

## Research Article

# MANOVA for Nested Designs with Unequal Cell Sizes and Unequal Cell Covariance Matrices

**Li-Wen Xu**

*College of Sciences, North China University of Technology, Beijing 100144, China*

Correspondence should be addressed to Li-Wen Xu; [xulw163@163.com](mailto:xulw163@163.com)

Received 7 May 2014; Accepted 3 July 2014; Published 16 July 2014

Academic Editor: William Zhu

Copyright © 2014 Li-Wen Xu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose and study parametric bootstrap (PB) tests for heteroscedastic two-factor MANOVA with nested designs. For the problem of testing “main effects” of both factors, we develop a flexible test based on a parametric bootstrap approach. The PB test is shown to be invariant under affine-transformations. Moreover, the PB test does not depend on the chosen weights used to define the parameters uniquely. The proposed test is compared with the approximate Hotelling  $T^2$  (AHT) test by the simulations. Simulation results indicate that the PB test performs satisfactorily for various cell sizes and parameter configurations and generally outperforms the AHT test in terms of controlling the nominal size. For the heteroscedastic cases, the PB test outperforms the AHT test in terms of power. In addition, the PB test does not lose too much power when the homogeneity assumption is actually valid.

## 1. Introduction

There are many situations where we record more than one response variable from each sampling or experimental unit and where these units are allocated to or occur in treatment groups. Ecologists often record the abundances of many species from each sampling or experimental unit and physiologists commonly measure more than one variable (e.g., blood pressure, heart rate, etc.) on experimental animals. With multiple response variables, we might be more interested in whether there are group differences on all the response variables considered simultaneously. This is the aim of multivariate analysis of variance (MANOVA), the analogue of univariate ANOVA when we have multiple response variables for each experimental or sampling unit.

The simplest design is single factor MANOVA (or one-way MANOVA) which tests the effect of a factor, having  $k$  levels. If the cell covariance matrices are assumed to be equal, then there are some popular tests available to test the equality of the mean vectors. The tests that are commonly used are the Wilks likelihood ratio (WLR), Lawley-Hotelling trace (LHT), Bartlett-Nanda-Pillai (BNP), and Roys largest root tests ([1], chap. 8, sec. 6). When there are some departures from the standard assumption, that is, unequal cell covariance matrices, these solutions were proposed by James [2],

Johansen [3], Gamage et al. [4], and Krishnamoorthy and Lu [5], among others. The more complex design is multifactor MANOVA especially when the homogeneity of the cell covariance matrices assumption is seriously violated. There has been a continuous interest in the heteroscedastic two-factor MANOVA which tests in checking the significance of the effects of two factors  $A$  and  $B$ , each having  $a$  and  $b$  levels, respectively, in a multivariate factorial experiment with crossed designs. Recently, Harrar and Bathke [6] attacked 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; see some details in Section 2.2 of Zhang and Xiao [7]. Zhang [8] proposed an approximate Hotelling  $T^2$  (AHT) test. A Wald-type test statistic is used. Its null distribution is approximated by a Hotelling  $T^2$ -distribution with one parameter estimated from the data. Some simulation studies conducted in Zhang [8] showed that the AHT test outperforms the modified LHT test of Harrar and Bathke [6].

Another useful two-factor design is the nested MANOVA [9]. Notice that each of the testing problems associated with

the three null hypotheses in the heteroscedastic two-way MANOVA can be equivalently expressed in the form of the general linear hypothesis testing (GLHT) problem as described in Zhang [8]. The GLHT problem is very general. It includes not only the main and interaction effect tests but also various post hoc and contrast tests as special cases. To construct the test for two-factor nested MANOVA, we may use the same Wald-type test statistic as in Zhang [8]. However, according to our simulation studies, the empirical sizes of the AHT test for the two-factor nested MANOVA model may far exceed the nominal level. On the other hand, the AHT test needs to consider the the selected weights when the cell sizes are unequal. Therefore, it is important to develop a test procedure for the nesting effects and the nested effects with satisfactory size and power regardless of number of factorial effects and the sample sizes.

Accordingly, the present paper will develop a parametric bootstrap (PB) test for heteroscedastic two-factor nested MANOVA. We use standardized effects sum of squares and a natural test statistic obtained by replacing cell covariance matrices by the corresponding sample cell covariance matrices. The PB test admits several nice properties: (1) it can be simply conducted by a routine Monte Carlo algorithm; (2) it is shown to be invariant under affine-transformations; (3) The PB test does not depend on choices of the weights used to identify the parameters; and (4) it works well. Simulation results reported in Section 4 indicate that the PB test performs satisfactorily for various cell sizes and parameter configurations when the homogeneity assumption is seriously violated and generally outperforms the AHT test in terms of power and controlling size. The main ideas of the proposed PB test are closely related to the work by Xu et al. [10]. The methodologies for the PB test are presented in Sections 2 and 3. Simulation results are presented in Section 4. Some concluding remarks are given in Section 5. Finally, some technical proofs of the main results are given in the Appendix.

## 2. Tests for Nested Effects

Consider a two-factor nested design model with factors  $A$  and  $B$ . Suppose that the nesting effect corresponds to the factor  $A$  having  $a$  factor levels, and the nested effect corresponds to the factor  $B$  having a total of  $b = \sum_{i=1}^a b_i$  levels with  $b_i$  levels of  $B$  nested within the  $i$ th level of the factor  $A$  ( $i = 1, \dots, a$ ). Suppose a  $r$ -variate random sample of size  $n_{ij}$  is available under  $(i, j)$ th level,  $i = 1, \dots, a$ ;  $j = 1, \dots, b_i$ . Let  $\mathbf{Y}_{ijk}$ ,  $i = 1, \dots, a$ ;  $j = 1, \dots, b_i$ ;  $k = 1, \dots, n_{ij}$  represent these random vectors and  $y_{ijk}$  represent their observed (sample) values. Assume that  $n_{ij} > r$  so that positive definite sample covariance matrices can be computed for each cell of the design. Suppose that  $\mathbf{Y}_{ijk}$  satisfy the following model:

$$\mathbf{Y}_{ijk} = \boldsymbol{\mu}_{ij} + \mathbf{e}_{ijk}, \quad \mathbf{e}_{ijk} \sim N_r(\mathbf{0}, \boldsymbol{\Sigma}_{ij}), \quad i = 1, \dots, a; \quad (1)$$

$$j = 1, \dots, b_i; \quad k = 1, \dots, n_{ij},$$

where  $\boldsymbol{\mu}_{ij} : r \times 1$  and  $\boldsymbol{\Sigma}_{ij} : r \times r$  are the cell mean vector and cell covariance matrix of the random sample at the  $(i, j)$ th cell and  $\mathbf{e}_{ijk}$  is an experimental error vector term. All these

samples are assumed to be independent with each other. In the two-factor nested model, the cell mean vectors  $\boldsymbol{\mu}_{ij}$  are usually decomposed into the form

$$\boldsymbol{\mu}_{ij} = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_{ij}, \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, b_i, \quad (2)$$

where  $\boldsymbol{\mu}_0$  is the grand mean vector,  $\boldsymbol{\alpha}_i$  is the effect vector of the  $i$ th level of  $A$  on each of the  $r$  variables in  $\mathbf{Y}_{ijk}$ , and  $\boldsymbol{\beta}_{ij}$  is the effect due to the  $j$ th level of the factor  $B$  nested within the  $i$ th level of the factor  $A$  so that (1) can be further written as the following two-factor multivariate nested model with unequal error covariance matrices:

$$\mathbf{Y}_{ijk} = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_{ij} + \mathbf{e}_{ijk}, \quad \mathbf{e}_{ijk} \sim N_r(\mathbf{0}, \boldsymbol{\Sigma}_{ij}), \quad (3)$$

$$i = 1, \dots, a; \quad j = 1, \dots, b_i; \quad k = 1, \dots, n_{ij}.$$

In order to have  $\boldsymbol{\mu}_0$ ,  $\boldsymbol{\alpha}_i$ , and  $\boldsymbol{\beta}_{ij}$  uniquely defined, we need to have additional constraints. Let  $u_1, \dots, u_a$  and  $v_{11}, \dots, v_{ib_i}$  be nonnegative weights such that  $\sum_{i=1}^a u_i > 0$  and  $\sum_{j=1}^{b_i} v_{ij} > 0$ . We consider the following constraints:

$$\sum_{i=1}^a u_i \boldsymbol{\alpha}_i = \mathbf{0}, \quad \sum_{j=1}^{b_i} v_{ij} \boldsymbol{\beta}_{ij} = \mathbf{0}, \quad (4)$$

where  $u_1, \dots, u_a$  and  $v_{11}, \dots, v_{ib_i}$  are nonnegative weights such that  $\sum_{i=1}^a u_i > 0$  and  $\sum_{j=1}^{b_i} v_{ij} > 0$  for each  $i$ .

In this section, we are interested in testing the null hypothesis,

$$H_{0\beta} : \boldsymbol{\beta}_{ij} = \mathbf{0}; \quad i = 1, \dots, a, \quad j = 1, \dots, b_i, \quad (5)$$

against its natural alternative hypothesis. The null hypothesis in (5) aims to test if the nested effect corresponding to the factor  $B$  is statistically significant.

