Modified Tests for Heteroscedastic Two-Way MANOVA

Shengning Xiao[1][*] and Jin-Ting Zhang[2]

[1]Department of Mathematics, School of Science, Tianjin University, China
Email: `xiaosn@tju.edu.cn`
[2]Department of Statistics and Applied Probability, National University of Singapore, Singapore

Abstract In this article we consider testing for main and interaction effects in heteroscedastic two-way MANOVA model. We express the model in the form of the general linear hypothesis testing (GLHT) problem and construct the sum of squares and cross products (SSCP) matrices due to hypothesis and error respectively. We modify the classical Wilks's Likelihood Ratio (WLR), Lawley-Hotelling Trace (LHT) and Bartlett-Nanda-Pillai trace (BNP) tests based on these two SSCP matrices with the approximate degrees of freedoms (ADF) obtained by matching the means and the total variations of the SSCP matrices and their respective approximating Wishart distributions. The resulting modified WLR, LHT, BNP tests are shown to be invariant under affine-transformations, different choices of the contrast matrix used to define the same hypothesis, and different labeling schemes of the cell mean vectors. Simulation studies presented in this paper also show that the proposed tests generally perform well and outperform one existing approach in terms of controlling the desired size and enhancing the powers. An example from a Smoking Cessation trial is given to illustrate the proposed methodologies.

Keywords: Two-way MANOVA, Heteroscedastic, General linear hypothesis test, Unbalanced, Wishart-approximation, Affine-invariant, Main-effect, Interaction-effect

1 Introduction

Two-way Multivariate Analysis of Variance (MANOVA) is an important component in multivariate data analysis. It is a procedure used to test the statistical significance of the effect of two categorical independent variables on a set of two or more continuous dependent variables. Motivations for two-way MANOVA in agriculture, biology, physics and other disciplines can be found in [1], [2], [3], [4], and among others.

Classical MANOVA models assume that the covariance matrices are homogeneous across experimental cells. This homogeneity assumption can be tested by the well-known Box's M test [5] which is similar to Levene's test [6] for ANOVA. When the homogeneity assumption is satisfied, four well-known multivariate tests: Wilks's Likelihood Ratio (WLR), Lawley-Hotelling Trace (LHT), Bartlett-Nanda-Pillai trace (BNP) and Roy Maximum Root, can be used for two-way MANOVA [7].

When there is a small departure from the homogeneity assumption but the numbers of observations in each experimental cell are equal, that is, the MANOVA model is balanced, the aforementioned tests can still be used, only with a very slight reduction in statistical power, see [8]. When the homogeneity assumption is seriously violated, none of these four tests are robust, their sizes will be either inflated or deflated greatly, which may lead to serious inappropriate conclusions. Hence, these four tests may not be directly used for heteroscedastic two-way MANOVA. However, in real data analysis, researchers often apply these classical MANOVA testing procedures blindly for the sake of simplicity even though they know that the homogeneity assumption may be violated. This is probably due to the fact that no

simple and efficient testing procedures are available in the literature which are robust for heteroscedastic two-way MANOVA.

Since the asymptotic theory for heteroscedastic two-way MANOVA tends to be much more involved, to our best knowledge, not much related research work has been carried out. [9] may be the first literature for solving this problem. They suggested modifications to the usual sum of squares and cross products (SSCP) matrices to define their modified WLR, LHT and BNP tests for heteroscedastic two-way MANOVA. The approximate degrees of freedom (ADF) of these modified MANOVA tests are estimated from the data. To approximate the null distributions of these modified MANOVA tests, [9] used some χ^2 and normal-based asymptotic expansions. Though simple to understand, their modified MANOVA tests admit several drawbacks: (1) the modified MANOVA tests are too conservative and less powerful, as indicated by the simulation results presented in Section 3; (2) very complex matrix operations and tedious calculations are involved in the derivation of the ADF formulas so that the resulting ADF formulas take very complicated forms and are difficult to be used; (3) the resulting modified MANOVA tests are not affine-invariant; and (4) [9] failed to introduce a weight scheme for defining the main and interaction effects in their modified MANOVA tests. Recently, the first three drawbacks have been addressed by [10] which proposed better formulas for computing the approximate degrees of freedom of the modified MANOVA tests and a method to make the modified MANOVA tests affine-invariant. Alternatively, [11] proposed an approximate Hotelling T^2-test for heteroscedastic two-way MANOVA.

In this article, we extend the modified MANOVA tests of [9] and [10] to test the so-called general linear hypothesis testing (GLHT) problem in heteroscedastic two-way MANOVA. All related tests under heteroscedastic two-way MANOVA, e.g., the main-effect, interaction-effect, post hoc, and contrast tests among others, can be unified under the framework of the GLHT problem. We first define two SSCP matrices, respectively due to hypothesis and error, and then define the associated modified WLR, LHT and BNP tests based on these two SSCP matrices with their approximate degrees of freedom estimated from the data. We show that the resulting modified WLR, LHT, BNP tests are invariant under affine-transformations, different choices of the contrast matrix used to define the same hypothesis, and different labeling schemes of the cell mean vectors. Unlike in the modified MANOVA tests of [9] and [10], a weight scheme for defining the main and interaction effects is naturally incorporated in our new modified MANOVA tests. Furthermore, the formulas for computing the approximate degrees of freedom of the SSCP matrices are easy to compute. Simulated results reported in Section 3 show that the proposed modified MANOVA tests perform very well and outperform those of [9]'s in terms of controlling the desired sizes and enhancing the powers.

This article is organized as follows. In Section 2, the methodologies for the new modified MANOVA tests are presented. Simulation studies are given in Section 3. In Section 4, the methodologies proposed in this article are illustrated using a real data example. Finally, some technical proofs of the main results are given in the Appendix.

2 Methodologies

2.1 Tests of Main and Interaction Effects

Consider the two-way MANOVA model with factors A and B, each with levels a and b, respectively, creating a design of ab treatment combinations or cells. Suppose the number of subjects in the (i, j)th cell is $n_{ij}, i = 1, \cdots, a; j = 1, \cdots, b$, and we measure p variables on each subject. Let $\boldsymbol{y}_{ijk}, k = 1, \cdots, n_{ij}$ be the observations associated with the (i, j)th cell, satisfying

$$\boldsymbol{y}_{ijk} = \boldsymbol{\mu}_{ij} + \boldsymbol{\epsilon}_{ijk}, \ \boldsymbol{\epsilon}_{ijk} \overset{i.i.d.}{\sim} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{ij}), k = 1, 2, \cdots, n_{ij}, \tag{2.1}$$

where $\boldsymbol{\mu}_{ij}$ is the mean vector of the (i, j)th cell and there is no knowledge about the equality of the covariance matrices $\boldsymbol{\Sigma}_{ij}, i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$. The cell mean vectors $\boldsymbol{\mu}_{ij}$ are often 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 main-effects due to the ith level of factor A and the jth level of factor B, $\boldsymbol{\gamma}_{ij}$ is the interaction-effect between the ith level of factor A and the jth level of factor B so that the model (2.1)

can be written into the following heteroscedastic two-way MANOVA model:

$$
\begin{aligned}
\boldsymbol{y}_{ijk} &= \boldsymbol{\mu}_0 + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij} + \boldsymbol{\epsilon}_{ijk}, \\
\boldsymbol{\epsilon}_{ijk} &\sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{ij}), \; ; \; k = 1, \cdots, n_{ij}; \; i = 1, \cdots, a, \; j = 1, \cdots, b.
\end{aligned}
\tag{2.2}
$$

For this heteroscedastic two-way MANOVA model, we are interested in testing the following hypotheses:

$$
\begin{aligned}
\mathcal{H}_0^{(A)} &: \; \boldsymbol{\alpha}_i = 0, \; i = 1, \cdots, a. \quad \text{(no main-effects of factor A),} \\
\mathcal{H}_0^{(B)} &: \; \boldsymbol{\beta}_j = 0, \; j = 1, \cdots, b. \quad \text{(no main-effects of factor B),} \\
\mathcal{H}_0^{(AB)} &: \; \boldsymbol{\gamma}_{ij} = 0, \; i = 1, \cdots, a; \; j = 1, \cdots, b. \quad \text{(no interaction-effects).}
\end{aligned}
\tag{2.3}
$$

Note that the model (2.2) is not identifiable since the parameters $\boldsymbol{\mu}_0, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}_{ij}$ can not be uniquely defined unless some side constraints are imposed. Given a sequence of positive weights $w_{ij}, i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$, following [11], we impose the following side constraints:

$$
\begin{aligned}
\sum_{i=1}^a w_{i\cdot} \boldsymbol{\alpha}_i = 0, \quad &\sum_{i=1}^a w_{ij} \boldsymbol{\gamma}_{ij} = 0, \quad j = 1, 2, \cdots, b-1, \\
\sum_{j=1}^b w_{\cdot j} \boldsymbol{\beta}_j = 0, \quad &\sum_{j=1}^b w_{ij} \boldsymbol{\gamma}_{ij} = 0, \quad i = 1, 2, \cdots, a-1, \\
&\sum_{i=1}^a \sum_{j=1}^b \boldsymbol{\gamma}_{ij} = 0,
\end{aligned}
\tag{2.4}
$$

where $w_{i\cdot} = \sum_{j=1}^b w_{ij}$ and $w_{\cdot j} = \sum_{i=1}^a w_{ij}$.

Remark 1. There are several methods which can be used to specify the weights $w_{ij}, i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$; see for example, [12]. Following [11], 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.

