

A modified Bartlett test for linear hypotheses in heteroscedastic one-way ANOVA*

JIN-TING ZHANG[†] AND XUEFENG LIU

In this paper, we propose and study a so-called modified Bartlett (MB) test for the general linear hypothesis testing (GLHT) problem in heteroscedastic one-way ANOVA. The MB test is easy to compute and implement via using the usual chi-squared distribution. The MB test is shown to be invariant under affine transformations, different choices of the contrast matrix used to define the same hypothesis and different labeling schemes of the population means. Simulation studies demonstrate that the MB test performs well and it outperforms or is comparable to some existing tests for the k -sample Behrens-Fisher problem, a special case of the GLHT problem. The MB test is illustrated using a real data example.

KEYWORDS AND PHRASES: Heteroscedastic one-way ANOVA, General linear hypothesis testing problem, k -Sample Behrens-Fisher problem, Modified Bartlett correction.

1. INTRODUCTION

For several decades, much attention has been paid for comparing k normal means under heteroscedasticity (Welch 1947, 1951; James 1951, 1954; Krutchkoff 1988; Wilcox 1988, 1989; Krishnamoorthy, Lu, and Mathew 2007 etc.). When only two normal means are involved, this problem is referred to as the Behrens-Fisher (BF) problem (Behrens 1929, Fisher 1935), and it has been well addressed in the literature. Among the tests proposed for the two-sample BF problem, Welch's (1947) approximate degrees of freedom (ADF) test is the most popular one. It has been well accepted and widely used in real data applications because of its simplicity and accuracy as argued by Krishnamoorthy, Lu, and Mathew (2007).

The problem for comparing k normal means under variance heteroscedasticity is usually referred to as the k -sample BF problem. A number of testing procedures have been proposed and studied on this problem, including Welch's (1951) ADF test, James' (1954) second-order test, Weerahandi's (1995) generalized F-test, and Krishnamoorthy,

Lu, and Mathew's (2007) parametric bootstrap (PB) test, etc. Although Welch's (1951) ADF test performs well when $k = 2$, its performance is unsatisfactory in terms of size controlling when k is large. In fact, Krishnamoorthy, Lu, and Mathew (2007) compared the Welch, generalized F , James' second order and their PB tests via intensive simulations and demonstrated that in terms of size controlling and power, their PB test generally performs the best, followed by James' (1954) second order test while the Welch and generalized F tests are sometimes very liberal when k is large. Since the PB test is time-consuming and James' (1954) second-order test has a very complicated form which prevents it from being widely used in real data analysis, it is still worthwhile to develop some simple test for the k -sample BF problem which is comparable to the PB test in terms of size controlling and power.

In this paper, we aim to propose and study such a simple and accurate test, namely, a modified Bartlett (MB) test, for the general linear hypothesis testing (GLHT) problem in heteroscedastic one-way ANOVA, which includes the k -sample BF problem as a special case. The MB test is obtained via an application of the modified Bartlett correction of Fujikoshi (2000) to a Wald-type statistic constructed for the GLHT problem. The MB test can be easily computed and implemented via the usual χ^2 -distribution. We show that the MB test is invariant under affine transformations, different choices of the contrast matrix used to define the same hypothesis and different labeling schemes of the population means. Simulation studies and real data applications show that the MB test outperforms the Welch test and is comparable to the PB test in terms of size controlling and power.

The rest of the paper is organized as follows. In Section 2, the MB test is developed and some of its important properties are discussed. Simulation studies are presented in Section 3. An application of the MB test to a real data set is given in Section 4. Some concluding remarks are given in Section 5. Technical proofs of the main results are outlined in the Appendix.

2. MAIN RESULTS

2.1 The MB test

Throughout this paper, let $N(\mu, \sigma^2)$ denote a normal distribution with mean μ and variance σ^2 . Given k independent normal samples $x_{ij}, j = 1, 2, \dots, n_i \sim N(\mu_i, \sigma_i^2)$,

*The work was supported by the NUS Academic Research Grant R-155-000-108-112. The authors thank the co-editor and two referees for their invaluable comments and suggestions which helped improve the paper substantially.

[†]Corresponding author.

$l = 1, 2, \dots, k$, the heteroscedastic one-way ANOVA is referred to the following testing problem:

$$(1) \quad H_0: \mu_1 = \mu_2 = \dots = \mu_k, \quad \text{versus} \quad H_1: H_0 \text{ is not true,}$$

without assuming the equality of variances σ_l^2 , $l = 1, 2, \dots, k$. This problem is also known as the k -sample BF problem. For this k -sample BF problem, several solutions are available in the literature, including Welch's (1951) ADF test and James' (1954) first and second order approximation solutions, Krutchkoff's (1988) modified F -test, and the PB test of Krishnamoorthy, Lu, and Mathew (2007) among others. The k -sample BF problem (1) can be written as a special case of the following GLHT problem:

$$(2) \quad H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{c}, \quad \text{vs} \quad H_1: \mathbf{C}\boldsymbol{\mu} \neq \mathbf{c},$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_k)^T$, $\mathbf{C}: q \times k$ is a known coefficient matrix with $\text{rank}(\mathbf{C}) = q$, and $\mathbf{c}: q \times 1$ is a known constant vector, often set to zero. In fact, the GLHT problem (2) reduces to the heteroscedastic one-way ANOVA (1) if we set $\mathbf{c} = \mathbf{0}$ and $\mathbf{C} = [\mathbf{I}_{k-1}, -\mathbf{1}_{k-1}]$ where \mathbf{I}_r and $\mathbf{1}_r$ denote the identity matrix of size r and the r -dimensional vector of ones respectively. Notice that \mathbf{C} is a contrast matrix and its choice is not unique for (1) and later we shall show that the MB test is invariant to different choices of \mathbf{C} . To propose and study the so-called MB test for the GLHT problem (2), for $l = 1, 2, \dots, k$, set

$$(3) \quad \hat{\mu}_l = \bar{x}_l = n_l^{-1} \sum_{j=1}^{n_l} x_{lj}, \\ \hat{\sigma}_l^2 = (n_l - 1)^{-1} \sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l)^2,$$

which are the sample mean and sample variance of the l -th sample respectively. Set $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_k)^T$, as an unbiased estimator of $\boldsymbol{\mu}$. Then we have $\hat{\boldsymbol{\mu}} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \text{diag}(\frac{\sigma_1^2}{n_1}, \frac{\sigma_2^2}{n_2}, \dots, \frac{\sigma_k^2}{n_k})$. It follows that $\mathbf{C}\hat{\boldsymbol{\mu}} - \mathbf{c} \sim N_q(\mathbf{C}\boldsymbol{\mu} - \mathbf{c}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)$. Thus, we can then construct the following Wald-type statistic:

$$(4) \quad T = (\mathbf{C}\hat{\boldsymbol{\mu}} - \mathbf{c})^T (\mathbf{C}\hat{\boldsymbol{\Sigma}}\mathbf{C}^T)^{-1} (\mathbf{C}\hat{\boldsymbol{\mu}} - \mathbf{c}),$$

where $\hat{\boldsymbol{\Sigma}} = \text{diag}(\frac{\hat{\sigma}_1^2}{n_1}, \frac{\hat{\sigma}_2^2}{n_2}, \dots, \frac{\hat{\sigma}_k^2}{n_k})$, an unbiased estimator of $\boldsymbol{\Sigma}$. We now re-express T as

$$(5) \quad T = \mathbf{z}^T \mathbf{W}^{-1} \mathbf{z},$$

where

$$(6) \quad \mathbf{z} = (\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)^{-1/2} (\mathbf{C}\hat{\boldsymbol{\mu}} - \mathbf{c}), \\ \mathbf{W} = \mathbf{H}\hat{\boldsymbol{\Sigma}}\mathbf{H}^T, \\ \mathbf{H} = (\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)^{-1/2} \mathbf{C}.$$

It is easy to see that $\mathbf{z} \sim N_q(\boldsymbol{\mu}_z, \mathbf{I}_q)$, where $\boldsymbol{\mu}_z = (\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)^{-1/2} (\mathbf{C}\boldsymbol{\mu} - \mathbf{c})$. For further investigation, let $n_{\min} = \min_{l=1}^k n_l$ and $n_{\max} = \max_{l=1}^k n_l$ denote the smallest and largest sample sizes. We impose the following condition:

$$(7) \quad \frac{n_l}{n_{\min}} \rightarrow r_l < \infty, \quad l = 1, 2, \dots, k, \quad \text{as } n_{\min} \rightarrow \infty.$$

This condition requires that the sample sizes n_1, n_2, \dots, n_k proportionally tend to ∞ , preventing the case when n_{\min} is too small compared with the other sample sizes. This guarantees that $n_{\min}(\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)$ tends to a non-singular matrix as $n_{\min} \rightarrow \infty$ so that we can write $(\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)^{-1} = O(n_{\min})$ and $\mathbf{H} = (\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)^{-1/2} \mathbf{C} = O(n_{\min}^{1/2})$. Let χ_m^2 denote a chi-square distribution with m degrees of freedom. We have the following result.