The sample mean vector and the sample covariance matrix of the  $(i, j)$ th cell are denoted by  $\bar{\mathbf{Y}}_{ij}$  and  $\mathbf{S}_{ij}$ , respectively, where

$$\bar{\mathbf{Y}}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} \mathbf{Y}_{ijk}, \quad (6)$$

$$\mathbf{S}_{ij} = \frac{1}{n_{ij} - 1} \sum_{k=1}^{n_{ij}} (\mathbf{Y}_{ijk} - \bar{\mathbf{Y}}_{ij})(\mathbf{Y}_{ijk} - \bar{\mathbf{Y}}_{ij})^T,$$

$i = 1, \dots, a$ ;  $j = 1, \dots, b_i$ , and  $\mathbf{M}^T$  denotes the transpose of the matrix  $\mathbf{M}$ . The observed values of these random variables are denoted as  $\bar{y}_{ij}$  and  $\mathbf{s}_{ij}$ , respectively,  $i = 1, \dots, a$ ,  $j = 1, \dots, b_i$ . We note that  $\bar{\mathbf{Y}}_{ij}$ 's and  $\mathbf{S}_{ij}$ 's are mutually independent with  $(n_{ij} - 1)\mathbf{S}_{ij} \sim W_r(n_{ij} - 1, \boldsymbol{\Sigma}_{ij})$  and the model for  $\bar{\mathbf{Y}}_{ij}$  as follows:

$$\bar{\mathbf{Y}}_{ij} = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_{ij} + \bar{\mathbf{e}}_{ij}, \quad \bar{\mathbf{e}}_{ij} \sim N_r\left(0, \frac{1}{n_{ij}} \boldsymbol{\Sigma}_{ij}\right), \quad (7)$$

$$i = 1, \dots, a; \quad j = 1, \dots, b_i,$$

where  $\bar{\mathbf{e}}_{ij} = (1/n_{ij}) \sum_{k=1}^{n_{ij}} \mathbf{e}_{ijk}$  and  $W_r(m, \Gamma)$  denotes the  $r$ -dimensional Wishart distribution with degrees of freedom

(df) =  $m$  and scale parameter matrix  $\Gamma$ . Writing  $\bar{\mathbf{Y}} = (\bar{\mathbf{Y}}_{11}^T, \dots, \bar{\mathbf{Y}}_{1b_1}^T, \dots, \bar{\mathbf{Y}}_{a1}^T, \dots, \bar{\mathbf{Y}}_{ab_a}^T)^T$  and  $\bar{\mathbf{e}} = (\bar{\mathbf{e}}_{11}^T, \dots, \bar{\mathbf{e}}_{1b_1}^T, \dots, \bar{\mathbf{e}}_{a1}^T, \dots, \bar{\mathbf{e}}_{ab_a}^T)^T$ , the model (7) can be written as

$$\bar{\mathbf{Y}} = (\mathbf{1}_b \otimes \mathbf{I}_r) \boldsymbol{\mu}_0 + (\text{diag}(\mathbf{1}_{b_1}, \dots, \mathbf{1}_{b_a}) \otimes \mathbf{I}_r) \boldsymbol{\alpha} + \boldsymbol{\beta} + \bar{\mathbf{e}}, \quad (8)$$

where  $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_a^T)^T$ ,  $\boldsymbol{\beta} = (\boldsymbol{\beta}_{11}^T, \dots, \boldsymbol{\beta}_{1b_1}^T, \dots, \boldsymbol{\beta}_{a1}^T, \dots, \boldsymbol{\beta}_{ab_a}^T)^T$ ,  $\bar{\mathbf{e}} \sim N_{br}(\mathbf{0}, \boldsymbol{\Sigma})$ ,  $\boldsymbol{\Sigma} = \text{diag}((1/n_{11})\boldsymbol{\Sigma}_{11}, \dots, (1/n_{1b_1})\boldsymbol{\Sigma}_{1b_1}, \dots, (1/n_{a1})\boldsymbol{\Sigma}_{a1}, \dots, (1/n_{ab_a})\boldsymbol{\Sigma}_{ab_a})$ ,  $\mathbf{1}_m$  denotes the  $m \times 1$  vector of ones,  $\mathbf{I}_m$  is an identity matrix with order  $m$ ,  $\mathbf{V} \otimes \mathbf{W}$  denotes the Kronecker product of matrices  $\mathbf{V}$  and  $\mathbf{W}$ , and  $\text{diag}(\mathbf{M}_1, \dots, \mathbf{M}_m)$  denotes a block-diagonal matrix with  $\mathbf{M}_1, \dots, \mathbf{M}_m$  along the blocks.

Define the standardized sum of squares due to the factor  $B$

$$\begin{aligned} \tilde{S}_\beta(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{ab_a}) \\ = \sum_{i=1}^a \sum_{j=1}^{b_i} (\bar{\mathbf{Y}}_{ij} - \hat{\boldsymbol{\mu}}_0 - \hat{\boldsymbol{\alpha}}_i)^T (n_{ij} \boldsymbol{\Sigma}_{ij}^{-1}) (\bar{\mathbf{Y}}_{ij} - \hat{\boldsymbol{\mu}}_0 - \hat{\boldsymbol{\alpha}}_i), \end{aligned} \quad (9)$$

where  $\hat{\boldsymbol{\mu}}_0$  and  $\hat{\boldsymbol{\alpha}}_i$  are solutions of  $\boldsymbol{\mu}_0$  and  $\boldsymbol{\alpha}_i$  that minimize the quadratic equation

$$\begin{aligned} \tilde{S}(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{ab_a}) \\ = \sum_{i=1}^a \sum_{j=1}^{b_i} (\bar{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\alpha}_i)^T (n_{ij} \boldsymbol{\Sigma}_{ij}^{-1}) (\bar{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\alpha}_i), \end{aligned} \quad (10)$$

subject to the constraints given in (4).

In fact, denoting  $\boldsymbol{\theta} = (\boldsymbol{\mu}_0^T, \boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_a^T)^T$  and  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\mu}}_0^T, \hat{\boldsymbol{\alpha}}_1^T, \dots, \hat{\boldsymbol{\alpha}}_a^T)^T$ , it follows from Theorem 5.2.5 in Wang and Chow [11] that

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\mu}}_0^T, \hat{\boldsymbol{\alpha}}_1^T, \dots, \hat{\boldsymbol{\alpha}}_a^T)^T = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{L}^T \mathbf{L})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \bar{\mathbf{Y}}, \quad (11)$$

where  $\mathbf{X} = (\mathbf{1}_b, \text{diag}(\mathbf{1}_{b_1}, \dots, \mathbf{1}_{b_a})) \otimes \mathbf{I}_r$  and  $\mathbf{L} = (0, \mathbf{u}_1, \dots, \mathbf{u}_a) \otimes \mathbf{I}_r$ . Then

$$\tilde{S}_\beta(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{ab_a}) = (\bar{\mathbf{Y}} - \mathbf{X} \hat{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{Y}} - \mathbf{X} \hat{\boldsymbol{\theta}}). \quad (12)$$

Notice that  $\hat{\boldsymbol{\theta}}$  indeed depends on  $\mathbf{L}$  of chosen weights. However, we can prove that the standardized sum of squares in (9) does not depend on  $\mathbf{L}$ . This is the following Theorem.

**Theorem 1.** *The standardized sum of squares  $\tilde{S}_\beta(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{ab_a})$  does not depend on  $\mathbf{L}$  of chosen weights, and*

$$\begin{aligned} \tilde{S}_\beta(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{ab_a}) \\ = \bar{\mathbf{Y}}^T \boldsymbol{\Sigma}^{-1/2} (\mathbf{I}_{br} - \boldsymbol{\Sigma}^{-1/2} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1/2}) \boldsymbol{\Sigma}^{-1/2} \bar{\mathbf{Y}}, \end{aligned} \quad (13)$$

where  $\mathbf{M}^-$  denotes any generalized inverse of matrix  $\mathbf{M}$ .

Note from Theorem 1 that  $\tilde{S}_\beta(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{ab_a})$  does not depend on  $\mathbf{L}$ . If  $\boldsymbol{\Sigma}_{ij}$  are known, then a natural statistic for testing (5) is  $\tilde{S}_\beta(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{ab_a})$ . In fact,  $\boldsymbol{\Sigma}^{-1/2} \bar{\mathbf{Y}} \sim N_{br}(\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\mu}, \mathbf{I}_{br})$ , and

$$\Delta = (\mathbf{I}_{br} - \boldsymbol{\Sigma}^{-1/2} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1/2}) \quad (14)$$

is an idempotent matrix with rank  $r(b-a)$ ; we have

$$\bar{\mathbf{Y}}^T \boldsymbol{\Sigma}^{-1/2} \Delta \boldsymbol{\Sigma}^{-1/2} \bar{\mathbf{Y}} \sim \chi_{r(b-a)}^2 (\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1/2} \Delta \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\mu}), \quad (15)$$

where  $\boldsymbol{\mu} = (\boldsymbol{\mu}_{11}^T, \dots, \boldsymbol{\mu}_{1b_1}^T, \dots, \boldsymbol{\mu}_{a1}^T, \dots, \boldsymbol{\mu}_{ab_a}^T)^T$  and  $\chi_m^2(\delta)$  denotes a noncentral chi-square random variable with degrees of freedom  $m$  and noncentrality parameter  $\delta$ . The noncentrality parameter  $\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1/2} \Delta \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\mu}$  is equal to zero when  $\boldsymbol{\beta}_{11} = \dots = \boldsymbol{\beta}_{ab_a}$ . Let  $\bar{\mathbf{y}} = (\bar{\mathbf{y}}_{11}^T, \dots, \bar{\mathbf{y}}_{1b_1}^T, \dots, \bar{\mathbf{y}}_{a1}^T, \dots, \bar{\mathbf{y}}_{ab_a}^T)^T$  be the observed value of  $\bar{\mathbf{Y}}$ . Then, the test that rejects  $H_{0\beta}$  in (5) whenever

$$\tilde{S}_\beta(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{ab_a}) > \chi_{r(b-a), \lambda}^2 \quad (16)$$

is a size  $\lambda$  test, where  $\chi_{m, \lambda}^2$  is the upper  $\lambda$ th quantile of a chi-square distribution with  $df = m$ .

In general, the covariance matrices  $\boldsymbol{\Sigma}_{ij}$  are unknown; in this case, a test statistic can be obtained by replacing  $\boldsymbol{\Sigma}_{ij}$  in (13) by  $\mathbf{S}_{ij}$ ,  $i = 1, \dots, a$ ,  $j = 1, \dots, b_i$  and is given by

$$\begin{aligned} \tilde{S}_\beta(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \mathbf{S}_{11}, \dots, \mathbf{S}_{ab_a}) \\ = \bar{\mathbf{Y}}^T \mathbf{S}^{-1/2} (\mathbf{I}_{br} - \mathbf{S}^{-1/2} \mathbf{X} (\mathbf{X}^T \mathbf{S}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}^{-1/2}) \mathbf{S}^{-1/2} \bar{\mathbf{Y}}, \end{aligned} \quad (17)$$

where  $\mathbf{S} = \text{diag}((1/n_{11})\mathbf{S}_{11}, \dots, (1/n_{1b_1})\mathbf{S}_{1b_1}, \dots, (1/n_{a1})\mathbf{S}_{a1}, \dots, (1/n_{ab_a})\mathbf{S}_{ab_a})$ .

