Ann Inst Stat Math (2012) 64:135–165
DOI 10.1007/s10463-010-0299-0

A modified two-factor multivariate analysis of variance: asymptotics and small sample approximations

Solomon W. Harrar · Arne C. Bathke

Received: 25 May 2009 / Revised: 23 February 2010 / Published online: 27 May 2010

Abstract In this paper, we present results for testing main, simple and interaction effects in heteroscedastic two factor MANOVA models. In particular, we suggest modifications to the MANOVA sum of squares and cross product matrices to account for heteroscedasticity. Based on these modified matrices, we define some multivariate test statistics and derive their asymptotic distributions under non-normality for the null as well as non-null cases. Derivation of these results relies on the perturbation method and limit theorems for independently distributed random matrices. Based on the asymptotic distributions, we devise small sample approximations for the quantiles of the null distributions. The numerical accuracy of the large sample as well as small sample approximations are favorable. A real data set from a Smoking Cessation Trial is analyzed to illustrate the application of the methods.

Keywords MANOVA · Perturbation method · Heteroscedasticity · Non-normality · Local alternatives · Multivariate tests

1 Introduction

Consider the multivariate linear model $\mathbf{Y}_{p\times n} = B_{p\times k}\mathbf{X}_{k\times n}+\mathcal{E}_{p\times n}$ where $\mathbf{X}$ is a known design matrix, B is a matrix of unknown parameters, and $\mathcal{E} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n)$ is a random

S. W. Harrar (✉)
Department of Mathematical Sciences, The University of Montana,
32 Campus Drive, Missoula, MT 59812, USA
e-mail: harrar@mso.umt.edu

A. C. Bathke
Department of Statistics, University of Kentucky,
875 Patterson Office Tower, Lexington, KY 40506, USA
e-mail: arne@uky.edu

matrix whose columns $\boldsymbol{\varepsilon}_i$ are independently distributed with mean $\mathbf{0}$ and positive definite variance-covariance matrix Σ_i. Consider a linear hypothesis $H_0 : BC = \mathbf{0}$ for some known full column rank matrix C. Assuming that $\boldsymbol{\varepsilon}_i$ has multivariate normal distribution and that $\Sigma_1 = \cdots = \Sigma_n$, the theory for testing the hypothesis H_0 is very well developed and documented in the multivariate statistics literature (Anderson 2003; Muirhead 1985). The most commonly used test statistics are the Likelihood Ratio (LR), Lawley–Hotelling (LH), and Bartlett–Nanda–Pillai (BNP) statistics. Exact null and non-null distributions of these statistics take quite complicated forms except in few special cases. Tabulations have been provided in some cases. Dempster (1958, 1960) devised a test statistic suited to the situation where the dimension is large and proposed an approximation to the null distribution by matching moments with that of the F-distribution.

On the other hand, these statistics are known to have an asymptotic chi-square distribution as the sample size tends to infinity. There are also satisfactory asymptotic expansions as a function of chi-square variables, see for example Anderson (2003) and Siotani et al. (1985). For the situation when the covariance matrices are not assumed equal, the earliest research appears to be by James (1954) and Ito (1969) who provided only approximate solutions for the one, two and k sample multivariate problems.

Under non-normality and equal covariance matrices, the null distributions of these statistics with the exception of Dempster's are known to converge to a chi-square limit distribution as the sample size tends to infinity, under certain restrictions on the design matrix (Huber 1973). There are also a few recent works on asymptotic expansions of the distributions of these statistics under non-normality (Fujikoshi 2002; Wakaki et al. 2002; Kakizawa 2009). Despite its theoretical as well as practical importance, the more general case without the assumption of constant covariance matrix has received only little attention. Part of the reason is that the asymptotic theory tends to be much more involved. Ito (1969) obtained asymptotic expansions for distributions of the multivariate statistics for testing equality of k mean vectors when the covariance matrices are unequal but normality holds. Later, Ito (1980) studied the robustness of one-way MANOVA problems when normality and homoscedasticity are violated one at a time. In these works, Ito noted that the effect of heteroscedasticity is substantially different on the different tests. Under non-normality, Kakizawa and Iwashita (2008) obtained asymptotic expansions for the null distribution of Hotelling's T^2-type counterpart of Welch's t-test statistic by including terms of order up to $1/n$. Kakizawa (2007) obtained asymptotic expansions of James (1954) statistic for the one-way MANOVA layout. One major problem with James' statistic is that its application is rather too complicated when there are two or more factors.

