Journal of Multivariate Analysis 115 (2013) 172–180

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A parametric bootstrap approach for two-way ANOVA in presence of possible interactions with unequal variances☆

Li-Wen Xu *, Fang-Qin Yang, Aji'erguli Abula, Shuang Qin

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

ARTICLE INFO

Article history:
Received 5 June 2012
Available online 22 October 2012

AMS 2000 subject classifications:
62F03
62F40
62J10

Keywords:
Bootstrap re-sampling
Generalized p-values
Heteroscedasticity
Unbalanced data

ABSTRACT

In this article we consider the Two-Way ANOVA model with unequal cell frequencies without the assumption of equal error variances. For the problem of testing no interaction effects and equal main effects, we propose a parametric bootstrap (PB) approach and compare it with existing the generalized F (GF) test. The Type I error rates and powers of the tests are evaluated using Monte Carlo simulation. Our studies show that the PB test performs better than the generalized F-test. The PB test performs very satisfactorily even for small samples while the GF test exhibits poor Type I error properties when the number of factorial combinations or treatments goes up.

1. Introduction

In a two-way ANOVA model with factors A and B, it is customary to assume that the cell variances are the same even when they are not. In fact, it is well known that without the assumption of equal error variances, under the conventional Neyman–Pearson theory, exact tests for testing the effects of factors A and B do not exist. When variances are unequal, classical F-tests which are calculated under the equal error variance assumption will provide only approximate solutions for testing the effects of factors A and B. The sizes of classical F-tests are fairly robust against the assumption of equal variances when the sample sizes are equal [4]. When the sample sizes are different, the sizes of F-tests can substantially exceed the intended size. Moreover, they suffer from serious lack of power even under moderate heteroscedasticity. The generalized F-test [1] is a recently developed solution which is based on an extended definition of the p-values [10]. However, [7] observed in the literature of ANOVA that some asymptotic procedures and the generalized F-test perform satisfactorily for a small number of treatments and/or moderate to large samples. For one-way ANOVA, they proposed a parametric bootstrap (PB) approach as a solution. The PB approach has been applied to solve a number of problems when conventional methods are difficult to apply or fail to provide exact solutions; see, for example, [8,9,6].

For testing the interaction effect, [3] carried out a simulation study to compare the performance of the generalized F-test and the classical F-test when the number of factorial combinations of factors A and B is small. In this case, the generalized F-test performs better than the classical F-test. As already pointed out, for a bigger number of factorial combinations, the type I error probability of the generalized F-test may far exceed the nominal level. Therefore, it is important to develop

☆ This work was supported in part by the National Natural Science Foundation of China (11171002), Beijing Natural Science Foundation (The Theory of Mixed Effects Models of Multivariate Complex Data and Its Applications; 1112008).

* Corresponding author.
E-mail address: xulw2000@yahoo.com.cn (L.-W. Xu).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.10.008

a test procedure for the interaction effect and the main effect with satisfactory Type-I error rate and power regardless of number of factorial combinations and the sample sizes. In the present paper, we will develop a parametric bootstrap (PB) approach. Bootstrap approach is a type of Monte Carlo method applied on observed data [5]. The bootstrap methods can be in either parametric or nonparametric settings. However, the problems addressed in this paper are in a strict parametric setting, namely the two-way ANOVA model with the usual normality assumptions. Therefore, we only propose a parametric bootstrap approach.

This article is organized as follows. For testing no interaction effect to the two-way ANOVA model with unequal cell frequencies unequal error variances in Section 2 and compare it with the generalized F-test. For the tests on main effects, we also propose a parametric bootstrap (PB) approach in Section 3. The methods are compared with respect to Type I error rates and powers using Monte Carlo simulation. Comparison studies in Section 4 show that the PB test performs better than generalized F-test. Some discussion and further remarks are provided in Section 5.

2. Tests for the interaction effects

Consider the two-way ANOVA model with factors A and B, with factor levels $A_1, \dots, A_a$ and $B_1, \dots, B_b$, respectively giving a total of ab factorial combinations or treatments. Suppose a random sample of size n_{ij} is available from ijth treatment, $i = 1, \dots, a;\ j = 1, \dots, b$. Let Y_{ijk}, $i = 1, \dots, a;\ j = 1, \dots, b;\ k = 1, \dots, n_{ij}$ represent these random variables and y_{ijk} represent their observed (sample) values. Assume that $n_{ij} > 1$ so that sample variances can be computed for each cell of the design. Sample mean and the sample variance of the ijth treatment are denoted by $\overline{Y}_{ij}$ and S^2_{ij}, $i = 1, \dots, a;\ j = 1, \dots, b$ respectively, where

$$\overline{Y}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk} \quad \text{and} \quad S^2_{ij} = \frac{1}{n_{ij}-1} \sum_{k=1}^{n_{ij}} (Y_{ijk} - \overline{Y}_{ij})^2.$$

The observed values of these random variables are denoted as $\bar{y}_{ij}$ and s^2_{ij}, $i = 1, \dots, a;\ j = 1, \dots, b$ respectively. Consider the two-way ANOVA model with unequal error variances:

$$\begin{aligned} &Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \\ &e_{ijk} \sim N(0, \sigma^2_{ij}), \quad i = 1, \dots, a;\ j = 1, \dots, b;\ k = 1, \dots, n_{ij}, \end{aligned} \tag{2.1}$$

where μ is the general mean, α_i is an effect due to the ith level of the factor A, β_j is an effect due to the jth level of the factor B, and γ_{ij} represents an effect due to the interaction of the factor level A_i and the factor level B_j. Writing $Y_{ij} = (Y_{ij1}, \dots, Y_{ijn_{ij}})'$, $Y = (Y'_{11}, \dots, Y'_{1b}, Y'_{21}, \dots, Y'_{2b}, \dots, Y'_{a1}, \dots, Y'_{ab})'$, $\alpha = (\alpha_1, \dots, \alpha_a)'$, $\beta = (\beta_1, \dots, \beta_b)'$, $\gamma = (\gamma_{11}, \dots, \gamma_{1b}, \gamma_{21}, \dots, \gamma_{2b}, \dots, \gamma_{a1}, \dots, \gamma_{ab})'$, the model (2.1) can be written as

$$Y = \mathbf{1}_{n_{\dots}} \mu + Z_1 \alpha + Z_2 \beta + Z_3 \gamma + e, \tag{2.2}$$

where $n_{\dots} = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}$ and e is defined similarly to Y. The design matrices Z_1, Z_2 and Z_3 are given by