As shown in Zhang [8], the testing problem associated with the null hypothesis (5) can also be equivalently expressed in the form of the general linear hypothesis testing (GLHT) problem as  $H_{0\beta} : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ , where  $\mathbf{C} = \mathbf{H}_\beta \mathbf{A}_\beta \otimes \mathbf{I}_r$ ,

$\mathbf{H}_\beta$

$$= \begin{pmatrix} (\mathbf{I}_{b_1-1}, -\mathbf{I}_{b_1-1}) & 0 & 0 & 0 \\ 0 & (\mathbf{I}_{b_2-1}, -\mathbf{I}_{b_2-1}) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & (\mathbf{I}_{b_a-1}, -\mathbf{I}_{b_a-1}) \end{pmatrix},$$

$$\mathbf{A}_\beta = \begin{pmatrix} \mathbf{I}_{b_1} - \mathbf{1}_{b_1} \mathbf{u}_1 \mathbf{v}_1^T & 0 & 0 & 0 \\ 0 & \mathbf{I}_{b_2} - \mathbf{1}_{b_2} \mathbf{u}_2 \mathbf{v}_2^T & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{I}_{b_a} - \mathbf{1}_{b_a} \mathbf{u}_a \mathbf{v}_a^T \end{pmatrix}, \quad (18)$$

and  $\mathbf{v}_i = (v_{i1}, \dots, v_{ib_i})^T$ ,  $i = 1, 2, \dots, a$ . Then the associated Wald-type test statistic is given as

$$T = (\mathbf{C}\hat{\boldsymbol{\mu}})^T (\mathbf{CSC}^T)^{-1} (\mathbf{C}\hat{\boldsymbol{\mu}}), \quad (19)$$

where  $\hat{\boldsymbol{\mu}} = \bar{\mathbf{Y}}$ , and  $\mathbf{S}$  is defined as in (17).

In the following, we shall describe two tests based on (17) and (19), respectively.

2.1. *The Approximate Hotelling  $T^2$  (AHT) Test.* Similar to Zhang [8], we proposed a test referred as the AHT test, which is based on the test statistic

$$\frac{(\hat{d} - q + 1)T}{q\hat{d}}, \quad (20)$$

where  $q = \text{rank}(\mathbf{C}) = r(b - a)$ ,

$$\hat{d} = \frac{q(q + 1)}{\sum_{i=1}^a \sum_{j=1}^{b_i} (n_{ij} - 1)^{-1} \left[ \text{tr}^2(\hat{\Omega}_{ij}) + \text{tr}(\hat{\Omega}_{ij}^2) \right]}, \quad (21)$$

and  $\hat{\Omega}_{ij} = n_{ij}^{-1} \hat{\mathbf{H}}_{ij} \mathbf{S}_{ij} \hat{\mathbf{H}}_{ij}^T$  and  $\hat{\mathbf{H}}_{ij} = (\mathbf{CSC}^T)^{-1/2} \mathbf{C}_{ij}$  with  $\mathbf{C} = [\mathbf{C}_{11}, \dots, \mathbf{C}_{1b_1}, \dots, \mathbf{C}_{a1}, \dots, \mathbf{C}_{ab_a}]$ , and  $\mathbf{C}_{ij} : q \times r$  is the  $(ij)$ th block matrix of  $\mathbf{C}$ . Zhang [8] showed that, under  $H_{0\beta}$ ,  $(\hat{d} - q + 1)T/q\hat{d}$  is approximately distributed as  $F_{q, \hat{d}-q+1}$  random variable. Thus, the AHT test rejects the null hypothesis in (5) when the critical value  $(q\hat{d}/(\hat{d} - q + 1))F_{q, \hat{d}-q+1}(1 - \lambda)$  for the nominal significance level  $\lambda$  is exceeded by the observed test statistic  $T$  in (20). The AHT test can also be conducted via computing the  $P$ -value based on the approximate null distribution.

2.2. *The Parametric Bootstrap (PB) Test.* The parametric bootstrap involves sampling from the estimated models. That is, samples or sample statistics are generated from parametric models with the parameters replaced by their estimates. Recall that under  $H_{0\beta}$  the vector  $\bar{\mathbf{Y}}$  have the mean  $\mathbf{X}\boldsymbol{\theta}$ , where  $\boldsymbol{\theta} = (\boldsymbol{\mu}_0^T, \boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_a^T)^T$ . As the test statistic in (17) is location invariant under the group of location transformations  $\mathcal{G} = \{\bar{\mathbf{Y}} \rightarrow \bar{\mathbf{Y}} + \mathbf{X}\boldsymbol{\eta}, \boldsymbol{\eta} \in \mathbb{R}^{r(a+1)}\}$ , without loss of generality, we can take  $\mathbf{X}\boldsymbol{\theta} = \mathbf{0}$ . Using these facts, the parametric bootstrap pivot variable can be developed as follows. For a given  $(\bar{\mathbf{y}}_{11}, \dots, \bar{\mathbf{y}}_{ab_a}; \mathbf{s}_{11}, \dots, \mathbf{s}_{ab_a})$ , let  $\bar{\mathbf{Y}}_{Bij} \sim N_r(0, (1/n_{ij})\mathbf{s}_{ij})$  and  $\mathbf{S}_{Bij} \sim W_r(n_{ij} - 1, (1/(n_{ij} - 1))\mathbf{s}_{ij})$ ,  $i = 1, \dots, a$ ,  $j = 1, \dots, b_i$ . Then the PB pivot variable based on the test statistic (17) is given by

$$\begin{aligned} & \tilde{\mathcal{S}}_{\beta B}(\bar{\mathbf{Y}}_{B11}, \dots, \bar{\mathbf{Y}}_{Bab_a}; \mathbf{S}_{B11}, \dots, \mathbf{S}_{Bab_a}) \\ &= \bar{\mathbf{Y}}_B^T \mathbf{S}_B^{-1/2} \left( \mathbf{I}_{br} - \mathbf{S}_B^{-1/2} \mathbf{X} (\mathbf{X}^T \mathbf{S}_B^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}_B^{-1/2} \right) \mathbf{S}_B^{-1/2} \bar{\mathbf{Y}}_B, \end{aligned} \quad (22)$$

where  $\bar{\mathbf{Y}}_B = (\bar{\mathbf{Y}}_{B11}^T, \dots, \bar{\mathbf{Y}}_{B1b_1}^T, \dots, \bar{\mathbf{Y}}_{Ba1}^T, \dots, \bar{\mathbf{Y}}_{Bab_a}^T)^T$  and  $\mathbf{S}_B = \text{diag}((1/n_{11})\mathbf{S}_{B11}, \dots, (1/n_{1b_1})\mathbf{S}_{B1b_1}, \dots, (1/n_{a1})\mathbf{S}_{Ba1}, \dots, (1/n_{ab_a})\mathbf{S}_{Bab_a})$ . For a given level  $\lambda$ , the PB test rejects  $H_{0\beta}$  in (5) when

$$P(\tilde{\mathcal{S}}_{\beta B}(\bar{\mathbf{Y}}_{B11}, \dots, \bar{\mathbf{Y}}_{Bab_a}; \mathbf{S}_{B11}, \dots, \mathbf{S}_{Bab_a}) > \tilde{s}_{\beta}) < \lambda, \quad (23)$$

where

$$\tilde{s}_{\beta} = \tilde{\mathcal{S}}_{\beta}(\bar{\mathbf{y}}_{11}, \dots, \bar{\mathbf{y}}_{ab_a}; \mathbf{s}_{11}, \dots, \mathbf{s}_{ab_a}) \quad (24)$$

is an observed value of  $\tilde{\mathcal{S}}_{\beta}(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \mathbf{S}_{11}, \dots, \mathbf{S}_{ab_a})$  in (17). For fixed  $(\bar{\mathbf{y}}_{11}, \dots, \bar{\mathbf{y}}_{ab_a}; \mathbf{s}_{11}, \dots, \mathbf{s}_{ab_a})$ , the above probability does not depend on any unknown parameters, and so it can be estimated using Monte Carlo simulation given in the following Algorithm 2.

*Algorithm 2.* For a given  $(n_{11}, \dots, n_{ab_a})$ ,  $(\bar{\mathbf{y}}_{11}, \dots, \bar{\mathbf{y}}_{ab_a})$ , and  $(\mathbf{s}_{11}, \dots, \mathbf{s}_{ab_a})$ ,

Compute  $\tilde{\mathcal{S}}_{\beta}(\bar{\mathbf{y}}_{11}, \dots, \bar{\mathbf{y}}_{ab_a}; \mathbf{s}_{11}, \dots, \mathbf{s}_{ab_a})$  in (17) and call it  $\tilde{s}_{\beta}$ .

For  $h = 1, \dots, m$ ,

Generate  $\bar{\mathbf{Y}}_{Bij} \sim N_r(0, (1/n_{ij})\mathbf{s}_{ij})$  and  $\mathbf{S}_{Bij} \sim W_r(n_{ij} - 1, (1/(n_{ij} - 1))\mathbf{s}_{ij})$ ,  $i = 1, \dots, a$ ,  $j = 1, \dots, b_i$ ,

Compute  $\tilde{\mathcal{S}}_{\beta B}(\bar{\mathbf{Y}}_{B11}, \dots, \bar{\mathbf{Y}}_{Bab_a}; \mathbf{S}_{B11}, \dots, \mathbf{S}_{Bab_a})$  using (22).

If  $\tilde{\mathcal{S}}_{\beta B}(\bar{\mathbf{Y}}_{B11}, \dots, \bar{\mathbf{Y}}_{Bab_a}; \mathbf{S}_{B11}, \dots, \mathbf{S}_{Bab_a}) > \tilde{s}_{\beta}$ , set  $Q_h = 1$ .

(End loop)

$(1/m) \sum_{h=1}^m Q_h$  is a Monte Carlo estimate of the  $P$  value in (23).