Theorem 1. *Under the condition (7) and H_0 , as $n_{\min} \rightarrow \infty$, we have that T converges to χ_q^2 in distribution.*

From the proof of Theorem 1 in the Appendix, it is seen that the convergence rate of T is of order $n_{\min}^{-1/2}$. This indicates that the null distribution of T approaches to χ_q^2 slowly. Alternatively speaking, the χ_q^2 -distribution can not give an accurate approximation to the null distribution of T when n_{\min} is too small. To overcome this difficulty, following Yanagihara and Yuan (2005), we apply the modified Bartlett correction of Fujikoshi (2000) to improve the convergence rate of T , resulting in the so-called modified Bartlett (MB) test. The MB test considered by Yanagihara and Yuan (2005) is for a multivariate two-sample BF problem. Let $\mathbf{h}_l = (\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)^{-1/2} \mathbf{c}_l$, $l = 1, 2, \dots, k$ where $\mathbf{c}_1, \dots, \mathbf{c}_k$ are the k columns of \mathbf{C} . To apply the modified Bartlett correction in the current context, we need the following result.

Theorem 2. *Under the condition (7) and H_0 , as $n_{\min} \rightarrow \infty$, we have*

$$(8) \quad E(T) = q \left(1 + \frac{\alpha_1}{n_{\min}} \right) + O(n_{\min}^{-2}), \\ E(T^2) = q(q+2) \left(1 + \frac{\alpha_2}{n_{\min}} \right) + O(n_{\min}^{-2}),$$

where $\alpha_1 = \frac{2n_{\min}}{q} \Delta$, $\alpha_2 = \frac{(14+4q)n_{\min}}{q(q+2)} \Delta$, and $\Delta = \sum_{l=1}^k \frac{[\frac{\sigma_l^2}{n_l} \mathbf{h}_l^T \mathbf{h}_l]^2}{(n_l - 1)}$. In addition, we have

$$(9) \quad \frac{q^2}{(n_{\max} - 1)k} \leq \Delta \leq \frac{q}{n_{\min} - 1}.$$

Notice that under the conditions of Theorem 2, the quantities α_1 and α_2 will tend to their finite limits respectively as $n_{\min} \rightarrow \infty$. Theorem 2 implies that $E(T) = q + O(n_{\min}^{-1})$ and $E(T^2) = q(q+2) + O(n_{\min}^{-1})$. The modified Bartlett correction of Fujikoshi (2000) aims to improve this convergence rate to a higher order, say, of order n_{\min}^{-2} using the log-transformation $T_{\text{MB}} = (n_{\min}\beta_1 + \beta_2) \log(1 + \frac{T}{n_{\min}\beta_1})$, where

$\beta_1 = \frac{2}{\alpha_2 - 2\alpha_1}$ and $\beta_2 = \frac{(q+2)\alpha_2 - 2(q+4)\alpha_1}{2(\alpha_2 - 2\alpha_1)}$. One can show that $E(T_{\text{MB}}) = q + O(n_{\text{min}}^{-2})$ and $E(T_{\text{MB}}^2) = q(q+2) + O(n_{\text{min}}^{-2})$; see some details in Yanagihara and Yuan (2005) and Fujikoshi (2000). It is then expected that T_{MB} converges to χ_q^2 with a faster rate than T does.

In real data application, β_1 and β_2 have to be replaced by their estimators. Proper estimators are obtained via replacing Δ by its estimator:

$$(10) \quad \hat{\Delta} = \sum_{l=1}^k \left[\frac{\hat{\sigma}_l^2}{n_l} \hat{\mathbf{h}}_l^T \hat{\mathbf{h}}_l \right]^2 / (n_l - 1) \\ = \sum_{l=1}^k \left[\frac{\hat{\sigma}_l^2}{n_l} \mathbf{c}_l^T (\mathbf{C}\hat{\Sigma}\mathbf{C}^T)^{-1} \mathbf{c}_l \right]^2 / (n_l - 1),$$

where $\hat{\mathbf{h}}_l = (\mathbf{C}\hat{\Sigma}\mathbf{C}^T)^{-1/2} \mathbf{c}_l$, $l = 1, 2, \dots, k$. The estimators $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2$ are then obtained accordingly so that

$$(11) \quad \hat{T}_{\text{MB}} = (n_{\text{min}}\hat{\beta}_1 + \hat{\beta}_2) \log \left(1 + \frac{T}{n_{\text{min}}\hat{\beta}_1} \right) \\ \sim \chi_q^2 \text{ approximately.}$$

Some simple algebra gives that

$$n_{\text{min}}\hat{\beta}_1 = \frac{q(q+2)}{3\hat{\Delta}}$$

and

$$n_{\text{min}}\hat{\beta}_1 + \hat{\beta}_2 = \frac{(q+2)(2q - \hat{\Delta})}{6\hat{\Delta}}.$$

From the proof of Theorem 2, one can easily see that the range of Δ given in (9) is also the range of $\hat{\Delta}$. Thus, provided $n_{\text{min}} \geq 2$, we always have $n_{\text{min}}\hat{\beta}_1 > 0$ and $n_{\text{min}}\hat{\beta}_1 + \hat{\beta}_2 > 0$. This guarantees that \hat{T}_{MB} is a nonnegative and monotonically increasing function of T .

The critical value of the MB test can be specified as $\chi_q^2(1 - \alpha)$ for any given significance level α . We reject the null hypothesis in (2) when this critical value is exceeded by \hat{T}_{MB} . The MB test can also be conducted via computing the P-value based on the χ_q^2 -distribution easily.

2.2 Properties of the MB test

As mentioned previously, the contrast matrix \mathbf{C} is not unique. It is known from Kshirsagar (1972, Ch. 5, Sec. 4) that for any two contrast matrices $\tilde{\mathbf{C}}$ and \mathbf{C} specifying the same hypothesis, there is a nonsingular matrix \mathbf{P} such that $\tilde{\mathbf{C}} = \mathbf{P}\mathbf{C}$. Theorem 3 below shows that the MB test is invariant to different choices of \mathbf{C} for the same hypothesis.

Theorem 3. *The MB test is invariant when \mathbf{C} and \mathbf{c} in (2) are replaced by*

$$(12) \quad \tilde{\mathbf{C}} = \mathbf{P}\mathbf{C} \quad \text{and} \quad \tilde{\mathbf{c}} = \mathbf{P}\mathbf{c},$$

respectively where \mathbf{P} is any nonsingular matrix.

In practice, the observed data often have to be re-scaled or re-centered before conducting a statistical inference. Data recentering and rescaling are two special cases of the following affine transformation:

$$(13) \quad \tilde{x}_{lj} = ax_{lj} + b, \quad j = 1, 2, \dots, n_l, \quad l = 1, 2, \dots, k,$$

where $a \neq 0$ and b are two given constants.

Theorem 4. *The MB test is invariant under the affine transformation (13).*

It is generally required that a good test is invariant under different labeling schemes of the k population means. The MB test has such a property as stated below.