Set $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_a]^T, \boldsymbol{\beta} = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_b]^T, \boldsymbol{\gamma} = [\boldsymbol{\gamma}_{11}, \cdots, \boldsymbol{\gamma}_{1b}, \cdots, \boldsymbol{\gamma}_{a1}, \cdots, \boldsymbol{\gamma}_{ab}]^T$. Notice that the effect matrices $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are of sizes $a \times p, b \times p$, and $(ab) \times p$ respectively. Then under the side conditions (2.4), the three null hypotheses in (2.3) can be further written as

$$
\begin{aligned}
\mathcal{H}_0^{(A)} &: \; \boldsymbol{S}_a \boldsymbol{\alpha} = 0, \; \text{with} \; \boldsymbol{S}_a = \left(\boldsymbol{I}_{a-1}, -\boldsymbol{1}_{a-1} \right), \\
\mathcal{H}_0^{(B)} &: \; \boldsymbol{S}_b \boldsymbol{\beta} = 0, \; \text{with} \; \boldsymbol{S}_b = \left(\boldsymbol{I}_{b-1}, -\boldsymbol{1}_{b-1} \right), \\
\mathcal{H}_0^{(AB)} &: \; \boldsymbol{S}_{ab} \boldsymbol{\gamma} = 0, \; \text{with} \; \boldsymbol{S}_{ab} = \left(\boldsymbol{I}_{a-1}, -\boldsymbol{1}_{a-1} \right) \otimes \left(\boldsymbol{I}_{b-1}, -\boldsymbol{1}_{b-1} \right),
\end{aligned}
\tag{2.5}
$$

where $\boldsymbol{I}_r$ 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 $\boldsymbol{S}_a, \boldsymbol{S}_b$, and $\boldsymbol{S}_{ab}$ are full rank contrast matrices whose rows sum up to 0, 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 $v_j > 0, \sum_{j=1}^b v_j = 1$ as in the equal-weight method and the size-adapted weight method mentioned above, 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 $\boldsymbol{M} = [\boldsymbol{\mu}_{11}, \cdots, \boldsymbol{\mu}_{1b}, \cdots, \boldsymbol{\mu}_{a1}, \cdots, \boldsymbol{\mu}_{ab}]^T$. Note that $\boldsymbol{M}$ is a matrix of size $(ab) \times p$ whose rows are the ab cell mean vectors. Denote an l-dimensional unit vector whose rth component is 1 and others are 0 as $\boldsymbol{e}_{r,l}$. Then we have

$$
\begin{aligned}
\boldsymbol{\mu}_0^T &= (\boldsymbol{u}^T \otimes \boldsymbol{v}^T) \boldsymbol{M}, & \boldsymbol{\alpha}_i^T &= [(\boldsymbol{e}_{i,a} - \boldsymbol{u})^T \otimes \boldsymbol{v}^T] \boldsymbol{M}, \\
\boldsymbol{\beta}_j^T &= [\boldsymbol{u}^T \otimes (\boldsymbol{e}_{j,b} - \boldsymbol{v})^T] \boldsymbol{M}, & \boldsymbol{\gamma}_{ij}^T &= [(\boldsymbol{e}_{i,a} - \boldsymbol{u})^T \otimes (\boldsymbol{e}_{j,b} - \boldsymbol{v})^T] \boldsymbol{M}, \\
i &= 1, 2, \cdots, a; \; j = 1, 2, \cdots, b.
\end{aligned}
\tag{2.6}
$$

It follows that the effect matrices $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be further written as

$$
\begin{aligned}
\boldsymbol{\alpha} &= \boldsymbol{A}_a \boldsymbol{M}, \ \text{with} \ \boldsymbol{A}_a = (\boldsymbol{I}_a - \boldsymbol{1}_a \boldsymbol{u}^T) \otimes \boldsymbol{v}^T, \\
\boldsymbol{\beta} &= \boldsymbol{A}_b \boldsymbol{M}, \ \text{with} \ \boldsymbol{A}_b = \boldsymbol{u}^T \otimes (\boldsymbol{I}_b - \boldsymbol{1}_b \boldsymbol{v}^T), \\
\boldsymbol{\gamma} &= \boldsymbol{A}_{ab} \boldsymbol{M}, \ \text{with} \ \boldsymbol{A}_{ab} = (\boldsymbol{I}_a - \boldsymbol{1}_a \boldsymbol{u}^T) \otimes (\boldsymbol{I}_b - \boldsymbol{1}_b \boldsymbol{v}^T),
\end{aligned}
\tag{2.7}
$$

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. It follows that each of the testing problems associated with the three null hypotheses (2.5) can then be equivalently expressed in the form of the GLHT problem (2.9) as defined in the next subsection with $\boldsymbol{C}_0 = \boldsymbol{0}$ and $\boldsymbol{C}$, respectively, being

$$
\begin{aligned}
\boldsymbol{C}_a &= \boldsymbol{S}_a \boldsymbol{A}_a = (\boldsymbol{I}_{a-1}, -\boldsymbol{1}_{a-1}) \otimes \boldsymbol{v}^T, \\
\boldsymbol{C}_b &= \boldsymbol{S}_b \boldsymbol{A}_b = \boldsymbol{u}^T \otimes (\boldsymbol{I}_{b-1}, -\boldsymbol{1}_{b-1}), \\
\boldsymbol{C}_{ab} &= \boldsymbol{S}_{ab} \boldsymbol{A}_{ab} = \left(\boldsymbol{I}_{a-1}, -\boldsymbol{1}_{a-1}\right) \otimes \left(\boldsymbol{I}_{b-1}, -\boldsymbol{1}_{b-1}\right).
\end{aligned}
\tag{2.8}
$$

Remark 2. It is interesting to notice from (2.8) that $\boldsymbol{C}_a$ does not depend on the weight vector $\boldsymbol{u}$ of the levels of factor A, $\boldsymbol{C}_b$ does not depend on the weight vector $\boldsymbol{v}$ of the levels of factor B, and $\boldsymbol{C}_{ab}$ does not depend on either of the weight vectors $\boldsymbol{u}$ and $\boldsymbol{v}$.

2.2   Modified MANOVA Tests for the GLHT Problem

Using the cell mean matrix $\boldsymbol{M}$ defined in the previous subsection, we can write a GLHT problem as:

$$
\mathcal{H}_0: \ \boldsymbol{CM} = \boldsymbol{C}_0, \quad \text{vs} \quad \mathcal{H}_1: \ \boldsymbol{CM} \neq \boldsymbol{C}_0,
\tag{2.9}
$$

where $\boldsymbol{C}: q \times (ab)$ is a known full rank matrix with $\text{rank}(\boldsymbol{C}) = q$, and $\boldsymbol{C}_0: q \times p$ is a known constant matrix, often specified as $\boldsymbol{0}$. For the three null hypotheses in (2.3), $\boldsymbol{C}_0 = \boldsymbol{0}$ and the associated $\boldsymbol{C}$-matrices are given in (2.8).

The GLHT problem (2.9) is very general. It includes not only the main and interaction effect tests but also various post hoc and contrast tests as special cases since any post hoc and contrast tests can be written in the form of (2.9). For example, when the main-effect test, as given in (2.3), of factor A in the two-way MANOVA model is rejected, it is of interest to further test, e.g., if $\boldsymbol{\alpha}_i = 5\boldsymbol{\alpha}_j$ or if a contrast $\boldsymbol{\alpha}_i - 4\boldsymbol{\alpha}_j + 3\boldsymbol{\alpha}_k = 0$. It is easy to check that these two testing problems can be written in the form of (2.9) with $\boldsymbol{C} = [\boldsymbol{e}_{i,a} - 5\boldsymbol{e}_{j,a}]^T \boldsymbol{A}_a$ and $\boldsymbol{C} = [\boldsymbol{e}_{i,a} - 4\boldsymbol{e}_{j,a} + 3\boldsymbol{e}_{k,a}]^T \boldsymbol{A}_a$ respectively and with $\boldsymbol{C}_0 = \boldsymbol{0}$.

For further development, denote the usual unbiased estimators of the cell mean vectors and cell covariance matrices as:

$$
\begin{aligned}
\hat{\boldsymbol{\mu}}_{ij} &= n_{ij}^{-1} \sum_{k=1}^{n_{ij}} \boldsymbol{y}_{ijk}, \quad \hat{\boldsymbol{\Sigma}}_{ij} = (n_{ij} - 1)^{-1} \sum_{k=1}^{n_{ij}} (\boldsymbol{y}_{ijk} - \hat{\boldsymbol{\mu}}_{ij})(\boldsymbol{y}_{ijk} - \hat{\boldsymbol{\mu}}_{ij})^T, \\
& i = 1, 2, \cdots, a; j = 1, 2, \cdots, b.
\end{aligned}
\tag{2.10}
$$

Set $\hat{\boldsymbol{M}} = [\hat{\boldsymbol{\mu}}_{11}, \cdots, \hat{\boldsymbol{\mu}}_{1b}, \cdots, \hat{\boldsymbol{\mu}}_{a1}, \cdots, \hat{\boldsymbol{\mu}}_{ab}]^T$ and $\boldsymbol{D} = \text{diag}(\frac{1}{n_{11}}, \cdots, \frac{1}{n_{1b}}, \cdots, \frac{1}{n_{a1}}, \cdots, \frac{1}{n_{ab}})$. Notice that $\hat{\boldsymbol{M}}$ is a random matrix whose rows are independent with each other. For testing (2.9), we construct the SSCP matrix due to hypothesis as

$$
\boldsymbol{H} = (\boldsymbol{C}\hat{\boldsymbol{M}} - \boldsymbol{C}_0)^T \left(\boldsymbol{CDC}^T\right)^{-1} (\boldsymbol{C}\hat{\boldsymbol{M}} - \boldsymbol{C}_0).
\tag{2.11}
$$

Notice that under $\mathcal{H}_0: \boldsymbol{CM} = \boldsymbol{C}_0$, the SSCP matrix $\boldsymbol{H}$ can be further expressed as

$$
\boldsymbol{H} = (\boldsymbol{C}\hat{\boldsymbol{M}} - \boldsymbol{CM})^T \left(\boldsymbol{CDC}^T\right)^{-1} (\boldsymbol{C}\hat{\boldsymbol{M}} - \boldsymbol{CM}) = \mathcal{E}^T \boldsymbol{W} \mathcal{E},
\tag{2.12}
$$

where $\boldsymbol{W} = \boldsymbol{C}^T \left(\boldsymbol{CDC}^T\right)^{-1} \boldsymbol{C} = (w_{ij,\alpha\beta}) : (ab) \times (ab)$ with the entries of $\boldsymbol{W}$ labeled by two-digit subscripts to be associated with the subscripts of the cell mean vectors $\boldsymbol{\mu}_{ij}, i = 1, \cdots, a; j = 1, \cdots, b$, and

$\mathcal{E} = \hat{\boldsymbol{M}} - \boldsymbol{M} = [\bar{\boldsymbol{\epsilon}}_{11.}, \cdots, \bar{\boldsymbol{\epsilon}}_{1b.}, \cdots, \bar{\boldsymbol{\epsilon}}_{a1.}, \cdots, \bar{\boldsymbol{\epsilon}}_{ab.}]^T$ with $\bar{\boldsymbol{\epsilon}}_{ij.} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} \boldsymbol{\epsilon}_{ijk}$. Notice that $\mathcal{E}$ is a random matrix whose rows are independent with each other and

$$\bar{\boldsymbol{\epsilon}}_{ij.} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{ij}/n_{ij}), i = 1, \cdots, a; j = 1, \cdots, b. \tag{2.13}$$

Let $\boldsymbol{\Omega}$ denote the expectation of $\boldsymbol{H}$ under the null hypothesis. In this case, applying Lemma 1 in the Appendix to $\boldsymbol{H}$ in the form of (2.12) yields that

$$\boldsymbol{\Omega} = \mathrm{E}(\boldsymbol{H}) = \sum_{i=1}^{a} \sum_{j=1}^{b} w_{ij,ij} \boldsymbol{\Sigma}_{ij}/n_{ij}. \tag{2.14}$$

