aligned} H_{0A} \ &: \ [\boldsymbol{H}_a \otimes \boldsymbol{I}_p] \boldsymbol{\alpha} = 0, \ \text{with} \ \boldsymbol{H}_a = \left(\boldsymbol{I}_{a-1}, -\boldsymbol{1}_{a-1}\right), \\ H_{0B} \ &: \ [\boldsymbol{H}_b \otimes \boldsymbol{I}_p] \boldsymbol{\beta} = 0, \ \text{with} \ \boldsymbol{H}_b = \left(\boldsymbol{I}_{b-1}, -\boldsymbol{1}_{b-1}\right), \\ H_{0AB} &: [\boldsymbol{H}_{ab} \otimes \boldsymbol{I}_p] \boldsymbol{\gamma} = 0, \ \text{with} \ \boldsymbol{H}_{ab} = \left(\boldsymbol{I}_{a-1}, -\boldsymbol{1}_{a-1}\right) \otimes \left(\boldsymbol{I}_{b-1}, -\boldsymbol{1}_{b-1}\right), \end{aligned} \tag{8}$$

where and throughout, $\boldsymbol{I}_r$ and $\boldsymbol{1}_r$ denote the identity matrix of size r and the r-dimensional vector of ones, respectively, and $\otimes$ denotes the Kronecker product operation. The matrices $\boldsymbol{H}_a, \boldsymbol{H}_b$, and $\boldsymbol{H}_{ab}$ are full rank contrast matrices, having ranks $(a - 1), (b - 1)$ and $(a - 1)(b - 1)$, respectively.

When the weights can be written as $w_{ij} = u_i v_j, \ i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$, such that $u_i > 0, \sum_{i=1}^{a} u_i = 1$ and $v_j > 0, \sum_{j=1}^{b} v_j = 1$, we can easily identify the parameters $\boldsymbol{\mu}_0, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}_{ij}$ as $\boldsymbol{\mu}_0 = \sum_{i=1}^{a} \sum_{j=1}^{b} u_i v_j \boldsymbol{\mu}_{ij}, \boldsymbol{\alpha}_i = \sum_{j=1}^{b} v_j \boldsymbol{\mu}_{ij} - \boldsymbol{\mu}_0, \boldsymbol{\beta}_j = \sum_{i=1}^{a} u_i \boldsymbol{\mu}_{ij} - \boldsymbol{\mu}_0$, and $\boldsymbol{\gamma}_{ij} = \boldsymbol{\mu}_{ij} - \boldsymbol{\alpha}_i - \boldsymbol{\beta}_j - \boldsymbol{\mu}_0$. Let

$\boldsymbol{u} = [u_1, \cdots, u_a]^T, \boldsymbol{v} = [v_1, \cdots, v_b]^T$, and $\boldsymbol{\mu} = [\boldsymbol{\mu}_{11}^T, \cdots, \boldsymbol{\mu}_{1b}^T, \cdots, \boldsymbol{\mu}_{a1}^T, \cdots, \boldsymbol{\mu}_{ab}^T]^T$. Denote an l-dimensional unit vector whose r-th component is 1 and others are 0 as $\boldsymbol{e}_{r,l}$. Then we have

$$\begin{aligned}
\boldsymbol{\mu}_0 &= [(\boldsymbol{u}^T \otimes \boldsymbol{v}^T) \otimes \boldsymbol{I}_p]\boldsymbol{\mu}, & \boldsymbol{\alpha}_i &= [(\boldsymbol{e}_{i,a} - \boldsymbol{u})^T \otimes \boldsymbol{v}^T \otimes \boldsymbol{I}_p]\boldsymbol{\mu}, \\
\boldsymbol{\beta}_j &= [\boldsymbol{u}^T \otimes (\boldsymbol{e}_{j,b} - \boldsymbol{v})^T \otimes \boldsymbol{I}_p]\boldsymbol{\mu}, & \boldsymbol{\gamma}_{ij} &= [(\boldsymbol{e}_{i,a} - \boldsymbol{u})^T \otimes (\boldsymbol{e}_{j,b} - \boldsymbol{v})^T \otimes \boldsymbol{I}_p]\boldsymbol{\mu}, \\
& i = 1, 2, \cdots, a; \; j = 1, 2, \cdots, b.
\end{aligned} \tag{9}$$

In matrix notation, we can further write

$$\begin{aligned}
\boldsymbol{\alpha} &= [\boldsymbol{A}_a \otimes \boldsymbol{I}_p]\boldsymbol{\mu}, \; \text{with} \; \boldsymbol{A}_a = (\boldsymbol{I}_a - \boldsymbol{1}_a \boldsymbol{u}^T) \otimes \boldsymbol{v}^T, \\
\boldsymbol{\beta} &= [\boldsymbol{A}_b \otimes \boldsymbol{I}_p]\boldsymbol{\mu}, \; \text{with} \; \boldsymbol{A}_b = \boldsymbol{u}^T \otimes (\boldsymbol{I}_b - \boldsymbol{1}_b \boldsymbol{v}^T), \\
\boldsymbol{\gamma} &= [\boldsymbol{A}_{ab} \otimes \boldsymbol{I}_p]\boldsymbol{\mu}, \; \text{with} \; \boldsymbol{A}_{ab} = (\boldsymbol{I}_a - \boldsymbol{1}_a \boldsymbol{u}^T) \otimes (\boldsymbol{I}_b - \boldsymbol{1}_b \boldsymbol{v}^T),
\end{aligned} \tag{10}$$

where the matrices $\boldsymbol{A}_a, \boldsymbol{A}_b$, and $\boldsymbol{A}_{ab}$ are not full rank matrices, having ranks $(a-1), (b-1)$, and $(a-1)(b-1)$, respectively.

Notice that each of the testing problems associated with the three null hypotheses (8) can then be equivalently expressed in the form of the general linear hypothesis testing (GLHT) problem (12) as described in next subsection with $\boldsymbol{C}$, respectively, being

$$\boldsymbol{C}_a = (\boldsymbol{H}_a \boldsymbol{A}_a) \otimes \boldsymbol{I}_p, \;\; \boldsymbol{C}_b = (\boldsymbol{H}_b \boldsymbol{A}_b) \otimes \boldsymbol{I}_p, \;\; \boldsymbol{C}_{ab} = (\boldsymbol{H}_{ab} \boldsymbol{A}_{ab}) \otimes \boldsymbol{I}_p. \tag{11}$$

There are a few methods which can be used to specify the weights $w_{ij}, i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$; see for example, [28]. In this paper, we use the following two simple methods: the equal-weight method and the size-adapted-weight method. Both methods specify the weights as $w_{ij} = u_i v_j, i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$, with the equal-weight method specifying $\boldsymbol{u}$ and $\boldsymbol{v}$ as $u_i = 1/a, v_j = 1/b, i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$, while the size-adapted-weight method specifying $\boldsymbol{u}$ and $\boldsymbol{v}$ as $u_i = \sum_{j=1}^b n_{ij}/N, i = 1, 2, \cdots, a$, and $v_j = \sum_{i=1}^a n_{ij}/N, j = 1, 2, \cdots, b$, where $N = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$. When the two-way MANOVA design is balanced, i.e., when all the cell sizes $n_{ij}, i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$, are the same, the size-adapted-weight method reduces to the equal-weight method. In practice, both weight methods can be used when the cell sizes are near the same. However, when the cell sizes are quite different, the size-adapted-weight method is recommended so that the effect of the cell sizes can be taken into account.

2.2 Wald-type Test Statistic for Linear Hypotheses

Using the cell mean vector $\boldsymbol{\mu}$ defined in the previous subsection, we can write the GLHT problem under the two-way MANOVA model (2) as

$$H_0: \; \boldsymbol{C}\boldsymbol{\mu} = \boldsymbol{c}, \quad \text{vs} \quad H_1: \; \boldsymbol{C}\boldsymbol{\mu} \neq \boldsymbol{c}, \tag{12}$$

where $\boldsymbol{C} = \boldsymbol{C}_0 \otimes \boldsymbol{I}_p : q \times (abp)$ is a known matrix of full rank with $\text{rank}(\boldsymbol{C}_0) = q_0$ and $q = q_0 p$, and $\boldsymbol{c} : q \times 1$ is a known constant vector, often specified as $\boldsymbol{0}$. For the three testing problems (3), the associated $\boldsymbol{C}$-matrices are given in (11).

To construct the test statistic for the GLHT problem (12), we denote the usual unbiased estimators of the cell mean vectors and cell covariance matrices of the random sample (1) as

$$\hat{\boldsymbol{\mu}}_{ij} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} \boldsymbol{y}_{ijk}, \;\; \hat{\boldsymbol{\Sigma}}_{ij} = (n_{ij} - 1)^{-1} \sum_{k=1}^{n_{ij}} (\boldsymbol{y}_{ijk} - \hat{\boldsymbol{\mu}}_{ij})(\boldsymbol{y}_{ijk} - \hat{\boldsymbol{\mu}}_{ij})^T,$$
$$i = 1, 2, \cdots, a; j = 1, 2, \cdots, b. \tag{13}$$

Set $\hat{\boldsymbol{\mu}} = [\hat{\boldsymbol{\mu}}_{11}^T, \cdots, \hat{\boldsymbol{\mu}}_{1b}^T, \cdots, \hat{\boldsymbol{\mu}}_{a1}^T, \cdots, \hat{\boldsymbol{\mu}}_{ab}^T]^T$ as the estimator of $\boldsymbol{\mu}$. Then $\hat{\boldsymbol{\mu}} \sim N_{abp}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \text{diag}(\frac{\boldsymbol{\Sigma}_{11}}{n_{11}}, \frac{\boldsymbol{\Sigma}_{12}}{n_{12}}, \ldots, \frac{\boldsymbol{\Sigma}_{1b}}{n_{1b}}, \ldots, \frac{\boldsymbol{\Sigma}_{a1}}{n_{a1}}, \frac{\boldsymbol{\Sigma}_{a2}}{n_{a2}}, \ldots, \frac{\boldsymbol{\Sigma}_{ab}}{n_{ab}})$. Since $\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c} \sim N_q(\boldsymbol{C}\boldsymbol{\mu} - \boldsymbol{c}, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^T)$, the associated Wald-type test statistic is

$$T = (\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})^T \left(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T \right)^{-1} (\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c}), \tag{14}$$

where $\hat{\boldsymbol{\Sigma}} = \text{diag}(\frac{\hat{\boldsymbol{\Sigma}}_{11}}{n_{11}}, \frac{\hat{\boldsymbol{\Sigma}}_{12}}{n_{12}}, \ldots, \frac{\hat{\boldsymbol{\Sigma}}_{1b}}{n_{1b}}, \ldots, \frac{\hat{\boldsymbol{\Sigma}}_{a1}}{n_{a1}}, \frac{\hat{\boldsymbol{\Sigma}}_{a2}}{n_{a2}}, \ldots, \frac{\hat{\boldsymbol{\Sigma}}_{ab}}{n_{ab}})$. Notice that the test statistic T is affine invariant with respect to the GLHT problem (12) in the sense that for any nonsingular $q \times q$ matrix $\boldsymbol{B}$, T is invariant if the constant matrix $\boldsymbol{C}$ and the constant vector $\boldsymbol{c}$ in (12) are replaced with $\boldsymbol{BC}$ and $\boldsymbol{Bc}$, respectively.