2.3. *Some Desirable Properties of the PB Test.* The PB test (23) has several desirable invariance properties. First of all, the PB test is affine-invariant. That is, it is invariant under the following affine-transformation:

$$\begin{aligned} \mathbf{Y}_{ijk}^* &= \mathbf{D}\mathbf{Y}_{ijk} + \boldsymbol{\xi}, \quad i = 1, \dots, a; \\ & j = 1, \dots, b_i; \quad k = 1, \dots, n_{ij}, \end{aligned} \quad (25)$$

where  $\mathbf{D}$  is any nonsingular matrix and  $\boldsymbol{\xi}$  is any given vector. This property is desirable since in practice, the cell responses  $\mathbf{Y}_{ijk}$  (1) are often recentered or rescaled before an inference is conducted. The recentering and rescaling transformations are special cases of (25).

**Theorem 3.** *The PB test (23) is affine-invariant in the sense that both the observed test statistic  $\tilde{s}_{\beta}$  (24) and the distribution of PB pivot variable  $\tilde{\mathcal{S}}_{\beta B}$  (22) are affine-invariant.*

The model (3) is identifiable when the constraints (4) are imposed. In many situations there are no natural weights to justify a particular test procedure. For this, we have the following result.

**Theorem 4.** *The PB test (23) is invariant on the different weight choices in constraints (4).*

Notice that the test statistic (17) does not depend on the chosen weights. The conclusion in Theorem 4 is obvious. However, we have the following property.

**Theorem 5.** *The test statistic (17)*

$$\begin{aligned} & \tilde{\mathcal{S}}_{\beta}(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \mathbf{S}_{11}, \dots, \mathbf{S}_{ab_a}) \\ &= \bar{\mathbf{Y}}^T \mathbf{S}^{-1/2} \left( \mathbf{I}_{br} - \mathbf{S}^{-1/2} \mathbf{X} (\mathbf{X}^T \mathbf{S}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}^{-1/2} \right) \mathbf{S}^{-1/2} \bar{\mathbf{Y}} \\ &= \bar{\mathbf{Y}}^T \mathbf{S}^{-1/2} \left( \mathbf{I}_{br} - \mathbf{S}^{-1/2} \mathbf{X} (\mathbf{X}^T \mathbf{S}^{-1} \mathbf{X} + \mathbf{L}^T \mathbf{L})^{-1} \mathbf{X}^T \mathbf{S}^{-1/2} \right) \mathbf{S}^{-1/2} \bar{\mathbf{Y}}. \end{aligned} \quad (26)$$

This property is desirable since in practice, in order to avoid computation of generalized inverse of matrices and enhance the accuracy of computation, we have taken a particular  $\mathbf{L}$  such as  $\mathbf{L} = (0, 1/a, \dots, 1/a) \otimes \mathbf{I}_r$  in the simulation study.

### 3. Tests for the Nesting Effects

When the nested effects are present, the nesting effect  $\alpha_i$  can not reflect the effect of  $A_i$  because it depends on which level of factor  $B$  it is in. As pointed in Searle [12], Chap. 7, in the case of  $r = 1$ , a popular solution to the problem is not quite a test for  $\alpha_i = 0$  in the presence of nested effects but rather to test the null hypothesis

$$H_{0\alpha|\beta} : \alpha_i + \beta_{ij} = \mathbf{0}, \quad i = 1, \dots, a, \quad j = 1, \dots, b_i. \quad (27)$$

We will see that the problem of testing (27) is the same case of the problem of comparing  $b = \sum_{i=1}^a b_i$  normal mean vectors when the cell covariance matrices are unknown and arbitrary [5]. In fact, if  $\Sigma_{ij}$  are known, then

$$\hat{\boldsymbol{\mu}}_0 = \left( \sum_{i=1}^a \sum_{j=1}^{b_i} n_{ij} \Sigma_{ij}^{-1} \right)^{-1} \sum_{i=1}^a \sum_{j=1}^{b_i} n_{ij} \Sigma_{ij}^{-1} \bar{\mathbf{Y}}_{ij} \quad (28)$$

is the best linear unbiased estimator of  $\boldsymbol{\mu}_0$ , and a natural statistic for testing (27) is

$$\begin{aligned} & \tilde{S}_\alpha (\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{ab_a}) \\ &= \sum_{i=1}^a \sum_{j=1}^{b_i} (\bar{\mathbf{Y}}_{ij} - \hat{\boldsymbol{\mu}}_0)^T (n_{ij} \Sigma_{ij}^{-1}) (\bar{\mathbf{Y}}_{ij} - \hat{\boldsymbol{\mu}}_0) \\ &= \bar{\mathbf{Y}}^T \boldsymbol{\Sigma}^{-1/2} \left( \mathbf{I}_{br} - \boldsymbol{\Sigma}^{-1/2} \mathbf{X}_1 (\mathbf{X}_1^T \boldsymbol{\Sigma}^{-1} \mathbf{X}_1)^{-1} \mathbf{X}_1^T \boldsymbol{\Sigma}^{-1/2} \right) \boldsymbol{\Sigma}^{-1/2} \bar{\mathbf{Y}}, \end{aligned} \quad (29)$$

where  $\mathbf{X}_1 = \mathbf{1}_b \otimes \mathbf{I}_r$ . Notice that  $\boldsymbol{\Sigma}^{-1/2} \bar{\mathbf{Y}} \sim N_{br}(\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\mu}, \mathbf{I}_{br})$ , and  $\boldsymbol{\Delta}_1 = (\mathbf{I}_{br} - \boldsymbol{\Sigma}^{-1/2} \mathbf{X}_1 (\mathbf{X}_1^T \boldsymbol{\Sigma}^{-1} \mathbf{X}_1)^{-1} \mathbf{X}_1^T \boldsymbol{\Sigma}^{-1/2})$  is an idempotent matrix with rank  $r(b-1)$ ; we have

$$\bar{\mathbf{Y}}^T \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Delta}_1 \boldsymbol{\Sigma}^{-1/2} \bar{\mathbf{Y}} \sim \chi_{r(b-1)}^2 \left( \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Delta}_1 \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\mu} \right). \quad (30)$$

The noncentrality parameter  $\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Delta}_1 \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\mu}$  is equal to zero when  $\alpha_i + \beta_{ij} = 0, i = 1, \dots, a, j = 1, \dots, b_i$ . Then, the test that rejects  $H_{0\alpha|\beta}$  in (27) whenever

$$\tilde{S}_\alpha (\bar{\mathbf{y}}_{11}, \dots, \bar{\mathbf{y}}_{ab_a}; \boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{ab_a}) > \chi_{r(b-1), \lambda}^2 \quad (31)$$

is a size  $\lambda$  test. In general, the covariance matrices  $\Sigma_{ij}$  are unknown; in this case, a test statistic can be obtained by replacing  $\Sigma_{ij}$  in (29) by  $S_{ij}, i = 1, \dots, a, j = 1, \dots, b_i$  and is given by

$$\begin{aligned} & \tilde{S}_\alpha (\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \mathbf{S}_{11}, \dots, \mathbf{S}_{ab_a}) \\ &= \bar{\mathbf{Y}}^T \mathbf{S}^{-1/2} \left( \mathbf{I}_{br} - \mathbf{S}^{-1/2} \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{S}^{-1} \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{S}^{-1/2} \right) \mathbf{S}^{-1/2} \bar{\mathbf{Y}}. \end{aligned} \quad (32)$$

In the following, we describe the PB test for  $H_{0\alpha|\beta}$  in (27). Recall that under  $H_{0\alpha|\beta}$  the vector  $\bar{\mathbf{Y}}$  have the mean  $\mathbf{X}_1 \boldsymbol{\mu}_0$ . As the test statistic in (32) is location invariant under the group of location transformations  $\mathcal{G}_1 = \{\bar{\mathbf{Y}} + \mathbf{X}_1 \boldsymbol{\eta}, \boldsymbol{\eta} \in \mathbb{R}^r\}$ , without loss of generality, we can take  $\mathbf{X}_1 \boldsymbol{\mu}_0 = \mathbf{0}$ . Using these facts, the parametric bootstrap pivot variable can be developed as follows. For a given  $(\bar{\mathbf{y}}_{11}, \dots, \bar{\mathbf{y}}_{ab_a}; \mathbf{s}_{11}, \dots, \mathbf{s}_{ab_a})$ , let  $\bar{\mathbf{Y}}_{Bij} \sim N_r(0, (1/n_{ij})\mathbf{s}_{ij})$  and  $\mathbf{S}_{Bij} \sim W_r(n_{ij} - 1, (1/(n_{ij} - 1))\mathbf{s}_{ij}), i = 1, \dots, a, j = 1, \dots, b_i$ . Then the PB pivot variable based on the test statistic (32) is given by

$$\begin{aligned} & \tilde{S}_{\alpha B} (\bar{\mathbf{Y}}_{B11}, \dots, \bar{\mathbf{Y}}_{Bab_a}; \mathbf{S}_{B11}, \dots, \mathbf{S}_{Bab_a}) \\ &= \bar{\mathbf{Y}}_B^T \mathbf{S}_B^{-1/2} \left( \mathbf{I}_{br} - \mathbf{S}_B^{-1/2} \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{S}_B^{-1} \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{S}_B^{-1/2} \right) \mathbf{S}_B^{-1/2} \bar{\mathbf{Y}}_B, \end{aligned} \quad (33)$$

where  $\bar{\mathbf{Y}}_B = (\bar{\mathbf{Y}}_{B11}^T, \dots, \bar{\mathbf{Y}}_{Bab_a}^T)^T$  and  $\mathbf{S}_B = \text{diag}((1/n_{11})\mathbf{S}_{B11}, \dots, (1/n_{ab_a})\mathbf{S}_{Bab_a})$ . For a given level  $\lambda$ , the PB test rejects  $H_{0\alpha|\beta}$  in (27) when

$$P(\tilde{S}_{\alpha B} (\bar{\mathbf{Y}}_{B11}, \dots, \bar{\mathbf{Y}}_{Bab_a}; \mathbf{S}_{B11}, \dots, \mathbf{S}_{Bab_a}) > \tilde{s}_\alpha) < \lambda, \quad (34)$$

where

$$\tilde{s}_\alpha = \tilde{S}_\alpha (\bar{\mathbf{y}}_{11}, \dots, \bar{\mathbf{y}}_{ab_a}; \mathbf{s}_{11}, \dots, \mathbf{s}_{ab_a}) \quad (35)$$

is an observed value of  $\tilde{S}_\alpha (\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \mathbf{S}_{11}, \dots, \mathbf{S}_{ab_a})$  in (32). For fixed  $(\bar{\mathbf{y}}_{11}, \dots, \bar{\mathbf{y}}_{ab_a}; \mathbf{s}_{11}, \dots, \mathbf{s}_{ab_a})$ , the above probability does not depend on any unknown parameters, and so it can be estimated using Monte Carlo simulation given in the following Algorithm 6.