It is easy to see that a natural unbiased estimator of $\boldsymbol{\Omega}$ is given by

$$\boldsymbol{G} = \hat{\boldsymbol{\Omega}} = \sum_{i=1}^{a} \sum_{j=1}^{b} w_{ij,ij} \hat{\boldsymbol{\Sigma}}_{ij}/n_{ij}, \tag{2.15}$$

where the sample covariance matrices $\hat{\boldsymbol{\Sigma}}_{ij}$ are given in (2.10). It is natural to count $\boldsymbol{G}$ as the SSCP matrix due to error. Since the sample cell mean vectors $\hat{\boldsymbol{\mu}}_{ij}, i = 1, \cdots, a; j = 1, \cdots, b$ are independent of the sample cell covariance matrices $\hat{\boldsymbol{\Sigma}}_{ij}, i = 1, \cdots, a; j = 1, \cdots, b$, the random matrices $\boldsymbol{H}$ and $\boldsymbol{G}$ are independent. Let $W_p(r, \boldsymbol{V})$ denote a Wishart-distribution with r degrees of freedom and covariance matrix $\boldsymbol{V}$. When the covariance homogeneity assumption holds, we have

$$\boldsymbol{\Sigma}_{ij} = \boldsymbol{\Sigma}, i = 1, \cdots, a; j = 1, \cdots, b, \tag{2.16}$$

where $\boldsymbol{\Sigma}$ is the common cell covariance matrix, which can be unbiasedly and effectively estimated by the following pooled sample covariance matrix

$$\hat{\boldsymbol{\Sigma}} = (N - ab)^{-1} \sum_{i=1}^{a} \sum_{j=1}^{b} (n_{ij} - 1) \hat{\boldsymbol{\Sigma}}_{ij}. \tag{2.17}$$

Remark 3. Under the homogeneity assumption (2.16) and the null hypothesis in (2.9), it is easy to show that

$$\boldsymbol{H} \sim W_p(q, \boldsymbol{\Omega}/q), \quad \boldsymbol{G} \sim W_p(N - ab, \boldsymbol{\Omega}/(N - ab)), \tag{2.18}$$

after the sample covariance matrices $\hat{\boldsymbol{\Sigma}}_{ij}$ in the expression of $\boldsymbol{G}$ are replaced with the pooled sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ (2.17) where $\boldsymbol{\Omega} = q\boldsymbol{\Sigma}$. In this case, set $\boldsymbol{R}_1 = q\boldsymbol{H}$ and $\boldsymbol{R}_2 = (N - ab)\boldsymbol{G}$. Then

$$\boldsymbol{R}_1 \sim W_p(q, \boldsymbol{\Omega}), \text{ and } \boldsymbol{R}_2 \sim W_p(N - ab, \boldsymbol{\Omega}), \tag{2.19}$$

so that one can easily use R_1 and R_2 to construct the classical WLR, LHT and BNP test statistics as in (2.22).

However, when the homogeneity assumption (2.16) is seriously violated, the distribution expressions in (2.18) are no longer valid. Nevertheless, we expect that the expressions in (2.18) hold approximately, provided that the associated degrees of freedom are adjusted properly. In other words, the distributions of $\boldsymbol{H}$ and $\boldsymbol{G}$ may be well approximated respectively by the distributions of the following Wishart random matrices:

$$\boldsymbol{H}_a \sim W_p(f_{\mathrm{H}}, \boldsymbol{\Omega}/f_{\mathrm{H}}) \text{ and } \boldsymbol{G}_a \sim W_p(f_{\mathrm{G}}, \boldsymbol{\Omega}/f_{\mathrm{G}}), \tag{2.20}$$

where f_{H} and f_{G} are the approximate degrees of freedom for $\boldsymbol{H}$ and $\boldsymbol{G}$ respectively. The approximate degrees of freedom f_{H} and f_{G} can be determined via matching the total variations of $\boldsymbol{H}$ and $\boldsymbol{H}_a$, and those of $\boldsymbol{G}$ and $\boldsymbol{G}_a$, respectively. As pointed out in [10], the resulting f_{H} and f_{G} are not affine-invariant so that the modified MANOVA tests based on $\boldsymbol{H}, \boldsymbol{G}, f_{\mathrm{H}}$, and f_{G}, are not affine-invariant as in [9].

To overcome this problem, notice that when (2.20) holds, we have $\tilde{\boldsymbol{H}}_a = \boldsymbol{\Omega}^{-1/2} \boldsymbol{H}_a \boldsymbol{\Omega}^{-1/2} \sim W_p(f_{\mathrm{H}}, \boldsymbol{I}_p/f_{\mathrm{H}})$ and $\tilde{\boldsymbol{G}}_a = \boldsymbol{\Omega}^{-1/2} \boldsymbol{G}_a \boldsymbol{\Omega}^{-1/2} \sim W_p(f_{\mathrm{G}}, \boldsymbol{I}_p/f_{\mathrm{G}})$. This implies that the approximate degrees

of freedom f_H and f_G of $\boldsymbol{H}$ and $\boldsymbol{G}$ are also the approximate degrees of freedom of $\tilde{\boldsymbol{H}} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{H}\boldsymbol{\Omega}^{-1/2}$ and $\tilde{\boldsymbol{G}} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{G}\boldsymbol{\Omega}^{-1/2}$. Therefore, f_H and f_G can also be determined via matching the total variations of $\tilde{\boldsymbol{H}}$ and $\tilde{\boldsymbol{H}}_a$, and those of $\tilde{\boldsymbol{G}}$ and $\tilde{\boldsymbol{G}}_a$, respectively. According to [10], the total variation of a random matrix $\boldsymbol{X} = (x_{ij})$ is defined as $V(\boldsymbol{X}) = \mathrm{Etr}(\boldsymbol{X} - \mathrm{E}\boldsymbol{X})^2 = \sum_{i,j} \mathrm{Var}(x_{ij})$, i.e., the sum of the variances of all the entries of $\boldsymbol{X}$. After some calculation, we have the following proposition.

Proposition 1. *The approximate degrees of freedom f_H and f_G obtained via matching the total variations of $\tilde{\boldsymbol{H}}$ and $\tilde{\boldsymbol{H}}_a$, and those of $\tilde{\boldsymbol{G}}$ and $\tilde{\boldsymbol{G}}_a$ respectively, are given by*

$$f_H = p(p+1)\left\{ \sum_{i,\alpha=1}^{a} \sum_{j,\beta=1}^{b} w_{ij,\alpha\beta}^2 [tr(\boldsymbol{V}_{ij})tr(\boldsymbol{V}_{\alpha\beta}) + tr(\boldsymbol{V}_{ij}\boldsymbol{V}_{\alpha\beta})] \right\}^{-1},$$

$$f_G = p(p+1)\left\{ \sum_{i=1}^{a} \sum_{j=1}^{b} w_{ij,ij}^2 (n_{ij}-1)^{-1} \left[tr^2(\boldsymbol{V}_{ij}) + tr(\boldsymbol{V}_{ij}^2) \right] \right\}^{-1},$$

where $tr(\boldsymbol{V})$ denotes the trace of $\boldsymbol{V}$ and $\boldsymbol{V}_{ij} = \boldsymbol{\Sigma}_{ij}\boldsymbol{\Omega}^{-1}/n_{ij}, i = 1, \cdots, a; j = 1, \cdots, b$. In addition, we have the following equalities:

$$1 \le f_H \le ab, \quad n_{\min} - 1 \le f_G \le ab(n_{\max} - 1), \qquad (2.21)$$

where $n_{\min} = \min_{i=1}^{a} \min_{j=1}^{b} n_{ij}$ and $n_{\max} = \max_{i=1}^{a} \max_{j=1}^{b} n_{ij}$.

Remark 4. Under the homogeneity assumption (2.16), it is easy to show that $f_H = q$ but $f_G \ne N - ab$. This is because the homogeneity assumption is not taken into account in the construction of $\boldsymbol{G}$ and the computation of the variation of $\boldsymbol{G}$; otherwise, $f_G = N - ab$ as indicated in Remark 3.

Set $\boldsymbol{R}_1 = f_H\boldsymbol{H}$ and $\boldsymbol{R}_2 = f_G\boldsymbol{G}$. Then $\boldsymbol{R}_1 \sim W_p(f_H, \boldsymbol{\Omega})$ and $\boldsymbol{R}_2 \sim W_p(f_G, \boldsymbol{\Omega})$ approximately so that the modified WLR, LHT and BNP test statistics can be respectively defined as:

$$T_{\mathrm{WLR}} = -\log(|\boldsymbol{R}_2|/|\boldsymbol{R}_1 + \boldsymbol{R}_2|), \; T_{\mathrm{LHT}} = tr(\boldsymbol{R}_1\boldsymbol{R}_2^{-1}), \; \text{and} \; T_{\mathrm{BNP}} = tr(\boldsymbol{R}_1(\boldsymbol{R}_1 + \boldsymbol{R}_2)^{-1}). \qquad (2.22)$$

Since $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are independent, the distributions of $T_{\mathrm{WLR}}, T_{\mathrm{LHT}}$ and T_{BNP} are approximately the distributions of the classical WLR, LHT and BNP test statistics with f_H and f_G degrees of freedom. The exact distributions of these test statistics under the null and non-null hypotheses have very complicated forms except for a few special cases. Fortunately, for given f_H and f_G, there are some satisfactory approximations for the distributions of these test statistics. In popular statistical software such as SAS, Splus and R, these distributions are usually approximated by the F-approximation method. Alternatively, [9] used the chi-squared based series expansion [7] and the normal based series expansion [13].

Remark 5. In real data analysis, the approximate degrees of freedom f_H and f_G can be obtained from Proposition 1, denoted as $\hat{f}_H$ and $\hat{f}_G$, via replacing $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}_{ij}, i = 1, \cdots, a; j = 1, \cdots, b$ by their unbiased estimators $\boldsymbol{G}$ and $\hat{\boldsymbol{\Sigma}}_{ij}, i = 1, \cdots, a; j = 1, \cdots, b$, respectively. The modified MANOVA test statistics (2.22) can then be calculated and their distributions can be approximated either by the F-approximation method, or by the chi-squared or normal based series expansion method.

2.3 Properties of the Proposed Modified MANOVA Tests

The proposed modified WLR, LHT and BNP tests have several desirable invariance properties. First of all, we can show that these modified MANOVA tests are affine-invariant.

Proposition 2. *The proposed modified WLR, LHT and BNP tests are invariant under the following affine-transformation:*

$$\boldsymbol{y}_{ijk}^0 = \boldsymbol{A}\boldsymbol{y}_{ijk} + \boldsymbol{\xi}, \; k = 1, \cdots, n_{ij}; i = 1, \cdots, a; j = 1, \cdots, b, \qquad (2.23)$$

where $\boldsymbol{A}$ is any nonsingular matrix and $\boldsymbol{\xi}$ is any constant vector.

The affine-invariance property is often considered as a natural requirement for a test in multivariate statistical inference. This is because in practice, the observed cell responses are often re-centered or re-scaled before any inference is conducted. Re-centering and re-scaling transformations are just special cases of the above affine-transformation. The modified MANOVA tests proposed in [9] are not affine-invariant.