$$\begin{aligned} &Z_1 = \mathrm{diag}(\mathbf{1}_{n_{1.}}, \dots, \mathbf{1}_{n_{a.}}), \\ &Z_2 = [\mathrm{diag}(\mathbf{1}'_{n_{11}}, \dots, \mathbf{1}'_{n_{1b}}), \mathrm{diag}(\mathbf{1}'_{n_{21}}, \dots, \mathbf{1}'_{n_{2b}}), \dots, \mathrm{diag}(\mathbf{1}'_{n_{a1}}, \dots, \mathbf{1}'_{n_{ab}})]', \\ &Z_3 = \mathrm{diag}(\mathbf{1}_{n_{11}}, \dots, \mathbf{1}_{n_{1b}}, \mathbf{1}_{n_{21}}, \dots, \mathbf{1}_{n_{2b}}, \dots, \mathbf{1}_{n_{a1}}, \dots, \mathbf{1}_{n_{ab}}), \end{aligned} \tag{2.3}$$

where $n_{i.} = \sum_{j=1}^{b} n_{ij}$, and $\mathbf{1}_k$ denotes the $k \times 1$ vector of ones, and $\mathrm{diag}(M_1, \dots, M_a)$ denotes a block-diagonal matrix with $M_1, \dots, M_a$ along the blocks.

In order to have μ, α_i, β_j, and γ_{ij} uniquely defined, we need to have additional constraints. Let $w_1, \dots, w_a$ and $v_1, \dots, v_b$ be nonnegative weights such that $\sum_{i=1}^{a} w_i > 0$ and $\sum_{j=1}^{b} v_j > 0$. We consider the following constraints

$$\sum_{i=1}^{a} w_i \alpha_i = 0, \qquad \sum_{j=1}^{b} v_j \beta_j = 0, \qquad \sum_{i=1}^{a} w_i \gamma_{ij} = 0, \qquad \sum_{j=1}^{b} v_j \gamma_{ij} = 0. \tag{2.4}$$

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

$$H_{0AB}:\ \gamma_{ij} = 0; \quad i = 1, \dots, a, j = 1, \dots, b \tag{2.5}$$

against its natural alternative hypothesis. From (2.1), the model for $\overline{Y}_{ij}$ is

$$\begin{aligned} &\overline{Y}_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \bar{e}_{ij}, \\ &\bar{e}_{ij} \sim N\left(0, \sigma^2_{ij}/n_{ij}\right), \quad i = 1, \dots, a;\ j = 1, \dots, b, \end{aligned} \tag{2.6}$$

where $\bar{e}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} e_{ijk}$. Writing $\overline{Y} = (\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab})'$, and $\bar{e} = (\bar{e}_{11}, \bar{e}_{12}, \ldots, \bar{e}_{ab})'$, the model (2.6) can be written as

$$\overline{Y} = \mathbf{1}_{ab}\mu + (I_g \otimes \mathbf{1}_b)\alpha + (\mathbf{1}_g \otimes I_b)\beta + \gamma + \bar{e}, \quad \bar{e} \sim N(0, \Sigma), \tag{2.7}$$

where $\Sigma = \operatorname{diag}\left(\sigma_{11}^2/n_{11}, \sigma_{12}^2/n_{12}, \ldots, \sigma_{ab}^2/n_{ab}\right)$, $\alpha = (\alpha_1, \ldots, \alpha_a)'$, $\beta = (\beta_1, \ldots, \beta_b)'$, $\gamma = (\gamma_{11}, \gamma_{12}, \ldots, \gamma_{ab})'$, I_k is an identity matrix with order k, $V \otimes W$ denotes the Kronecker product of matrices V and W. Define the standardized interaction sum of squares

$$\tilde{S}_I\left(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; \sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{ab}^2\right) = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{n_{ij}}{\sigma_{ij}^2}\left(\overline{Y}_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j\right)^2, \tag{2.8}$$

where $\hat{\mu}$, $\hat{\alpha}_i$ and $\hat{\beta}_j$ are solutions of μ, α_i and β_j that minimize the quadratic equation

$$\tilde{S}\left(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; \sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{ab}^2\right) = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{n_{ij}}{\sigma_{ij}^2}\left(\overline{Y}_{ij} - \mu - \alpha_i - \beta_j\right)^2 \tag{2.9}$$

subject to the constraints given in Eq. (2.4). In fact, denoting $\theta = (\mu, \alpha_1, \ldots, \alpha_a, \beta_1, \ldots, \beta_b)'$ and $\hat{\theta} = (\hat{\mu}, \hat{\alpha}_1, \ldots, \hat{\alpha}_a, \hat{\beta}_1, \ldots, \hat{\beta}_b)'$, it follows from Theorem 5.2.5 in [11] that

$$\hat{\theta} = (\hat{\mu}, \hat{\alpha}_1, \ldots, \hat{\alpha}_a, \hat{\beta}_1, \ldots, \hat{\beta}_b)' = (X'\Sigma^{-1}X + L'L)^{-1}X'\Sigma^{-1}\overline{Y}, \tag{2.10}$$

where $X = (\mathbf{1}_{ab}, I_a \otimes \mathbf{1}_b, \mathbf{1}_a \otimes I_b)$, $L = (l_1', l_2')'$, $l_1 = (0, w_1, \ldots, w_a, 0, \ldots, 0)$, and $l_2 = (0, 0, \ldots, 0, v_1, \ldots, v_b)$. It may be shown that

$$\begin{aligned}\tilde{S}_I\left(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; \sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{ab}^2\right) &= \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{n_{ij}}{\sigma_{ij}^2}\left(\overline{Y}_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j\right)^2 \\ &= \overline{Y}'\Sigma^{-1/2}\left(I - \Sigma^{-1/2}X(X'\Sigma^{-1}X)^{-}X'\Sigma^{-1/2}\right)\Sigma^{-1/2}\overline{Y}.\end{aligned} \tag{2.11}$$

Namely, $\tilde{S}_I\left(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; \sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{ab}^2\right)$ does not depend on L of chosen weights. If σ_{ij}^2's are known, then a natural statistic for testing (2.5) is $\tilde{S}_I\left(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; \sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{ab}^2\right)$. In fact, $\Sigma^{-1/2}\overline{Y} \sim N\left(\Sigma^{-1/2}\boldsymbol{\mu}, I_{ab}\right)$, and $C = \left(I - \Sigma^{-1/2}X(X'\Sigma^{-1}X)^{-}X'\Sigma^{-1/2}\right)$ is an idempotent matrix with rank $(a-1)(b-1)$, we have

$$\overline{Y}'\Sigma^{-1/2}C\Sigma^{-1/2}\overline{Y} \sim \chi^2_{(a-1)(b-1)}\left(\boldsymbol{\mu}'\Sigma^{-1/2}C\Sigma^{-1/2}\boldsymbol{\mu}\right),$$

where $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \ldots, \mu_{ab})'$, $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, $\chi_m^2(\delta)$ denotes a noncentral chi-square random variable with degrees of freedom m and noncentrality parameter δ. The noncentrality parameter

$$\boldsymbol{\mu}'\Sigma^{-1/2}C\Sigma^{-1/2}\boldsymbol{\mu}$$

is equal to zero when $\gamma_{11} = \gamma_{12} = \cdots = \gamma_{ab}$. Let $\bar{y} = (\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab})'$ be the observed value of $\overline{Y}$. Then, the test that rejects H_{0AB} in (2.5) whenever

$$\tilde{S}_I\left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; \sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{ab}^2\right) > \chi^2_{(a-1)(b-1),\lambda}$$