2.3   MB Test for Heteroscedastic Two-way MANOVA Models

To construct the MB test, following [18], we re-express T as

$$T = \boldsymbol{z}^T \boldsymbol{W}^{-1} \boldsymbol{z}, \tag{15}$$

where

$$\boldsymbol{z} = (\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^T)^{-1/2}(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c}), \quad \boldsymbol{W} = (\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^T)^{-1/2}(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T)(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^T)^{-1/2}. \tag{16}$$

Notice that $\boldsymbol{z} \sim N_q(\boldsymbol{\mu}_z, \boldsymbol{I}_q)$, where $\boldsymbol{\mu}_z = (\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^T)^{-1/2}(\boldsymbol{C}\boldsymbol{\mu} - \boldsymbol{c})$, and under the null hypothesis, $\boldsymbol{z}$ is a standard q-dimensional normal random vector, i.e., $\boldsymbol{z} \sim N_q(\boldsymbol{0}, \boldsymbol{I}_q)$. Let $n_{\min} = \min_{i=1}^a \min_{j=1}^b n_{ij}$ denote the minimum cell size. To study the asymptotic distribution of T, we impose the following condition:

$$\text{As } n_{\min} \to \infty, \ \frac{n_{ij}}{n_{\min}} \to r_{ij} < \infty, \ \ i = 1, 2, \cdots, a; j = 1, 2, \cdots, b. \tag{17}$$

This condition indicates that all the cell sizes tend to infinity at about the same rate so that $n_{\min}(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^T)$ tend to a non-singular matrix as $n_{\min} \to \infty$. We then have the following result.

Theorem 1 *Under the condition (17) and H_0, as $n_{\min} \to \infty$, we have*

$$T \xrightarrow{L} \chi_q^2.$$

The above theorem shows that T asymptotically follows a χ^2-distribution with q degrees of freedom. Similar results for some special $\boldsymbol{C}$ must have appeared in the literature but we here show that it is true for the GLHT problem (12). Based on Theorem 1, a χ^2-test for the GLHT problem can be constructed easily. However, it is well known that a χ^2-test for (12) usually converges very slowly. In fact, from our proof of Theorem 1 in the Appendix, we can see that the convergence rate of T to χ_q^2 is of $O_p(n_{\min}^{-1/2})$. This indicates that the χ^2-test directly based on T's asymptotic distribution can be very inaccurate when $n_{\min}$ is small or moderate. In this case, the convergence of the mean and variance of T to those of the limit distribution χ_q^2 is also very slow. To overcome this problem, we may use the well-known Bartlett correction which corrects the mean of T to order $O(n_{\min}^{-2})$ to improve the convergence rate, but a better choice would be the modified Bartlett (MB) correction proposed by [2] which corrects both the mean and variance of T to order $O(n_{\min}^{-2})$ to improve the convergence rate further. Set $\boldsymbol{\Omega}_{ij} = n_{ij}^{-1}\boldsymbol{H}_{ij}\boldsymbol{\Sigma}_{ij}\boldsymbol{H}_{ij}^T, \ \ i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$, where $\boldsymbol{H}_{ij} = (\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^T)^{-1/2}\boldsymbol{C}_{ij}, \ \ i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$ with $\boldsymbol{C}_{11}, \boldsymbol{C}_{12}, \cdots, \boldsymbol{C}_{ab}$ being the (ab)-th column of $\boldsymbol{C}$. To apply the MB correction to T and propose the so called MB test, we need the following result.

Theorem 2 *Under the condition (17) and H_0, as $n_{\min} \to \infty$, we have*

$$E(T) = q(1 + \frac{\alpha_1}{n_{\min}}) + O(n_{\min}^{-2}) \ and \ E(T^2) = q(q+2)(1 + \frac{\alpha_2}{n_{\min}}) + O(n_{\min}^{-2}),$$

where

$$
\begin{aligned}
&\alpha_1 = n_{\min}(\Delta_1 + \Delta_2)/q, \\
&\alpha_2 = n_{\min}[(2q+8)\Delta_1 + (2q+6)\Delta_2]/[q(q+2)], \\
&\Delta_1 = \sum_{i=1}^a \sum_{j=1}^b tr(\boldsymbol{\Omega}_{ij}^2)/(n_{ij}-1), \\
&\Delta_2 = \sum_{i=1}^a \sum_{j=1}^b tr^2(\boldsymbol{\Omega}_{ij})/(n_{ij}-1).
\end{aligned}
$$

Furthermore, we have

$$\frac{q^2}{(n_{\max}-1)abp} \leq \Delta_1 \leq \frac{q}{n_{\min}-1} \ \ and \ \ \frac{q^2}{(n_{\max}-1)ab} \leq \Delta_2 < \frac{pq}{(n_{\min}-1)}. \tag{18}$$

Notice that under the conditions of Theorem 2, α_1 and α_2 will tend to their finite limits as $n_{\min} \to \infty$. Based on Theorems 1 and 2, we can apply the modified Bartlett correction to T through the log-transformation

$$T_{\text{MB}} = (n_{\min}\beta_1 + \beta_2)log(1 + \frac{T}{n_{\min}\beta_1}),$$

where $\beta_1 = \frac{2}{\alpha_2 - 2\alpha_1}$ and $\beta_2 = \frac{(q+2)\alpha_2 - 2(q+4)\alpha_1}{2(\alpha_2 - 2\alpha_1)}$.

By [2], we have $T_{MB} \xrightarrow{L} \chi_q^2$ and $\mathrm{E}(T_{\mathrm{MB}}) = q + O(n_{\min}^{-2})$ and $\mathrm{E}(T_{\mathrm{MB}}^2) = q(q+2) + O(n_{\min}^{-2})$ as $n_{\min} \to \infty$. On contrast, as seen from Theorem 2, we only have $\mathrm{E}(T) = q + O(n_{\min}^{-1})$ and $\mathrm{E}(T^2) = q(q+2) + O(n_{\min}^{-1})$. It follows from [2] that T_{MB} converges to χ_q^2 much faster than T does.

In real data analysis, β_1 and β_2 have to be estimated from the data. Proper estimators can be obtained via replacing $\boldsymbol{\Omega}_{ij}, \ i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$ with their estimators:

$$\hat{\boldsymbol{\Omega}}_{ij} = n_{ij}^{-1}(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T)^{-1/2}\boldsymbol{C}_{ij}\hat{\boldsymbol{\Sigma}}_{ij}\boldsymbol{C}_{ij}^T(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T)^{-1/2}, \ \ i = 1, 2, \cdots, a; j = 1, 2, \cdots, b. \tag{19}$$

Thus,

$$\hat{\Delta}_1 = \sum_{i=1}^{a}\sum_{j=1}^{b}\mathrm{tr}(\hat{\boldsymbol{\Omega}}_{ij}^2)/(n_{ij} - 1) \ \ \text{and} \ \ \hat{\Delta}_2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\mathrm{tr}^2(\hat{\boldsymbol{\Omega}}_{ij})/(n_{ij} - 1). \tag{20}$$

The estimators $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1$ and $\hat{\beta}_2$ are then obtained accordingly. By the law of large numbers, it is classical to show that as $n_{\min} \to \infty$, $\frac{\hat{\alpha}_1}{\alpha_1}, \frac{\hat{\alpha}_2}{\alpha_2}, \frac{\hat{\beta}_1}{\beta_1}$ and $\frac{\hat{\beta}_2}{\beta_2}$ all tend to 1 almost surely, so that as $n_{\min} \to \infty$,

$$\hat{T}_{\mathrm{MB}} = (n_{\min}\hat{\beta}_1 + \hat{\beta}_2)log(1 + \frac{T}{n_{\min}\hat{\beta}_1}) \xrightarrow{L} \chi_q^2. \tag{21}$$

Some simple algebra leads to $n_{\min}\hat{\beta}_1 = \frac{q(q+2)}{2\hat{\Delta}_1 + \hat{\Delta}_2}$ and $n_{\min}\hat{\beta}_1 + \hat{\beta}_2 = \frac{(q+2)(2q - \hat{\Delta}_2)}{4\hat{\Delta}_1 + 2\hat{\Delta}_2}$. From the proof of Theorem 2, we can see that the ranges of Δ_1 and Δ_2 as given in (18) are also the ranges of $\hat{\Delta}_1$ and $\hat{\Delta}_2$ respectively. Thus, provided $n_{\min} \geq p + 1$, we always have $n_{\min}\hat{\beta}_1 > 0$ and $n_{\min}\hat{\beta}_1 + \hat{\beta}_2 > 0$. This guarantees that $\hat{T}_{\mathrm{MB}} \geq 0$ and it is a monotonically increasing function of T. The critical value of the MB test can then be specified as $\chi_q^2(1 - \alpha)$ for any given significance level α. We reject the null hypothesis in (12) when this critical value is exceeded by $\hat{T}_{\mathrm{MB}}$. The MB test can also be conducted via computing the P-value based on χ_q^2. Thus, the MB test can be conducted easily via using the usual χ^2-table.