Theorem 5. *The MB test is invariant under different labeling schemes of the population means μ_l , $l = 1, 2, \dots, k$.*

2.3 MB test for one-way random-effect models

One-way random-effect models are very important in the analysis of inter-laboratory data. In this subsection, we would also like to mention that like the Welch test and the PB test (Krishnamoorthy, Lu, and Mathew 2007), the MB test is also appropriate for one-way random-effect models. Let x_{lj} denote the j -th observation at the l -th lab, where $j = 1, 2, \dots, n_l$; $l = 1, 2, \dots, k$. A one-way random-effect model can be written as

$$(14) \quad x_{lj} = \mu_0 + \tau_l + \epsilon_{lj}, \quad j = 1, 2, \dots, n_l; \quad l = 1, 2, \dots, k,$$

where μ_0 is a fixed-effect (i.e. grand mean), τ_l , $l = 1, \dots, k$ are random-effects, and ϵ_{lj} are measurement errors. We assume that $\tau_l \sim N(0, \sigma_\tau^2)$, $l = 1, \dots, k$ and $\epsilon_{lj} \sim N(0, \sigma_l^2)$, $j = 1, \dots, n_l$; $l = 1, \dots, k$ and they all are independent. To check whether the inter-laboratory effect is significant is equivalent to test if the variance component σ_τ^2 equals 0. That is, we want to test the following problem:

$$(15) \quad H_0 : \sigma_\tau^2 = 0, \quad \text{vs} \quad H_1 : \sigma_\tau^2 > 0.$$

Firstly, we show that the test statistic (4) with some \mathbf{C} and \mathbf{c} can be used to test (15). For this purpose, notice that the best linear unbiased predictors for τ_l , $l = 1, \dots, k$ are $\hat{\tau}_l = \hat{\mu}_l - \hat{\mu}_0$, $l = 1, \dots, k$ where $\hat{\mu}_0 = \sum_{l=1}^k \sum_{j=1}^{n_l} x_{lj} / N$ is the sample grand mean, $\hat{\mu}_l = \sum_{j=1}^{n_l} x_{lj} / n_l$, $l = 1, \dots, k$ are the usual group means, and $N = \sum_{l=1}^k n_l$ is the total sample size. It follows that $\hat{\tau}_l - \hat{\tau}_k = \hat{\mu}_l - \hat{\mu}_k$, $l = 1, \dots, k-1$. That is, we have $\mathbf{C}\hat{\boldsymbol{\tau}} = \mathbf{C}\hat{\boldsymbol{\mu}}$ where $\mathbf{C} = [\mathbf{I}_{k-1}, -\mathbf{1}_{k-1}]$, $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \dots, \hat{\tau}_k)^T$, and $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \dots, \hat{\mu}_k)^T$. Under the null hypothesis in (15), we have $\mathbf{C}\hat{\boldsymbol{\mu}} \sim N_{k-1}(\mathbf{0}, \mathbf{C}\hat{\Sigma}\mathbf{C}^T)$ where $\hat{\Sigma} = \text{diag}(\sigma_1^2/n_1, \dots, \sigma_k^2/n_k)$. Therefore, it is natural to use the following Wald-type test statistic

$$T = (\mathbf{C}\hat{\boldsymbol{\mu}})^T (\mathbf{C}\hat{\Sigma}\mathbf{C}^T)^{-1} (\mathbf{C}\hat{\boldsymbol{\mu}}) = \hat{\boldsymbol{\mu}}^T \left[\mathbf{C}^T (\mathbf{C}\hat{\Sigma}\mathbf{C}^T)^{-1} \mathbf{C} \right] \hat{\boldsymbol{\mu}},$$

for testing (15). This shows that T is in the form of (4) with $\mathbf{C} = [\mathbf{I}_{k-1}, -\mathbf{1}_{k-1}]$ and $\mathbf{c} = \mathbf{0}$. It is seen that given $\hat{\Sigma}$, T is a positive definite quadratic form in $\hat{\boldsymbol{\mu}}$, showing that T has a distribution which is stochastically increasing in σ_τ^2 (Dajani and Mathew 2003). That is, conditional on $\hat{\Sigma}$, the values of T are also stochastically increasing in σ_τ^2 and this stochastic monotonicity also holds for T unconditionally. Secondly, we show that the null distribution of T can be approximated using the MB correction. In fact, under the null hypothesis in (15), Theorems 1 and 2 are still valid and Theorems 3, 4 and 5 can also be verified. Thus, the MB correction can still be used to approximate the null distribution of T . Therefore, we have showed that the MB test can be used for the one-way random-effect testing problem (15).

3. SIMULATION STUDIES

In this section, we assess the performance of the MB test via five simulation studies. In the first three simulation studies, we compare the MB test against Welch's (1951) ADF test and Krishnamoorthy, Lu, and Mathew's (2007) PB test. The reasons for our choosing the Welch and PB tests as competitors include, as mentioned in the introduction section, that Welch's test is the most popular testing procedure used in the literature, and Krishnamoorthy, Lu, and Mathew (2007) showed via intensive simulation studies that the PB test is so far the most accurate testing procedure for the k -sample BF problem (1) in terms of size controlling. The last two simulation studies aim to study the performance of the MB test for two contrast tests.

First of all, we briefly describe the Welch and PB tests as follows. For the k -sample BF problem (1), the Welch test statistic is $T = \frac{\sum_{l=1}^k w_l (\hat{\mu}_l - \hat{\mu})^2 / (k-1)}{1 + \frac{2(k-2)}{k^2-1} \sum_{l=1}^k \frac{1}{n_l-1} (1-w_l/w)^2}$ where $w_l = n_l / \hat{\sigma}_l^2$, $w = \sum_{l=1}^k w_l$ and $\hat{\mu} = \sum_{l=1}^k w_l \hat{\mu}_l / w$. The null hypothesis of (1) is rejected when T exceeds the upper (100α) -th percentile of $F_{k-1, d}$ where $d = [\frac{2}{k^2-1} \sum_{l=1}^k \frac{1}{n_l-1} (1-w_l/w)^2]^{-1}$. The null hypothesis of (1) is rejected by the PB test when $P[\sum_{l=1}^k z_l^2 / u_l^2 - \frac{[\sum_{l=1}^k w_l^{1/2} z_l / u_l^2]^2}{\sum_{l=1}^k w_l / u_l^2}] > \sum_{l=1}^k w_l (\hat{\mu}_l - \hat{\mu})^2] < \alpha$, where $u_l^2 \sim \chi_{n_l-1}^2 / (n_l - 1)$ and $z_l \sim N(0, 1)$, $l = 1, 2, \dots, k$ are independent. The left-hand side probability has to be evaluated using Monte Carlo via simulating (z_l, u_l^2) , $l = 1, 2, \dots, k$ a large number of times.

For a given sample size vector $\mathbf{n} = (n_1, n_2, \dots, n_k)$, a mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_k)^T$ and a variance vector $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)$ (for easy presentation, row vectors are used for \mathbf{n} and $\boldsymbol{\sigma}^2$ in this section), we first generate k sample means $\bar{x}_1, \dots, \bar{x}_k$ and k sample variances $\hat{\sigma}_1^2, \dots, \hat{\sigma}_k^2$ by $\bar{x}_l \sim N(\mu_l, \sigma_l^2 / n_l)$ and $\hat{\sigma}_l^2 \sim \frac{\sigma_l^2}{n_l-1} \chi_{n_l-1}^2$, $l = 1, 2, \dots, k$. For simplicity, the entries of $\boldsymbol{\mu}$ are defined as $\mu_l = 2.1 + \delta \Delta_l$, $l = 1, 2, \dots, k$ where $\Delta_1 = .5, \Delta_2 = -.3, \Delta_k = 1.1, \Delta_l = 0$, $l = 3, \dots, k-1$ and δ is a tuning parameter controlling the validity of the null hypothesis in the k -sample BF problem (1). The three tests are then applied to $\hat{\mu}_l = \bar{x}_l$,

$l = 1, \dots, k$ and $\hat{\sigma}_l^2$, $l = 1, \dots, k$ respectively and their P-values are recorded. The P-values of the PB test are obtained via 10,000 inner runs. This process is repeated 10,000 times. The empirical sizes (when $\delta = 0$) and powers (when $\delta > 0$) of the tests are the proportions of rejecting the null hypothesis, i.e., when the P-values of the tests are less than the nominal significance level α . In all the simulations conducted, we used $\alpha = 5\%$ for simplicity.