Notice that for the hypotheses in (2.5), the contrast matrices $\boldsymbol{S}_a, \boldsymbol{S}_b$ and $\boldsymbol{S}_{ab}$ are not unique. For example, $\tilde{\boldsymbol{S}}_a = (-\mathbf{1}_{a-1}, \boldsymbol{I}_{a-1})$ is also a contrast matrix for the first hypothesis in (2.5). It is known from [14](Ch. 5, Sec. 4) that for any two contrast matrices $\tilde{\boldsymbol{S}}$ and $\boldsymbol{S}$ specifying the same hypothesis, there is a nonsingular matrix $\boldsymbol{P}$ such that $\tilde{\boldsymbol{S}} = \boldsymbol{P}\boldsymbol{S}$. Let the C-matrices associated with the contrast matrices $\boldsymbol{S}_*$ and $\tilde{\boldsymbol{S}}_*$ are $\boldsymbol{C}_*$ and $\tilde{\boldsymbol{C}}_*$ respectively where $*$ can be a, b or ab. Since there is a nonsingular matrix $\boldsymbol{P}$ such that $\tilde{\boldsymbol{S}}_* = \boldsymbol{P}\boldsymbol{S}_*$, by (2.8), we have $\tilde{\boldsymbol{C}}_* = \tilde{\boldsymbol{S}}_*\boldsymbol{A}_* = \boldsymbol{P}\boldsymbol{S}_*\boldsymbol{A}_* = \boldsymbol{P}\boldsymbol{C}_*$. Based on this, the following proposition implies that the proposed modified WLR, LHT and BNP tests are invariant to the choice of the contrast matrices for the same hypothesis.

Proposition 3. *The proposed modified WLR, LHT and BNP tests are invariant when the coefficient matrix $\boldsymbol{C}$ and the constant matrix $\boldsymbol{C}_0$ in (2.9) are replaced with*

$$\tilde{\boldsymbol{C}} = \boldsymbol{P}\boldsymbol{C}, \quad and \quad \tilde{\boldsymbol{C}}_0 = \boldsymbol{P}\boldsymbol{C}_0, \tag{2.24}$$

respectively where $\boldsymbol{P}$ is any nonsingular matrix of size $q \times q$.

Finally, we have the following simple result.

Proposition 4. *The proposed modified WLR, LHT and BNP tests are invariant under different labeling schemes of the cell mean vectors $\boldsymbol{\mu}_{ij}, i = 1, \cdots, a; \; j = 1, \cdots, b$.*

3   Simulation Studies

In the introduction section, we pointed out that the proposed modified MANOVA tests are extensions of [10]'s test. Not limited to main and interaction effects testing, the proposed modified MANOVA tests can conduct all linear hypothesis tests under heteroscedastic two-way MANOVA in the unified framework of the GLHT problem. Besides, the weight scheme for defining the main and interaction effects is naturally incorporated in the proposed modified MANOVA test to take the unequal cell sizes into account. This is obviously another improvement to [10]'s test. Furthermore, we expect that the proposed modified MANOVA tests will also perform well in terms of size controlling and power.

In this section, we shall present simulation studies to compare the proposed modified MANOVA tests with three existing MANOVA testing procedures for heteroscedastic two-way MANOVA in terms of maintaining the nominal size and power. [9] showed that their modified WLR, LHT and BNP tests performed similarly. [10] also showed that their affine-invariant modified WLR, LHT and BNP tests were overall comparable. By a pre-simulation study, we also found that the modified WLR, LHT and BNP tests proposed in the previous section were quite similar in terms of maintaining the nominal size and power. In light of these facts, in this section, our simulation studies will compare the modified LHT test (namely LHT_3) proposed in the previous section against the modified LHT test of [9] (namely LHT_1), the affine-invariant modified LHT test of [10] (namely LHT_2), and the approximate Hotelling T^2-test (namely AHT) of [11]. In all the simulation studies, the null distributions of the modified MANOVA tests are approximated by the F-approximation method and only the equal-weight method is used. We do this to achieve a fair comparison. This is because the size-adapted-weight method has not been considered in either the modified LHT test of [9] or the affine-invariant modified LHT test of [10]. Furthermore, as pointed out in Remark 2, the coefficient matrix for interaction effect tests does not depend on either of the weight vectors of the levels of factor A and those of factor B. In other words, the proposed modified LHT test will have exactly the same performance for interaction effect tests under both weighting methods.

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

$$\boldsymbol{y}_{ijk} = \boldsymbol{\mu}_{ij} + \boldsymbol{\Sigma}_{ij}^{1/2}\boldsymbol{\epsilon}_{ijk}, k = 1, 2, \cdots, n_{ij}, \tag{3.1}$$

Table 1. Empirical sizes and powers of interaction-effect test for two-way MANOVA ($p = 2$).

$a = 2, \quad b = 20, \quad \boldsymbol{\Sigma}_{1j} = \boldsymbol{I}_2, \quad \boldsymbol{\Sigma}_{2j} = \mathrm{diag}(\boldsymbol{\lambda}), \quad j = 1, 2, \cdots, 20.$

			$\delta = 0$				$\delta = 3.6$				$\delta = 5.4$			
Error	$\boldsymbol{\lambda}$	$\boldsymbol{n}$	LHT_1	LHT_2	LHT_3	AHT	LHT_1	LHT_2	LHT_3	AHT	LHT_1	LHT_2	LHT_3	AHT
$N(0,1)$	$\boldsymbol{\lambda}_1$	$\boldsymbol{n}_1$	.0321	.0390	.0390	.0425	.4722	.5045	.5045	.3986	.9281	.9357	.9357	.8616
		$\boldsymbol{n}_2$	.0388	.0432	.0432	.0470	.7318	.7469	.7469	.6832	.9950	.9957	.9957	.9891
		$\boldsymbol{n}_3$	.0411	.0443	.0448	.0484	.9715	.9729	.9772	.9686	1.000	1.000	1.000	1.000
		$\boldsymbol{n}_4$	.0445	.0486	.0463	.0517	.9862	.9874	.9971	.9950	1.000	1.000	1.000	1.000
	$\boldsymbol{\lambda}_2$	$\boldsymbol{n}_1$	.0321	.0505	.0505	.0434	.1373	.1876	.1876	.2407	.3960	.4745	.4745	.6246
		$\boldsymbol{n}_2$	.0350	.0514	.0514	.0492	.2329	.2884	.2884	.4357	.6611	.7265	.7265	.8963
		$\boldsymbol{n}_3$	.0376	.0496	.0488	.0495	.5720	.6208	.5873	.7813	.9771	.9845	.9751	.9973
		$\boldsymbol{n}_4$	.0354	.0519	.0538	.0503	.4892	.5571	.7453	.9652	.9545	.9683	.9980	1.000
	$\boldsymbol{\lambda}_3$	$\boldsymbol{n}_1$	.0271	.0493	.0493	.0429	.0854	.1385	.1385	.2177	.2295	.3253	.3253	.5771
		$\boldsymbol{n}_2$	.0304	.0519	.0519	.0496	.1410	.2002	.2002	.4085	.4173	.5199	.5199	.8718
		$\boldsymbol{n}_3$	.0361	.0569	.0544	.0501	.3469	.4267	.4021	.7483	.8579	.9052	.8809	.9951
		$\boldsymbol{n}_4$	.0321	.0504	.0510	.0501	.3087	.3886	.5320	.9630	.8073	.8625	.9656	1.000
	ARE		29.62	6.500	6.933	4.950								
$t_4/\sqrt{2}$	$\boldsymbol{\lambda}_1$	$\boldsymbol{n}_1$	.0177	.0267	.0267	.0231	.3793	.4456	.4456	.4579	.8669	.9063	.9063	.9320
		$\boldsymbol{n}_2$	.0235	.0312	.0312	.0373	.6543	.6975	.6975	.7660	.9774	.9862	.9862	.9972
		$\boldsymbol{n}_3$	.0315	.0373	.0377	.0412	.9496	.9615	.9670	.9849	.9988	.9999	1.000	1.000
		$\boldsymbol{n}_4$	.0363	.0426	.0405	.0414	.9723	.9803	.9920	.9980	.9987	.9996	.9997	1.000
	$\boldsymbol{\lambda}_2$	$\boldsymbol{n}_1$	.0183	.0307	.0307	.0262	.1026	.1528	.1528	.2594	.3391	.4372	.4372	.7240
		$\boldsymbol{n}_2$	.0201	.0329	.0329	.0321	.2078	.2701	.2701	.5177	.6092	.6935	.6935	.9467
		$\boldsymbol{n}_3$	.0321	.0430	.0422	.0396	.5385	.6030	.5645	.8509	.9581	.9731	.9626	.9992
		$\boldsymbol{n}_4$	.0268	.0417	.0415	.0391	.4518	.5354	.7254	.9810	.9258	.9550	.9933	1.000
	$\boldsymbol{\lambda}_3$	$\boldsymbol{n}_1$	.0150	.0324	.0324	.0250	.0584	.1002	.1002	.2443	.1844	.2793	.2793	.6728
		$\boldsymbol{n}_2$	.0207	.0377	.0377	.0323	.1141	.1744	.1744	.4797	.3612	.4781	.4781	.9254
		$\boldsymbol{n}_3$	.0278	.0486	.0468	.0421	.3237	.4152	.3910	.8117	.8189	.8754	.8504	.9987
		$\boldsymbol{n}_4$	.0244	.0429	.0481	.0443	.2714	.3642	.5078	.9734	.7625	.8349	.9461	1.000
	ARE		50.97	25.38	25.27	29.38								

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

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 $\mathrm{E}(\boldsymbol{\epsilon}_{ijk}) = \boldsymbol{0}$ and $\mathrm{Cov}(\boldsymbol{\epsilon}_{ijk}) = \boldsymbol{I}_p$. This means that (3.1) 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 $\boldsymbol{h}_0 = [1, 2, \cdots, p]'$ for any given dimension p and $\|\boldsymbol{h}_0\|$ denotes the usual L^2-norm of $\boldsymbol{h}_0$. We then apply the four tests to the generated data, and record their P-values. This process is repeated $N = 10,000$ times. The empirical sizes (when $\delta = 0$) and powers (when $\delta > 0$) of the four tests are the proportions of rejecting the null hypothesis, i.e.,

Table 2. Empirical sizes and powers of interaction-effect test for two-way MANOVA ($p = 3$).