2.4 Some Desirable Properties of the MB Test

Notice for hypotheses (8), the contrast matrices $\boldsymbol{H}_a, \boldsymbol{H}_b$ and $\boldsymbol{H}_{ab}$ which are used to specify the main effect test and interaction effect test respectively are not unique. For example, $\tilde{\mathbf{H}}_a = (-\boldsymbol{1}_{a-1}, \boldsymbol{I}_{a-1})$ is also a contrast matrix for the first hypothesis in (8). It is known from [29] (Ch. 5, Sec. 4) that for any two contrast matrices $\tilde{\mathbf{H}}_*$ and $\boldsymbol{H}_*$ which specify the same hypothesis, there is a nonsingular matrix $\boldsymbol{P}$ such that $\tilde{\mathbf{H}}_* = \boldsymbol{P}\boldsymbol{H}_*$, where $*$ may be replaced with a, b or ab. By (11), the $\boldsymbol{C}$-matrix associated with the contrast matrix $\boldsymbol{H}_*$ can be expressed as $\boldsymbol{C}_* = (\boldsymbol{H}_*\boldsymbol{A}_*) \otimes \boldsymbol{I}_p$. Let $\tilde{\mathbf{C}}_*$ be the $\boldsymbol{C}$-matrix associated with the contrast matrix $\tilde{\mathbf{H}}_*$. Then we have $\tilde{\mathbf{C}}_* = (\tilde{\mathbf{H}}_*\boldsymbol{A}_*) \otimes \boldsymbol{I}_p = (\boldsymbol{P}\boldsymbol{H}_*\boldsymbol{A}_*) \otimes \boldsymbol{I}_p = (\boldsymbol{P} \otimes \boldsymbol{I}_p)\boldsymbol{C}_*$. Theorem 3 shows that the MB test is invariant to different choices of the contrast matrix for the same hypothesis.

Theorem 3 *The MB test is invariant when the coefficient matrix $\boldsymbol{C}$ and the constant vector $\boldsymbol{c}$ in (12) are replaced with*

$$\tilde{\mathbf{C}} = (\boldsymbol{P} \otimes \boldsymbol{I}_p)\boldsymbol{C} \ \text{and} \ \tilde{\mathbf{c}} = (\boldsymbol{P} \otimes \boldsymbol{I}_p)\boldsymbol{c}, \tag{22}$$

respectively where $\boldsymbol{P}$ is any nonsingular matrix.

The MB test is also affine-invariant. That is, it is invariant under the following affine-transformation:

$$\tilde{\mathbf{y}}_{ijk} = \boldsymbol{B}\boldsymbol{y}_{ijk} + \boldsymbol{\xi}, k = 1, 2, \cdots, n_{ij}; \ \ i = 1, 2, \cdots, a; j = 1, 2, \cdots, b, \tag{23}$$

where $\boldsymbol{B}$ is any nonsingular matrix and $\boldsymbol{\xi}$ is any given vector. This property is desirable since in practice, the observed cell responses $\boldsymbol{y}_{ij}$ (1) are often re-centered or re-scaled before an inference is conducted. The re-centering and re-scaling transformations are special cases of (23).

Theorem 4 *The MB test is invariant under the affine transformation (23).*

Finally, we have the following result.

Theorem 5 *The MB test is invariant under different labeling schemes of the cell mean vectors $\boldsymbol{\mu}_{ij}, \ i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$.*

3 Simulation Studies

It is well known ([1]) that for homogeneous data, the WLR, LHT and BNP tests are asymptotically equivalent and they perform similarly for finite samples with the LHT test outperforming the other two in many situations. [3] showed that for heteroscedastic data, the large-a asymptotics of the three tests are also equivalent. Based on some simulations, [3] further showed that their modified WLR, LHT and BNP tests also perform similarly for finite samples, with the modified LHT test, namely LHT_m, slightly outperforming the BNP test. Therefore, in this section, we only need to compare the MB test with the LHT and LHT_m tests via comparing their empirical sizes (Type I error rates) and powers for the main and interaction effects of two factors in two-way MANOVA models via simulations.

Let the two factors be A and B with a and b levels respectively. Let $\boldsymbol{n} = [n_{11}, n_{12}, \cdots, n_{ab}]$ denote the vector of cell sizes. For given $\boldsymbol{n}$ and covariance matrices $\boldsymbol{\Sigma}_{ij}$, $i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$, we first generate ab multivariate random samples as

$$\boldsymbol{y}_{ijk} = \boldsymbol{\mu}_{ij} + \boldsymbol{\Sigma}_{ij}^{1/2} \boldsymbol{\epsilon}_{ijk}, k = 1, 2, \cdots, n_{ij}, \tag{24}$$

where the cell mean vectors $\boldsymbol{\mu}_{ij} = \boldsymbol{\mu}_{11} + ij\delta\boldsymbol{h}/(ab)$ with $\boldsymbol{\mu}_{11}$ being the first cell mean vector, $\boldsymbol{h}$ a constant unit vector specifying the direction of the cell mean differences, and δ a tuning parameter controlling the amount of the cell mean differences. We independently generate the p entries of the error terms $\boldsymbol{\epsilon}_{ijk}$ using two schemes: (1) from the $N(0,1)$ distribution and (2) from the $t_4/\sqrt{2}$ distribution, so that we always have $\text{E}(\boldsymbol{\epsilon}_{ijk}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{\epsilon}_{ijk}) = \boldsymbol{I}_p$. This means that (24) will generate the (ij)-th multivariate normal or non-normal sample $\boldsymbol{y}_{ijk}, k = 1, 2, \cdots, n_{ij}$ with the given mean vector $\boldsymbol{\mu}_{ij}$ and covariance matrix $\boldsymbol{\Sigma}_{ij}$. Without loss of generality, we specify $\boldsymbol{\mu}_{11}$ as $\boldsymbol{0}$ and $\boldsymbol{h}$ as $\boldsymbol{h}_0/\|\boldsymbol{h}_0\|$ where $\boldsymbol{h}_0 = [1, 2, \cdots, p]^T$ for any given dimension p and $\|\boldsymbol{h}_0\|$ denotes the usual L^2-norm of $\boldsymbol{h}_0$. We then apply the three tests to the generated data, and record their P-values. This process is repeated $N = 10000$ times. The empirical sizes (when $\delta = 0$) and powers (when $\delta > 0$) of the three tests are the proportions of rejecting the null hypothesis, i.e., when their P-values are less than the nominal significance level α. In all the simulations conducted, we used $\alpha = 5\%$ for simplicity.

For space saving, here we just report the simulation results for interaction effect tests. Similar conclusions can be drawn from the simulation results for main effect tests. We used the equal-weight method to specify the weights of the LHT and MB tests so that their simulation results are comparable with those of the LHT_m test. The empirical sizes and powers of the three tests for interaction effect tests, together with the associated tuning parameters, are presented in Tables $1 - 3$, in the columns labeled with LHT, LHT_m, and MB under "$\delta = 0$" and "$\delta > 0$" respectively. As seen from the three tables, three sets of the tuning parameters for the cell covariance matrices are examined, with the first set specifying the homogeneous cases; four sets of the cell sizes are specified, with the first two sets specifying the balanced cell size cases; and the two error schemes are considered. To measure the overall performance of a test in terms of maintaining the nominal size α, we define the average relative error as $\text{ARE} = M^{-1}\sum_{j=1}^{M}|\hat{\alpha}_j - \alpha|/\alpha \times 100$ where $\hat{\alpha}_j$ denotes the j-th empirical size for $j = 1, 2, \cdots, M$, $\alpha = .05$ and M is the number of empirical sizes under consideration. The smaller ARE value indicates the better overall performance of the associated test. Usually, when $\text{ARE} \le 10$, the test performs very well; when $10 < \text{ARE} \le 20$, the test performs reasonably well; and when $\text{ARE} > 20$, the test does not perform well since its empirical sizes are either too liberal or too conservative and hence may be unacceptable. Notice that for a good test, the larger the cell sizes, the smaller the ARE values. The ARE values of the three tests under the two error schemes are also presented in these three tables. Notice that for simplicity, in the specification of the covariance and size tuning parameters, we often use "$(\boldsymbol{u}_r)$" to denote "$\boldsymbol{u}$ repeats r times". For simplicity and space saving, following [30], the cell covariance matrices and the cell sizes were specified the same as for the b levels of factor B but they may be different for the a levels of factor A. That is, for each $i = 1, 2, \cdots, a$, we have $\boldsymbol{\Sigma}_{ij} = \boldsymbol{\Sigma}_{i1}, n_{ij} = n_{i1}, j = 1, 2, \cdots, b$. The above method for specifying the cell covariance matrices and the cell sizes will have no effect on our methodologies and conclusions on general designs. Table 1 shows the empirical sizes and powers of the three tests for a bivariate case with $a = 2$ and $b = 20$. With $b = 20$, one may be able to check how the three tests behave when one of the factors has a large number of levels. Tables 2 and 3 show the empirical sizes and powers of the three tests for a 3-variate case with $a = 3$ and $b = 10$ and a 10-variate case with $a = 3$ and $b = 5$

Table 1. Empirical sizes and powers of the three tests for interaction effect tests for bivariate two-way MANOVA.