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

In general, the variances σ_{ij}^2's are unknown; in this case, a test statistic can be obtained by replacing σ_{ij}^2 in (2.11) by S_{ij}^2, $i = 1, \ldots, a$, $j = 1, \ldots, b$, and is given by

$$\tilde{S}_I\left(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; S_{11}^2, S_{12}^2, \ldots, S_{ab}^2\right) = \overline{Y}'\mathbf{S}^{-1/2}\left(I - \mathbf{S}^{-1/2}X(X'\mathbf{S}^{-1}X)^{-}X'\mathbf{S}^{-1/2}\right)\mathbf{S}^{-1/2}\overline{Y}, \tag{2.12}$$

where $\mathbf{S} = \operatorname{diag}\left(S_{11}^2/n_{11}, S_{12}^2/n_{12}, \ldots, S_{ab}^2/n_{ab}\right)$.

In the following, we describe the generalized F test due to [1] and the PB test.

2.1. The generalized F (GF) test

We shall now describe Ananda and Weerahandi's [1] generalized F test. A generalized test variable is given by

$$\begin{aligned}\mathrm{GV} &= \frac{\tilde{S}_I\left(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; \sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{ab}^2\right)}{\tilde{S}_I\left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; \sigma_{11}^2 s_{11}^2/S_{11}^2, \sigma_{12}^2 s_{12}^2/S_{12}^2, \ldots, \sigma_{ab}^2 s_{ab}^2/S_{ab}^2\right)} \\ &= \frac{\tilde{S}_I\left(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; \sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{ab}^2\right)}{\tilde{S}_I\left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; (n_{11}-1)s_{11}^2/U_{11}, (n_{12}-1)s_{12}^2/U_{12}, \ldots, (n_{ab}-1)s_{ab}^2/U_{ab}\right)},\end{aligned}$$

where $U_{11}, U_{12}, \ldots, U_{ab}$ are independent random variables with $U_{ij} = (n_{ij}-1)S_{ij}^2/\sigma_{ij}^2 \sim \chi^2_{n_{ij}-1}, i = 1, \ldots, a, j = 1, \ldots, b$. Furthermore, $\tilde{S}_I\left(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; \sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{ab}^2\right) \sim \chi^2_{(a-1)(b-1)}$ independently of $U_{11}, U_{12}, \ldots, U_{ab}$. The "observed value" of GV is defined as the value of GV at $\left(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; S_{11}^2, S_{12}^2, \ldots, S_{ab}^2\right) = \left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; s_{11}^2, s_{12}^2, \ldots, s_{ab}^2\right)$, and this observed value is 1. Therefore, for a given $\left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; s_{11}^2, s_{12}^2, \ldots, s_{ab}^2\right)$, the generalized p-value is given by

$$\begin{aligned} p &= P\left(\mathrm{GV} > 1 | H_{0AB}\right) \\ &= P\left(\frac{\chi^2_{(a-1)(b-1)}}{\tilde{S}_I\left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; (n_{11}-1)s_{11}^2/U_{11}, (n_{12}-1)s_{12}^2/U_{12}, \ldots, (n_{ab}-1)s_{ab}^2/U_{ab}\right)} > 1\right). \end{aligned} \tag{2.13}$$

The GF test rejects the null hypothesis in (2.5) whenever the generalized p-value in (2.13) is less than a given nominal level λ. Notice that, for a given $\left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; s_{11}^2, s_{12}^2, \ldots, s_{ab}^2\right)$, the probability in (2.13) does not depend on any unknown parameters, so it can be estimated using Monte Carlo simulation or computed using the integral expression given in [1]. For further details on the generalized p-value idea, along with a number of examples, we refer to [1].

It should be noted that in general, the distribution of the generalized p-value may not be uniform (0, 1), and hence the generalized F test is not exact in the classical sense and its properties should be evaluated using Monte Carlo simulation.

2.2. The 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_{0AB} the vector $\overline{Y}$ has the mean $X\theta$, where $\theta = (\mu, \alpha_1, \ldots, \alpha_a, \beta_1, \ldots, \beta_b)'$. As the test statistic in (2.12) is location invariant under the group of location transformations $\mathcal{G} = \{\overline{Y} + X\eta, \eta \in \mathbb{R}^{a+b+1}\}$, without loss of generality, we can take $X\theta = 0$. Using these facts, the parametric bootstrap pivot variable can be developed as follows. For a given $\left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; s_{11}^2, s_{12}^2, \ldots, s_{ab}^2\right)$, let

$$\overline{Y}_{Bij} \sim N\left(0, s_{ij}^2/n_{ij}\right) \quad \text{and} \quad S_{Bij}^2 \sim s_{ij}^2\chi^2_{n_{ij}-1}/(n_{ij}-1), i = 1, \ldots, a, j = 1, \ldots, b.$$

Then the PB pivot variable based on the test statistic (2.12) is given by

$$\tilde{S}_{IB}\left(\overline{Y}_{B11}, \overline{Y}_{B12}, \ldots, \overline{Y}_{Bab}; S_{B11}^2, S_{B12}^2, \ldots, S_{Bab}^2\right) = \overline{Y}_B'\mathbf{S}_B^{-1/2}\left(I - \mathbf{S}_B^{-1/2}X(X'\mathbf{S}_B^{-1}X)^{-}X'\mathbf{S}_B^{-1/2}\right)\mathbf{S}_B^{-1/2}\overline{Y}_B, \tag{2.14}$$

where $\overline{Y}_B = (\overline{Y}_{B11}, \overline{Y}_{B12}, \ldots, \overline{Y}_{Bab})'$, and $\mathbf{S}_B = \mathrm{diag}\left(S_{B11}^2/n_{11}, S_{B12}^2/n_{12}, \ldots, S_{Bab}^2/n_{ab}\right)$. For a given level λ, the PB test rejects H_{0AB} in (2.5) when

$$P\left(\tilde{S}_{IB}\left(\overline{Y}_{B11}, \overline{Y}_{B12}, \ldots, \overline{Y}_{Bab}; S_{B11}^2, S_{B12}^2, \ldots, S_{Bab}^2\right) > \tilde{s}_I\right) < \lambda, \tag{2.15}$$

where

$$\tilde{s}_I = \tilde{S}_I\left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; s_{11}^2, s_{12}^2, \ldots, s_{ab}^2\right)$$

is an observed value of $\tilde{S}_I(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; S_{11}^2, S_{12}^2, \ldots, S_{ab}^2)$ in (2.12). For fixed $(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; s_{11}^2, s_{12}^2, \ldots, s_{ab}^2)$, the above probability does not depend on any unknown parameters, and so it can be estimated using Monte Carlo simulation given in Algorithm 1.