			$\delta = 0$				$\delta = 3.6$				$\delta = 5.4$			
Error	$(\boldsymbol{\lambda}, \rho)$	$\boldsymbol{n}$	LHT$_1$	LHT$_2$	LHT$_3$	AHT	LHT$_1$	LHT$_2$	LHT$_3$	AHT	LHT$_1$	LHT$_2$	LHT$_3$	AHT
$N(0,1)$	$(\boldsymbol{\lambda}_1, \rho_1)$	$\boldsymbol{n}_1$	.0254	.0417	.0417	.0307	.2567	.3257	.3257	.1849	.7345	.7946	.7946	.5377
		$\boldsymbol{n}_2$	.0300	.0428	.0428	.0408	.4943	.5441	.5441	.4310	.9518	.9650	.9650	.8979
		$\boldsymbol{n}_3$	.0331	.0467	.0434	.0607	.8032	.8349	.9285	.8875	.9992	.9994	.9999	.9998
		$\boldsymbol{n}_4$	.0410	.0491	.0492	.0534	.9284	.9380	.9524	.9331	1.000	1.000	1.000	.9999
	$(\boldsymbol{\lambda}_2, \rho_2)$	$\boldsymbol{n}_1$	.0222	.0445	.0445	.0303	.2366	.3499	.3499	.2543	.7067	.8112	.8112	.6941
		$\boldsymbol{n}_2$	.0314	.0530	.0530	.0484	.4641	.5617	.5617	.5539	.9440	.9666	.9666	.9645
		$\boldsymbol{n}_3$	.0436	.0545	.0535	.0691	.8294	.8525	.9506	.9801	.9996	.9999	1.000	1.000
		$\boldsymbol{n}_4$	.0374	.0550	.0557	.0532	.9262	.9474	.9626	.9838	1.000	1.000	1.000	1.000
	$(\boldsymbol{\lambda}_3, \rho_3)$	$\boldsymbol{n}_1$	.0299	.0474	.0474	.0304	.2787	.3647	.3647	.2561	.7596	.8312	.8312	.6899
		$\boldsymbol{n}_2$	.0302	.0438	.0438	.0423	.5247	.5913	.5913	.5641	.9598	.9729	.9729	.9652
		$\boldsymbol{n}_3$	.0412	.0487	.0502	.0675	.8521	.8714	.9606	.9781	.9999	1.000	1.000	1.000
		$\boldsymbol{n}_4$	.0419	.0522	.0513	.0499	.9497	.9601	.9717	.9845	1.000	1.000	1.000	1.000
	ARE		32.12	8.333	8.483	21.85								
$t_4/\sqrt{2}$	$(\boldsymbol{\lambda}_1, \rho_1)$	$\boldsymbol{n}_1$	.0109	.0268	.0268	.0145	.1833	.2861	.2861	.2123	.6098	.7451	.7451	.6580
		$\boldsymbol{n}_2$	.0194	.0337	.0337	.0296	.3963	.4993	.4993	.5099	.9022	.9484	.9484	.9553
		$\boldsymbol{n}_3$	.0182	.0335	.0344	.0462	.7369	.8181	.9166	.9386	.9859	.9967	.9994	.9999
		$\boldsymbol{n}_4$	.0246	.0359	.0359	.0397	.8892	.9242	.9392	.9609	.9975	.9998	1.000	1.000
	$(\boldsymbol{\lambda}_2, \rho_2)$	$\boldsymbol{n}_1$	.0116	.0302	.0302	.0183	.1739	.3000	.3000	.3121	.6077	.7661	.7661	.8174
		$\boldsymbol{n}_2$	.0191	.0385	.0385	.0285	.4070	.5440	.5440	.6678	.8873	.9465	.9465	.9900
		$\boldsymbol{n}_3$	.0300	.0413	.0425	.0464	.7742	.8289	.9321	.9916	.9881	.9952	.9985	1.000
		$\boldsymbol{n}_4$	.0258	.0431	.0436	.0423	.8948	.9331	.9493	.9938	.9980	.9993	.9996	1.000
	$(\boldsymbol{\lambda}_3, \rho_3)$	$\boldsymbol{n}_1$	.0131	.0332	.0332	.0158	.2108	.3314	.3314	.3180	.6538	.7871	.7871	.8172
		$\boldsymbol{n}_2$	.0198	.0369	.0369	.0324	.4400	.5603	.5603	.6673	.9088	.9551	.9551	.9882
		$\boldsymbol{n}_3$	.0236	.0350	.0360	.0485	.7913	.8489	.9481	.9908	.9896	.9965	.9993	1.000
		$\boldsymbol{n}_4$	.0280	.0400	.0401	.0405	.9159	.9437	.9571	.9937	.9978	.9995	.9996	1.000
	ARE		59.32	28.65	28.03	32.88								

The table header note: $a = 3, \quad b = 10, \quad \boldsymbol{\Sigma}_{1j} = \boldsymbol{I}_3, \quad \boldsymbol{\Sigma}_{2j} = \text{diag}(\boldsymbol{\lambda}), \quad \boldsymbol{\Sigma}_{3j} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}, \quad j = 1, 2, \cdots, 10.$

$(\boldsymbol{\lambda}_1, \rho_1) = (1_3, 0)$, $(\boldsymbol{\lambda}_2, \rho_2) = (1, 5, 0.1, 0.05)$ and $(\boldsymbol{\lambda}_3, \rho_3) = (1, 3, 0.1, 0.09)$. $\boldsymbol{n}_1 = (7, 7, 7)_{10}$, $\boldsymbol{n}_2 = (10, 10, 10)_{10}$, $\boldsymbol{n}_3 = (10, 20, 40)_{10}$ and $\boldsymbol{n}_4 = (30, 20, 15)_{10}$.

when their P-values are less than the nominal significance level α. In all the simulations conducted, we used $\alpha = 5\%$ for simplicity.

To save space, here we just report the simulation results for interaction effect tests. Similar conclusions may be drawn from the simulation results for main effect tests. We used the equal weight method to specify the weights of the LHT$_3$ and AHT tests so that their simulation results are comparable with those of the LHT$_1$ and LHT$_2$ tests. The empirical sizes and powers of the four tests for interaction effect tests, together with the associated tuning parameters, are presented in Tables 1– 3, in the columns labeled with LHT$_1$, LHT$_2$, LHT$_3$, and AHT 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

Table 3. Empirical sizes and powers of interaction-effect test for two-way MANOVA ($p = 10$).

			$a = 3,$	$b = 5,$	$\boldsymbol{\Sigma}_{1j} = \boldsymbol{I}_{10},$	$\boldsymbol{\Sigma}_{2j} = \mathrm{diag}(\boldsymbol{\lambda}),$	$\boldsymbol{\Sigma}_{3j} = \mathrm{diag}(\boldsymbol{\eta}),$	$j = 1, 2, ..., 5.$				
			$\delta = 0$				$\delta = 4.2$				$\delta = 6.3$		
Error	$(\boldsymbol{\lambda}, \boldsymbol{\eta})$	$\boldsymbol{n}$	LHT$_1$	LHT$_2$	LHT$_3$	AHT	LHT$_1$	LHT$_2$	LHT$_3$	AHT	LHT$_1$	LHT$_2$	LHT$_3$	AHT
$N(0,1)$	$(\boldsymbol{\lambda}_1, \boldsymbol{\eta}_1)$	$\boldsymbol{n}_1$	.0278	.0469	.0469	.0421	.6215	.6950	.6950	.5866	.9894	.9937	.9937	.9789
		$\boldsymbol{n}_2$	.0447	.0519	.0519	.0541	.9818	.9850	.9850	.9763	1.000	1.000	1.000	1.000
		$\boldsymbol{n}_3$	.0329	.0491	.0495	.0507	.8196	.8653	.9015	.8358	.9997	.9998	1.000	.9996
		$\boldsymbol{n}_4$	.0351	.0500	.0512	.0523	.9509	.9635	.9745	.9541	1.000	1.000	1.000	1.000
	$(\boldsymbol{\lambda}_2, \boldsymbol{\eta}_2)$	$\boldsymbol{n}_1$	.0313	.0655	.0655	.0608	.1287	.2142	.2142	.3159	.3733	.5172	.5172	.7327
		$\boldsymbol{n}_2$	.0451	.0671	.0671	.0552	.3915	.4739	.4739	.7342	.8913	.9297	.9297	.9971
		$\boldsymbol{n}_3$	.0349	.0634	.0639	.0544	.2004	.2808	.3055	.4462	.5734	.6797	.7305	.9139
		$\boldsymbol{n}_4$	.0347	.0635	.0672	.0543	.3699	.4865	.4607	.7161	.8916	.9382	.9166	.9951
	$(\boldsymbol{\lambda}_3, \boldsymbol{\eta}_3)$	$\boldsymbol{n}_1$	.0284	.0512	.0512	.0607	.0467	.0762	.0762	.1109	.0687	.1099	.1099	.1858
		$\boldsymbol{n}_2$	.0396	.0517	.0517	.0536	.0867	.1048	.1048	.1778	.1662	.1986	.1986	.4405
		$\boldsymbol{n}_3$	.0351	.0498	.0510	.0517	.0713	.0924	.0898	.1248	.1432	.1786	.1720	.2801
		$\boldsymbol{n}_4$	.0322	.0493	.0484	.0680	.0515	.0737	.0816	.1823	.0931	.1249	.1489	.3913
	ARE		29.70	11.53	12.65	12.28								
$t_4/\sqrt{2}$	$(\boldsymbol{\lambda}_1, \boldsymbol{\eta}_1)$	$\boldsymbol{n}_1$	.0191	.0458	.0458	.0402	.5617	.7003	.7003	.6468	.9761	.9938	.9938	.9877
		$\boldsymbol{n}_2$	.0311	.0451	.0451	.0426	.9744	.9856	.9856	.9852	1.000	1.000	1.000	1.000
		$\boldsymbol{n}_3$	.0226	.0461	.0453	.0436	.7752	.8632	.8974	.8744	.9965	.9996	1.000	.9998
		$\boldsymbol{n}_4$	.0253	.0486	.0471	.0496	.9302	.9613	.9739	.9699	.9991	1.000	1.000	1.000
	$(\boldsymbol{\lambda}_2, \boldsymbol{\eta}_2)$	$\boldsymbol{n}_1$	.0185	.0557	.0557	.0454	.0967	.2168	.2168	.3517	.3218	.5363	.5363	.8054
		$\boldsymbol{n}_2$	.0303	.0585	.0585	.0489	.3539	.4807	.4807	.7763	.8681	.9288	.9288	.9981
		$\boldsymbol{n}_3$	.0287	.0597	.0589	.0452	.1708	.2879	.3107	.4982	.5339	.6857	.7346	.9449
		$\boldsymbol{n}_4$	.0234	.0592	.0615	.0451	.3398	.4977	.4657	.7645	.8641	.9377	.9186	.9969
	$(\boldsymbol{\lambda}_3, \boldsymbol{\eta}_3)$	$\boldsymbol{n}_1$	.0183	.0461	.0461	.0566	.0300	.0676	.0676	.0972	.0445	.1029	.1029	.1905
		$\boldsymbol{n}_2$	.0334	.0514	.0514	.0461	.0699	.1009	.1009	.1865	.1517	.2043	.2043	.4777
		$\boldsymbol{n}_3$	.0280	.0466	.0478	.0414	.0611	.0940	.0926	.1313	.1214	.1737	.1667	.2971
		$\boldsymbol{n}_4$	.0235	.0489	.0483	.0678	.0404	.0755	.0821	.1844	.0682	.1199	.1496	.4290
	ARE		49.63	9.550	10.08	12.72								

$\boldsymbol{\lambda}_1 = (1_{10})_5$, $\boldsymbol{\eta}_1 = (1_{10})_5$; $\boldsymbol{\lambda}_2 = (12_3, 1_3, 24_3, 1)_5$, $\boldsymbol{\eta}_2 = (1_3, 0.1_3, 2_2, 24, 21)_5$ and $\boldsymbol{\lambda}_3 = (1_3, 3_3, 9_3, 20)_5$, $\boldsymbol{\eta}_3 = (5_3, 15_3, 45_3, 50)_5$. $\boldsymbol{n}_1 = (25_3)_5$, $\boldsymbol{n}_2 = (50_3)_5$, $\boldsymbol{n}_3 = (25, 35, 45)_5$ and $\boldsymbol{n}_4 = (70, 50, 30)_5$.