			$a=2,$	$b=20,$		$\boldsymbol{\Sigma}_{1j}=\boldsymbol{I}_2,$	$\boldsymbol{\Sigma}_{2j}=\mathrm{diag}(\boldsymbol{\lambda}),$		$j=1,2,\cdots,20.$			
			$\delta=0$			$\delta=1.8$			$\delta=3.6$			$\delta=5.4$	
Error	$\boldsymbol{\lambda}$	$\boldsymbol{n}$	LHT	LHT$_m$	MB	LHT	LHT$_m$	MB	LHT	LHT$_m$	MB	LHT	LHT$_m$	MB
$N(0,1)$	$\boldsymbol{\lambda}_1$	$\boldsymbol{n}_1$	.053	.036	.042	.130	.093	.093	.532	.457	.381	.948	.923	.857
		$\boldsymbol{n}_2$	.053	.041	.047	.182	.151	.148	.755	.718	.668	.996	.994	.988
		$\boldsymbol{n}_3$	.044	.030	.041	.155	.114	.122	.656	.557	.533	.983	.968	.957
		$\boldsymbol{n}_4$	.054	.049	.052	.437	.330	.402	.996	.986	.994	1.00	1.00	1.00
	$\boldsymbol{\lambda}_2$	$\boldsymbol{n}_1$	.066	.028	.042	-	.050	.078	-	.134	.235	-	.391	.621
		$\boldsymbol{n}_2$	.068	.032	.046	-	.063	.108	-	.237	.436	-	.658	.895
		$\boldsymbol{n}_3$	.039	.036	.046	-	.062	.089	-	.221	.309	-	.595	.739
		$\boldsymbol{n}_4$	.194	.035	.051	-	.095	.303	-	.480	.968	-	.958	1.00
	$\boldsymbol{\lambda}_3$	$\boldsymbol{n}_1$	.078	.027	.044	-	.038	.068	-	.078	.218	-	.222	.574
		$\boldsymbol{n}_2$	.076	.029	.045	-	.049	.104	-	.139	.403	-	.418	.864
		$\boldsymbol{n}_3$	.039	.033	.043	-	.049	.082	-	.131	.278	-	.369	.688
		$\boldsymbol{n}_4$	.253	.031	.049	-	.074	.287	-	.311	.958	-	.808	1.00
	ARE		79.2	31.4	9.02									
$t_4/\sqrt{2}$	$\boldsymbol{\lambda}_1$	$\boldsymbol{n}_1$	.047	.018	.024	.132	.062	.078	.553	.379	.464	.949	.860	.933
		$\boldsymbol{n}_2$	.050	.024	.032	.182	.112	.150	.770	.650	.762	.995	.978	.996
		$\boldsymbol{n}_3$	.049	.022	.029	.161	.076	.108	.673	.475	.624	.982	.924	.983
		$\boldsymbol{n}_4$	.051	.036	.040	.442	.289	.444	.995	.973	.997	1.00	.999	1.00
	$\boldsymbol{\lambda}_2$	$\boldsymbol{n}_1$	.064	.017	.023	-	.031	.054	-	.104	.264	-	.337	.728
		$\boldsymbol{n}_2$	.064	.021	.031	-	.046	.092	-	.202	.504	-	.595	.945
		$\boldsymbol{n}_3$	.032	.018	.027	-	.041	.075	-	.170	.354	-	.536	.830
		$\boldsymbol{n}_4$	.191	.029	.041	-	.073	.313	-	.441	.978	-	.926	1.00
	$\boldsymbol{\lambda}_3$	$\boldsymbol{n}_1$	.072	.017	.024	-	.022	.055	-	.058	.240	-	.178	.684
		$\boldsymbol{n}_2$	.071	.018	.029	-	.032	.090	-	.107	.467	-	.363	.921
		$\boldsymbol{n}_3$	.037	.022	.030	-	.037	.075	-	.105	.325	-	.324	.795
		$\boldsymbol{n}_4$	.252	.023	.038	-	.057	.305	-	.274	.968	-	.773	1.00
	ARE		75.2	55.1	37.7									

$\boldsymbol{\lambda}_1=(1,1)$, $\boldsymbol{\lambda}_2=(1,5)$, and $\boldsymbol{\lambda}_3=(1,10)$. $\boldsymbol{n}_1=(7,7)_{20}$, $\boldsymbol{n}_2=(10,10)_{20}$, $\boldsymbol{n}_3=(7,10)_{20}$, and $\boldsymbol{n}_4=(30,15)_{20}$.

respectively. These two tables allow us to compare the three tests for higher-dimensional normal and non-normal data.

First of all, let us compare the LHT and MB test via examining their empirical sizes and powers. It is seen from the three tables that under homogeneous cases, the LHT test generally outperforms the MB test under both $N(0,1)$ and $t_4/\sqrt{2}$ error schemes. Its empirical sizes are closer to 5% and powers are bigger. The MB test also performs reasonably well, but not so well as the LHT test. It is not surprising since the MB test does not take the homogeneity assumption into account while the LHT test does. On the other hand, for the heteroscedastic cases, the LHT test no longer performs well. Under the $N(0,1)$ error scheme, its empirical sizes are either too conservative or too liberal, ranging from 3.9% to 25.3% in Table 1, from 4.5% to 15.4% in Table 2 and from 0% to 75.0% in Table 3 as listed in Table 4. However, the MB test still maintains the nominal sizes quite well, with the empirical sizes ranging from 4.1% to 5.2% in Table 1, from 4.3% to 6.9% in Table 2 and from 4.6% to 7.2% in Table 3 as given in Table 4. Since the empirical sizes of the LHT and MB tests are very different, it does not make too much sense to compare their powers for the heteroscedastic cases. That is why we replaced the power values of the LHT test with "-" in these cases. Under $t_4/\sqrt{2}$ error scheme, the MB test also outperforms the LHT test with

Table 2. Empirical sizes and powers of the three tests for interaction effect tests for 3-variate two-way MANOVA.

	$a = 3,$	$b = 10,$	$\Sigma_{1j} = I_3,$		$\Sigma_{2j} = \mathrm{diag}(\lambda),$			$\Sigma_{3j} = \left(\begin{smallmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{smallmatrix}\right),$		$j = 1, 2, \cdots, 10.$		
			$\delta = 0$			$\delta = 1.8$			$\delta = 3.6$			$\delta = 5.4$	
Error	(λ, ρ)	n	LHT	LHT$_m$	MB	LHT	LHT$_m$	MB	LHT	LHT$_m$	MB	LHT	LHT$_m$	MB
$N(0,1)$	(λ_1, ρ_1)	n_1	.051	.031	.043	.138	.095	.105	.577	.490	.435	.968	.951	.906
		n_2	.050	.036	.045	.199	.165	.173	.830	.792	.760	1.00	.999	.997
		n_3	.045	.029	.060	.269	.152	.238	.945	.807	.890	1.00	.999	1.00
		n_4	.053	.035	.064	.260	.150	.238	.936	.806	.881	1.00	.999	1.00
	(λ_2, ρ_2)	n_1	.077	.020	.053	-	.054	.124	-	.316	.538	-	.868	.958
		n_2	.069	.021	.049	-	.099	.208	-	.630	.847	-	.993	1.00
		n_3	.104	.036	.066	-	.135	.320	-	.736	.976	-	.998	1.00
		n_4	.103	.037	.069	-	.134	.317	-	.710	.975	-	.997	1.00
	(λ_3, ρ_3)	n_1	.074	.025	.048	-	.059	.126	-	.319	.514	-	.844	.947
		n_2	.075	.030	.053	-	.110	.199	-	.616	.824	-	.988	.999
		n_3	.154	.037	.066	-	.143	.301	-	.743	.967	-	.998	1.00
		n_4	.093	.038	.062	-	.134	.308	-	.676	.970	-	.994	1.00
	ARE		60.7	36.5	18.6									
$t_4/\sqrt{2}$	(λ_1, ρ_1)	n_1	.049	.017	.031	.140	.061	.101	.591	.399	.528	.969	.895	.957
		n_2	.049	.022	.035	.203	.126	.176	.835	.729	.833	.998	.988	.998
		n_3	.054	.022	.045	.281	.113	.259	.942	.736	.935	1.00	.986	1.00
		n_4	.054	.020	.045	.265	.116	.243	.936	.734	.936	1.00	.986	1.00
	(λ_2, ρ_2)	n_1	.070	.012	.033	-	.035	.131	-	.262	.654	-	.798	.985
		n_2	.070	.017	.035	-	.076	.229	-	.577	.912	-	.976	1.00
		n_3	.106	.024	.050	-	.109	.357	-	.696	.989	-	.986	1.00
		n_4	.109	.026	.045	-	.102	.354	-	.668	.987	-	.980	1.00
	(λ_3, ρ_3)	n_1	.066	.012	.029	-	.039	.116	-	.264	.631	-	.773	.988
		n_2	.071	.017	.035	-	.079	.211	-	.571	.896	-	.971	1.00
		n_3	.166	.028	.052	-	.115	.328	-	.698	.982	-	.985	1.00
		n_4	.095	.025	.048	-	.110	.337	-	.636	.985	-	.975	1.00
	ARE		61.3	59.1	19.8									

$(\lambda_1, \rho_1) = (1_3, 0)$, $(\lambda_2, \rho_2) = (1, 15, .1, .1)$, and $(\lambda_3, \rho_3) = (1, 10, .1, .5)$. $n_1 = (10, 10, 10)_{10}$, $n_2 = (15, 15, 15)_{10}$, $n_3 = (10, 20, 40)_{10}$, and $n_4 = (40, 20, 10)_{10}$.

the LHT test's performance really unacceptable. This also shows that without the normality assumption, the classical LHT test is much more sensitive to the homogeneity assumption, that is why we need to propose new procedures which can work well under heteroscedastic cases.

We now compare the LHT$_m$ and MB tests via examining their empirical sizes and powers under the two error schemes. In terms of size controlling, the MB test generally outperforms the LHT$_m$ test for all the cases under consideration as shown by their empirical sizes and associated ARE values presented in the three tables. The LHT$_m$ test is generally too conservative, especially under the $t_4/\sqrt{2}$ error scheme. It is probably due to the fact that the LHT$_m$ test is not affine-invariant and its degrees of freedom can not be accurately estimated using the method proposed by [3]. In terms of power, the MB test outperforms the LHT$_m$ test for almost all the cases except for some homogeneous cases under the $N(0,1)$ scheme. The LHT$_m$ test can have higher powers in these cases probably due to the fact that the LHT$_m$ test uses the similar test statistic to the one used by the LHT test.

Table 3. Empirical sizes and powers of the three tests for interaction effect tests for 10-variate two-way MANOVA.