In Tables 1–3, the first 2 columns list the tuning parameters for the sample sizes and population variances under consideration. For simplicity, we sometimes use \mathbf{a}_r to denote “ \mathbf{a} repeats r times”, e.g., $(4_3, 2_2) = (4, 4, 4, 2, 2)$ and $(1, 2, 3)_2 = (1, 2, 3, 1, 2, 3)$. The columns labeled with “Welch”, “PB” and “MB” display the empirical sizes or powers of the Welch, PB and MB tests respectively. When the null distribution of the MB test is χ_q^2 , the column labeled with “ $\hat{\chi}_q^2(.05)$ ” lists the bootstrapped critical values of the MB test obtained via 10,000 bootstrap replicates of \hat{T}_{MB} . If the MB test works well, the bootstrapped critical values should be close to the associated theoretical critical value of the MB test as given in the associated table caption. This offers another way to compare the MB test against the PB test. To measure the overall performance of a test in terms of maintaining the nominal size α , we define the average relative error as $\text{ARE} = M^{-1} \sum_{j=1}^M |\hat{\alpha}_j - \alpha| / \alpha \times 100$ where $\hat{\alpha}_j$ denotes the j -th empirical size of the test for $j = 1, 2, \dots, M$. The smaller ARE value indicates the better overall performance of the associated test. Usually, when $\text{ARE} < 10$, the test performs very well and when $10 < \text{ARE} < 20$, the test is acceptable. When $\text{ARE} > 20$, the test may be quite liberal or quite conservative.

Table 1 displays the simulation results for 3-sample BF problem (1). Four cases of \mathbf{n} are considered with the total sample size $N = 27$ being the same with n_{\min} increasing from 2 to 9. Seven cases of $\boldsymbol{\sigma}^2$ are considered with the first case having homogeneous variances. It is seen that when n_{\min} is too small, e.g., for those cases with $n_{\min} = 2$, none of the Welch, PB and MB tests performed well in terms of size controlling: the ARE values of the three tests are more than 90; their empirical sizes are around 10%, much larger than the nominal size 5%; and the bootstrapped critical values of the MB test are much larger than the associated theoretical critical value $\chi_2^2(.05) = 5.99$. However, with increasing n_{\min} , the performances of the three tests are getting better and better: the ARE values of the three tests are now less than 20; their empirical sizes are now around 5%; and the bootstrapped critical values of the MB test are getting closer and closer to the associated theoretical critical value. In particular, the three tests performed very well for those cases with $\mathbf{n} = (9, 9, 9)$. Notice also that the powers of the three tests are comparable and they are getting larger and larger with increasing n_{\min} from 2 to 9 although the powers of the MB test are slightly smaller than those of the other two tests when $n_{\min} = 2$ and 4. Overall speaking, in this

Table 1. Simulation results for 3-sample BF problem (1) [$\chi_2^2(.05) = 5.99$]

\mathbf{n}	σ^2	$\delta = 0$				$\delta = 1$		
		Welch	PB	MB	$\chi_2^2(.05)$	Welch	PB	MB
(2, 6, 19)	(1, 1, 1)	.108	.120	.101	9.36	.333	.379	.257
	(1, 1, 2)	.105	.112	.098	8.90	.288	.331	.236
	(1, 1, 4)	.122	.139	.109	11.0	.204	.229	.160
	(1, 2, 3)	.086	.089	.081	7.71	.216	.246	.189
	(2, 1, 1)	.121	.135	.111	11.0	.285	.303	.212
	(4, 1, 1)	.118	.134	.105	11.5	.257	.256	.175
	(3, 2, 1)	.089	.095	.084	7.81	.238	.269	.209
	ARE	114	136	97.4				
(4, 6, 17)	(1, 1, 1)	.058	.059	.054	6.19	.564	.569	.534
	(1, 1, 2)	.054	.055	.050	5.99	.480	.482	.460
	(1, 1, 4)	.061	.062	.055	6.19	.336	.334	.305
	(1, 2, 3)	.050	.050	.046	5.84	.281	.285	.269
	(2, 1, 1)	.064	.063	.058	6.33	.539	.534	.506
	(4, 1, 1)	.057	.055	.051	6.03	.520	.502	.481
	(3, 2, 1)	.051	.052	.049	5.94	.361	.361	.346
	ARE	13.0	13.5	6.71				
(6, 6, 15)	(1, 1, 1)	.049	.050	.046	5.85	.609	.611	.596
	(1, 1, 2)	.048	.048	.046	5.84	.507	.510	.497
	(1, 1, 4)	.053	.053	.049	5.95	.385	.384	.368
	(1, 2, 3)	.050	.051	.048	5.91	.316	.317	.308
	(2, 1, 1)	.054	.054	.050	6.00	.597	.599	.582
	(4, 1, 1)	.057	.055	.053	6.09	.581	.580	.565
	(3, 2, 1)	.050	.050	.048	5.91	.396	.396	.389
	ARE	5.17	4.34	4.54				
(9, 9, 9)	(1, 1, 1)	.050	.049	.048	5.93	.677	.676	.672
	(1, 1, 2)	.052	.051	.049	5.97	.529	.529	.524
	(1, 1, 4)	.048	.048	.047	5.83	.485	.483	.479
	(1, 2, 3)	.048	.048	.046	5.80	.330	.328	.324
	(2, 1, 1)	.051	.050	.050	5.97	.662	.662	.657
	(4, 1, 1)	.051	.050	.049	5.95	.658	.656	.653
	(3, 2, 1)	.050	.049	.049	5.91	.419	.416	.413
	ARE	2.40	2.03	3.23				

simulation study, the three tests are roughly comparable in terms of size controlling and power.

Table 2 displays the simulation results for a 10-sample BF problem (1). Four cases of \mathbf{n} and five cases of σ^2 are considered. In this simulation study, we would like to emphasize that we compared $k = 10$ normal means. It is seen that the Welch test became very liberal, with its empirical sizes around 8% except when $\mathbf{n} = (15_{10})$ but the PB and MB tests still performed well and were comparable with each other in terms of size controlling and power. For all the cases under consideration, the bootstrapped critical values of the MB test are close to their theoretical critical value $\chi_9^2(.05) = 16.90$.

Table 3 displays the simulation results for a 20-sample BF problem (1). Three cases of \mathbf{n} and five cases of σ^2 are considered. It is seen that the Welch test became more liberal when $\mathbf{n} = (5_{20})$, with its empirical sizes around 12%. In terms of size controlling, the PB test slightly outperformed the MB test while in terms of power, the MB test slightly outperformed the PB test. Therefore, in this simulation study, the

PB and MB tests were still comparable with each other and both of them performed better than the Welch test. For all the cases under consideration, the bootstrapped critical values of the MB test are close to their theoretical critical value $\chi_{19}^2(.05) = 30.1$.

As mentioned in the previous section, the GLHT problem (2) includes not only the k -sample BF problem (1) but also contrast tests as special cases. To see if the MB test also works well for contrast tests, we conducted two simulation studies with contrast tests. The Welch and PB tests were not included since they were developed only for the k -sample BF problem unless some further work is done. Tables 4 and 5 show the results of the two simulation studies for the GLHT problem (2) with $\mathbf{C} = (-2, -1, 0, 1, 2)$, $\boldsymbol{\mu}^T = (1, 2, 0, 2, 1)$ and with $\mathbf{C} = (1, 1, -2)$, $\boldsymbol{\mu}^T = (1, 3, 2)$, respectively. It is seen that the MB test indeed performed well in these two simulation studies with all the ARE values below 10 and all the bootstrapped critical values close to the theoretical critical value, $\chi_1^2(.05) = 3.84$, of the MB test.