*Algorithm 6.* For a given  $(n_{11}, \dots, n_{ab_a}), (\bar{\mathbf{y}}_{11}, \dots, \bar{\mathbf{y}}_{ab_a})$ , and  $(\mathbf{s}_{11}, \dots, \mathbf{s}_{ab_a})$ ,

Compute  $\tilde{S}_\alpha (\bar{\mathbf{y}}_{11}, \dots, \bar{\mathbf{y}}_{ab_a}; \mathbf{s}_{11}, \dots, \mathbf{s}_{ab_a})$  in (32) and call it  $\tilde{s}_\alpha$ .

For  $h = 1, \dots, m$ ,

Generate  $\bar{\mathbf{Y}}_{Bij} \sim N_r(0, (1/n_{ij})\mathbf{s}_{ij})$  and  $\mathbf{S}_{Bij} \sim W_r(n_{ij} - 1, (1/(n_{ij} - 1))\mathbf{s}_{ij})$ ,

$i = 1, \dots, a, j = 1, \dots, b_i$ ,

Compute  $\tilde{S}_{\alpha B} (\bar{\mathbf{Y}}_{B11}, \dots, \bar{\mathbf{Y}}_{Bab_a}; \mathbf{S}_{B11}, \dots, \mathbf{S}_{Bab_a})$  using (33).

If  $\tilde{S}_{\alpha B} (\bar{\mathbf{Y}}_{B11}, \dots, \bar{\mathbf{Y}}_{Bab_a}; \mathbf{S}_{B11}, \dots, \mathbf{S}_{Bab_a}) > \tilde{s}_\alpha$ , set  $Q_h = 1$ .

(End loop)

$(1/m) \sum_{h=1}^m Q_h$  is a Monte Carlo estimate of the  $P$  value in (34).

Note that the PB test provided in this section has the same invariance properties as in Section 2.

### 4. Monte Carlo Studies

In this section, we need to compare the PB test with the AHT test via comparing their empirical sizes (Type I error



rates) and powers for two-factor nested MANOVA models via simulations. As pointed out in Section 3, the problem of testing (27) is the same case of the problem of comparing  $b = \sum_{i=1}^a b_i$  normal mean vectors when the cell covariance matrices are unknown and arbitrary. Due to this reason, we will look only at the nested effects for comparisons.

Let the two factors be  $A$  and  $B$ , respectively. Suppose that the nesting effect of the factor  $A$  has  $a$  factor levels, and the nested effect of the factor  $B$  has a total of  $b = \sum_{i=1}^a b_i$  levels with  $b_i$  levels of  $B$  nested within the  $i$ th level of the factor  $A$ . Let  $\mathbf{n} = [n_{11}, n_{12}, \dots, n_{ab}]$  denote the vector of cell sizes. For given  $\mathbf{n}$  and covariance matrices  $\Sigma_{ij}$ ,  $i = 1, 2, \dots, a$ ,  $j = 1, 2, \dots, b_i$ , we first generate  $\sum_{i=1}^a \sum_{j=1}^{b_i} n_{ij}$  multivariate samples as

$$\mathbf{y}_{ijk} = \boldsymbol{\mu}_{ij} + \Sigma_{ij}^{1/2} \mathbf{e}_{ijk}, \quad k = 1, 2, \dots, n_{ij}, \quad (36)$$

where the cell mean vectors  $\boldsymbol{\mu}_{ij} = \boldsymbol{\mu}_{11} + ij\delta\mathbf{h}/b$  with  $\boldsymbol{\mu}_{11}$  being the first cell mean vector,  $\mathbf{h}$  a constant unit vector specifying the direction of the cell mean differences, and  $\delta$  a tuning parameter controlling the amount of the cell mean differences. We independently generate the  $r$  entries of the error terms  $\mathbf{e}_{ijk}$  from the  $N(0, 1)$  distribution so that we always have  $E(\mathbf{e}_{ijk}) = \mathbf{0}$  and  $\text{Cov}(\mathbf{e}_{ijk}) = \mathbf{I}_r$ . This means that (36) will generate the  $(ij)$ th multivariate normal sample  $\mathbf{y}_{ijk}$ ,  $k = 1, 2, \dots, n_{ij}$  with the given mean vector  $\boldsymbol{\mu}_{ij}$  and covariance matrix  $\Sigma_{ij}$ . Without loss of generality, we specify  $\boldsymbol{\mu}_{11}$  as  $\mathbf{0}$  and  $\mathbf{h}$  as  $\mathbf{h}_0/\|\mathbf{h}_0\|$  where  $\mathbf{h}_0 = [1, 2, \dots, r]^T$  for any given dimension  $r$  and  $\|\mathbf{h}_0\|$  denotes the usual  $L^2$ -norm of  $\mathbf{h}_0$ . To estimate the sizes and powers of the AHT test, we used simulation consisting of 10,000 runs and recorded corresponding  $P$  values. Notice that two nested “do loops” are required to estimate the sizes and powers of the PB test, we used 2500 runs for outer “do loops” (for generating the data) and 5000 runs for inner “do loops” for estimating the bootstrap  $P$  values. The empirical sizes (when  $\delta = 0$ ) and powers (when  $\delta > 0$ ) of the two tests are the proportions of rejecting the null hypothesis, that is, when their  $P$  values are less than the nominal significance level  $\lambda$ . In all the simulations conducted, we used  $\lambda = 5\%$  for simplicity.

Notice that the PB test does not depend on chosen weights. Here we report only the comparative studies for the equal-weight method to specify the weights of the AHT test so that the AHT test in Zhang [8] may be used for showing properties of the PB test. The empirical sizes and powers of the two tests for nested effect tests, together with the associated tuning parameters, are presented in Tables 1–3, in the columns labeled with AHT and PB 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. To measure the overall performance of a test in terms of maintaining the nominal size  $\lambda$ , we use the average relative error defined in Zhang [8] as  $\text{ARE} = M^{-1} \sum_{m=1}^M |\hat{\lambda}_m - \lambda|/\lambda \times 100$  where  $\hat{\lambda}_m$  denotes the  $m$ th empirical size for  $m = 1, 2, \dots, M$ ,  $\lambda = 0.05$ , 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  $\text{ARE} \leq 10$ , the test performs very well; when  $10 < \text{ARE} \leq 20$ , 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 two 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  $(\mathbf{u}_t)$  to denote “ $\mathbf{u}$  repeats  $t$  times.” Table 1 shows the empirical sizes and powers of the two tests for a bivariate case with  $a = 2$  and  $b_1 = 16$ ,  $b_2 = 20$ . With  $b = b_1 + b_2 = 36$ , one may be able to check how the two 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 two tests for a three-variate case with  $a = 3$  and  $b_1 = 8$ ,  $b_2 = 10$ ,  $b_3 = 12$ , and a 10-variate case with  $a = 3$  and  $b_1 = 3$ ,  $b_2 = 5$ ,  $b_3 = 7$ , respectively. These two tables allow us to compare the two tests for higher-dimensional data.

In the following, let us compare the AHT and PB tests via examining their empirical sizes and powers. It is seen from the three tables that for all the cases, the PB test generally outperforms the AHT test for all the cases under consideration as shown by their empirical sizes and the associated ARE values presented in the three tables. The AHT test performs reasonably well only for 10-dimensional data. In the homogeneous cases, the AHT test appears to be more powerful than the PB tests because of its inflated empirical sizes exceeding the nominal level considerably. For the heteroscedastic cases, we once again observe from Tables 2 and 3 that the PB test performs superior to the AHT test in terms of controlling nominal level and powers.

We conclude this simulation section via summarizing all the simulation studies conducted. In terms of size, the overall conclusion is that the PB test is a flexible procedure that performs satisfactorily, regardless of the cell sizes, values of the cell covariance matrices, and the number of effects being compared. In terms of power, the PB test generally outperforms the AHT test for most heteroscedastic cases under consideration. Thus, one may recommend to use the PB test as a good alternative in practical applications because of its simplicity and accuracy.

## 5. Concluding Remarks

The available classical tests for the two-factor MANOVA model with crossed designs and heteroscedastic cell covariance matrices have serious Type I error problems that have been overlooked. To address this serious problem, Zhang [8] proposed the AHT test. In this paper, we proposed and studied the PB test and the AHT test for the two-factor MANOVA model with nested designs and heteroscedastic cell covariance matrices. We showed that the PB test is invariant under affine-transformations and different choices of weights used to define the parameters uniquely. We demonstrated via intensive simulations that the PB test generally performs well and outperforms the AHT test in terms of size and power for most cell sizes and parameter configurations.

TABLE 1: Empirical sizes and powers of the two tests for nested effects for bivariate two-factor multivariate nested designs [ $a = 2, b_1 = 16, b_2 = 20, (\Sigma_{11}, \Sigma_{12}, \dots, \Sigma_{ab_a}) = \mathbf{1}_{18}^T \otimes (\mathbf{I}_2, \text{diag}(\lambda))$ ].