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 $\mathrm{ARE} = 100 M^{-1} \sum_{j=1}^{M} |\hat{\alpha}_j - \alpha|/\alpha$ where $\hat{\alpha}_j$ denotes the j-th empirical size for $j = 1, 2, \cdots, M$, $\alpha = 5\%$ and M is the number of empirical sizes under consideration. The smaller ARE value indicates the better overall performance of the associated test, and for a good test, the larger the cell sizes, the smaller the ARE values. The ARE values of the four tests under the two error schemes are also presented in these three tables. Notice that for simplicity, in the specification of the cell covariance and size tuning parameters, we often use $(\boldsymbol{u}_r)$ to denote "$\boldsymbol{u}$ repeats r times ". For simplicity and space saving, following

[15], the cell covariance matrices and the cell sizes were specified as the same 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 $\Sigma_{ij} = \Sigma_{i1}, n_{ij} = n_{i1}, j = 1, 2, \cdots, b$. 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 four tests for a bivariate case with $a = 2$ and $b = 20$. With $b = 20$, one may be able to check how the four 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 four tests for a 3-variate case with $a = 3$ and $b = 10$ and a 10-variate case with $a = 3$ and $b = 5$ respectively. These two tables allow us to compare the four tests for higher-dimensional normal and non-normal data.

We first compare the $\text{LHT}_1, \text{LHT}_2, \text{LHT}_3$, and AHT tests under the $N(0, 1)$ error scheme. Overall speaking, the LHT_1 test is outperformed by the other three tests in terms of size controlling and power. In terms of size controlling, the LHT_1 test is generally too conservative while the other three tests maintain the nominal size very well as summarized by the ARE values of the four tests. In terms of power, the LHT_1 test generally has lower powers than the other three tests. This is probably due to the fact that the empirical sizes of the LHT_1 test are generally smaller than those of the other three tests. We now compare the last three tests in more details. Although these three tests are generally comparable in terms of size controlling and power, slight differences still can be found. We notice that the LHT_2 and LHT_3 tests perform quite similarly and they are even identical for homogeneous and balanced cases. This is not a surprise since they are constructed in a similar way and use the same approaches to compute the approximate degrees of freedom. We also notice that in terms of size controlling, the LHT_2 and LHT_3 tests generally outperform the AHT test as indicated by the ARE values of the three tests. This is probably due to the fact that the LHT_2 and LHT_3 tests use two tuning parameters while the AHT test uses only one. In terms of power, the LHT_2 and LHT_3 tests generally have higher powers than the AHT test for homogeneous cases. This is probably due to the fact that the LHT_2 and LHT_3 tests are constructed in a similar way as the way the classical LHT test is constructed for homogeneous cases. For heteroscedastic cases, however, the AHT test generally has higher powers than the LHT_2 and LHT_3 tests.

Under the $t_4/\sqrt{2}$ error scheme, we can draw similar conclusions as mentioned above except that the empirical sizes for the four tests are now slightly smaller. It is reasonable since all these modified MANOVA tests have not taken the data non-normality into account.

4 A Real Data Example

In this section, a real data set from a smoking cessation trial is used to illustrate and compare the $\text{LHT}_1, \text{LHT}_2, \text{LHT}_3$, AHT tests and the classical LHT test as well. The smoking cessation trial is part of a large clinical trial about a behavioral intervention to promote smoking cessation among college student smokers [16]. The subjects are students with a low and high degree of depression and from 20 individual fraternity or sorority chapters (Greek houses) of the University of Missouri-Colombia. The researchers believed that the level of depression of each subject is highly associated with the nicotine dependence of the subject and they also wanted to know whether the nicotine dependence of the subjects depended on the chapter they came from. The nicotine dependence can be measured by three well-known scales, namely, the Fagerstöm Test for Nicotine Dependence [17], the Hooked on Nicotine Checklist [18], and the Minnesota Tobacco Withdrawal Scale [19]. The resulting data may be referred to as the smoking cessation data which can be analyzed by a heteroscedastic two-way MANOVA model as in [9], [10], and [11].

We applied the $\text{LHT}_1, \text{LHT}_2, \text{LHT}_3$, AHT tests and the classical LHT test as well to the smoking cessation data for checking the significance of the main and interaction effects of the two factors: Chapter and Depression. The test results are listed in Table 4. Both the equal-weight method and the size-adapted-weight method, as described in Section 2.1, were considered. The F-values, P-values and the associated degrees of freedom of all the tests were computed using the F-approximation method.

We first compare the results of the five tests under the equal-weight method. As seen from Table 4, all of the five tests lead to the same conclusion about the effect of Depression. However, different conclusions about the main-effect of Chapter and the interaction-effect between Chapter and Depression are obtained. The P-values of $\text{LHT}_r, r = 1, 2, 3$ and AHT tests all suggest that the main-effect of Chapter and the interaction-effect between Chapter and Depression are not significant, while the classical LHT test concludes that the main-effect of Chapter is significant at 5%, and the interaction-effect between Chapter

Table 4. Test results for the smoking cessation data.

Method	Test	Chapter				Depression				Chapter×Depression			
		F-value	P-value	df_1	df_2	F-value	P-value	df_1	df_2	F-value	P-value	df_1	df_2
Equal-weight	LHT	1.47	.015	57	911	21.8	0	3	305	1.29	.075	57	911
	LHT_1	1.08	.350	43.0	253	15.7	0	3	85.8	1.01	.465	43.0	253
	LHT_2	1.06	.384	34.4	182	15.6	0	3	62.0	.993	.487	34.4	182
	LHT_3	1.15	.261	39.2	247	15.6	0	3	62.0	1.00	.476	39.2	247
	AHT	1.16	.257	57	90.5	15.6	0	3	62.0	.896	.669	57	90.5
Size-adapted-weight	LHT	1.23	.124	57	911	23.0	0	3	305	1.29	.075	57	911
	LHT_1	-	-	-	-	-	-	-	-	-	-	-	-
	LHT_2	-	-	-	-	-	-	-	-	-	-	-	-
	LHT_3	1.17	.225	41.1	360	16.8	0	3	86.8	1.00	.476	39.2	247
	AHT	1.17	.232	57	118	16.8	0	3	86.8	.896	.669	57	90.5

and Depression is significant at 10%, which is opposite to the ones obtained by $LHT_r, r = 1, 2, 3$ and AHT tests. [9] verified via using the univariate Levene test [6] that the cell covariance matrices homogeneity assumption for the smoking cessation data is unlikely to be satisfied and a logarithm-transformation could not effectively solve the problem. In light of this fact, the conclusions made by the $LHT_r, r = 1, 2, 3$ and AHT tests are more trustworthy than those made by the classical LHT test since all the modified LHT tests and AHT test take into account the heteroscedasticity of the data. The misleading conclusions made by applying the classical LHT test in this example is a consequence of unreasonable homogeneity assumption, which demonstrates that our modified MANOVA tests are really essential and necessary when the problem of heteroscedasticity is serious. We would also like to point out that compared with the classical LHT test, there exists a drastic drop in the second degrees of freedom, df_2, of the approximate F-distributions of the $LHT_r, r = 1, 2, 3$ and AHT tests. As seen from Table 4, the df_2's values of the classical LHT test are about $3 \sim 5$ times of those of its modified versions, and about 10 times of those of the AHT test, which indicates a serious impact of the heteroscedasticity on the classical MANOVA tests. Since the lower the degrees of freedom, the larger the value of F needed to be significant, this may also provide some explanation why the modified $LHT_r, r = 1, 2, 3$ and AHT tests get different conclusions on the main-effect of Chapter and the interaction-effect. For the comparison among the $LHT_r, r = 1, 2, 3$ and AHT tests, generally speaking, all these tests made consistent conclusions about the effects of the two factors. On the other hand, differences in F-values, P-values and degrees of freedom of these tests can also be seen clearly.

We next examine the test results under the size-adapted method. Notice that [9]'s LHT_1 test and [10]'s LHT_2 test are not defined for the size-adapted method. The test results of our new LHT test and [11]'s AHT 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 classical 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 classical LHT test.

References

1. G. A. F. Seber, *Multivariate observations.* John Wiley & sons, New York, 1984.
2. R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 5th ed. Prentice Hall, Englewood Cliffs, NJ, USA, 2002.
3. N. H. Timm, *Applied multivariate analysis.* Springer, New York, 2002.

4. C. A. Tsai and J. J. Chen, "Multivariate analysis of variance test for gene set analysis," *Bioinformatics*, vol. 25, pp. 897–903, 2009.

5. G. E. P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, pp. 317–346, 1949.

6. H. Levene, "Robust tests for equality of variances," in *Contributions to probability and statistics: Essays in honor of Harold Hotelling*, I. Olkin, S. G. Churye, W. Hoeffding, W. G. Madow, and H. B. Mann, Eds. Stanford University Press, 1960, pp. 278–292.

7. T. W. Anderson, *An introduction to multivariate statistical analysis*, 3rd ed. John Wiley & Sons, New York, 2003.

8. J. Stevens, *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Erlbaum, 1986.

9. 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.

10. J. T. Zhang and S. Xiao, "A note on the modified two-way MANOVA tests," *Statistics and Probability Letters*, vol. 82, no. 3, pp. 519–527, 2012.

11. J. T. Zhang, "Two-way MANOVA with unequal cell sizes and unequal cell covariance matrices," *Technometrics*, vol. 53, no. 4, pp. 426–439, 2011.

12. Y. Fujikoshi, "Two-way ANOVA models with unbalanced data," *Discrete Mathematics*, vol. 116, pp. 315–334, 1993.

13. ——, "Asymptotic formulas for the non-null distributions of three statistics for multivariate linear hypothesis," *Annals of the Institute of Statistical Mathematics*, vol. 27, pp. 99–108, 1975.

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

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

16. K. J. Harris, D. Catley, G. E. Good, N. J. Cronk, S. W. Harrar, and K. B. Williams, "Motivational interviewing for smoking cessation in college students: A group randomized trial," *Preventive Medicine*, vol. 51, no. 5, pp. 387–393, 2010.

17. 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.

18. R. J. Wellman, J. R. DiFranza, J. A. Savageau, S. Godiwala, K. Friedman, and J. Hazelton, "Measuring adults' loss of autonomy over cigarette use: The hooked on nicotine checklist," *Nicotine and Tobacco Research*, vol. 7, no. 1, pp. 157–161, 2005.

19. D. K. Hatsukami, J. R. Hughes, R. W. Pickens, and D. Svikis, "Tobacco withdrawal symptoms: An experimental analysis," *Psychopharmacology*, vol. 84, no. 2, pp. 231–236, 1984.

20. D. G. Nel and C. A. Van Der Merwe, "A solution to the multivariate Behrens-Fisher problem," *Communications in Statistics - Theory and Methods*, vol. 15, no. 12, pp. 3719–3735, 1986.