Table 2. Simulation results for 10-sample BF problem (1) [$\chi_9^2(.05) = 16.90$]

n	σ^2	$\delta = 0$				$\delta = 1$		
		Welch	PB	MB	$\hat{\chi}_9^2(.05)$	Welch	PB	MB
(5 ₁₀)	(1, 1, 1, 1, 1) ₂	.081	.044	.046	16.6	.301	.197	.201
	(1, 1, 1, 1, 5) ₂	.080	.044	.047	16.7	.166	.099	.104
	(1, 2, 3, 4, 5) ₂	.080	.046	.050	16.8	.148	.087	.093
	(5, 1, 1, 1, 1) ₂	.077	.038	.041	16.3	.272	.175	.183
	(5, 4, 3, 2, 1) ₂	.083	.046	.048	16.7	.229	.141	.150
	ARE	60.8	12.9	6.84				
(15 ₁₀)	(1, 1, 1, 1, 1) ₂	.056	.051	.052	17.0	.852	.843	.844
	(1, 1, 1, 1, 5) ₂	.054	.050	.051	16.9	.424	.410	.412
	(1, 2, 3, 4, 5) ₂	.056	.052	.052	17.1	.345	.332	.334
	(5, 1, 1, 1, 1) ₂	.052	.048	.049	16.8	.797	.787	.789
	(5, 4, 3, 2, 1) ₂	.053	.048	.048	16.8	.691	.678	.680
	ARE	8.28	3.00	3.44				
(4 ₃ , 5 ₄ , 6 ₃)	(1, 1, 1, 1, 1) ₂	.086	.043	.047	16.7	.311	.196	.208
	(1, 1, 1, 1, 5) ₂	.090	.046	.051	16.9	.164	.093	.101
	(1, 2, 3, 4, 5) ₂	.083	.042	.047	16.7	.144	.082	.091
	(5, 1, 1, 1, 1) ₂	.091	.046	.052	17.1	.291	.169	.190
	(5, 4, 3, 2, 1) ₂	.097	.053	.060	17.6	.241	.134	.153
	ARE	79.5	10.3	7.20				
(5 ₈ , 15 ₂)	(1, 1, 1, 1, 1) ₂	.082	.049	.056	17.4	.551	.402	.439
	(1, 1, 1, 1, 5) ₂	.084	.051	.058	17.5	.211	.130	.147
	(1, 2, 3, 4, 5) ₂	.081	.046	.054	17.2	.183	.109	.130
	(5, 1, 1, 1, 1) ₂	.089	.050	.060	17.5	.528	.365	.407
	(5, 4, 3, 2, 1) ₂	.088	.047	.056	17.3	.412	.265	.303
	ARE	70.3	3.48	14.1				

From the first three simulation studies, it is seen that overall speaking, the Welch test is very liberal when k is large and n is small. This observation is in agreement with the one observed by Krishnamoorthy, Lu, and Mathew (2007). In addition, we see that the MB test is generally comparable to the PB test in terms of size controlling and power for the k -sample BF problem (1) for various sample sizes and parameter configurations. From the last two simulation studies, we see that the MB test also works well for contrast tests. In view of the above and the computational effort required by the PB test, we generally recommend the MB test in real data analysis.

4. APPLICATIONS TO THE PTSD DATA

The study by Foa, Rothbaum, Riggs, and Murdock (1991) involved 45 subjects (rape victims) who were randomly assigned to one of four groups treated by four different treatments: (1) Stress Inoculation Therapy (SIT) in which subjects were taught a variety of coping skills; (2) Prolonged Exposure (PE) in which subjects went over the rape in their mind repeatedly for seven sessions; (3) Supportive Counseling (SC) which was a standard therapy control group; and (4) a Waiting List (WL) control. In the actual study, pre- and post-treatment measures were taken on a number of variables. Here, however, we only look at the post traumatic stress disorder (PTSD) data (the total number of

symptoms endorsed by the subject). The data are available at http://www.uvm.edu/~dhowell/StatPages/FoaFolder/Foa_Anova.html where David C. Howell presented a standard ANOVA analysis, assuming homogeneity of the variances. However, the sample variances of the PTSD data are quite different, ranging from 15.6 to 124, as seen from Table 6.

In this section, we make use of this PTSD data to illustrate the MB test and to study the impacts of the variance heteroscedasticity on the results of the standard ANOVA, Welch (1951), PB (Krishnamoorthy, Lu, and Mathew 2007) and MB tests. For this purpose, we consider all the four, three and two-group mean comparisons as listed in Column 1 of Table 7. It is seen that the homogeneity assumption is violated in various degrees in these comparisons with Case (2) most seriously and Case (11) most lightly. Overall speaking, the degrees of heteroscedasticity for Cases (1)–(3) are much higher than those for Cases (4)–(11).

We, first of all, look at the P-values of the Welch, PB and MB tests as listed in Columns 2–5 of Table 7 respectively. The P-values of the PB test were obtained based on 100,000 bootstrap replicates. Since the sample sizes are relatively large, all the three tests performed well and they gave about the same P-values for all 11 cases although the computational cost of the MB test and the Welch test is much less than that of the PB test.

Table 3. Simulation results for 20-sample BF problem (1) [$\chi_{19}^2(.05) = 30.1$].

n	σ^2	$\delta = 0$				$\delta = 1$		
		Welch	PB	MB	$\hat{\chi}_{19}^2(.05)$	Welch	PB	MB
(5 ₂₀)	(1, 1, 1, 1) ₅	.129	.044	.061	31.1	.307	.138	.177
	(1, 1, 1, 4) ₅	.121	.050	.067	30.7	.202	.075	.101
	(1, 2, 3, 4) ₅	.127	.045	.061	31.1	.192	.071	.098
	(4, 1, 1, 1) ₅	.129	.043	.060	31.2	.274	.118	.156
	(4, 3, 2, 1) ₅	.126	.044	.061	31.1	.267	.114	.147
	ARE	151	9.60	23.8				
(15 ₂₀)	(1, 1, 1, 1) ₅	.059	.051	.054	30.4	.753	.729	.736
	(1, 1, 1, 4) ₅	.054	.052	.055	29.9	.344	.317	.325
	(1, 2, 3, 4) ₅	.055	.052	.054	30.1	.299	.275	.281
	(4, 1, 1, 1) ₅	.058	.049	.051	30.3	.679	.654	.660
	(4, 3, 2, 1) ₅	.058	.054	.057	30.3	.626	.595	.604
	ARE	19.9	4.00	8.08				
(10 ₁₀ , 15 ₁₀)	(1, 1, 1, 1) ₅	.064	.051	.055	30.6	.675	.631	.645
	(1, 1, 1, 4) ₅	.062	.050	.055	30.3	.290	.245	.259
	(1, 2, 3, 4) ₅	.064	.049	.053	30.5	.255	.215	.229
	(4, 1, 1, 1) ₅	.062	.052	.056	30.5	.621	.578	.593
	(4, 3, 2, 1) ₅	.066	.051	.056	30.5	.584	.531	.551
	ARE	30.0	2.08	9.84				

Table 4. Simulation results for the GLHT problem (2) with $C = (-2, -1, 0, 1, 2)$ and $\mu^T = (1, 2, 0, 2, 1)$ [$\chi_1^2(.05) = 3.84$]

n	(5 ₅)		(10 ₅)		(5 ₃ , 10 ₂)		(4 ₂ , 5, 6 ₂)	
	MB	$\hat{\chi}_1^2(.05)$	MB	$\hat{\chi}_1^2(.05)$	MB	$\hat{\chi}_1^2(.05)$	MB	$\hat{\chi}_1^2(.05)$
(1 ₅)	.048	3.76	.049	3.80	.050	3.85	.050	3.83
(1, 1, 1, 1, 5)	.047	3.73	.048	3.77	.051	3.89	.048	3.79
(1, 2, 3, 4, 5)	.050	3.83	.050	3.82	.050	3.85	.047	3.72
(5, 1, 1, 1, 1)	.053	3.97	.053	3.93	.055	4.03	.058	4.11
(5, 4, 3, 2, 1)	.045	3.68	.051	3.85	.052	3.93	.052	3.89
ARE	5.64		3.08		3.72		6.28	

We now check the impact of the heteroscedasticity on the results via looking at the P-value discrepancies between the standard ANOVA and the MB test for all the 11 cases. For Cases (4)–(11) where the degrees of heteroscedasticity are low, the P-value discrepancies between the standard ANOVA and the MB test are not large enough to yield inconsistent conclusions. For Cases (1)–(3) where the degrees of heteroscedasticity is high, however, the P-values of the standard ANOVA are about 2 to 5 times larger than those of the MB test. Depending on the specified nominal significance level α , the conclusions made by the standard ANOVA may be opposite to those made by the MB test. For example, when $\alpha = 1\%$, the standard ANOVA accepts (while the MB test rejects) the null hypotheses of Cases (1)–(3); when $\alpha = 5\%$, the standard ANOVA accepts (while the MB test rejects) the null hypothesis of Case (2). Since the MB test assumes much weaker conditions than the standard ANOVA does, the conclusions made by it are generally more reliable than those made by the standard ANOVA. Thus, when the degree of heteroscedasticity is high, application of the standard ANOVA may yield misleading conclusions.