			$a=3,$	$b=5,$	$\boldsymbol{\Sigma}_{1j}=\boldsymbol{I}_{10},$			$\boldsymbol{\Sigma}_{2j}=\mathrm{diag}(\boldsymbol{\lambda}),$			$\boldsymbol{\Sigma}_{3j}=\mathrm{diag}(\boldsymbol{\eta}),$		$j=1,2,...,5.$	
			$\delta=0$			$\delta=2.1$			$\delta=4.2$			$\delta=6.3$		
Error	$(\boldsymbol{\lambda},\boldsymbol{\eta})$	$\boldsymbol{n}$	LHT	LHT$_m$	MB	LHT	LHT$_m$	MB	LHT	LHT$_m$	MB	LHT	LHT$_m$	MB
$N(0,1)$	$(\boldsymbol{\lambda}_1,\boldsymbol{\eta}_1)$	$\boldsymbol{n}_1$	.052	.031	.046	.162	.111	.137	.701	.617	.597	.994	.989	.977
		$\boldsymbol{n}_2$	.049	.037	.048	.334	.300	.304	.987	.983	.978	1.00	1.00	1.00
		$\boldsymbol{n}_3$	.051	.033	.056	.234	.152	.208	.916	.819	.853	1.00	1.00	1.00
		$\boldsymbol{n}_4$	.051	.038	.052	.271	.234	.251	.961	.950	.936	1.00	1.00	1.00
	$(\boldsymbol{\lambda}_2,\boldsymbol{\eta}_2)$	$\boldsymbol{n}_1$	.095	.030	.065	-	.044	.113	-	.130	.336	-	.386	.760
		$\boldsymbol{n}_2$	.086	.043	.054	-	.089	.166	-	.380	.727	-	.888	.996
		$\boldsymbol{n}_3$	.138	.036	.052	-	.064	.128	-	.202	.478	-	.590	.923
		$\boldsymbol{n}_4$	.324	.037	.054	-	.082	.153	-	.334	.658	-	.841	.989
	$(\boldsymbol{\lambda}_3,\boldsymbol{\eta}_3)$	$\boldsymbol{n}_1$	.106	.028	.072	-	.035	.088	-	.042	.120	-	.064	.204
		$\boldsymbol{n}_2$	.094	.040	.053	-	.048	.082	-	.074	.179	-	.159	.448
		$\boldsymbol{n}_3$	.000	.033	.048	-	.043	.066	-	.073	.131	-	.143	.284
		$\boldsymbol{n}_4$	.750	.036	.066	-	.041	.077	-	.057	.158	-	.107	.347
	ARE		217	29.0	14.0									
$t_4/\sqrt{2}$	$(\boldsymbol{\lambda}_1,\boldsymbol{\eta}_1)$	$\boldsymbol{n}_1$	.048	.017	.039	.159	.077	.141	.720	.565	.658	.994	.977	.988
		$\boldsymbol{n}_2$	.050	.034	.047	.346	.275	.326	.988	.980	.986	1.00	1.00	1.00
		$\boldsymbol{n}_3$	.051	.022	.048	.238	.122	.214	.924	.792	.896	1.00	.998	1.00
		$\boldsymbol{n}_4$	.052	.030	.052	.276	.203	.268	.962	.933	.954	1.00	1.00	1.00
	$(\boldsymbol{\lambda}_2,\boldsymbol{\eta}_2)$	$\boldsymbol{n}_1$	.091	.019	.054	-	.026	.102	-	.098	.362	-	.333	.817
		$\boldsymbol{n}_2$	.083	.032	.048	-	.074	.171	-	.366	.780	-	.869	.998
		$\boldsymbol{n}_3$	.142	.029	.051	-	.048	.121	-	.179	.526	-	.550	.951
		$\boldsymbol{n}_4$	.312	.027	.048	-	.063	.164	-	.305	.707	-	.809	.992
	$(\boldsymbol{\lambda}_3,\boldsymbol{\eta}_3)$	$\boldsymbol{n}_1$	.102	.021	.064	-	.022	.076	-	.030	.109	-	.046	.212
		$\boldsymbol{n}_2$	.086	.028	.047	-	.038	.077	-	.065	.187	-	.134	.468
		$\boldsymbol{n}_3$	.000	.029	.043	-	.036	.058	-	.057	.128	-	.122	.305
		$\boldsymbol{n}_4$	.759	.031	.060	-	.030	.081	-	.049	.163	-	.084	.377
	ARE		214	46.0	9.97									

$\boldsymbol{\lambda}_1=(1_{10})_5,\ \boldsymbol{\eta}_1=(1_{10})_5;\ \boldsymbol{\lambda}_2=(12_3,1_3,24_3,1)_5,\ \boldsymbol{\eta}_2=(1_3,0.1_3,2_2,24,21)_5;$ and $\boldsymbol{\lambda}_3=(1_3,3_3,9_3,20)_5,\ \boldsymbol{\eta}_3=(5_3,15_3,45_3,100)_5.$
$\boldsymbol{n}_1=(25_3)_5,\boldsymbol{n}_2=(50_3)_5,\boldsymbol{n}_3=(25,35,50)_5,$ and $\boldsymbol{n}_4=(70,40,35)_5.$

We also notice that our MB test may perform relatively worse when the cell sizes are very different from each other and when $n_{\min}$ is too small. It is probably because the MB test is based on the test statistic's asymptotic properties and Theorem 1 requires all cell sizes tend to infinity proportionally.

In summary, in terms of size controlling, overall speaking, the MB test generally outperforms the LHT and LHT$_m$ tests as shown by the ARE values listed in the three tables under the two error schemes. In terms of power, the MB test generally outperforms the LHT$_m$ test for almost all the heteroscedastic cases under consideration. Thus, one may recommend to use the MB test in real data analysis.

4   An Example

In this section, for illustration and comparison, the MB test, together with the LHT and LHT$_m$ tests, is applied to a real multivariate data set collected from a smoking cessation trial conducted by Dr. Kari J. Harris in her Greek Health Project. The project aimed to assess the efficacy of a motivational interviewing

Table 4. The empirical size ranges (in percentage) of the three tests, taken from Tables 1–3.

	LHT			LHT_m			MB		
Error	Tab1	Tab2	Tab3	Tab1	Tab2	Tab3	Tab1	Tab2	Tab3
$N(0,1)$	3.9–25.3	4.5–15.4	0–75.0	2.7–4.9	2.0–3.8	2.8–4.3	4.1–5.2	4.3–6.9	4.6–7.2
$t_4/\sqrt{2}$	3.2–25.2	4.9–16.6	0–75.9	1.7–3.6	1.2–2.8	1.7–3.4	2.3–4.1	2.9–5.2	3.9–6.4

versus an attention matched control on smoking quit rate. The subjects for the research are students from 20 individual fraternity or sorority chapters (Greek houses) of the University of Missouri-Colombia and with 2 levels (low and high) of depression. The researchers believed that the level of depression of each subject is associated with the nicotine dependence of the subject and they also wanted to know if the nicotine dependence of the subjects depended on the chapter they came from. The nicotine dependence of the subjects was measured by three well-known scales, namely, the Fagerstr´om Test for Nicotine Dependence ([31]), the Hooked on Nicotine Checklist ([32]), and the Minnesota Tobacco Withdrawal Scale ([33]). The resulting data may be referred to as the smoking cessation data. For the detailed description about the background of the smoking cessation trial and the interpretation of the variables, the reader is referred to [3] who analyzed the data using their modified WLR, LHT and BNP tests.

Table 5 shows the test results of applications of the LHT, LHT_m and MB tests to the smoking cessation data for checking the significance of the main and interaction effects of the two factors "Chapter" and "Depression". Both the equal-weight and the size-adapted-weight methods, as described in Section 2.1, were considered. Note that the P-values of the LHT and LHT_m tests were computed using the F-approximation method widely adopted in SAS and SPSS.

Table 5. P-values for the smoking cessation data.

	LHT	LHT_m	MB
Equal-weight method			
Chapter	.015	.350	.211
Depression	.000	.000	.000
Chapter× Depression	.075	.465	.601
Size-adapted-weight method			
Chapter	.124	-	.204
Depression	.000	-	.000
Chapter× Depression	.075	-	.601

We first examine the test results of the three tests under the equal-weight method. It is seen that all the three tests suggest that the main effect of "Depression" is highly significant. However, both MB and LHT_m tests suggest that the main effect of "Chapter" and the interaction effect between "Chapter" and "Depression" are not significant, while LHT test indicates that "Chapter" is significant at 5% level and the interaction is significant at 10% level. Application of the Box's M test ([7]) to the smoking cessation data suggests that the cell covariance matrices for the two-way MANOVA model are significantly different. In this case, the conclusion made by the LHT_m and MB tests is more credible than that made by the LHT test since both LHT_m and MB tests take the data heteroscedasticity into account.

The LHT_m is not defined for the size-adapted method in [3]. The test results of the MB test for the main and interaction effects of "Chapter" and "Depression" are consistent under both weight methods.

However, it is not the case for the LHT test. Actually, for the main effect of "Chapter", the conclusion made by the LHT test under the equal-weight method is opposite to the one under the size-adapted-weight method, showing some impact of the cell covariance matrices heteroscedasticity on the LHT test. It is well known ([28]) that if the cell covariance matrices were homogeneous, the test results of the LHT test would not be affected by the weight method used. Thus, the LHT test's inconsistent results under different weight methods indicate a serious impact of the cell covariance matrices heteroscedasticity on the LHT test.

References

1. T. W. Anderson, *An introduction to multivariate statistical analysis, 3rd Edition.* Wiley New York, 2003.
2. Y. Fujikoshi, "Transformations with improved chi-squared approximations," *Journal of Multivariate Analysis*, vol. 72, no. 2, pp. 249–263, 2000.
3. S. W. Harrar and A. C. Bathke, "A modified two-factor multivariate analysis of variance: asymptotics and small sample approximations," *Annals of the Institute of Statistical Mathematics*, vol. 64, no. 1, pp. 135–165, 2012.
4. R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis, 6th Edition.* Upper Saddle River, N.J. : Prentice Hall, 2008.
5. J. Xu and X. Cui, "Robustified MANOVA with applications in detecting differentially expressed genes from oligonucleotide arrays," *Bioinformatics*, vol. 24, no. 8, pp. 1056–1062, 2008.
6. C.-A. Tsai and J. J. Chen, "Multivariate analysis of variance test for gene set analysis," *Bioinformatics*, vol. 25, no. 7, pp. 897–903, 2009.
7. G. E. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, no. 3/4, pp. 317–346, 1949.
8. J.-T. Zhang, "Two-way MANOVA with unequal cell sizes and unequal cell covariance matrices," *Technometrics*, vol. 53, no. 4, pp. 426–439, 2011.
9. J.-T. Zhang and S. Xiao, "A note on the modified two-way MANOVA tests," *Statistics & Probability Letters*, vol. 82, no. 3, pp. 519–527, 2012.
10. J.-T. Zhang, "An approximate degrees of freedom test for heteroscedastic two-way ANOVA," *Journal of Statistical Planning and Inference*, vol. 142, no. 1, pp. 336–346, 2012.
11. ——, "An approximate Hotelling T^2-Test for heteroscedastic one-way MANOVA," *Open Journal of Statistics*, vol. 2, no. 1, pp. 1–11, 2012.
12. Y. Yao, "An approximate degrees of freedom solution to the multivariate Behrens-Fisher problem," *Biometrika*, vol. 52, no. 1/2, pp. 139–147, 1965.
13. S. Johansen, "The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression," *Biometrika*, vol. 67, no. 1, pp. 85–92, 1980.
14. D. Nel and C. V. D. Merwe, "A solution to the multivariate Behrens-Fisher problem," *Communications in Statistics - Theory and Methods*, vol. 15, no. 12, pp. 3719–3735, 1986.
15. S.-J. Kim, "A practical solution to the multivariate Behrens-Fisher problem," *Biometrika*, vol. 79, no. 1, pp. 171–176, 1992.
16. W. F. Christensen and A. C. Rencher, "A comparison of type I error rates and power levels for seven solutions to the multivariate Behrens-Fisher problem," *Communications in Statistics - Simulation and Computation*, vol. 26, no. 4, pp. 1251–1273, 1997.
17. K. Krishnamoorthy and J. Yu, "Modified Nel and Van der Merwe test for the multivariate Behrens-Fisher problem," *Statistics & Probability Letters*, vol. 66, no. 2, pp. 161–169, 2004.
18. H. Yanagihara and K.-H. Yuan, "Three approximate solutions to the multivariate Behrens-Fisher problem," *Communications in Statistics - Simulation and Computation*, vol. 34, no. 4, pp. 975–988, 2005.
19. A. Belloni and G. Didier, "On the Behrens-Fisher problem: A globally convergent algorithm and a finite-sample study of the Wald, LR and LM tests," *The Annals of Statistics*, vol. 36, no. 5, pp. 2377–2408, 2008.
20. Ł. Smaga, "Wald-type statistics using 2-inverses for hypothesis testing in general factorial designs," *Statistics & Probability Letters*, vol. 107, pp. 215–220, 2015.
21. R. Krutchkoff, "Two-way fixed effects analysis of variance when the error variances may be unequal," *Journal of Statistical Computation and Simulation*, vol. 32, no. 3, pp. 177–183, 1989.
22. M. M. Ananda and S. Weerahandi, "Two-way ANOVA with unequal cell frequencies and unequal variances," *Statistica Sinica*, vol. 7, no. 3, pp. 631–646, 1997.
23. P. Bao and M. M. A. Ananda, "Performance of two-way ANOVA procedures when cell frequencies and variances are unequal," *Communications in Statistics - Simulation and Computation*, vol. 30, no. 4, pp. 805–829, 2001.