Algorithm 1. For a given $(n_{11}, n_{12}, \ldots, n_{ab})$, $(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab})$, and $(s_{11}^2, s_{12}^2, \ldots, s_{ab}^2)$:
compute $\tilde{S}_I\left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; s_{11}^2, s_{12}^2, \ldots, s_{ab}^2\right)$ in (2.12) and call it $\tilde{s}_I$
For $k = 1, \ldots, m$
generate $\overline{Y}_{Bij} \sim N\left(0, s_{ij}^2/n_{ij}\right)$ and $S_{Bij}^2 \sim s_{ij}^2\chi^2_{n_{ij}-1}/(n_{ij}-1), i = 1, \ldots, a, j = 1, \ldots, b$
compute $\tilde{S}_{IB}\left(\overline{Y}_{B11}, \overline{Y}_{B12}, \ldots, \overline{Y}_{Bab}; S_{B11}^2, S_{B12}^2, \ldots, S_{Bab}^2\right)$ using (2.14)
if $\tilde{S}_{IB}\left(\overline{Y}_{B11}, \overline{Y}_{B12}, \ldots, \overline{Y}_{Bab}; S_{B11}^2, S_{B12}^2, \ldots, S_{Bab}^2\right) > \tilde{s}_I$, set $Q_k = 1$
(end loop)
$(1/m)\sum_{k=1}^m Q_k$ is a Monte Carlo estimate of the p-value in (2.15).

3. Tests for the main effects

For the two-way classification model with no interaction, the hypotheses of interest are $H_{0A}: \alpha_1 = \alpha_2 = \cdots = \alpha_a$ and $H_{0B}: \beta_1 = \beta_2 = \cdots = \beta_b$, respectively. When the interaction between factors A and B is present, the main effect α_i can not reflect the effect of A_i because it depends on which level of factor B it is in. A popular solution to the problem (Searle (1971), Chapter 7), [2, Chapter 7] is not quite a test for $\alpha_i = 0$ in the presence of interactions, but rather to test the null hypothesis

$$H_{0A*}:\ \alpha_i+\gamma_{ij}=0,\quad i=1,\ldots,a,j=1,\ldots,b \tag{3.1}$$

subject to the constraint on β in (2.4). Similarly, in the case of the presence of interactions, we also want to test the null hypothesis

$$H_{0B*}:\ \beta_j+\gamma_{ij}=0,\quad i=1,\ldots,a,j=1,\ldots,b \tag{3.2}$$

subject to the constraint on α in (2.4). We only consider the test for H_{0A*} in (3.1), testing procedures for H_{0B*}, H_{0A} and H_{0B} can be derived similarly.

Define the standardized sum of squares due to factor A and the interaction

$$\tilde{S}_A\left(\overline{Y}_{11},\overline{Y}_{12},\ldots,\overline{Y}_{ab};\sigma_{11}^2,\sigma_{12}^2,\ldots,\sigma_{ab}^2\right)=\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{ij}}{\sigma_{ij}^2}\left(\overline{Y}_{ij}-\hat{\mu}-\hat{\beta}_j\right)^2, \tag{3.3}$$

where $\hat{\mu}$ and $\hat{\beta}_j$ are solutions of μ and β_j that minimize the quadratic equation

$$\tilde{S}_1\left(\overline{Y}_{11},\overline{Y}_{12},\ldots,\overline{Y}_{ab};\sigma_{11}^2,\sigma_{12}^2,\ldots,\sigma_{ab}^2\right)=\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{ij}}{\sigma_{ij}^2}\left(\overline{Y}_{ij}-\mu-\beta_j\right)^2 \tag{3.4}$$

subject to the constraints given in Eq. (2.4). In fact, denoting $\theta_1=(\mu,\beta_1,\ldots,\beta_b)'$ and $\hat{\theta}_1=(\hat{\mu},\hat{\beta}_1,\ldots,\hat{\beta}_b)'$, it follows from Theorem 5.2.5 in [11] that

$$\hat{\theta}_1=(\hat{\mu},\hat{\beta}_1,\ldots,\hat{\beta}_b)'=(X_1'\Sigma^{-1}X_1+L_1'L_1)^{-1}X_1'\Sigma^{-1}\overline{Y}, \tag{3.5}$$

where $X_1=(\mathbf{1}_{ab},\mathbf{1}_a\otimes I_b)$, and $L_1=(0,v_1,\ldots,v_b)$. It can be shown that

$$\begin{aligned}\tilde{S}_A\left(\overline{Y}_{11},\overline{Y}_{12},\ldots,\overline{Y}_{ab};\sigma_{11}^2,\sigma_{12}^2,\ldots,\sigma_{ab}^2\right)&=\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{ij}}{\sigma_{ij}^2}\left(\overline{Y}_{ij}-\hat{\mu}-\hat{\beta}_j\right)^2\\&=\overline{Y}'\Sigma^{-1/2}\left(I-\Sigma^{-1/2}X_1(X_1'\Sigma^{-1}X_1)^{-}X_1'\Sigma^{-1/2}\right)\Sigma^{-1/2}\overline{Y}.\end{aligned} \tag{3.6}$$

Namely, $\tilde{S}_A\left(\overline{Y}_{11},\overline{Y}_{12},\ldots,\overline{Y}_{ab};\sigma_{11}^2,\sigma_{12}^2,\ldots,\sigma_{ab}^2\right)$ does not depend on L_1. If σ_{ij}^2's are known, then a natural statistic for testing (3.1) is $\tilde{S}_A\left(\overline{Y}_{11},\overline{Y}_{12},\ldots,\overline{Y}_{ab};\sigma_{11}^2,\sigma_{12}^2,\ldots,\sigma_{ab}^2\right)$. In fact, $\Sigma^{-1/2}\overline{Y}\sim N\left(\Sigma^{-1/2}\boldsymbol{\mu},I_{ab}\right)$, and $C_1=(I-\Sigma^{-1/2}X_1(X_1'\Sigma^{-1}X_1)^{-}X_1'\Sigma^{-1/2})$ is an idempotent matrix with rank $(a-1)b$, we have

$$\overline{Y}'\Sigma^{-1/2}C_1\Sigma^{-1/2}\overline{Y}\sim\chi^2_{(a-1)b}\left(\boldsymbol{\mu}'\Sigma^{-1/2}C_1\Sigma^{-1/2}\boldsymbol{\mu}\right),$$