5. CONCLUDING REMARKS

In this paper, we proposed and studied the so-called MB test for the GLHT problem in heteroscedastic one-way ANOVA, making use of the modified Bartlett correction of Fujikoshi (2000) to improve the convergence rate of the Wald-type statistic T which approaches to a χ^2 -distribution slowly. It allows a unified treatment of the overall, post hoc and contrast tests in heteroscedastic one-way ANOVA about the population means and hence has wide applications. Simulation studies and real data applications show that the MB test generally outperforms the Welch (1951) test and is comparable to the PB test of Krishnamoorthy, Lu, and Mathew (2007) for the k -sample BF problem. The modified Bartlett correction has wide applications. We here provide another useful example following Yanagihara and Yuan (2005).

Notice that the condition (7) we imposed for Theorems 1 and 2 is based on n_{\min} . For large samples, this condition is equivalent to the ordinary condition " $\frac{n_l}{N} = O(1)$, $l = 1, 2, \dots, k$ " imposed by some other authors, e.g., Yanagihara and Yuan (2005), where $N = \sum_{l=1}^k n_l$ denotes the total sample size. We used n_{\min} rather than N based on the

Table 5. Simulation results for the GLHT problem (2) with $C = (1, 1, -2)$ and $\mu^T = (1, 3, 2)$ [$\chi_1^2(.05) = 3.84$]

n	(4, 6, 19)		(6, 6, 15)		(8, 8, 11)		(9, 9, 9)	
	MB	$\hat{\chi}_1^2(.05)$	MB	$\hat{\chi}_1^2(.05)$	MB	$\hat{\chi}_1^2(.05)$	MB	$\hat{\chi}_1^2(.05)$
(1, 1, 1)	.047	3.75	.048	3.75	.049	3.81	.050	3.84
(1, 1, 4)	.042	3.61	.050	3.82	.049	3.81	.051	3.87
(1, 2, 3)	.050	3.84	.050	3.84	.049	3.79	.048	3.77
(4, 1, 1)	.059	4.17	.053	3.94	.044	3.67	.044	3.66
(3, 2, 1)	.046	3.73	.048	3.77	.047	3.75	.046	3.68
ARE	9.44		2.88		4.76		5.36	

Table 6. Summary statistics for the PTSD data

Group	Size	Sample Mean	Sample Variance
1 (SIT)	14	11.1	15.6
2 (PE)	10	15.4	124
3 (SC)	11	18.1	50.9
4 (WL)	10	19.5	50.5

Table 7. Group mean comparisons for the PSTD data

Case (Null Hypothesis)	P-values $\times 100$			
	ANOVA	Welch	PB	MB
(1) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$	3.94	0.75	0.78	0.83
(2) $H_0 : \mu_1 = \mu_2 = \mu_3$	7.85	2.95	3.04	3.06
(3) $H_0 : \mu_1 = \mu_2 = \mu_4$	3.71	1.36	1.37	1.43
(4) $H_0 : \mu_1 = \mu_3 = \mu_4$	0.31	0.33	0.30	0.35
(5) $H_0 : \mu_2 = \mu_3 = \mu_4$	56.3	63.4	63.6	63.4
(6) $H_0 : \mu_1 = \mu_2$	19.0	26.4	26.4	26.4
(7) $H_0 : \mu_1 = \mu_3$	0.47	1.05	1.03	1.06
(8) $H_0 : \mu_1 = \mu_4$	0.12	0.48	0.42	0.49
(9) $H_0 : \mu_2 = \mu_3$	51.3	52.4	52.4	52.4
(10) $H_0 : \mu_2 = \mu_4$	33.9	34.1	34.1	34.1
(11) $H_0 : \mu_3 = \mu_4$	65.6	65.6	65.3	65.6

following reasons. First of all, n_{\min} plays a central role in stating Theorems 1 and 2 and their proofs. For example, from the proofs of Theorems 1 and 2, one can see that it is natural and accurate to use n_{\min} rather than N to define the convergence rates of the random matrices $\mathbf{H}, \mathbf{W}, \mathbf{\Sigma}, \mathbf{R}$ among others. Secondly, it is well known that for homogeneous one-way ANOVA, the convergence rate of T tending to a χ^2 -distribution is of order $N^{-1/2}$. By the proof of Theorem 1, it is seen that the convergence rate of T for heteroscedastic one-way ANOVA is of order $n_{\min}^{-1/2}$. Therefore, the convergence rate of T for homogeneous one-way ANOVA is much faster than that for heteroscedastic one-way ANOVA. This is intuitively understood since for homogeneous one-way ANOVA, the N observations are pooled to estimate the common variance while for heteroscedastic one-way ANOVA, each of the population variances has to be estimated separately using only the observations in the sample. Finally, from the simulation results in Table 1, it is seen that n_{\min} has a strong impact on the performances of the Welch, PB and MB tests: when n_{\min} is too small, none

of the three tests performed well even when N is sufficiently large.

In this paper, the normality assumption was made when we developed the MB test. In particular, Theorem 2 was derived under the normality assumption. Therefore, the performance of the MB test proposed in this paper is not guaranteed for non-normal data. To overcome this difficulty, a normality test may be needed to check if the data are nearly normally distributed. If the normality test is rejected, the well-known Box-Cox power transformation may be properly applied so that the MB test can be applied to the resulting data which are nearly normally distributed.

APPENDIX A. APPENDIX SECTION

Proof of Theorem 1. Under the given conditions, we have

$$(A.1) \quad \hat{\sigma}_l^2 \stackrel{d}{=} \sigma_l^2 \chi_{n_l-1}^2 / (n_l - 1), \quad l = 1, 2, \dots, k,$$

where $X \stackrel{d}{=} Y$ means X and Y have the same distribution. It follows that $(\hat{\sigma}_l^2 - \sigma_l^2) / n_l = O_p(n_l^{-3/2})$, $l = 1, 2, \dots, k$. Thus $\hat{\Sigma} - \Sigma = O_p(n_{\min}^{-3/2})$. Noticing that $\Sigma = O(n_{\min}^{-1})$, we further have

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

where \mathbf{H} is defined in (6) and $\mathbf{H} = O(n_{\min}^{1/2})$. This implies that $\mathbf{W} = \mathbf{I}_q + \mathbf{H}(\hat{\Sigma} - \Sigma)\mathbf{H}^T = \mathbf{I}_q + \mathbf{R} = \mathbf{I}_q + O_p(n_{\min}^{-1/2})$. Theorem 1 follows by Slutsky's theorem and noticing that under H_0 , $\mathbf{z}^T \mathbf{z} \sim \chi_q^2$. \square

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

$$(A.3) \quad \begin{aligned} E(T) &= E \operatorname{tr}(\mathbf{W}^{-1}) \quad \text{and} \\ E(T^2) &= 2E \operatorname{tr}(\mathbf{W}^{-2}) + E \operatorname{tr}^2(\mathbf{W}^{-1}). \end{aligned}$$

where and throughout, $E^2(\mathbf{X}) = [E(\mathbf{X})]^2$, $\operatorname{tr}^2(\mathbf{Y}) = [\operatorname{tr}(\mathbf{Y})]^2$ and $\operatorname{tr}(\mathbf{Y})$ denotes the trace of \mathbf{Y} , i.e., the sum of the diagonal entries of \mathbf{Y} . From the proof of Theorem 1, we have that $\mathbf{W} = \mathbf{I}_q + \mathbf{R}$ with $\mathbf{R} = O_p(n_{\min}^{-1/2})$; see (A.2). Then we have $\mathbf{W}^{-1} = (\mathbf{I}_q + \mathbf{R})^{-1} = \mathbf{I}_q - \mathbf{R} + \mathbf{R}^2 - \mathbf{R}^3 + O_p(n_{\min}^{-2})$ and $\mathbf{W}^{-2} = (\mathbf{I}_q + \mathbf{R})^{-2} = \mathbf{I}_q - 2\mathbf{R} + 3\mathbf{R}^2 - 4\mathbf{R}^3 + O_p(n_{\min}^{-2})$.