24. W. Mu, S. Xiong, and X. Xu, "Generalized confidence regions of fixed effects in the two-way ANOVA," *Journal of Systems Science and Complexity*, vol. 21, no. 2, pp. 276–282, 2008.

25. L.-W. Xu, F.-Q. Yang, A. Abula, and S. Qin, "A parametric bootstrap approach for two-way ANOVA in presence of possible interactions with unequal variances," *Journal of Multivariate Analysis*, vol. 115, pp. 172–180, 2013.

26. F. Konietschke, A. C. Bathke, S. W. Harrar, and M. Pauly, "Parametric and nonparametric bootstrap methods for general MANOVA," *Journal of Multivariate Analysis*, vol. 140, pp. 291–301, 2015.

27. M. Pauly, E. Brunner, and F. Konietschke, "Asymptotic permutation tests in general factorial designs," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 77, no. 2, pp. 461–473, 2015.

28. Y. Fujikoshi, "Two-way ANOVA models with unbalanced data," *Discrete Mathematics*, vol. 116, no. 1, pp. 315–334, 1993.

29. A. M. Kshirsagar, *Multivariate Analysis.* New York: Marcel Decker, 1972.

30. R. R. Wilcox, "Adjusting for unequal variances when comparing means in one-way and two-way fixed effects ANOVA models," *Journal of Educational and Behavioral Statistics*, vol. 14, no. 3, pp. 269–278, 1989.

31. T. F. Heatherton, L. T. Kozlowski, R. C. Frecker, and K.-O. Fagerström, "The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire," *British Journal of Addiction*, vol. 86, no. 9, pp. 1119–1127, 1991.

32. R. J. Wellman, J. A. Savageau, S. Godiwala, N. Savageau, K. Friedman, J. Hazelton, and J. R. DiFranza, "A comparison of the hooked on nicotine checklist and the Fagerström test for nicotine dependence in adult smokers," *Nicotine & Tobacco Research*, vol. 8, no. 4, pp. 575–580, 2006.

33. J. R. Hughes, D. K. Hatsukami, R. W. Pickens, D. Krahn, S. Malin, and A. Luknic, "Effect of nicotine on the tobacco withdrawal syndrome," *Psychopharmacology*, vol. 83, no. 1, pp. 82–87, 1984.

34. G. Letac and H. Massam, "All invariant moments of the Wishart distribution," *Scandinavian Journal of Statistics*, vol. 31, no. 2, pp. 295–318, 2004.

APPENDIX: Technical Proofs

Proof of Theorem 1 Under the given conditions, we have

$$\hat{\boldsymbol{\Sigma}}_{ij} \sim W_q(n_{ij} - 1, \boldsymbol{\Sigma}_{ij}/(n_{ij} - 1)), \quad i = 1, 2, \cdots, a; j = 1, 2, \cdots, b, \tag{A.1}$$

where $W_q(m, \boldsymbol{V})$ denotes a q-dimensional Wishart distribution with m degree of freedom and covariance matrix $\boldsymbol{V}$. It follows that $(\hat{\boldsymbol{\Sigma}}_{ij} - \boldsymbol{\Sigma}_{ij})/n_{ij} = O_p(n_{ij}^{-3/2}), \quad i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$. Thus $\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} = O_p(n_{\min}^{-3/2})$. Notice that $\boldsymbol{\Sigma} = O(n_{\min}^{-1})$, we further have

$$\boldsymbol{R} = \boldsymbol{H}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{H}^T = O_p(n_{\min}^{-1/2}), \tag{A.2}$$

where $\boldsymbol{H}$ is defined in (16) and $\boldsymbol{H} = O(n_{\min}^{1/2})$. This implies that

$$\boldsymbol{W} = \boldsymbol{I}_q + \boldsymbol{H}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{H}^T = \boldsymbol{I}_q + \boldsymbol{R} = \boldsymbol{I}_q + O_p(n_{\min}^{-1/2}). \tag{A.3}$$

Theorem 1 holds from Slutsky's theorem and the fact that under H_0, $\boldsymbol{z}^T\boldsymbol{z} \sim \chi_q^2$.

Proof of Theorem 2 Notice that under H_0, we have $\boldsymbol{z} \sim N_q(\boldsymbol{0}, \boldsymbol{I}_q)$. Applying the conditional expectation rule, some simple algebra leads to

$$\mathrm{E}(T) = \mathrm{E}\,\mathrm{tr}(\boldsymbol{W}^{-1}) \text{ and } \mathrm{E}(T^2) = 2\mathrm{E}\,\mathrm{tr}(\boldsymbol{W}^{-2}) + \mathrm{E}\,\mathrm{tr}^2(\boldsymbol{W}^{-1}). \tag{A.4}$$

From the proof of Theorem 1, we have that $\boldsymbol{W} = \boldsymbol{I}_q + \boldsymbol{R}$ with $\boldsymbol{R} = O_p(n_{\min}^{-1/2})$; see (A.2). Then we have

$$\boldsymbol{W}^{-1} = (\boldsymbol{I}_q + \boldsymbol{R})^{-1} = \boldsymbol{I}_q - \boldsymbol{R} + \boldsymbol{R}^2 - \boldsymbol{R}^3 + O_p(n_{\min}^{-2}),$$
$$\boldsymbol{W}^{-2} = (\boldsymbol{I}_q + \boldsymbol{R})^{-2} = \boldsymbol{I}_q - 2\boldsymbol{R} + 3\boldsymbol{R}^2 - 4\boldsymbol{R}^3 + O_p(n_{\min}^{-2}).$$

It is easy to see from (A.2) that $\mathrm{E}(\boldsymbol{R}) = 0$ and $\mathrm{Etr}(\boldsymbol{R}) = 0$. Thus

$$\begin{aligned}
\mathrm{Etr}(\boldsymbol{W}^{-1}) &= q + \mathrm{Etr}(\boldsymbol{R}^2) - \mathrm{Etr}(\boldsymbol{R}^3) + O(n_{\min}^{-2}), \\
\mathrm{Etr}(\boldsymbol{W}^{-2}) &= q + 3\mathrm{Etr}(\boldsymbol{R}^2) - 4\mathrm{Etr}(\boldsymbol{R}^3) + O(n_{\min}^{-2}), \\
\mathrm{Etr}^2(\boldsymbol{W}^{-1}) &= q^2 + \mathrm{Etr}^2(\boldsymbol{R}) + 2q\mathrm{Etr}(\boldsymbol{R}^2) - 2q\mathrm{Etr}(\boldsymbol{R}^3) - 2\mathrm{Etr}(\boldsymbol{R})\mathrm{tr}(\boldsymbol{R}^2) + O(n_{\min}^{-2}).
\end{aligned} \tag{A.5}$$

To find $\text{Etr}(\boldsymbol{R}^2)$ and $\text{Etr}^2(\boldsymbol{R})$ among others, we need some results from [34]. They showed that if $\boldsymbol{Y} \sim W_q(m, \boldsymbol{V})$, then

$$
\begin{aligned}
&\text{Etr}^2[\boldsymbol{Y} - \text{E}(\boldsymbol{Y})] = 2m\text{tr}(\boldsymbol{V}^2), \quad \text{Etr}[\boldsymbol{Y} - \text{E}(\boldsymbol{Y})]^2 = m[\text{tr}(\boldsymbol{V}^2) + \text{tr}^2(\boldsymbol{V})], \\
&\text{Etr}[\boldsymbol{Y} - \text{E}(\boldsymbol{Y})]^3 = m\text{tr}^3(\boldsymbol{V}) + 3m\text{tr}(\boldsymbol{V})\text{tr}(\boldsymbol{V}^2) + 4m\text{tr}(\boldsymbol{V}^3), \\
&\text{Etr}[\boldsymbol{Y} - \text{E}(\boldsymbol{Y})]\text{tr}[\boldsymbol{Y} - \text{E}(\boldsymbol{Y})]^2 = 4m\text{tr}(\boldsymbol{V})\text{tr}(\boldsymbol{V}^2) + 4m\text{tr}(\boldsymbol{V}^3).
\end{aligned} \tag{A.6}
$$