where $\boldsymbol{\mu}=(\mu_{11},\mu_{12},\ldots,\mu_{ab})'$, $\mu_{ij}=\mu+\alpha_i+\beta_j+\gamma_{ij}$. The noncentrality parameter $\boldsymbol{\mu}'\Sigma^{-1/2}C_1\Sigma^{-1/2}\boldsymbol{\mu}$ is equal to zero when $\alpha_i+\gamma_{ij}=0,i=1,\ldots,a,j=1,\ldots,b$. Let $\overline{y}=(\overline{y}_{11},\overline{y}_{12},\ldots,\overline{y}_{ab})'$ be the observed value of $\overline{Y}$. Then, the test that rejects H_{0A*} in (3.1) whenever

$$\tilde{S}_A\left(\overline{y}_{11},\overline{y}_{12},\ldots,\overline{y}_{ab};\sigma_{11}^2,\sigma_{12}^2,\ldots,\sigma_{ab}^2\right)>\chi^2_{(a-1)b,\lambda}$$

is a size λ test. In general, the variances σ_{ij}^2's are unknown; in this case, a test statistic can be obtained by replacing σ_{ij}^2 in (3.6) by S_{ij}^2, $i=1,\ldots,a,j=1,\ldots,b$, and is given by

$$\tilde{S}_A\left(\overline{Y}_{11},\overline{Y}_{12},\ldots,\overline{Y}_{ab};S_{11}^2,S_{12}^2,\ldots,S_{ab}^2\right)=\overline{Y}'\mathbf{S}^{-1/2}\left(I-\mathbf{S}^{-1/2}X_1(X_1'\mathbf{S}^{-1}X_1)^{-}X_1'\mathbf{S}^{-1/2}\right)\mathbf{S}^{-1/2}\overline{Y}, \tag{3.7}$$

where $\mathbf{S}=\operatorname{diag}\left(S_{11}^2/n_{11},S_{12}^2/n_{12},\ldots,S_{ab}^2/n_{ab}\right)$.

In the following, we describe the PB test for H_{0A*} in (3.1). Recall that under H_{0A*} the vector $\overline{Y}$ have the mean $X_1\theta_1$, where $\theta_1=(\mu,\beta_1,\ldots,\beta_b)'$. As the test statistic in (3.7) is location invariant under the group of location transformations $\mathcal{G}=\{\overline{Y}+X_1\eta,\eta\in\mathbb{R}^{b+1}\}$, without loss of generality, we can take $X_1\theta_1=0$. Using these facts, the parametric bootstrap pivot variable can be developed as follows. For a given $\left(\overline{y}_{11},\overline{y}_{12},\ldots,\overline{y}_{ab};s_{11}^2,s_{12}^2,\ldots,s_{ab}^2\right)$, let

$$\overline{Y}_{Bij}\sim N\left(0,s_{ij}^2/n_{ij}\right)\quad\text{and}\quad S_{Bij}^2\sim s_{ij}^2\chi^2_{n_{ij}-1}/(n_{ij}-1),i=1,\ldots,a,j=1,\ldots,b.$$

Then the PB pivot variable based on the test statistic (3.7) is given by

$$\tilde{S}_{AB}\left(\overline{Y}_{B11}, \overline{Y}_{B12}, \ldots, \overline{Y}_{Bab}; S^2_{B11}, S^2_{B12}, \ldots, S^2_{Bab}\right) = \overline{Y}'_B \mathbf{S}_B^{-1/2}\left(I - \mathbf{S}_B^{-1/2} X_1 (X_1' \mathbf{S}_B^{-1} X_1)^{-} X_1' \mathbf{S}_B^{-1/2}\right) \mathbf{S}_B^{-1/2} \overline{Y}_B, \tag{3.8}$$

where $\overline{Y}_B = (\overline{Y}_{B11}, \overline{Y}_{B12}, \ldots, \overline{Y}_{Bab})'$, and $\mathbf{S}_B = \text{diag}\left(S^2_{B11}/n_{11}, S^2_{B12}/n_{12}, \ldots, S^2_{Bab}/n_{ab}\right)$. For a given level λ, the PB test rejects H_{0A*} in (3.1) when

$$P\left(\tilde{S}_{AB}\left(\overline{Y}_{B11}, \overline{Y}_{B12}, \ldots, \overline{Y}_{Bab}; S^2_{B11}, S^2_{B12}, \ldots, S^2_{Bab}\right) > \tilde{s}_A\right) < \lambda, \tag{3.9}$$

where

$$\tilde{s}_A = \tilde{S}_A\left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; s^2_{11}, s^2_{12}, \ldots, s^2_{ab}\right)$$

is an observed value of $\tilde{S}_A(\overline{Y}_{11}, \overline{Y}_{12}, \ldots, \overline{Y}_{ab}; S^2_{11}, S^2_{12}, \ldots, S^2_{ab})$ in (3.7). For fixed $(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; s^2_{11}, s^2_{12}, \ldots, s^2_{ab})$, the above probability does not depend on any unknown parameters, and so it can be estimated using Monte Carlo simulation given in Algorithm 2.

Algorithm 2. For a given $(n_{11}, n_{12}, \ldots, n_{ab})$, $(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab})$, and $(s^2_{11}, s^2_{12}, \ldots, s^2_{ab})$:
compute $\tilde{S}_A\left(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; s^2_{11}, s^2_{12}, \ldots, s^2_{ab}\right)$ in (3.7) and call it $\tilde{s}_A$
For $k = 1, \ldots, m$
generate $\overline{Y}_{Bij} \sim N\left(0, s^2_{ij}/n_{ij}\right)$ and $S^2_{Bij} \sim s^2_{ij}\chi^2_{n_{ij}-1}/(n_{ij}-1)$, $i = 1, \ldots, a, j = 1, \ldots, b$
compute $\tilde{S}_{AB}\left(\overline{Y}_{B11}, \overline{Y}_{B12}, \ldots, \overline{Y}_{Bab}; S^2_{B11}, S^2_{B12}, \ldots, S^2_{Bab}\right)$ using (3.8)
if $\tilde{S}_{AB}\left(\overline{Y}_{B11}, \overline{Y}_{B12}, \ldots, \overline{Y}_{Bab}; S^2_{B11}, S^2_{B12}, \ldots, S^2_{Bab}\right) > \tilde{s}_A$, set $Q_k = 1$
(end loop)
$(1/m)\sum_{k=1}^{m} Q_k$ is a Monte Carlo estimate of the p-value in (3.9).