$\lambda$	$\mathbf{n}$	$\delta = 0$		$\delta = 1$		$\delta = 2$		$\delta = 3$	
		AHT	PB	AHT	PB	AHT	PB	AHT	PB
$\lambda_1$	$\mathbf{n}_1$	0.056	0.046	0.210	0.162	0.831	0.793	0.999	0.999
	$\mathbf{n}_2$	0.057	0.050	0.394	0.372	0.995	0.994	1.000	1.000
	$\mathbf{n}_3$	0.070	0.052	0.295	0.232	0.957	0.930	1.000	1.000
	$\mathbf{n}_4$	0.057	0.054	0.926	0.923	1.000	1.000	1.000	1.000
$\lambda_2$	$\mathbf{n}_1$	0.058	0.055	0.108	0.124	0.379	0.551	0.826	0.964
	$\mathbf{n}_2$	0.068	0.044	0.173	0.232	0.718	0.930	0.995	1.000
	$\mathbf{n}_3$	0.071	0.053	0.152	0.150	0.549	0.717	0.958	0.995
	$\mathbf{n}_4$	0.059	0.051	0.475	0.776	0.999	1.000	1.000	1.000
$\lambda_3$	$\mathbf{n}_1$	0.056	0.045	0.093	0.114	0.301	0.507	0.722	0.943
	$\mathbf{n}_2$	0.058	0.046	0.151	0.204	0.614	0.878	0.977	0.999
	$\mathbf{n}_3$	0.070	0.048	0.131	0.145	0.458	0.654	0.903	0.990
	$\mathbf{n}_4$	0.061	0.055	0.397	0.755	0.994	1.000	1.000	1.000
ARE		24.03	6.83						

$\lambda_1 = (1, 1), \lambda_2 = (1, 5),$  and  $\lambda_3 = (1, 10).$   $\mathbf{n}_1 = (7, 7)_{20}, \mathbf{n}_2 = (10, 10)_{20}, \mathbf{n}_3 = (7, 10)_{20},$  and  $\mathbf{n}_4 = (30, 15)_{20}.$

TABLE 2: Empirical sizes and powers of the two tests for nested effects for 3-variate two-factor multivariate nested designs [ $a = 3, b_1 = 8, b_2 = 10, b_3 = 12, (\Sigma_{11}, \Sigma_{12}, \dots, \Sigma_{ab_a}) = \mathbf{1}_{10}^T \otimes (\mathbf{I}_3, \text{diag}(\lambda), (1 - \rho)\mathbf{I}_3 + \rho\mathbf{1}_3\mathbf{1}_3^T)$ ].

$(\lambda, \rho)$	$\mathbf{n}$	$\delta = 0$		$\delta = 0.6$		$\delta = 1.2$		$\delta = 1.8$	
		AHT	PB	AHT	PB	AHT	PB	AHT	PB
$(\lambda_1, \rho_1)$	$\mathbf{n}_1$	0.052	0.054	0.099	0.103	0.369	0.353	0.840	0.815
	$\mathbf{n}_2$	0.056	0.041	0.164	0.152	0.705	0.691	0.994	0.991
	$\mathbf{n}_3$	0.092	0.041	0.302	0.194	0.944	0.891	1.000	1.000
	$\mathbf{n}_4$	0.096	0.044	0.288	0.185	0.927	0.847	1.000	1.000
$(\lambda_2, \rho_2)$	$\mathbf{n}_1$	0.054	0.049	0.181	0.213	0.756	0.860	0.997	1.000
	$\mathbf{n}_2$	0.055	0.048	0.335	0.422	0.985	0.997	1.000	1.000
	$\mathbf{n}_3$	0.090	0.050	0.629	0.519	1.000	1.000	1.000	1.000
	$\mathbf{n}_4$	0.089	0.061	0.522	0.545	0.999	1.000	1.000	1.000
$(\lambda_3, \rho_3)$	$\mathbf{n}_1$	0.049	0.050	0.271	0.194	0.727	0.852	0.995	1.000
	$\mathbf{n}_2$	0.052	0.056	0.308	0.427	0.978	0.998	1.000	1.000
	$\mathbf{n}_3$	0.094	0.060	0.593	0.519	1.000	0.999	1.000	1.000
	$\mathbf{n}_4$	0.098	0.057	0.486	0.551	0.999	1.000	1.000	1.000
ARE		46.67	10.83						

$(\lambda_1, \rho_1) = (1, 3, 0), (\lambda_2, \rho_2) = (1, 15, 0.1, 0.1),$  and  $(\lambda_3, \rho_3) = (1, 10, 0.1, 0.5).$   $\mathbf{n}_1 = (10, 10, 10)_{10}, \mathbf{n}_2 = (15, 15, 15)_{10}, \mathbf{n}_3 = (10, 20, 40)_{10},$  and  $\mathbf{n}_4 = (40, 20, 10)_{10}.$

### Appendix

*Proof of Theorem 1.* When  $H_{0\beta}$  is true, the model (8) can be written as the following reduced model:

$$\bar{\mathbf{Y}} = \mathbf{X}\boldsymbol{\theta} + \bar{\mathbf{e}}, \quad \bar{\mathbf{e}} \sim N_{br}(\mathbf{0}, \Sigma). \tag{A.1}$$

Premultiplying  $\Sigma^{-1/2}$  in model (A.1), we have the following transformed model:

$$\Sigma^{-1/2}\bar{\mathbf{Y}} = \Sigma^{-1/2}\mathbf{X}\boldsymbol{\theta} + \Sigma^{-1/2}\bar{\mathbf{e}}, \quad \Sigma^{-1/2}\bar{\mathbf{e}} \sim N_{br}(\mathbf{0}, \mathbf{I}). \tag{A.2}$$

Under the constraint  $\mathbf{L}\boldsymbol{\theta} = \mathbf{0}$ , the model (A.2) becomes

$$\Sigma^{-1/2}\bar{\mathbf{Y}} = \Sigma^{-1/2}\mathbf{X}\boldsymbol{\theta} + \Sigma^{-1/2}\bar{\mathbf{e}}, \quad \Sigma^{-1/2}\bar{\mathbf{e}} \sim N_{br}(\mathbf{0}, \mathbf{I}), \tag{A.3}$$

$$\mathbf{L}\boldsymbol{\theta} = \mathbf{0}.$$

In view of Theorem 5.2.5 in Wang and Chow [11], P. 177, we shall prove that  $\mathbf{L}\boldsymbol{\theta} = \mathbf{0}$  is a side condition on the matrix  $\Sigma^{-1/2}\mathbf{X}$ ; that is,  $\mathbf{L}$  satisfies conditions

$$\mathcal{R}\left((\Sigma^{-1/2}\mathbf{X})^T\right) \cap \mathcal{R}(\mathbf{L}^T) = \{\mathbf{0}\}, \tag{A.4}$$

$$\text{rank}\left(\begin{matrix} \Sigma^{-1/2}\mathbf{X} \\ \mathbf{L} \end{matrix}\right) = (a + 1)r,$$

where  $\mathcal{R}(\mathbf{M})$  denotes the subspace spanned by the column of matrix  $\mathbf{M}$ . Denote  $\mathbf{X}_0 = [\mathbf{1}_b, \text{diag}(\mathbf{1}_{b_1}, \dots, \mathbf{1}_{b_r})]$  and  $\mathbf{L}_0 = (0, u_1, \dots, u_a)$ , respectively. Then  $\mathbf{X} = \mathbf{X}_0 \otimes \mathbf{I}_r$ , and  $\mathbf{L} = \mathbf{L}_0 \otimes \mathbf{I}_r$ . Note that  $u_1, \dots, u_a$  are nonnegative weights such that

TABLE 3: Empirical sizes and powers of the two tests for nested effects for 10-variate two-factor multivariate nested designs [ $a = 3, b_1 = 3, b_2 = 5, b_3 = 7, (\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \dots, \boldsymbol{\Sigma}_{ab_a}) = \mathbf{1}_{10}^T \otimes (\mathbf{I}_3, \text{diag}(\boldsymbol{\lambda}), \text{diag}(\boldsymbol{\eta}))$ ].

$(\boldsymbol{\lambda}, \boldsymbol{\eta})$	$\mathbf{n}$	$\delta = 0$		$\delta = 1$		$\delta = 2$		$\delta = 3$	
		AHT	PB	AHT	PB	AHT	PB	AHT	PB
$(\boldsymbol{\lambda}_1, \boldsymbol{\eta}_1)$	$\mathbf{n}_1$	0.046	0.042	0.438	0.415	0.999	0.998	1.000	1.000
	$\mathbf{n}_2$	0.052	0.050	0.937	0.932	1.000	1.000	1.000	1.000
	$\mathbf{n}_3$	0.062	0.046	0.788	0.762	1.000	1.000	1.000	1.000
	$\mathbf{n}_4$	0.058	0.056	0.871	0.846	1.000	1.000	1.000	1.000
$(\boldsymbol{\lambda}_2, \boldsymbol{\eta}_2)$	$\mathbf{n}_1$	0.052	0.044	0.842	0.635	1.000	1.000	1.000	1.000
	$\mathbf{n}_2$	0.052	0.047	1.000	0.996	1.000	1.000	1.000	1.000
	$\mathbf{n}_3$	0.067	0.049	0.995	0.965	1.000	1.000	1.000	1.000
	$\mathbf{n}_4$	0.059	0.046	0.999	0.972	1.000	1.000	1.000	1.000
$(\boldsymbol{\lambda}_3, \boldsymbol{\eta}_3)$	$\mathbf{n}_1$	0.049	0.042	0.066	0.101	0.115	0.401	0.235	0.877
	$\mathbf{n}_2$	0.050	0.054	0.090	0.236	0.255	0.930	0.660	1.000
	$\mathbf{n}_3$	0.071	0.045	0.089	0.145	0.183	0.622	0.442	0.982
	$\mathbf{n}_4$	0.055	0.047	0.095	0.236	0.260	0.961	0.662	1.000
ARE		13.78	8.67						

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

$\sum_{i=1}^a u_i > 0$ . Then it can be easily verified that  $\mathbf{X}_0$  and  $\mathbf{L}_0$  satisfy the following conditions:

$$\begin{aligned} \mathcal{R}(\mathbf{X}_0^T) \cap \mathcal{R}(\mathbf{L}_0^T) &= \{\mathbf{0}\}, \\ \text{rank}\left(\begin{pmatrix} \mathbf{X}_0 \\ \mathbf{L}_0 \end{pmatrix}\right) &= a + 1. \end{aligned} \quad (\text{A.5})$$

Assume that  $\mathbf{d} \in \mathcal{R}(\mathbf{X}^T) \cap \mathcal{R}(\mathbf{L}^T)$ . Then there exist two vectors  $\mathbf{g}$  and  $\mathbf{h}$  such that

$$\mathbf{d} = \mathbf{X}^T \mathbf{g} = \mathbf{L}^T \mathbf{h}. \quad (\text{A.6})$$