Appendix: Proofs

In this Appendix, we first state a useful lemma which was given and proved by [10] and then proceed to give the proofs of the main results presented in this article.

Lemma 1. *Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_m$ be mutually independent with $\boldsymbol{u}_i \sim N_p(\boldsymbol{0}, \boldsymbol{V}_i), i = 1, 2, \cdots, m$. Let $\boldsymbol{Q} = \boldsymbol{U}^T \boldsymbol{R} \boldsymbol{U}$ be a quadratic form of $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_m]^T$ where $\boldsymbol{R} = (r_{ij})$ is an $m \times m$ non-negative symmetric matrix. Then $E(\boldsymbol{Q}) = \sum_{i=1}^m r_{ii} \boldsymbol{V}_i$ and $V(\boldsymbol{Q}) = \sum_{i=1}^m \sum_{j=1}^m r_{ij}^2 [tr(\boldsymbol{V}_i)tr(\boldsymbol{V}_j) + tr(\boldsymbol{V}_i\boldsymbol{V}_j)]$ where $V(\boldsymbol{Q})$ denotes the total variation of $\boldsymbol{Q}$.*

Proof of Proposition 1. We first find f_{H} so that $\tilde{\boldsymbol{H}} \sim W_p(f_{\mathrm{H}}, \boldsymbol{I}_p/f_{\mathrm{H}})$ approximately via matching the total variations of $\tilde{\boldsymbol{H}}$ and $\tilde{\boldsymbol{H}}_a$ where $\tilde{\boldsymbol{H}}_a \sim W_p(f_{\mathrm{H}}, \boldsymbol{I}_p/f_{\mathrm{H}})$. It is easy to show that $V(\tilde{\boldsymbol{H}}_a) = p(p+1)/f_{\mathrm{H}}$. To find $V(\tilde{\boldsymbol{H}})$, notice that $\tilde{\boldsymbol{H}} = \boldsymbol{\Omega}^{-1/2}\mathcal{E}^T\boldsymbol{W}\mathcal{E}\boldsymbol{\Omega}^{-1/2} = \tilde{\mathcal{E}}^T\boldsymbol{W}\tilde{\mathcal{E}}$ where $\tilde{\mathcal{E}} = [\tilde{\bar{\boldsymbol{\epsilon}}}_{11\cdot}, \cdots, \tilde{\bar{\boldsymbol{\epsilon}}}_{1b\cdot}, \cdots, \tilde{\bar{\boldsymbol{\epsilon}}}_{a1\cdot}, \cdots, \tilde{\bar{\boldsymbol{\epsilon}}}_{ab\cdot}]^T$ with $\tilde{\bar{\boldsymbol{\epsilon}}}_{ij\cdot} = \boldsymbol{\Omega}^{-1/2}\bar{\boldsymbol{\epsilon}}_{ij\cdot} \sim N_p(\boldsymbol{0}, \boldsymbol{\Omega}^{-1/2}\boldsymbol{\Sigma}_{ij}\boldsymbol{\Omega}^{-1/2}/n_{ij})$. Hence $\tilde{\boldsymbol{H}}$ is a quadratic form of $\tilde{\mathcal{E}}$. Applying Lemma 1, together with the fact that $\mathrm{tr}(\boldsymbol{V}_1\boldsymbol{V}_2) = \mathrm{tr}(\boldsymbol{V}_2\boldsymbol{V}_1)$, we have

$$V(\tilde{\boldsymbol{H}}) = \sum_{i,\alpha=1}^a \sum_{j,\beta=1}^b w_{ij,\alpha\beta}^2 [\mathrm{tr}(\boldsymbol{V}_{ij})\mathrm{tr}(\boldsymbol{V}_{\alpha\beta}) + \mathrm{tr}(\boldsymbol{V}_{ij}\boldsymbol{V}_{\alpha\beta})],$$

where $\boldsymbol{V}_{ij} = \boldsymbol{\Sigma}_{ij}\boldsymbol{\Omega}^{-1}/n_{ij}, i = 1, \cdots, a; j = 1, \cdots, b$. Equating $V(\tilde{\boldsymbol{H}}_a)$ and $V(\tilde{\boldsymbol{H}})$ leads to

$$f_{\mathrm{H}} = p(p+1)\left\{\sum_{i,\alpha=1}^{a}\sum_{j,\beta=1}^{b} w_{ij,\alpha\beta}^2[\mathrm{tr}(\boldsymbol{V}_{ij})\mathrm{tr}(\boldsymbol{V}_{\alpha\beta}) + \mathrm{tr}(\boldsymbol{V}_{ij}\boldsymbol{V}_{\alpha\beta})]\right\}^{-1}.$$

We now proceed to find f_{G} so that $\tilde{\boldsymbol{G}} \sim W_p(f_{\mathrm{G}}, \boldsymbol{I}_p/f_{\mathrm{G}})$ approximately via matching the total variations of $\tilde{\boldsymbol{G}}$ and $\tilde{\boldsymbol{G}}_a \sim W_p(f_{\mathrm{G}}, \boldsymbol{I}_p/f_{\mathrm{G}})$. Again, we have $V(\tilde{\boldsymbol{G}}_a) = p(p+1)/f_{\mathrm{G}}$. To find $V(\tilde{\boldsymbol{G}})$, notice that under the model (2.2), we have $\boldsymbol{\Omega}^{-1/2}\hat{\boldsymbol{\Sigma}}_{ij}\boldsymbol{\Omega}^{-1/2} \sim W_p\left(n_{ij} - 1, \boldsymbol{\Omega}^{-1/2}\boldsymbol{\Sigma}_{ij}\boldsymbol{\Omega}^{-1/2}/(n_{ij} - 1)\right)$. It follows that $\tilde{\boldsymbol{G}} = (ab)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}^{-1}\boldsymbol{\Omega}^{-1/2}\hat{\boldsymbol{\Sigma}}_{ij}\boldsymbol{\Omega}^{-1/2}$ is a Wishart-mixture [20]. Some simple algebra leads to $V(\tilde{\boldsymbol{G}}) = \sum_{i=1}^{a}\sum_{j=1}^{b} w_{ij,ij}^2(n_{ij} - 1)^{-1}[\mathrm{tr}^2(\boldsymbol{V}_{ij}) + \mathrm{tr}(\boldsymbol{V}_{ij}^2)]$. Then equating $V(\tilde{\boldsymbol{G}}_a)$ and $V(\tilde{\boldsymbol{G}})$ leads to

$$f_{\mathrm{G}} = p(p+1)\left\{\sum_{i=1}^{a}\sum_{j=1}^{b} w_{ij,ij}^2(n_{ij} - 1)^{-1}[\mathrm{tr}^2(\boldsymbol{V}_{ij}) + \mathrm{tr}(\boldsymbol{V}_{ij}^2)]\right\}^{-1},$$

as desired. We now move to find the lower and upper bounds of f_{H} and f_{G}. To find the lower and upper bounds of f_{H}, set $d_{\mathrm{H}} = \sum_{i,\alpha=1}^{a}\sum_{j,\beta=1}^{b} w_{ij,\alpha\beta}^2[\mathrm{tr}(\boldsymbol{V}_{ij})\mathrm{tr}(\boldsymbol{V}_{\alpha\beta}) + \mathrm{tr}(\boldsymbol{V}_{ij}\boldsymbol{V}_{\alpha\beta})]$ and set $\tilde{\boldsymbol{V}}_{ij} = w_{ij,ij}\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Sigma}_{ij}\boldsymbol{\Omega}^{-1/2}/n_{ij}$ which is nonnegative definite. Then we have

$$\sum_{i=1}^{a}\sum_{j=1}^{b} \tilde{\boldsymbol{V}}_{ij} = \sum_{i=1}^{a}\sum_{j=1}^{b} w_{ij,ij}\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Sigma}_{ij}\boldsymbol{\Omega}^{-1/2}/n_{ij} = \boldsymbol{I}_p$$

so that

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}) = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{p} \lambda_{ijk} = p, \tag{A.1}$$

where $\lambda_{ijk}, k = 1, 2, \cdots, p$ are the eigenvalues of $\tilde{\boldsymbol{V}}_{ij}$. It is easy to see that $\boldsymbol{I}_p - \tilde{\boldsymbol{V}}_{ij}$ is also nonnegative definite so that $0 \le \lambda_{ijk} \le 1$. It follows that $\mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}) = \sum_{k=1}^{p} \lambda_{ijk} \le p$ and $\mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}^2) = \sum_{k=1}^{p} \lambda_{ijk}^2 \le \sum_{k=1}^{p} \lambda_{ijk} = \mathrm{tr}(\tilde{\boldsymbol{V}}_{ij})$. By (A.1), we have

$$\sum_{i=1}^{a}\sum_{j=1}^{b}[\mathrm{tr}^2(\tilde{\boldsymbol{V}}_{ij}) + \mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}^2)] \le \sum_{i=1}^{a}\sum_{j=1}^{b}[p\mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}) + \mathrm{tr}(\tilde{\boldsymbol{V}}_{ij})] = p(p+1). \tag{A.2}$$

Notice that for any m nonnegative numbers $a_1, a_2, \cdots, a_m$, we have $\sum_{i=1}^{m} a_i^2 \ge (\sum_{i=1}^{m} a_i)^2/m$. Then by (A.1) again, we have

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\mathrm{tr}^2(\tilde{\boldsymbol{V}}_{ij}) = \sum_{i=1}^{a}\sum_{j=1}^{b}(\sum_{k=1}^{p} \lambda_{ijk})^2 \ge [\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{p} \lambda_{ijk}]^2/(ab) = p^2/(ab).$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}^2) = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{p} \lambda_{ijk}^2 \ge [\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{p} \lambda_{ijk}]^2/(abp) = p/(ab).$$

It follows that

$$\sum_{i=1}^{a}\sum_{j=1}^{b}[\mathrm{tr}^2(\tilde{\boldsymbol{V}}_{ij}) + \mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}^2)] \ge p(p+1)/(ab). \tag{A.3}$$

Notice that we can write

$$\boldsymbol{W} = \boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}^T)^{-1}\boldsymbol{C} = \boldsymbol{A}^T\boldsymbol{A}$$

where $\boldsymbol{A} = (\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}^T)^{-1/2}\boldsymbol{C} = (\boldsymbol{a}_{11}, \cdots, \boldsymbol{a}_{1b}, \cdots, \boldsymbol{a}_{a1}, \cdots, \boldsymbol{a}_{ab})^T$ so that we have $w_{ij,\alpha\beta} = \boldsymbol{a}_{ij}^T\boldsymbol{a}_{\alpha\beta}$. It follows from the Cauchy-Schwarz inequality that $w_{ij,\alpha\beta}^2 \le w_{ij,ij}w_{\alpha\beta,\alpha\beta}$. Therefore, $0 \le \rho_{ij,\alpha\beta} \le 1$

where $\rho_{ij,\alpha\beta} = \frac{w_{ij,\alpha\beta}}{\sqrt{w_{ij,ij}w_{\alpha\beta,\alpha\beta}}}$. Since $d_{\mathrm{H}} = \sum_{i,\alpha=1}^{a}\sum_{j,\beta=1}^{b}\rho_{ij,\alpha\beta}^{2}[\mathrm{tr}(\tilde{\boldsymbol{V}}_{ij})\mathrm{tr}(\tilde{\boldsymbol{V}}_{\alpha\beta}) + \mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}\tilde{\boldsymbol{V}}_{\alpha\beta})]$, we have $d_{\mathrm{H}} \leq \sum_{i,\alpha=1}^{a}\sum_{j,\beta=1}^{b}[\mathrm{tr}(\tilde{\boldsymbol{V}}_{ij})\mathrm{tr}(\tilde{\boldsymbol{V}}_{\alpha\beta}) + \mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}\tilde{\boldsymbol{V}}_{\alpha\beta})] = \mathrm{tr}^2(\boldsymbol{I}_p) + \mathrm{tr}(\boldsymbol{I}_p^2) = p(p+1)$. It follows that $f_{\mathrm{H}} \geq 1$. To find the upper bound of f_{H}, notice that $\tilde{\boldsymbol{V}}_{ij}$ is nonnegative definite so that $\mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}) \geq 0$ and $\mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}\tilde{\boldsymbol{V}}_{\alpha\beta}) \geq 0$. Thus, $d_{\mathrm{H}} \geq \sum_{i=1}^{a}\sum_{j=1}^{b}[\mathrm{tr}^2(\tilde{\boldsymbol{V}}_{ij}) + \mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}^{2})]$. By (A.2), we have $d_{\mathrm{H}} \geq p(p+1)/(ab)$ so that $f_{\mathrm{H}} \leq ab$.