By (A.2), $\boldsymbol{R} = \sum_{i=1}^a \sum_{j=1}^b (\boldsymbol{W}_{ij} - \boldsymbol{\Omega}_{ij}) = \sum_{i=1}^a \sum_{j=1}^b \boldsymbol{R}_{ij}$ where $\boldsymbol{R}_{ij} = \boldsymbol{W}_{ij} - \text{E}\boldsymbol{W}_{ij}$ with $\boldsymbol{W}_{ij} = n_{ij}^{-1}\boldsymbol{H}_{ij}\hat{\boldsymbol{\Sigma}}_{ij}\boldsymbol{H}_{ij}^T$, $i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$. Since $\boldsymbol{W}_{11}, \cdots, \boldsymbol{W}_{ab}$ are independent and $\text{E}\boldsymbol{R}_{ij} = 0$, $i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$, we have

$$
\begin{aligned}
\text{Etr}(\boldsymbol{R}^2) &= \sum_{i=1}^a \sum_{j=1}^b \text{Etr}(\boldsymbol{R}_{ij}^2), \\
\text{Etr}^2(\boldsymbol{R}) &= \sum_{i=1}^a \sum_{j=1}^b \text{Etr}^2(\boldsymbol{R}_{ij}), \\
\text{Etr}^3(\boldsymbol{R}) &= \sum_{i=1}^a \sum_{j=1}^b \text{Etr}^3(\boldsymbol{R}_{ij}), \\
\text{Etr}(\boldsymbol{R})\text{tr}(\boldsymbol{R}^2) &= \sum_{i=1}^a \sum_{j=1}^b \text{Etr}(\boldsymbol{R}_{ij})\text{tr}(\boldsymbol{R}_{ij}^2).
\end{aligned}
$$

By (A.1) and applying (A.6), we have

$$
\begin{aligned}
\text{Etr}(\boldsymbol{R}^2) &= \sum_{i=1}^a \sum_{j=1}^b [\text{tr}(\boldsymbol{\Omega}_{ij}^2) + \text{tr}^2(\boldsymbol{\Omega}_{ij})]/(n_{ij} - 1) = \Delta_1 + \Delta_2, \\
\text{Etr}^2(\boldsymbol{R}) &= 2\sum_{i=1}^a \sum_{j=1}^b \text{tr}(\boldsymbol{\Omega}_{ij}^2)/(n_{ij} - 1) = 2\Delta_1, \\
\text{Etr}^3(\boldsymbol{R}) &= \sum_{i=1}^a \sum_{j=1}^b [\text{tr}^3(\boldsymbol{\Omega}_{ij}) + 3\text{tr}(\boldsymbol{\Omega}_{ij})\text{tr}(\boldsymbol{\Omega}_{ij}^2) + 4\text{tr}(\boldsymbol{\Omega}_{ij}^3)]/(n_{ij} - 1)^2 = O(n_{\min}^{-2}), \\
\text{Etr}(\boldsymbol{R})\text{tr}(\boldsymbol{R}^2) &= \sum_{i=1}^a \sum_{j=1}^b [4\text{tr}(\boldsymbol{\Omega}_{ij})\text{tr}(\boldsymbol{\Omega}_{ij}^2) + 4\text{tr}(\boldsymbol{\Omega}_{ij}^3)]/(n_{ij} - 1)^2 = O(n_{\min}^{-2}),
\end{aligned} \tag{A.7}
$$

where Δ_1 and Δ_2 are as defined in Theorem 2 and we have used the fact that $0 \leq \text{tr}(\boldsymbol{\Omega}_{ij}) < q$ since $\sum_{i=1}^a \sum_{j=1}^b \text{tr}(\boldsymbol{\Omega}_{ij}) = q$. Combining (A.5) and (A.7) gives that

$$
\begin{aligned}
\text{Etr}(\boldsymbol{W}^{-1}) &= q + \Delta_1 + \Delta_2 + O(n_{\min}^{-2}), \\
\text{Etr}(\boldsymbol{W}^{-2}) &= q + 3(\Delta_1 + \Delta_2) + O(n_{\min}^{-2}), \\
\text{Etr}^2(\boldsymbol{W}^{-1}) &= q^2 + (2q + 1)\Delta_1 + 2q\Delta_2 + O(n_{\min}^{-2}).
\end{aligned} \tag{A.8}
$$

These, together with (A.4), yield that $\text{E}(T) = q\left[1 + \frac{\alpha_1}{n_{\min}}\right] + O(n_{\min}^{-2})$ and $\text{E}(T^2) = q(q + 2)\left[1 + \frac{\alpha_2}{n_{\min}}\right] + O(n_{\min}^{-2})$ where $\alpha_1 = n_{\min}(\Delta_1 + \Delta_2)/q$ and $\alpha_2 = n_{\min}[(2q + 8)\Delta_1 + (2q + 6)\Delta_2]/[q(q + 2)]$ as desired.

We now find the lower and upper bounds of Δ_1 and Δ_2 as given in (18). For $i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$, set $\boldsymbol{B}_{ij} = n_{ij}^{-1/2}\boldsymbol{H}_{ij}\boldsymbol{\Sigma}_{ij}^{1/2}$, a $q \times p$ full rank matrix so that $\boldsymbol{\Omega}_{ij} = \boldsymbol{B}_{ij}\boldsymbol{B}_{ij}^T$. It follows that $\boldsymbol{\Omega}_{ij}$'s are nonnegative, so are their eigenvalues. Notice that $\boldsymbol{\Omega}_{ij}$ and $\boldsymbol{Q}_{ij} = \boldsymbol{B}_{ij}^T\boldsymbol{B}_{ij} : p \times p$ have the same nonzero eigenvalues. Thus, $\boldsymbol{\Omega}_{ij}$ has at most p nonzero eigenvalues. Denote the largest p eigenvalues of $\boldsymbol{\Omega}_{ij}$ by $\lambda_{ij,r}, r = 1, 2, \cdots, p$ which include all the nonzero eigenvalues of $\boldsymbol{\Omega}_{ij}$. It is easy to verify that $\sum_{i=1}^a \sum_{j=1}^b \boldsymbol{\Omega}_{ij} = \boldsymbol{I}_q$. Therefore, we have $\sum_{i=1}^a \sum_{j=1}^b \text{tr}(\boldsymbol{\Omega}_{ij}) = q$ and $\boldsymbol{I}_q - \boldsymbol{\Omega}_{ij} = \sum_{s \neq i} \sum_{t \neq j} \boldsymbol{\Omega}_{st}$. Therefore $\boldsymbol{I}_q - \boldsymbol{\Omega}_{ij}$ is nonnegative, showing that the eigenvalues of $\boldsymbol{\Omega}_{ij}$ are less than 1. It follows that $\text{tr}(\boldsymbol{\Omega}_{ij}^2) = \sum_{r=1}^p \lambda_{ij,r}^2 \leq \sum_{r=1}^p \lambda_{ij,r} = \text{tr}(\boldsymbol{\Omega}_l)$ and $\text{tr}(\boldsymbol{\Omega}_{ij}) = \sum_{r=1}^p \lambda_{ij,r} \leq p$. These, together with $\sum_{i=1}^a \sum_{j=1}^b \text{tr}(\boldsymbol{\Omega}_{ij}) = q$, imply that $\Delta_1 = \sum_{i=1}^a \sum_{j=1}^b \text{tr}(\boldsymbol{\Omega}_{ij}^2)/(n_{ij} - 1) \leq q/(n_{\min} - 1)$ and $\Delta_2 = \sum_{i=1}^a \sum_{j=1}^b \text{tr}^2(\boldsymbol{\Omega}_{ij})/(n_{ij} - 1) \leq pq/(n_{\min} - 1)$.

Notice that for any nonnegative numbers $a_1, a_2, \cdots, a_m$, we always have

$$
\sum_{l=1}^m a_l^2 \geq (\sum_{l=1}^m a_l)^2/m. \tag{A.9}
$$

It follows that $\text{tr}(\boldsymbol{\Omega}_{ij}^2) = \sum_{r=1}^p \lambda_{ij,r}^2 \geq (\sum_{r=1}^p \lambda_{ij,r})^2/p = \text{tr}^2(\boldsymbol{\Omega}_{ij})/p$. So $\Delta_1 \geq p^{-1}\sum_{i=1}^a \sum_{j=1}^b \text{tr}^2(\boldsymbol{\Omega}_{ij})/(n_{ij} - 1) = \Delta_2/p$. Using (A.9) again and the fact that $\sum_{i=1}^a \sum_{j=1}^b \text{tr}(\boldsymbol{\Omega}_{ij}) = q$, we have $\Delta_2 \geq [\sum_{i=1}^a \sum_{j=1}^b \text{tr}(\boldsymbol{\Omega}_{ij})]^2/[(n_{\max} - 1)ab] = \frac{q^2}{(n_{\max}-1)ab}$. It follows that $\Delta_1 \geq \frac{q^2}{(n_{\max}-1)abp}$. The theorem is then proved.

Proof of Theorem 3 From the definition (14) of T, it is easy to see that T is invariant under the transformation (22). Then by (21), we only need to show that $\hat{\beta}_1$ and $\hat{\beta}_2$, or equivalently, $\hat{\Delta}_1$ and $\hat{\Delta}_2$ are invariant under the transformation (22).

Under (22), we have $\tilde{\mathbf{C}} = (\boldsymbol{P} \otimes \boldsymbol{I}_p)\boldsymbol{C}$. Define $\boldsymbol{C}_{11}, \cdots, \boldsymbol{C}_{ab}$ as the ab matrices of size $q \times p$ so that $\boldsymbol{C} = [\boldsymbol{C}_{11}, \boldsymbol{C}_{12}, \cdots, \boldsymbol{C}_{ab}]$. Define $\tilde{\mathbf{C}}_{11}, \cdots, \tilde{\mathbf{C}}_{ab}$ similarly so that $\tilde{\mathbf{C}} = [\tilde{\mathbf{C}}_{11}, \cdots, \tilde{\mathbf{C}}_{ab}]$. It follows that $\tilde{\mathbf{C}}_{ij} = (\boldsymbol{P} \otimes \boldsymbol{I}_p)\boldsymbol{C}_{ab}, \quad i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$. Set $\boldsymbol{G}_{ij} = \boldsymbol{C}_{ij}^T (\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T)^{-1}\boldsymbol{C}_{ij}, \quad i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$. Then it follows that $\tilde{\mathbf{G}}_{ij} = \tilde{\mathbf{C}}_{ij}^T (\tilde{\mathbf{C}}\hat{\boldsymbol{\Sigma}}\tilde{\mathbf{C}}^T)^{-1}\tilde{\mathbf{C}}_{ij} = \boldsymbol{C}_{ij}^T (\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T)^{-1}\boldsymbol{C}_{ij} = \boldsymbol{G}_{ij}, \quad i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$ are invariant under (22). By (19) and (20), we have

$$\hat{\Delta}_1 = \sum_{i=1}^{a}\sum_{j=1}^{b} \text{tr}(\hat{\boldsymbol{\Omega}}_{ij}^2)/(n_{ij} - 1) = \sum_{i=1}^{a}\sum_{j=1}^{b} \text{tr}([n_{ij}^{-1}\boldsymbol{G}_{ij}\hat{\boldsymbol{\Sigma}}_{ij}]^2)/(n_{ij} - 1)$$

and

$$\hat{\Delta}_2 = \sum_{i=1}^{a}\sum_{j=1}^{b} \text{tr}^2(\hat{\boldsymbol{\Omega}}_{ij})/(n_{ij} - 1) = \sum_{i=1}^{a}\sum_{j=1}^{b} \text{tr}^2(n_{ij}^{-1}\boldsymbol{G}_{ij}\hat{\boldsymbol{\Sigma}}_{ij})/(n_{ij} - 1),$$

showing that $\hat{\Delta}_1$ and $\hat{\Delta}_2$ are also invariant under (22). Theorem 3 is then proved.