4. Numerical results

As pointed out in [1], testing the hypothesis H_{0A} or H_{0B} does depend on the chosen weights in the presence of possible interactions [2]. The literature on two-way unbalanced models provided several procedures (under homoscedasticity) for testing the main effects in the presence of possible interactions. Some methods tested main effects ignoring the presence of possible interactions. There was no common agreement (Fujikoshi (1993)) about the circumstances under which these alternative testing procedures should be used. The controversy was not about the derivation of the testing procedures, but about the appropriate weights. In many situations there are no natural weights to justify a particular procedure. Arnold [2] gave an excellent coverage of this problem and controversies behind it. Due to these reasons, we will look only at the interaction effect for comparisons as it does not depend on chosen weights.

We have observed in [3] that the generalized F-test and the classical F-test were evaluated for their validity for small number of treatments, and the generalized F-test has better size and power performance than the classical F-test. Hence, it is of interest to study the properties of the proposed test for larger number of treatments by an extensive simulation study including the behavior of the type I error rates and powers, respectively. In this section we perform the size and power comparison for the generalized F-test and PB test.

The Type I error rates of the two-way ANOVA tests are estimated using Monte Carlo simulation. It is easy to be seen that both the generalized F-test and PB test are location-scale invariant, and so we can take, without loss of generality, that $X\theta = 0$, $\gamma_{11} = \cdots = \gamma_{ab} = 0$, and $0 < \sigma^2_{ij} \leq 1$, $i = 1, \ldots, a; j = 1, \ldots, b$, in our simulation studies. Thus the sample statistics $\bar{y}_{ij}$ and s^2_{ij} will be generated independently as $\bar{y}_{ij} \sim N(0, \sigma^2_{ij})$ and $s^2_{ij} \sim \sigma^2_{ij}\chi^2_{n_{ij}-1}/(n_{ij}-1)$, $0 < \sigma^2_{ij} \leq 1$, $i = 1, \ldots, a; j = 1, \ldots, b$.

To estimate the Type I error rates of the GF and PB tests, we have used a two-step simulation. The Monte Carlo method used for estimating the Type I error rates of the PB test is as follows. For a given sample size and parameter configuration, we generated 1000 observed vectors $(\bar{y}_{11}, \bar{y}_{12}, \ldots, \bar{y}_{ab}; s^2_{11}, s^2_{12}, \ldots, s^2_{ab})$, and the observed value $\tilde{s}_I$ in (2.12) was computed for each of the generated vectors. For each of the generated $\tilde{s}_A$'s, we used 1000 runs to estimate the p-value in (2.15). Finally, the Type I error rate of the PB test was estimated by the proportion of the 1000 p-values that are less than the nominal level λ. The Type I error rates of the GF test were similarly estimated. Each test was carried out at the nominal level of $\lambda = 0.05$.

To estimate the powers of the GF and PB tests, we have used a similar simulation except for taking different interaction effect vector γ's. The computations were realized in the MATLAB environment.

In Table 1, we present the estimates of Type I error rates for $a = 2, b = 3$; $a = 6, b = 3$; and $a = 10, b = 3$, and sample sizes ranging from small to moderate. We observe the following from the numerical results in Table 1.

1. For $a = 2, b = 3$, it appears that the generalized F-test seems to be more conservative than the PB test for the unbalanced case. For $a = 6, b = 3$, the GF test and the PB test have similar Type I error rates. In the worst cases, the Type I error rates of both tests are around 0.06 when the nominal level is 0.05.

Table 1
Simulated Type I error rates of the tests.

$a = 2, b = 3$	$\mathbf{n} = (5, 5, 5, 5, 5, 5)$		$\mathbf{n} = (10, 10, 10, 10, 10, 10)$		$\mathbf{n} = (3, 3, 4, 5, 6, 6)$		$\mathbf{n} = (4, 6, 8, 12, 16, 20)$	
$(\sigma_{11}^2, \ldots, \sigma_{23}^2)$	PB	GF	PB	GF	PB	GF	PB	GF
(1.0, 1.0, 1.0, 1.0, 1.0, 1.0)	0.05	0.02	0.05	0.04	0.05	0.01	0.04	0.03
(0.1, 0.1, 0.1, 0.5, 0.5, 0.5)	0.05	0.03	0.05	0.03	0.04	0.01	0.04	0.03
(1.0, 1.0, 1.0, 0.5, 0.5, 0.5)	0.04	0.02	0.04	0.02	0.05	0.01	0.06	0.04
(0.1, 0.2, 0.3, 0.4, 0.5, 1.0)	0.04	0.02	0.04	0.05	0.03	0.01	0.06	0.03
(0.3, 0.9, 0.4, 0.7, 0.5, 1.0)	0.04	0.01	0.05	0.02	0.05	0.01	0.04	0.02
(0.01, 0.1, 0.1, 0.1, 0.1, 1)	0.06	0.02	0.04	0.04	0.05	0.02	0.06	0.04
$a = 6, b = 3$	$\mathbf{n_1}$		$\mathbf{n_2}$		$\mathbf{n_3}$		$\mathbf{n_4}$	
$(\sigma_{11}^2, \ldots, \sigma_{63}^2)$	PB	GF	PB	GF	PB	GF	PB	GF
$\boldsymbol{\sigma_1^2}$	0.04	0.05	0.04	0.06	0.04	0.05	0.04	0.06
$\boldsymbol{\sigma_2^2}$	0.05	0.05	0.05	0.04	0.03	0.05	0.06	0.03
$\boldsymbol{\sigma_3^2}$	0.05	0.05	0.06	0.05	0.05	0.05	0.06	0.05
$\boldsymbol{\sigma_4^2}$	0.05	0.03	0.05	0.05	0.03	0.04	0.04	0.05
$\boldsymbol{\sigma_5^2}$	0.05	0.06	0.05	0.06	0.05	0.04	0.05	0.06
$\boldsymbol{\sigma_6^2}$	0.05	0.04	0.05	0.04	0.05	0.05	0.04	0.03
$a = 10, b = 3$	$\mathbf{n_5}$		$\mathbf{n_6}$		$\mathbf{n_7}$		$\mathbf{n_8}$	
$(\sigma_{11}^2, \ldots, \sigma_{10,3}^2)$	PB	GF	PB	GF	PB	GF	PB	GF
$\boldsymbol{\sigma_7^2}$	0.05	0.06	0.05	0.07	0.03	0.07	0.04	0.07
$\boldsymbol{\sigma_8^2}$	0.04	0.08	0.05	0.07	0.03	0.08	0.05	0.06
$\boldsymbol{\sigma_9^2}$	0.04	0.08	0.05	0.08	0.03	0.08	0.05	0.07
$\boldsymbol{\sigma_{10}^2}$	0.04	0.07	0.04	0.07	0.03	0.07	0.05	0.06
$\boldsymbol{\sigma_{11}^2}$	0.04	0.10	0.04	0.06	0.03	0.10	0.04	0.07
$\boldsymbol{\sigma_{12}^2}$	0.04	0.07	0.05	0.07	0.03	0.07	0.05	0.05