We now find the lower and upper bounds of f_{G}. Set $d_{\mathrm{G}} = \sum_{i=1}^{a}\sum_{j=1}^{b}w_{ij,ij}^{2}(n_{ij}-1)^{-1}[\mathrm{tr}^2(\boldsymbol{V}_{ij}) + \mathrm{tr}(\boldsymbol{V}_{ij}^{2})]$. Using the definition of $\tilde{\boldsymbol{V}}_{ij}$, we have $d_{\mathrm{G}} = \sum_{i=1}^{a}\sum_{j=1}^{b}(n_{ij}-1)^{-1}[\mathrm{tr}^2(\tilde{\boldsymbol{V}}_{ij}) + \mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}^{2})]$. By (A.2), we have

$$d_{\mathrm{G}} \leq (n_{\min}-1)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b}[\mathrm{tr}^2(\tilde{\boldsymbol{V}}_{ij}) + \mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}^{2})] \leq (n_{\min}-1)^{-1}p(p+1).$$

Thus, $f_{\mathrm{G}} \geq n_{\min} - 1$. Now by (A.3), we have

$$d_{\mathrm{G}} \geq (n_{\max}-1)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b}[\mathrm{tr}^2(\tilde{\boldsymbol{V}}_{ij}) + \mathrm{tr}(\tilde{\boldsymbol{V}}_{ij}^{2})] \geq (n_{\max}-1)^{-1}p(p+1)/(ab).$$

Thus, $f_{\mathrm{G}} \leq ab(n_{\max}-1)$. The proposition is proved.

Proof of Proposition 2. Let the parameters and their estimators associated with the new data resulted from the affine-transformation (2.23) be labeled with a superscript "0". For example, $\boldsymbol{\mu}_{ij}^{0}$ and $\hat{\boldsymbol{\mu}}_{ij}^{0}$ denote the (ij)th cell mean vector and its estimator based on the new data. Then under the affine-transformation (2.23), we have $\boldsymbol{\mu}_{ij}^{0} = \boldsymbol{A}\boldsymbol{\mu}_{ij} + \boldsymbol{\xi}$, $\boldsymbol{\Sigma}_{ij}^{0} = \boldsymbol{A}\boldsymbol{\Sigma}_{ij}\boldsymbol{A}^{T}$, $\hat{\boldsymbol{\mu}}_{ij}^{0} = \boldsymbol{A}\hat{\boldsymbol{\mu}}_{ij} + \boldsymbol{\xi}$, $\hat{\boldsymbol{\Sigma}}_{ij}^{0} = \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{ij}\boldsymbol{A}^{T}$. Thus, $\boldsymbol{M}^{0} = \boldsymbol{M}\boldsymbol{A}^{T} + \boldsymbol{1}_{ab}\boldsymbol{\xi}^{T}$ and $\hat{\boldsymbol{M}}^{0} = \hat{\boldsymbol{M}}\boldsymbol{A}^{T} + \boldsymbol{1}_{ab}\boldsymbol{\xi}^{T}$. It follows that the GLHT problem (2.9) associated with the new data now becomes $\mathcal{H}_0^{0} : \boldsymbol{C}\boldsymbol{M}^{0} = \boldsymbol{C}_0^{0}$, vs $\mathcal{H}_1^{0} : \boldsymbol{C}\boldsymbol{M}^{0} \neq \boldsymbol{C}_0^{0}$, where $\boldsymbol{C}_0^{0} = \boldsymbol{C}_0\boldsymbol{A}^{T} + \boldsymbol{C}\boldsymbol{1}_{ab}\boldsymbol{\xi}^{T}$. Then $\boldsymbol{C}\hat{\boldsymbol{M}}^{0} - \boldsymbol{C}_0^{0} = (\boldsymbol{C}\hat{\boldsymbol{M}} - \boldsymbol{C}_0)\boldsymbol{A}^{T}$. Based on the above results and the fact that $\boldsymbol{W} = \boldsymbol{C}^{T}(\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}^{T})^{-1}\boldsymbol{C}$ has nothing to do with the affine-transformation (2.23), after some simple algebra, we can show that $\boldsymbol{H}^{0} = \boldsymbol{A}\boldsymbol{H}\boldsymbol{A}^{T}$, $\boldsymbol{G}^{0} = \boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^{T}$, $\boldsymbol{\Omega}^{0} = \boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}^{T}$, and $\boldsymbol{V}_{ij}^{0} = \boldsymbol{A}\boldsymbol{V}_{ij}\boldsymbol{A}^{-1}$, $\hat{\boldsymbol{V}}_{ij}^{0} = \boldsymbol{A}\hat{\boldsymbol{V}}_{ij}\boldsymbol{A}^{-1}$, $i = 1,\cdots,a; j = 1,\cdots,b$. These results imply that under the affine-transformation (2.23), the modified MANOVA test statistics defined in (2.22) are invariant and the approximate degrees of freedom f_{H} and f_{G} given in Proposition 1 and their estimators are also invariant. The proposition is proved.

Proof of Proposition 3. It is sufficient to show that $\boldsymbol{H}, \boldsymbol{G}, f_{\mathrm{H}}$ and f_{G} are invariant under (2.24). Under (2.24), we have $\tilde{\boldsymbol{C}}\hat{\boldsymbol{M}} - \tilde{\boldsymbol{C}}_0 = \boldsymbol{P}(\boldsymbol{C}\hat{\boldsymbol{M}} - \boldsymbol{C}_0)$, and $(\tilde{\boldsymbol{C}}\boldsymbol{D}\tilde{\boldsymbol{C}}^{T})^{-1} = (\boldsymbol{P}^{-1})^{T}(\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}^{T})^{-1}\boldsymbol{P}^{-1}$. It follows that $\tilde{\boldsymbol{W}} = \tilde{\boldsymbol{C}}^{T}[\tilde{\boldsymbol{C}}\boldsymbol{D}\tilde{\boldsymbol{C}}^{T}]^{-1}\tilde{\boldsymbol{C}} = \boldsymbol{W}$. Then the invariance of the SSCP matrices $\boldsymbol{H}$ and $\boldsymbol{G}$ follows from (2.11) and (2.15) immediately and the invariance of f_{H} and f_{G} follows from the formulas of f_{H} and f_{G} given in Proposition 1 and the fact that every entry of $\boldsymbol{W}$ is invariant under (2.24). The proposition is then proved.

Proof of Proposition 4. It is sufficient to show that $\boldsymbol{H}, \boldsymbol{G}, f_{\mathrm{H}}$ and f_{G} are invariant under different labeling schemes of the cell mean vectors. Let $i_1, i_2, \cdots, i_a$ and $j_1, j_2, \cdots, j_b$ be any permutations of $1, 2, \cdots, a$ and $1, 2, \cdots, b$ respectively. Write $\boldsymbol{C} = [\boldsymbol{c}_{11}, \boldsymbol{c}_{12}, \cdots, \boldsymbol{c}_{ab}]$ where $\boldsymbol{c}_{ij}$ denotes the (ij)th column of $\boldsymbol{C}$. Now $\boldsymbol{C}\hat{\boldsymbol{M}} = \sum_{i=1}^{a}\sum_{j=1}^{b}\boldsymbol{c}_{ij}\hat{\boldsymbol{\mu}}_{ij}^{T} = \sum_{u=1}^{a}\sum_{v=1}^{b}\boldsymbol{c}_{i_u j_v}\hat{\boldsymbol{\mu}}_{i_u j_v}^{T}$ and $\boldsymbol{G} = \sum_{i=1}^{a}\sum_{j=1}^{b}w_{ij,ij}\hat{\Sigma}_{ij}/n_{ij} = \sum_{u=1}^{a}\sum_{v=1}^{b}w_{i_u j_v,i_u j_v}\hat{\Sigma}_{i_u j_v}/n_{i_u j_v}$, showing that the SSCP matrices $\boldsymbol{H}$ and $\boldsymbol{G}$ are invariant under different labeling schemes of the cell mean vectors.

To show that the approximate degrees of freedom f_{H} and f_{G} are invariant under different labeling schemes of the cell mean vectors, by Proposition 1, it is sufficient to show that the denominators of f_{H} and f_{G} have such a property. This is actually the case by noticing that the denominator of f_{H}:

$$\sum_{i,\alpha=1}^{a}\sum_{j,\beta=1}^{b}w_{ij,\alpha\beta}^{2}[\mathrm{tr}(\boldsymbol{V}_{ij})\mathrm{tr}(\boldsymbol{V}_{\alpha\beta}) + \mathrm{tr}(\boldsymbol{V}_{ij}\boldsymbol{V}_{\alpha\beta})] =$$

$$\sum_{i_u,\alpha_u=1}^{a}\sum_{j_v,\beta_v=1}^{b}w_{i_u j_v,\alpha_u \beta_v}^{2}[\mathrm{tr}(\boldsymbol{V}_{i_u j_v})\mathrm{tr}(\boldsymbol{V}_{\alpha_u \beta_v}) + \mathrm{tr}(\boldsymbol{V}_{i_u j_v}\boldsymbol{V}_{\alpha_u \beta_v})],$$

where we have used the fact that $\boldsymbol{V}_{ij} = \boldsymbol{\Sigma}_{ij}\boldsymbol{\Omega}^{-1}/n_{ij}$ and $\boldsymbol{\Omega}$ is invariant under different labeling schemes of the cell mean vectors. Similarly, we can show that the denominator of f_{G} has such a property too. This completes the proof of the proposition.