Proof of Theorem 4 The theorem will be proved if we can show that $T, \hat{\Delta}_1$ and $\hat{\Delta}_2$ are affine invariant. Let $\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}$ and $\tilde{\boldsymbol{\mu}}_{ij}, \tilde{\boldsymbol{\Sigma}}_{ij}$ denote the mean vectors and covariance matrices of the responses $\boldsymbol{x}_{ijk}, k = 1, \cdots, n_{ij}$ and the affine-transformed responses $\tilde{\boldsymbol{x}}_{ijk}, k = 1, 2, \cdots, n_{ij}$ respectively. Then we have $\tilde{\boldsymbol{\mu}}_{ij} = \boldsymbol{B}\boldsymbol{\mu}_{ij} + \boldsymbol{\xi}$ and $\tilde{\boldsymbol{\Sigma}}_{ij} = \boldsymbol{B}\boldsymbol{\Sigma}_{ij}\boldsymbol{B}^T$. It follows that $\boldsymbol{\mu}_{ij} = \boldsymbol{B}^{-1}(\tilde{\boldsymbol{\mu}}_{ij} - \boldsymbol{b})$. As we defined the long mean vector $\boldsymbol{\mu}$ and the big covariance $\boldsymbol{\Sigma}$ in Section 2, we define $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$ similarly. Then we have $\boldsymbol{\mu} = \tilde{\mathbf{B}}^{-1}(\tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\xi}})$ and $\tilde{\boldsymbol{\Sigma}} = \tilde{\mathbf{B}}\boldsymbol{\Sigma}\tilde{\mathbf{B}}^T$ where $\tilde{\mathbf{B}} = \boldsymbol{I}_{ab} \otimes \boldsymbol{B}$ and $\tilde{\boldsymbol{\xi}} = \mathbf{1}_{ab} \otimes \boldsymbol{\xi}$. It follows that the GLHT problem (12) can be equivalently expressed as $\tilde{H}_0 : \tilde{\mathbf{C}}\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\xi}}, \quad \text{vs} \quad \tilde{H}_1 : \tilde{\mathbf{C}}\tilde{\boldsymbol{\mu}} \neq \tilde{\mathbf{b}}$, where $\tilde{\mathbf{C}} = \boldsymbol{C}\tilde{\mathbf{B}}^{-1}$ and $\tilde{\mathbf{c}} = \boldsymbol{C}\tilde{\mathbf{B}}^{-1}\tilde{\boldsymbol{\xi}} + \boldsymbol{c}$.

Since $\hat{\boldsymbol{\mu}}_{ij}$ and $\hat{\boldsymbol{\Sigma}}_{ij}$ denote the unbiased estimators of $\boldsymbol{\mu}_{ij}$ and $\boldsymbol{\Sigma}_{ij}$ for the original responses $\boldsymbol{x}_{ijk}, k = 1, 2, \cdots, n_{ij}$, we denote $\hat{\tilde{\boldsymbol{\mu}}}_{ij}$ and $\hat{\tilde{\boldsymbol{\Sigma}}}_{ij}$ as the unbiased estimators of $\tilde{\boldsymbol{\mu}}_{ij}$ and $\tilde{\boldsymbol{\Sigma}}_{ij}$ for the affine-transformed responses $\tilde{\boldsymbol{x}}_{ijk}, k = 1, 2, \cdots, n_{ij}$. Then by the affine-transformation (23), it is easy to see that $\hat{\tilde{\boldsymbol{\mu}}}_{ij} = \boldsymbol{B}\hat{\boldsymbol{\mu}}_{ij} + \boldsymbol{\xi}$ and $\hat{\tilde{\boldsymbol{\Sigma}}}_{ij} = \boldsymbol{B}\hat{\boldsymbol{\Sigma}}_{ij}\boldsymbol{B}^T$. It follows that $\hat{\tilde{\boldsymbol{\mu}}} = \tilde{\mathbf{B}}\hat{\boldsymbol{\mu}} + \tilde{\boldsymbol{\xi}}$ and $\hat{\tilde{\boldsymbol{\Sigma}}} = \tilde{\mathbf{B}}\hat{\boldsymbol{\Sigma}}\tilde{\mathbf{B}}^T$. Using the above, we have $\tilde{\mathbf{C}}\hat{\tilde{\boldsymbol{\mu}}} - \tilde{\mathbf{c}} = \boldsymbol{C}\tilde{\mathbf{B}}^{-1}(\tilde{\mathbf{B}}\hat{\boldsymbol{\mu}} + \tilde{\boldsymbol{\xi}}) - (\boldsymbol{C}\tilde{\mathbf{B}}^{-1}\tilde{\boldsymbol{\xi}} + \boldsymbol{c}) = \boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c}$ and $\tilde{\mathbf{C}}\hat{\tilde{\boldsymbol{\Sigma}}}\tilde{\mathbf{C}}^T = \boldsymbol{C}\tilde{\mathbf{B}}^{-1}\tilde{\mathbf{B}}\hat{\boldsymbol{\Sigma}}\tilde{\mathbf{B}}^T(\boldsymbol{C}\tilde{\mathbf{B}}^{-1})^T = \boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T$. Thus, both $\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c}$ and $\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T$ are affine-invariant. It follows that T in (14) is affine-invariant.

We now turn to show that $\hat{\Delta}_1$ and $\hat{\Delta}_2$ are affine-invariant. It is sufficient to show that $\text{tr}(\hat{\boldsymbol{\Omega}}_{ij})$ and $\text{tr}(\hat{\boldsymbol{\Omega}}_{ij}^2)$ are affine-invariant. Since we have showed that $\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T$ is affine-invariant, we only need to show that $n_{ij}^{-1}\boldsymbol{C}_{ij}\hat{\boldsymbol{\Sigma}}_{ij}\boldsymbol{C}_{ij}^T, \quad i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$ are affine-invariant. This is obvious since $\tilde{\mathbf{C}} = \boldsymbol{C}\tilde{\mathbf{B}}^{-1}$ implies $\tilde{\mathbf{C}}_{ij} = \boldsymbol{C}_{ij}\boldsymbol{B}^{-1}, \quad i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$ and $\hat{\tilde{\boldsymbol{\Sigma}}} = \tilde{\mathbf{B}}\hat{\boldsymbol{\Sigma}}\tilde{\mathbf{B}}^T$ implies $\hat{\tilde{\boldsymbol{\Sigma}}}_{ij} = \boldsymbol{B}\hat{\boldsymbol{\Sigma}}_{ij}\boldsymbol{B}^T, \quad i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$. The theorem is then proved.

Proof of Theorem 5 To show this theorem, it is sufficient to show that $T, \hat{\Delta}_1$ and $\hat{\Delta}_2$ are invariant under different labeling schemes of the mean vectors $\boldsymbol{\mu}_{ij}, \quad i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$. Let $l_1, l_2, \cdots, l_a$ be any permutation of $1, 2, \cdots, a$, and $r_1, r_2, \cdots, r_b$ be any permutation of $1, 2, \cdots, b$. Then it is easy to see that $\sum_{i=1}^{a}\sum_{j=1}^{b}\boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_{ij} = \sum_{u=1}^{a}\sum_{v=1}^{b}\boldsymbol{C}_{l_u r_v}\hat{\boldsymbol{\mu}}_{l_u r_v}$, and $\sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}^{-1}\boldsymbol{C}_{ij}\hat{\boldsymbol{\Sigma}}_{ij}\boldsymbol{C}_{ij}^T = \sum_{u=1}^{a}\sum_{v=1}^{b}n_{l_u r_v}^{-1}\boldsymbol{C}_{l_u r_v}\hat{\boldsymbol{\Sigma}}_{l_u r_v}\boldsymbol{C}_{l_u r_v}^T$, showing that

$$\boldsymbol{C}\hat{\boldsymbol{\mu}} = \sum_{i=1}^{a}\sum_{j=1}^{b}\boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_{ij} \text{ and } \boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T = \sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}^{-1}\boldsymbol{C}_{ij}\hat{\boldsymbol{\Sigma}}_{ij}\boldsymbol{C}_{ij}^T$$

are invariant under different labeling schemes of the mean vectors and so is T.

We now show that $\hat{\Delta}_1$ and $\hat{\Delta}_2$ are invariant under different labeling schemes of the mean vectors. Set $S_{ij} = n_{ij}^{-1}\boldsymbol{C}_{ij}\hat{\boldsymbol{\Sigma}}_{ij}\boldsymbol{C}_{ij}^T, \quad i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$ and $\boldsymbol{S} = \boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^T$. By (20), we have $\hat{\Delta}_1 = \sum_{i=1}^{a}\sum_{j=1}^{b}\text{tr}([\boldsymbol{S}_{ij}\boldsymbol{S}^{-1}]^2)$ and $\hat{\Delta}_2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\text{tr}^2(\boldsymbol{S}_{ij}\boldsymbol{S}^{-1})$. Since $\boldsymbol{S}$ is previously shown to be invariant under different labeling schemes of the mean vectors, so are $\hat{\Delta}_1$ and $\hat{\Delta}_2$. This completes the proof of the theorem.