It is easy to see from (A.2) that $E(\mathbf{R}) = \mathbf{0}$ and $E\text{tr}(\mathbf{R}) = \mathbf{0}$. Thus

$$\begin{aligned}
& E\text{tr}(\mathbf{W}^{-1}) \\
&= q + E\text{tr}(\mathbf{R}^2) - E\text{tr}(\mathbf{R}^3) + O(n_{\min}^{-2}), \\
& E\text{tr}(\mathbf{W}^{-2}) \\
(A.4) \quad &= q + 3E\text{tr}(\mathbf{R}^2) - 4E\text{tr}(\mathbf{R}^3) + O(n_{\min}^{-2}), \\
& E\text{tr}^2(\mathbf{W}^{-1}) \\
&= q^2 + E\text{tr}^2(\mathbf{R}) + 2qE\text{tr}(\mathbf{R}^2) - 2qE\text{tr}(\mathbf{R}^3) \\
&\quad - 2E\text{tr}(\mathbf{R})\text{tr}(\mathbf{R}^2) + O(n_{\min}^{-2}).
\end{aligned}$$

We now find $E\text{tr}(\mathbf{R}^2)$ and $E\text{tr}^2(\mathbf{R})$ among others. By (A.2), we have $\mathbf{R} = \sum_{l=1}^k \mathbf{h}_l \mathbf{h}_l^T u_l$, where \mathbf{h}_l is the l -th column of \mathbf{H} and $u_l = \frac{\hat{\sigma}_l^2 - \sigma_l^2}{n_l}$, for $l = 1, 2, \dots, k$. By (A.1), we have

$$\begin{aligned}
(A.5) \quad & E(u_l) = 0, \quad E(u_l^2) = \frac{2\sigma_l^4}{n_l^2(n_l - 1)} \quad \text{and} \\
& E(u_l^3) = \frac{8\sigma_l^6}{n_l^3(n_l - 1)^2}, \quad l = 1, 2, \dots, k,
\end{aligned}$$

where we use the fact $E(\chi_d^2/d - 1)^3 = 8/d^2$. Noticing that u_1, u_2, \dots, u_k are independent and by (A.5), we have

$$\begin{aligned}
& E\text{tr}(\mathbf{R}^2) = E\text{tr}^2(\mathbf{R}) \\
&= \sum_{l=1}^k (\mathbf{h}_l^T \mathbf{h}_l)^2 E(u_l^2) = 2\Delta, \\
& E\text{tr}(\mathbf{R}^3) = E\text{tr}(\mathbf{R})\text{tr}(\mathbf{R}^2) \\
(A.6) \quad &= \sum_{l=1}^k (\mathbf{h}_l \mathbf{h}_l^T)^3 E(u_l^3) \\
&= 8 \sum_{l=1}^k \left[\frac{\sigma_l^2}{n_l} \mathbf{h}_l^T \mathbf{h}_l \right]^3 (n_l - 1)^{-2} \\
&= O(n_{\min}^{-2}),
\end{aligned}$$

where $\Delta = \sum_{l=1}^k [\frac{\sigma_l^2}{n_l} \mathbf{h}_l^T \mathbf{h}_l]^2 / (n_l - 1)$. Combining (A.4) and (A.6) gives that $E\text{tr}(\mathbf{W}^{-1}) = q + 2\Delta + O(n_{\min}^{-2})$, $E\text{tr}(\mathbf{W}^{-2}) = q + 6\Delta + O(n_{\min}^{-2})$, and $E\text{tr}^2(\mathbf{W}^{-1}) = q^2 + (8q + 2)\Delta + O(n_{\min}^{-2})$. These, together with (A.3), yield that $E(T) = q[1 + \frac{\alpha_1}{n_{\min}}] + O(n_{\min}^{-2})$ and $E(T^2) = q(q + 2)[1 + \frac{\alpha_2}{n_{\min}}] + O(n_{\min}^{-2})$ where $\alpha_1 = \frac{2n_{\min}}{q}\Delta$ and $\alpha_2 = \frac{(14+4q)n_{\min}}{q(q+2)}\Delta$ as desired.

We now find the lower and upper bounds of Δ as given in (9). For $l = 1, 2, \dots, k$, set $\mathbf{S}_l = \frac{\sigma_l^2}{n_l} \mathbf{h}_l \mathbf{h}_l^T$. It is easy to see that \mathbf{S}_l is nonnegative and it has only one nonzero eigenvalue $\delta_l = \frac{\sigma_l^2}{n_l} \mathbf{h}_l^T \mathbf{h}_l$. It is easy to verify that $\sum_{l=1}^k \mathbf{S}_l = \mathbf{I}_q$ so that $\sum_{l=1}^k \delta_l = \sum_{l=1}^k \frac{\sigma_l^2}{n_l} \mathbf{h}_l^T \mathbf{h}_l = \text{tr}(\sum_{l=1}^k \mathbf{S}_l) = q$ and $\mathbf{I}_q - \mathbf{S}_l = \sum_{r \neq l} \mathbf{S}_r$. Therefore $\mathbf{I}_q - \mathbf{S}_l$ is nonnegative, showing that the only nonzero eigenvalue δ_l of \mathbf{S}_l is less than 1. It follows that $\Delta = \sum_{l=1}^k \delta_l^2 / (n_l - 1) \leq q / (n_{\min} - 1)$ and $\Delta = \sum_{l=1}^k \delta_l^2 / (n_l - 1) \geq \sum_{l=1}^k \delta_l^2 / (n_{\max} - 1) \geq q^2 / [(n_{\max} - 1)k]$ where we have used the fact that for any nonnegative numbers a_1, a_2, \dots, a_m , we always have $\sum_{l=1}^m a_l^2 \geq (\sum_{l=1}^m a_l)^2 / m$. The theorem is then proved. \square

Proof of Theorem 3. From the definition (4) of T , it is easy to see that T is invariant under the transformation (12). Then by (11), we only need to show that $\hat{\Delta}$ is invariant under the transformation (12). For $l = 1, 2, \dots, k$, let \mathbf{c}_l and $\tilde{\mathbf{c}}_l$ be the l -th columns of the contrast matrices \mathbf{C} and $\tilde{\mathbf{C}} = \mathbf{P}\mathbf{C}$ respectively. Then it is easy to see that $\tilde{\mathbf{c}}_l = \mathbf{P}\mathbf{c}_l, l = 1, 2, \dots, k$. It follows from (10) that $\hat{\Delta} = \sum_{l=1}^k [\frac{\hat{\sigma}_l^2}{n_l} \mathbf{c}_l^T (\tilde{\mathbf{C}}\hat{\Sigma}\tilde{\mathbf{C}}^T)^{-1} \mathbf{c}_l]^2 / (n_l - 1) = \sum_{l=1}^k [\frac{\hat{\sigma}_l^2}{n_l} \mathbf{c}_l^T \mathbf{P}^T (\mathbf{P}\mathbf{C}\hat{\Sigma}\mathbf{C}^T \mathbf{P}^T)^{-1} \mathbf{P}\mathbf{c}_l]^2 / (n_l - 1) = \sum_{l=1}^k [\frac{\hat{\sigma}_l^2}{n_l} \mathbf{c}_l^T (\mathbf{C}\hat{\Sigma}\mathbf{C}^T)^{-1} \mathbf{c}_l]^2 / (n_l - 1) = \hat{\Delta}$, as desired. The theorem is proved. \square

Proof of Theorem 4. This theorem is proved if we can show that both T and $\hat{\Delta}$ are invariant under the affine transformation (13). Let μ_l, σ_l^2 and $\tilde{\mu}_l, \tilde{\sigma}_l^2$ respectively denote the mean and variance of x_{lj} before and after the affine transformation (13). Then we have $\tilde{\mu}_l = a\mu_l + b$ and $\tilde{\sigma}_l^2 = a^2\sigma_l^2$. It follows that $\mu_l = a^{-1}(\tilde{\mu}_l - b)$. As we define the mean vector $\boldsymbol{\mu}$ and the variance matrix $\boldsymbol{\Sigma}$ in Section 2, we define $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$ similarly. Then we have $\tilde{\boldsymbol{\mu}} = a\boldsymbol{\mu} + \mathbf{b}$ and $\tilde{\boldsymbol{\Sigma}} = a^2\boldsymbol{\Sigma}$, where $\mathbf{b} = \mathbf{b}\mathbf{1}_k$. It follows that the GLHT problem (2) can be equivalently expressed as $\tilde{H}_0 : \tilde{\mathbf{C}}\tilde{\boldsymbol{\mu}} = \tilde{\mathbf{c}}, \quad \text{vs} \quad \tilde{H}_1 : \tilde{\mathbf{C}}\tilde{\boldsymbol{\mu}} \neq \tilde{\mathbf{c}}$, where $\tilde{\mathbf{C}} = a^{-1}\mathbf{C}$ and $\tilde{\mathbf{c}} = a^{-1}\mathbf{C}\mathbf{b} + \mathbf{c}$.