Denote

$$\mathbf{X}_0^T = \begin{pmatrix} x_{11} & \cdots & x_{1b} \\ \vdots & \dots & \vdots \\ x_{a+1,1} & \cdots & x_{a+1,b} \end{pmatrix}, \quad \mathbf{L}_0^T = \begin{pmatrix} l_{11} \\ \vdots \\ l_{a+1,1} \end{pmatrix}, \quad (\text{A.7})$$

and  $\mathbf{g} = (\mathbf{g}_1^T, \mathbf{g}_2^T, \dots, \mathbf{g}_b^T)^T$ , where  $\mathbf{g}_i = (g_{i1}, \dots, g_{ir})^T$  and  $\mathbf{h} = (h_1, \dots, h_r)^T$  are  $r \times 1$  vectors, respectively,  $i = 1, 2, \dots, b$ . Then (A.6) may be rewritten as

$$\mathbf{X}^T \mathbf{g} = \begin{pmatrix} x_{11} \mathbf{g}_1 + \cdots + x_{1b} \mathbf{g}_b \\ \vdots \\ x_{a+1,1} \mathbf{g}_1 + \cdots + x_{a+1,b} \mathbf{g}_b \end{pmatrix} = \begin{pmatrix} l_{11} \mathbf{h} \\ \vdots \\ l_{a+1,1} \mathbf{h} \end{pmatrix} = \mathbf{L}^T \mathbf{h}, \quad (\text{A.8})$$

which is equivalent to  $\mathbf{X}_0^T \mathbf{g}_k^* = \mathbf{L}_0^T h_k$ , where  $\mathbf{g}_k^* = (g_{1k}, g_{2k}, \dots, g_{bk})^T, k = 1, \dots, r$ . It follows from (A.5) that

$$\mathbf{X}_0^T \mathbf{g}_k^* = \mathbf{L}_0^T h_k, \quad k = 1, \dots, r, \quad (\text{A.9})$$

and hence

$$\begin{aligned} \mathbf{d} = \mathbf{X}^T \mathbf{g} &= \begin{pmatrix} x_{11} \mathbf{g}_1 + \cdots + x_{1b} \mathbf{g}_b \\ \vdots \\ x_{a+1,1} \mathbf{g}_1 + \cdots + x_{a+1,b} \mathbf{g}_b \end{pmatrix} = \begin{pmatrix} l_{11} \mathbf{h} \\ \vdots \\ l_{a+1,1} \mathbf{h} \end{pmatrix} \\ &= \mathbf{L}^T \mathbf{h} = \mathbf{0}. \end{aligned} \quad (\text{A.10})$$

This means that

$$\mathcal{R}(\mathbf{X}^T) \cap \mathcal{R}(\mathbf{L}^T) = \{\mathbf{0}\}. \quad (\text{A.11})$$

It follows from  $\mathcal{R}(\mathbf{X}^T) = \mathcal{R}(\mathbf{X}^T \boldsymbol{\Sigma}^{-1/2})$  that

$$\mathcal{R}\left(\left(\boldsymbol{\Sigma}^{-1/2} \mathbf{X}\right)^T\right) \cap \mathcal{R}(\mathbf{L}^T) = \{\mathbf{0}\}. \quad (\text{A.12})$$

On the other hand,

$$\begin{aligned} \text{rank}\left(\begin{pmatrix} \mathbf{X} \\ \mathbf{L} \end{pmatrix}\right) &= \text{rank}\left(\left(\begin{pmatrix} \mathbf{X}_0 \\ \mathbf{L}_0 \end{pmatrix} \otimes \mathbf{I}_r\right)\right) \\ &= \text{rank}\left(\begin{pmatrix} \mathbf{X}_0 \\ \mathbf{L}_0 \end{pmatrix}\right) \times \text{rank}(\mathbf{I}_r) = (a + 1)r. \end{aligned} \quad (\text{A.13})$$

The second equation of (A.13) may be derived easily from Theorem 2.2.1 in Wang and Chow [11]. It follows from (A.13) that

$$\begin{aligned} \text{rank}\left(\begin{pmatrix} \boldsymbol{\Sigma}^{-1/2} \mathbf{X} \\ \mathbf{L} \end{pmatrix}\right) &= \text{rank}\left(\left(\begin{pmatrix} \boldsymbol{\Sigma}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{L} \end{pmatrix}\right)\right) \\ &= \text{rank}\left(\begin{pmatrix} \mathbf{X} \\ \mathbf{L} \end{pmatrix}\right) = (a + 1)r. \end{aligned} \quad (\text{A.14})$$

Combining (A.12) with (A.14), we obtain that  $\mathbf{L}\boldsymbol{\theta} = \mathbf{0}$  is a side condition on the matrix  $\boldsymbol{\Sigma}^{-1/2} \mathbf{X}$ . Thus it follows from Lemma 5.2.2 in Wang and Chow [11] that

$$\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{L}^T \mathbf{L})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1/2} = \mathbf{X}^T \boldsymbol{\Sigma}^{-1/2}. \quad (\text{A.15})$$



Thus, we obtain that  $\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{L}^T \mathbf{L})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} = \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}$ . This imply that  $(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{L}^T \mathbf{L})^{-1}$  is a generalized inverse of  $\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}$ . Notice from Corollary 2.2.3 in Wang and Chow [11] that  $\boldsymbol{\Sigma}^{-1/2} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1/2}$  is invariant with respect to the choice of the involved generalized  $(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1}$ , which leads to  $\boldsymbol{\Sigma}^{-1/2} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1/2} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{L}^T \mathbf{L})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1/2}$ . Therefore, It follows that

$$\begin{aligned} \mathbf{X} \hat{\boldsymbol{\theta}} &= \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{L}^T \mathbf{L})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \bar{\mathbf{Y}} \\ &= \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Sigma}^{-1/2} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{L}^T \mathbf{L})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Sigma}^{-1/2} \bar{\mathbf{Y}} \\ &= \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Sigma}^{-1/2} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Sigma}^{-1/2} \bar{\mathbf{Y}} \\ &= \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \bar{\mathbf{Y}}. \end{aligned} \quad (\text{A.16})$$

Then we have

$$\begin{aligned} \tilde{\mathbf{s}}_{\beta} &(\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{ab_a}) \\ &= (\bar{\mathbf{Y}} - \mathbf{X} \hat{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{Y}} - \mathbf{X} \hat{\boldsymbol{\theta}}) \\ &= \bar{\mathbf{Y}}^T \boldsymbol{\Sigma}^{-1/2} (\mathbf{I}_{br} - \boldsymbol{\Sigma}^{-1/2} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1/2}) \boldsymbol{\Sigma}^{-1/2} \bar{\mathbf{Y}}, \end{aligned} \quad (\text{A.17})$$

as desired. Theorem 1 is proved.  $\square$

*Proof of Theorem 3.* Since the observed values of  $\mathbf{Y}_{ijk}$ ,  $k = 1, 2, \dots, n_{ij}$ , are denoted as  $\mathbf{y}_{ijk}$ , we let  $\mathbf{y}_{ijk}^*$  denote the observed values of the affine-transformed cell responses  $\mathbf{Y}_{ijk}^*$ ,  $k = 1, 2, \dots, n_{ij}$ , given by (25). Then we have  $\mathbf{y}_{ijk}^* = \mathbf{D} \mathbf{y}_{ijk} + \boldsymbol{\xi}$ . On the other hand, we obtain that the sample mean vector and the sample covariance matrix of the  $(i, j)$ th affine-transformed cell are  $\bar{\mathbf{Y}}_{ij}^* = \mathbf{D} \bar{\mathbf{Y}}_{ij} + \boldsymbol{\xi}$  and  $\mathbf{S}_{ij}^* = \mathbf{D} \mathbf{S}_{ij} \mathbf{D}^T$ , respectively. The observed values of these random variables are  $\bar{\mathbf{y}}_{ij}^* = \mathbf{D} \bar{\mathbf{y}}_{ij} + \boldsymbol{\xi}$  and  $\mathbf{s}_{ij}^* = \mathbf{D} \mathbf{s}_{ij} \mathbf{D}^T$ , respectively,  $i = 1, \dots, a$ ,  $j = 1, \dots, b_i$ . Let  $\bar{\mathbf{Y}}^* = (\bar{\mathbf{Y}}_{11}^*, \dots, \bar{\mathbf{Y}}_{1b_1}^*, \dots, \bar{\mathbf{Y}}_{a1}^*, \dots, \bar{\mathbf{Y}}_{ab_a}^*)^T$  and  $\bar{\mathbf{y}}^* = (\bar{\mathbf{y}}_{11}^*, \dots, \bar{\mathbf{y}}_{1b_1}^*, \dots, \bar{\mathbf{y}}_{a1}^*, \dots, \bar{\mathbf{y}}_{ab_a}^*)^T$ ; from the reduced model (A.1), we have the following affine-transformed reduced model:

$$\bar{\mathbf{Y}}^* - \boldsymbol{\xi}_b = \mathbf{D}_b \mathbf{X} \boldsymbol{\theta} + \mathbf{D}_b \bar{\boldsymbol{\epsilon}}, \quad \mathbf{D}_b \bar{\boldsymbol{\epsilon}} \sim N_{br}(\mathbf{0}, \mathbf{D}_b \boldsymbol{\Sigma} \mathbf{D}_b^T), \quad (\text{A.18})$$

where  $\mathbf{D}_b = \mathbf{I}_b \otimes \mathbf{D}$ ,  $\boldsymbol{\xi}_b = \mathbf{I}_b \otimes \boldsymbol{\xi}$ .