PB—parametric bootstrap; GF—generalized F test; $\mathbf{n_1} = (5, \ldots, 5)$; $\mathbf{n_2} = (10, \ldots, 10)$; $\mathbf{n_3} = (3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 6)$; $\mathbf{n_4} = (4, 4, 4, 6, 6, 6, 8, 8, 8, 12, 12, 12, 16, 16, 16, 20, 20, 20)$; $\mathbf{n_5} = (5, \ldots, 5)$; $\mathbf{n_6} = (10, \ldots, 10)$; $\mathbf{n_7} = (3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6)$; $\mathbf{n_8} = (4, 4, 4, 4, 4, 6, 6, 6, 6, 6, 8, 8, 8, 8, 8, 12, 12, 12, 12, 12, 16, 16, 16, 16, 16, 20, 20, 20, 20, 20)$; $\boldsymbol{\sigma_1^2} = (1, \ldots, 1)$; $\boldsymbol{\sigma_2^2} = (0.1, 0.1, 0.2, 0.2, 0.3, 0.3, 0.4, 0.4, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.8, 0.8, 0.9, 0.9)$; $\boldsymbol{\sigma_3^2} = (0.1, 0.2, 0.3, 0.4, 0.5, 0.1, 0.2, 0.3, 0.4, 0.5, 0.1, 0.2, 0.3, 0.4, 0.5, 0.1, 0.2, 1.0)$; $\boldsymbol{\sigma_4^2} = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 1)$; $\boldsymbol{\sigma_5^2} = (0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 1)$; $\boldsymbol{\sigma_6^2} = (0.01, 0.01, 0.01, 0.05, 0.05, 0.05, 0.1, 0.1, 0.1, 0.5, 0.5, 0.5, 0.6, 0.6, 0.6, 0.8, 0.8, 1)$; $\boldsymbol{\sigma_7^2} = (1, \ldots, 1)$; $\boldsymbol{\sigma_8^2} = (0.1, 0.1, 0.1, 0.1, 0.2, 0.2, 0.2, 0.2, 0.3, 0.3, 0.3, 0.4, 0.4, 0.4, 0.5, 0.5, 0.6, 0.6, 0.6, 0.7, 0.7, 0.7, 0.8, 0.8, 0.8, 0.8, 0.9, 0.9, 0.9, 0.9)$; $\boldsymbol{\sigma_9^2} = (0.1, 0.2, 0.3, 0.4, 0.5, 1, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 0.1, 0.2, 0.3, 0.4, 0.5, 1)$; $\boldsymbol{\sigma_{10}^2} = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 1)$; $\boldsymbol{\sigma_{11}^2} = (1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1)$; $\boldsymbol{\sigma_{12}^2} = (0.01, 0.01, 0.01, 0.01, 0.01, 0.05, 0.05, 0.05, 0.05, 0.05, 0.1, 0.1, 0.1, 0.1, 0.1, 0.5, 0.5, 0.5, 0.5, 0.5, 0.6, 0.6, 0.6, 0.6, 0.6, 0.8, 0.8, 0.8, 0.8, 1)$.

2. We see from the reported Type I error rates for $a = 10$, $b = 3$ that the PB test controls the Type I errors satisfactorily. On the other hand, we see that the Type I errors of the GF test can be as large as 0.1 when $\lambda = 0.05$; this test, in general, appears to be liberal for moderate values of g.

In Table 2, we provide the powers of the two tests for $a = 2$, $b = 3$, $a = 6$, $b = 3$ and $a = 10$, $b = 3$. We once again observe from this table that the PB test controls the Type I errors very well. In one case, where the GF test was conservative under the null hypothesis, the PB test appears to be more powerful than the GF test, and the GF test appears to have the higher loss of power for moderate inequality of interaction effects. Both tests exhibit similar power properties provided the Type I error rates are close to each other. In some cases, the GF test appears to be more powerful than the PB tests because of its inflated Type I error rates exceeding the intended level of 0.05.

5. Discussion and further remarks

When heteroscedasticity is present or the number of treatments of the factors goes up, the classical F test and the generalized F test for the two-way ANOVA model with heteroscedastic error variances have serious Type I error problems that have been overlooked. This article develops an interesting and flexible bootstrap (PB) approach in order to arrive at a test procedure, and have compared the PB test with the generalized F test. Due to the reasons pointed in Section 4, the simulation study was restricted to the hypothesis about interactions. In view of the fact that the generalized F-test was evaluated for their validity for small number of treatments, we have studied the properties of the proposed test not only for small number of treatments but also for larger number of treatments by an extensive simulation study including the behavior of the type I error rates and powers, respectively. In terms of controlling the Type I error rate, the overall conclusion is that the PB test is better procedure that performs satisfactorily, regardless of the sample sizes, values of the error variances, and the number of levels being compared.

Table 2
Simulated powers of the tests.

$a=2, b=3$						
$\mathbf{n}_1$		$\boldsymbol{\gamma}=(\gamma_{11},\ldots,\gamma_{23})$				
$(\sigma_{11}^2,\ldots,\sigma_{23}^2)$	Tests	$\mathbf{0}$	$\boldsymbol{\gamma}_1$	$2\boldsymbol{\gamma}_1$	$3\boldsymbol{\gamma}_1$	$4\boldsymbol{\gamma}_1$
$\boldsymbol{\sigma}_1^2$	PB	0.06	0.21	0.69	0.96	1
	GF	0.02	0.09	0.52	0.92	0.99
$\boldsymbol{\sigma}_2^2$	PB	0.05	0.17	0.57	0.92	0.99
	GF	0.02	0.08	0.42	0.81	0.97
$\mathbf{n}_2$						
$\boldsymbol{\sigma}_1^2$	PB	0.05	0.25	0.76	0.98	1
	GF	0.02	0.14	0.58	0.95	1
$\boldsymbol{\sigma}_2^2$	PB	0.05	0.20	0.63	0.95	1
	GF	0.01	0.09	0.45	0.87	0.99
$a=6, b=3$						
$\mathbf{n}_3$		$\boldsymbol{\gamma}=(\gamma_{11},\ldots,\gamma_{63})$				
$(\sigma_{11}^2,\ldots,\sigma_{63}^2)$		$\mathbf{0}$	$\boldsymbol{\gamma}_2$	$2\boldsymbol{\gamma}_2$	$3\boldsymbol{\gamma}_2$	$4\boldsymbol{\gamma}_2$
$\boldsymbol{\sigma}_3^2$	PB	0.05	0.15	0.51	0.91	1
	GF	0.05	0.15	0.49	0.89	1
$\boldsymbol{\sigma}_4^2$	PB	0.05	0.08	0.27	0.53	0.85
	GF	0.04	0.10	0.25	0.57	0.87
$\mathbf{n}_4$						
$\boldsymbol{\sigma}_3^2$	PB	0.05	0.15	0.61	0.95	1
	GF	0.04	0.12	0.58	0.95	1
$\boldsymbol{\sigma}_4^2$	PB	0.06	0.08	0.28	0.66	0.92
	GF	0.06	0.12	0.29	0.66	0.91
$a=10, b=3$						
$\mathbf{n}_5$		$\boldsymbol{\gamma}=(\gamma_{11},\ldots,\gamma_{10,3})$				
$(\sigma_{11}^2,\ldots,\sigma_{10,3}^2)$		$\mathbf{0}$	$\boldsymbol{\gamma}_3$	$2\boldsymbol{\gamma}_3$	$3\boldsymbol{\gamma}_3$	$4\boldsymbol{\gamma}_3$
$\boldsymbol{\sigma}_5^2$	PB	0.05	0.08	0.25	0.60	0.90
	GF	0.05	0.11	0.25	0.62	0.90
$\boldsymbol{\sigma}_6^2$	PB	0.05	0.08	0.20	0.50	0.84
	GF	0.06	0.10	0.24	0.51	0.86
$\mathbf{n}_6$						
$\boldsymbol{\sigma}_5^2$	PB	0.04	0.09	0.29	0.71	0.94
	GF	0.04	0.10	0.31	0.70	0.96
$\boldsymbol{\sigma}_6^2$	PB	0.05	0.08	0.23	0.60	0.90
	GF	0.06	0.09	0.26	0.63	0.92