Let $\hat{\mu}_l, \hat{\sigma}_l^2$ and $\hat{\tilde{\mu}}_l, \hat{\tilde{\sigma}}_l^2$ respectively denote the unbiased estimators of the mean and variance of x_{lj} before and after the affine transformation (13). Then it is easy to see that $\hat{\tilde{\mu}}_l = a\hat{\mu}_l + b$ and $\hat{\tilde{\sigma}}_l^2 = a^2\hat{\sigma}_l^2$. It follows that $\hat{\tilde{\boldsymbol{\mu}}} = a\hat{\boldsymbol{\mu}} + \mathbf{b}$ and $\hat{\tilde{\boldsymbol{\Sigma}}} = a^2\hat{\boldsymbol{\Sigma}}$. Using the above, we have $\tilde{\mathbf{C}}\hat{\tilde{\boldsymbol{\mu}}} - \tilde{\mathbf{c}} = a^{-1}\mathbf{C}(a\hat{\boldsymbol{\mu}} + \mathbf{b}) - (a^{-1}\mathbf{C}\mathbf{b} + \mathbf{c}) = \mathbf{C}\hat{\boldsymbol{\mu}} - \mathbf{c}$ and $\tilde{\mathbf{C}}\hat{\tilde{\boldsymbol{\Sigma}}}\tilde{\mathbf{C}}^T = a^{-1}\mathbf{C}(a^2\hat{\boldsymbol{\Sigma}})(a^{-1}\mathbf{C})^T = \mathbf{C}\hat{\boldsymbol{\Sigma}}\mathbf{C}^T$. Thus, both $\mathbf{C}\hat{\boldsymbol{\mu}} - \mathbf{c}$ and $\mathbf{C}\hat{\boldsymbol{\Sigma}}\mathbf{C}^T$ are affine-invariant. So is T .

We now turn to show that $\hat{\Delta}$ is invariant under the affine transformation (13). Let $\hat{\tilde{\Delta}}$ be the affine-transformed $\hat{\Delta}$. Then by (10), we have

$$\begin{aligned}
\hat{\tilde{\Delta}} &= \sum_{l=1}^k \left[\frac{\hat{\tilde{\sigma}}_l^2}{n_l} \tilde{\mathbf{c}}_l^T (\tilde{\mathbf{C}}\hat{\tilde{\boldsymbol{\Sigma}}}\tilde{\mathbf{C}}^T)^{-1} \tilde{\mathbf{c}}_l \right]^2 / (n_l - 1) \\
&= \sum_{l=1}^k \left[a^2 \frac{\hat{\sigma}_l^2}{n_l} a^{-1} \mathbf{c}_l^T (\mathbf{C}\hat{\boldsymbol{\Sigma}}\mathbf{C}^T)^{-1} a^{-1} \mathbf{c}_l \right]^2 / (n_l - 1) \\
&= \sum_{l=1}^k \left[\frac{\hat{\sigma}_l^2}{n_l} \mathbf{c}_l^T (\mathbf{C}\hat{\boldsymbol{\Sigma}}\mathbf{C}^T)^{-1} \mathbf{c}_l \right]^2 / (n_l - 1) = \hat{\Delta},
\end{aligned}$$

as desired. The theorem is then proved. \square

Proof of Theorem 5. The theorem is proved if we can show that both T and $\hat{\Delta}$ are invariant under different labeling schemes of the means μ_1, \dots, μ_k . Let l_1, l_2, \dots, l_k be any permutation of $1, 2, \dots, k$. Then it is easy to see that $\mathbf{C}\hat{\boldsymbol{\mu}} = \sum_{l=1}^k \mathbf{c}_l \hat{\mu}_l = \sum_{u=1}^k \mathbf{c}_{l_u} \hat{\mu}_{l_u}$ and $\mathbf{C}\hat{\boldsymbol{\Sigma}}\mathbf{C}^T = \sum_{l=1}^k n_l^{-1} \hat{\sigma}_l^2 \mathbf{c}_l \mathbf{c}_l^T = \sum_{u=1}^k n_{l_u}^{-1} \hat{\sigma}_{l_u}^2 \mathbf{c}_{l_u} \mathbf{c}_{l_u}^T$, showing that $\mathbf{C}\hat{\boldsymbol{\mu}}$

and $C\hat{\Sigma}C^T$ are invariant under different labeling schemes of the population means and so is T .

The invariance of $\hat{\Delta}$ under different labeling schemes of the population means follows from (10) directly by noticing that

$$\begin{aligned}\hat{\Delta} &= \sum_{l=1}^k (n_l - 1)^{-1} \left[\frac{\hat{\sigma}_l^2}{n_l} \mathbf{c}_l^T (C\hat{\Sigma}C^T)^{-1} \mathbf{c}_l \right]^2 \\ &= \sum_{u=1}^k (n_u - 1)^{-1} \left[\frac{\hat{\sigma}_{l_u}^2}{n_{l_u}} \mathbf{c}_{l_u}^T (C\hat{\Sigma}C^T)^{-1} \mathbf{c}_{l_u} \right]^2,\end{aligned}$$

and that $C\hat{\Sigma}C^T$ is invariant under different labeling schemes of the population means. This completes the proof of the proposition. \square

Received 10 June 2011

REFERENCES

- BEHRENS, B. V. (1929). Ein Beitrag zur Fehlerberechnung bei wenige Beobachtungen. *Landwirtschaftliches Jahrbuch* **68** 807–837.
- DAJANI, A. N. and MATHEW, T. (2003). Comparison of some tests in the one-way ANOVA with unequal error variance. In *ASA Proceeding of the Joint Statistical Meetings*, 1149–1155.
- FISHER, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugenics* **11** 141–172.
- FOA, E. B., ROTHBAUM, B. O., RIGGS, D. S. and MURDOCK, T. B. (1991). Treatment of posttraumatic stress disorder in rape victims: A comparison between cognitive-behavioral procedures and counseling. *J. Consult. & Clin. Psych.* **59** 715–723.
- FUJIKOSHI, Y. (2000). Transformations with improved chi-squared approximations. *J. Multivar. Anal.* **72** 249–263. [MR1740633](#)
- JAMES, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika* **38** 324–329. [MR0046616](#)
- JAMES, G. S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika* **41** 19–43. [MR0068189](#)
- KRISHNAMOORTHY, K., LU, F. and MATHEW, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. *Comput. Statist. & Data Anal.* **51** 5731–5742. [MR2407673](#)
- KRUTCHKOFF, R. G. (1988). One-way fixed effects analysis of variance when the error variances may be unequal. *J. Statist. Comput. Simul.* **30** 259–271.
- KSHIRSAGAR, A. M. (1972). *Multivariate Analysis*. Marcel Decker, New York. [MR0343478](#)
- WEERAHANDI, S. (1995). *Exact Statistical Methods in Data Analysis*. Springer-Verlag, New York. [MR1316663](#)
- WELCH, B. L. (1947). The generalization of student's problem when several different population variances are involved. *Biometrika* **34** 28–35. [MR0019277](#)
- WELCH, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika* **38** 330–336. [MR0046617](#)
- WILCOX, R. R. (1988). A new alternative to the ANOVA F and new results on James's second order method. *Brit. J. Math. & Statist. Psychology* **41** 109–117. [MR0952886](#)
- WILCOX, R. R. (1989). Adjusting for unequal variances when comparing means in one-way and two-way fixed effects ANOVA model. *J. Educ. Statist.* **14** 269–278.
- YANAGIHARA, H. and YUAN, K. H. (2005). Three approximate solutions to the multivariate Behrens-Fisher problem. *Comm. Statist. Simul. Comput.* **34** 975–988. [MR2224272](#)

Jin-Ting Zhang
Department of Statistics and Applied Probability
National University of Singapore
Singapore 117546
Singapore
E-mail address: stazjt@nus.edu.sg

Xuefeng Liu
Department of Statistics and Applied Probability
National University of Singapore
Singapore 117546
Singapore
E-mail address: g0900756@nus.edu.sg