Denote  $\mathbf{s}^* = \text{diag}((1/n_{11})\mathbf{s}_{11}^*, \dots, (1/n_{1b_1})\mathbf{s}_{1b_1}^*, \dots, (1/n_{a1})\mathbf{s}_{a1}^*, \dots, (1/n_{ab_a})\mathbf{s}_{ab_a}^*)$ . Let  $\tilde{\mathbf{s}}_{\beta}^*$  be affine-transformed observed test statistic of  $\tilde{\mathbf{s}}_{\beta}$  (24). Then, by direct calculations, we have

$$\begin{aligned} \tilde{\mathbf{s}}_{\beta}^* &= \tilde{\mathbf{S}}_{\beta} (\bar{\mathbf{Y}}_{11}^* - \boldsymbol{\xi}, \dots, \bar{\mathbf{Y}}_{ab_a}^* - \boldsymbol{\xi}; \mathbf{s}_{11}^*, \dots, \mathbf{s}_{ab_a}^*) \\ &= (\bar{\mathbf{y}}^* - \boldsymbol{\xi}_b)^T \mathbf{s}^{*-1/2} (\mathbf{I}_{br} - \mathbf{s}^{*-1/2} \mathbf{D}_b \mathbf{X} (\mathbf{X}^T \mathbf{D}_b^T \mathbf{s}^{*-1} \mathbf{D}_b \mathbf{X})^{-1} \\ &\quad \times \mathbf{X}^T \mathbf{D}_b^T \mathbf{s}^{*-1/2}) \mathbf{s}^{*-1/2} (\bar{\mathbf{y}}^* - \boldsymbol{\xi}_b) \\ &= \bar{\mathbf{y}}^{*T} \mathbf{s}^{-1/2} (\mathbf{I}_{br} - \mathbf{s}^{-1/2} \mathbf{X} (\mathbf{X}^T \mathbf{s}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{s}^{-1/2}) \mathbf{s}^{-1/2} \bar{\mathbf{y}} = \tilde{\mathbf{s}}_{\beta}, \end{aligned} \quad (\text{A.19})$$

where  $\mathbf{s} = \text{diag}((1/n_{11})\mathbf{s}_{11}, \dots, (1/n_{1b_1})\mathbf{s}_{1b_1}, \dots, (1/n_{a1})\mathbf{s}_{a1}, \dots, (1/n_{ab_a})\mathbf{s}_{ab_a})$  is the observed values of  $\mathbf{S}$ . The affine-invariance of the observed test statistic  $\tilde{\mathbf{s}}_{\beta}$  (24) then follows.

Let

$$\bar{\mathbf{Y}}_{Bij}^* \sim N_r \left( 0, \frac{1}{n_{ij}} \mathbf{s}_{ij}^* \right), \quad \mathbf{S}_{Bij}^* \sim W_r \left( n_{ij} - 1, \frac{1}{n_{ij} - 1} \mathbf{s}_{ij}^* \right) \quad (\text{A.20})$$

be mutually independent,  $i = 1, \dots, a$ ,  $j = 1, \dots, b_i$ . Then  $\bar{\mathbf{Y}}_{Bij}^* \stackrel{d}{=} \mathbf{D} \bar{\mathbf{Y}}_{Bij}$  and  $\mathbf{S}_{Bij}^* \stackrel{d}{=} \mathbf{D} \mathbf{S}_{Bij} \mathbf{D}^T$ , where  $\mathbf{H} \stackrel{d}{=} \mathbf{G}$  means that  $\mathbf{H}$  and  $\mathbf{G}$  have the same distribution. Using these facts, the affine-transformed parametric bootstrap pivot variable can be given by

$$\begin{aligned} &\tilde{\mathbf{S}}_{\beta B}^* (\bar{\mathbf{Y}}_{B11}^*, \dots, \bar{\mathbf{Y}}_{Bab_a}^*; \mathbf{S}_{B11}^*, \dots, \mathbf{S}_{Bab_a}^*) \\ &= \bar{\mathbf{Y}}_B^{*T} \mathbf{S}_B^{*-1/2} (\mathbf{I}_{br} - \mathbf{S}_B^{*-1/2} \mathbf{D}_b \mathbf{X} (\mathbf{X}^T \mathbf{D}_b^T \mathbf{S}_B^{*-1} \mathbf{D}_b \mathbf{X})^{-1} \\ &\quad \times \mathbf{X}^T \mathbf{D}_b^T \mathbf{S}_B^{*-1/2}) \mathbf{S}_B^{*-1/2} \bar{\mathbf{Y}}_B^* \\ &\stackrel{d}{=} \bar{\mathbf{Y}}_B^T \mathbf{D}_b^T (\mathbf{D}_b \mathbf{S}_B \mathbf{D}_b^T)^{-1/2} (\mathbf{I}_{br} - (\mathbf{D}_b \mathbf{S}_B \mathbf{D}_b^T)^{-1/2} \mathbf{D}_b \mathbf{X} \\ &\quad \times (\mathbf{X}^T \mathbf{D}_b^T (\mathbf{D}_b \mathbf{S}_B \mathbf{D}_b^T)^{-1} \mathbf{D}_b \mathbf{X})^{-1} \\ &\quad \times \mathbf{X}^T \mathbf{D}_b^T (\mathbf{D}_b \mathbf{S}_B \mathbf{D}_b^T)^{-1/2}) \\ &\quad \times (\mathbf{D}_b \mathbf{S}_B \mathbf{D}_b^T)^{-1/2} \mathbf{D}_b \bar{\mathbf{Y}}_B \\ &= \bar{\mathbf{Y}}_B^T \mathbf{S}_B^{-1/2} (\mathbf{I}_{br} - \mathbf{S}_B^{-1/2} \mathbf{X} (\mathbf{X}^T \mathbf{S}_B^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}_B^{-1/2}) \mathbf{S}_B^{-1/2} \bar{\mathbf{Y}}_B \\ &= \tilde{\mathbf{S}}_{\beta B} (\bar{\mathbf{Y}}_{B11}, \dots, \bar{\mathbf{Y}}_{Bab_a}; \mathbf{S}_{B11}, \dots, \mathbf{S}_{Bab_a}), \end{aligned} \quad (\text{A.21})$$

where  $\bar{\mathbf{Y}}_B^* = (\bar{\mathbf{Y}}_{B11}^*, \dots, \bar{\mathbf{Y}}_{B1b_1}^*, \dots, \bar{\mathbf{Y}}_{Ba1}^*, \dots, \bar{\mathbf{Y}}_{Bab_a}^*)^T$  and  $\mathbf{S}_B^* = \text{diag}((1/n_{11})\mathbf{S}_{B11}^*, \dots, (1/n_{1b_1})\mathbf{S}_{B1b_1}^*, \dots, (1/n_{a1})\mathbf{S}_{Ba1}^*, \dots, (1/n_{ab_a})\mathbf{S}_{Bab_a}^*)$ . The affine-invariance of the distribution of PB pivot variable  $\tilde{\mathbf{s}}_{\beta B}$  (22) follows. Theorem 3 is then proved.  $\square$

*Proof of Theorem 4.* The proof of Theorem 4 is obvious and hence omitted here.  $\square$

*Proof of Theorem 5.* Using the same proof as that of Theorem 1, it follows that  $\mathbf{L} \boldsymbol{\theta} = \mathbf{0}$  is a side condition on the matrix  $\mathbf{S}^{-1/2} \mathbf{X}$ ; that is,  $\mathbf{L}$  satisfies conditions

$$\begin{aligned} &\mathcal{R} \left( (\mathbf{S}^{-1/2} \mathbf{X})^T \right) \cap \mathcal{R} (\mathbf{L}^T) = \{\mathbf{0}\}, \quad \text{a.e.}, \\ &\text{rank} \begin{pmatrix} \mathbf{S}^{-1/2} \mathbf{X} \\ \mathbf{L} \end{pmatrix} = (a+1)r, \quad \text{a.e.}, \end{aligned} \quad (\text{A.22})$$

where a.e. denotes ‘‘almost everywhere.’’ Thus it follows from the similar proof of Theorem 1 that

$$\mathbf{X} (\mathbf{X}^T \mathbf{S}^{-1} \mathbf{X} + \mathbf{L}^T \mathbf{L})^{-1} \mathbf{X}^T \mathbf{S}^{-1/2} = \mathbf{X} (\mathbf{X}^T \mathbf{S}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}^{-1/2}, \quad \text{a.e.} \quad (\text{A.23})$$

Then we obtain from (A.23) that the test statistic (17)

$$\begin{aligned} & \bar{S}_\beta (\bar{\mathbf{Y}}_{11}, \dots, \bar{\mathbf{Y}}_{ab_a}; \mathbf{S}_{11}, \dots, \mathbf{S}_{ab_a}) \\ &= \bar{\mathbf{Y}}^T \mathbf{S}^{-1/2} \left( \mathbf{I}_{br} - \mathbf{S}^{-1/2} \mathbf{X} (\mathbf{X}^T \mathbf{S}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}^{-1/2} \right) \mathbf{S}^{-1/2} \bar{\mathbf{Y}} \\ &= \bar{\mathbf{Y}}^T \mathbf{S}^{-1/2} \left( \mathbf{I}_{br} - \mathbf{S}^{-1/2} \mathbf{X} (\mathbf{X}^T \mathbf{S}^{-1} \mathbf{X} + \mathbf{L}^T \mathbf{L})^{-1} \mathbf{X}^T \mathbf{S}^{-1/2} \right) \mathbf{S}^{-1/2} \bar{\mathbf{Y}} \end{aligned} \quad (\text{A.24})$$

as desired. Theorem 5 is proved.  $\square$

## Conflict of Interests

The author declares that there is not any conflict of interests regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (11171002) and the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions (CIT&TCD201404002).

## References

- [1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, NY, USA, 3rd edition, 2003.
- [2] G. S. James, "Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown," *Biometrika*, vol. 41, pp. 19–43, 1954.
- [3] 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.
- [4] J. Gamage, T. Mathew, and S. Weerahandi, "Generalized  $p$ -values and generalized confidence regions for the multivariate Behrens-Fisher problem and MANOVA," *Journal of Multivariate Analysis*, vol. 88, no. 1, pp. 177–189, 2004.
- [5] K. Krishnamoorthy and F. Lu, "A parametric bootstrap solution to the MANOVA under heteroscedasticity," *Journal of Statistical Computation and Simulation*, vol. 80, no. 7–8, pp. 873–887, 2010.
- [6] 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.
- [7] J. Zhang and S. Xiao, "A note on the modified two-way MANOVA tests," *Statistics & Probability Letters*, vol. 82, no. 3, pp. 519–527, 2012.
- [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] S. Weerahandi, *Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York, NY, USA, 2004.
- [10] 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.
- [11] S. G. Wang and S. C. Chow, *Advanced Linear Models*, vol. 141, Marcel Dekker, New York, NY, USA, 1994.

- [12] S. R. Searle, *Linear Models*, John Wiley, New York, NY, USA, 1971.