$\mathbf{n}_1$ = (15, 15, 20, 20, 25, 25); $\mathbf{n}_2$ = (15, 18, 21, 24, 27, 30); $\mathbf{n}_3$ = (15, 15, 15, 15, 15, 15, 20, 20, 20, 20, 20, 20, 25, 25, 25, 25, 25, 25); $\mathbf{n}_4$ = (15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32); $\mathbf{n}_5$ = (15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 25, 25, 25, 25, 25, 25, 25, 25, 25, 25); $\mathbf{n}_6$ = (15, 15, 16, 16, 17, 17, 18, 18, 19, 19, 20, 20, 21, 22, 23, 24, 25, 26, 27, 27, 28, 28, 29, 29, 30, 30, 31, 31, 32, 32); $\boldsymbol{\sigma}_1^2$ = (0.1, 0.2, 0.3, 0.4, 0.5, 1.0); $\boldsymbol{\sigma}_2^2$ = (0.3, 0.9, 0.4, 0.7, 0.5, 1.0); $\boldsymbol{\sigma}_3^2$ = (0.1, 0.1, 0.2, 0.2, 0.3, 0.3, 0.4, 0.4, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.8, 0.8, 0.9, 1.0); $\boldsymbol{\sigma}_4^2$ = (0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1); $\boldsymbol{\sigma}_5^2$ = (0.1, 0.1, 0.1, 0.1, 0.2, 0.2, 0.2, 0.2, 0.3, 0.3, 0.3, 0.4, 0.4, 0.4, 0.5, 0.5, 0.5, 0.6, 0.6, 0.6, 0.7, 0.7, 0.7, 0.8, 0.8, 0.8, 0.8, 0.9, 0.9, 0.9, 0.9); $\boldsymbol{\sigma}_6^2$ = (1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1); $\boldsymbol{\gamma}_1$ = (0, 0, −0.1, 0.1, 0.2, 0.4); $\boldsymbol{\gamma}_2$ = (0, 0, −0.1, −0.1, 0.1, 0.1, 0.2, 0.2, 0.3, 0.3, 0.4, 0.4, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7); $\boldsymbol{\gamma}_3$ = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, −0.1, −0.1, 0.1, 0.1, 0.2, 0.2, 0.3, 0.3, 0.4, 0.4, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.8, 0.8).

Regarding the parametric bootstrap methodology that we have proposed here, note that the bootstrap can obviously be carried out both parametrically and nonparametrically. However, the problems addressed in the present paper are in a strict parametric setting, namely the two-way fixed model with the usual normality assumptions, and heterogeneous error variances. Thus we have chosen to do the bootstrap parametrically. If the model assumptions are approximately correct, the robustness of parametric bootstrap results and the nonparametric bootstrap are worth evaluating.

In the present paper, we have considered the two-way ANOVA model in which all effects are fixed treatment effects. A future direction is to extend the above results to higher-way layout. In many applications involving the use of a mixed effects model, testing hypotheses concerning the unknown variance components is an important part of data analysis. Then, for testing hypotheses concerning variance components in setups in which exact F tests do not exist, how to provide an appropriate parametric bootstrap procedure will be a meaningful problem.

Acknowledgments

The authors are grateful to the referees and the Editors for the helpful comments and clarifying suggestions resulting in the present version.

References

[1] M.M.A. Ananda, S. Weerahandi, Two-way ANOVA with unequal cell frequencies and unequal variances, Statist. Sinica 7 (1997) 631–646.
[2] S.F. Arnold, The Theory of Linear Models and Multivariate Analysis, John Wiley, New York, 1981.
[3] P. Bao, M.M.A. Ananda, Performance of two-way ANOVA procedures when cell frequency and variances are unequal, Comm. Statist. Simulation Comput. 30 (4) (2001) 805–829.
[4] A.N. Dajani, T. Mathew, Comparison of some tests in the one-way ANOVA with unequal error variances, in: ASA Proceedings of the Joint Statistical Meetings, 2003, pp. 1149–1155.
[5] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, London, 1993.
[6] K. Krishnamoorthy, F. Lu, A parametric bootstrap solution to the MANOVA under heteroscedasticity, J. Stat. Comput. Simul. 80 (2010) 873–887.
[7] K. Krishnamoorthy, F. Lu, T. Mathew, A parametric bootstrap approach for ANOVA with unequal variances: fixed and random models, Comput. Statist. Data Anal. 51 (2007) 5731–5742.
[8] C.X. Ma, L.L. Tian, A parametric bootstrap approach for testing equality of inverse Gaussian means under heterogeneity, Comm. Statist. Simulation Comput. 38 (2009) 1153–1160.
[9] L.L. Tian, C.X. Ma, A. Vexler, A parametric bootstrap test for comparing heteroscedastic regression models, Comm. Statist. Simulation Comput. 38 (2009) 1026–1036.
[10] K.W. Tusi, S. Weerahandi, Generalized *p*-values in significance testing of hypotheses in the presence of nuisance parameters, J. Amer. Statist. Assoc. 84 (1989) 602–607.
[11] S.G. Wang, S.C. Chow, Advanced Linear Models, Marcel Dekker, New York, 1994.