Besides focusing on the one-way layout problem, all the above asymptotic works assume that the hypothesis degrees of freedom remain fixed or small. Allowing the hypothesis degrees of freedom to go to infinity at the same rate as the error degrees of freedom, Fujikoshi (1975) derived asymptotic formulas for the null and non-null distributions of some multivariate statistics under normality and equal covariance matrices. In this paper, we are concerned with testing linear hypotheses in a multivariate factorial design setup. For brevity, we consider the two-way cross classified design, without making the assumptions of normality and equality of covariance matrices. Our asymptotic framework is that the replication sizes are fixed but the number of levels of one of the factors is large. It was shown in Harrar and Bathke (2008) that the distributions of the test statistics are sensitive to non-normality when the covariances and sample sizes per treatment are not constant. The underlying cause of this problem appears to be the weighting scheme in pulling the data together to get estimates of the within and between variabilities. Mindful of that, we redefine the estimates of these variabilities by using a suitable weighting scheme. Then the comparison of these measures of variabilities is done via the Dempster, LR, LH and BNP criteria to construct tests of significance for the main, simple, and interaction effects.

The asymptotic setup considered in this paper is becoming increasingly popular in view of the recent inventions of high throughput diagnostics and other biotechnologies such as fMRI and Microarrays which generate massive amounts of data to be analyzed. Another practical example in a Smoking Cessation Trial is discussed in Sect. 4.

More motivations for this type of asymptotics in agriculture, health sciences, and other disciplines are found in Boos and Brownie (1995), Akritas and Arnold (2000), Bathke (2002, 2004), Harrar and Gupta (2007), in univariate settings, and Gupta et al. (2006, 2008), Bathke and Harrar (2008), and Harrar and Bathke (2008) in the multivariate setting. Whereas Gupta et al. (2006, 2008) are restricted to the equal covariance case, Bathke and Harrar (2008) and Harrar and Bathke (2008) consider the single factor nonparametric situation. For a recent treatise on high-dimensional multivariate approximations, see also Fujikoshi et al. (2010).

In this paper, $\mathbf{0}$ will denote the vector $(0, \ldots, 0)'$, the dimension will be clear from the context, and $\mathbf{1}_n$ denotes an n-dimensional vector $(1, \ldots, 1)'$ consisting of ones. The matrix I_n is the identity matrix, and J_n, P_n are defined as $J_n = \mathbf{1}_n \cdot \mathbf{1}_n'$ and $P_n = \mathrm{I}_n - n^{-1} J_n$, respectively. The symbol $\overset{\mathcal{L}}{\rightarrow}$ stands as an abbreviation for "converges in law to" and $\overset{p}{\rightarrow}$ for "converges in probability to". Throughout the paper, we will extensively use the Kronecker (or direct) product $A \otimes B$ of matrices, the direct sum $A \oplus B$ of matrices, the vec operator that stacks columns of a matrix on top of each other, and the commutation matrix $K_{m,n}$. See Magnus and Neudecker (1979) for properties of the commutation matrix.

The paper is organized as follows. Section 2 presents the modified test statistics and their distributions under the null hypothesis as well as local alternatives. Also in Sect. 2, the numerical accuracy of the asymptotic distributions are investigated. Then Sect. 3 devises finite sample approximations based on the robustness results obtained in Sect. 2 and some existing results under normality and constant covariance matrices. The analysis of a real data set is presented in Sect. 4. An R-Script which implements the methods of this paper can be obtained from the authors by request. We will eventually develop a self contained R-package and make it available via CRAN. Section 5 contains discussions and some concluding remarks. For the sake of clarity and efficient presentation of the ideas, proofs and other technical details are placed in the Appendix.

2 Heteroscedastic MANOVA

2.1 Preliminaries

Let $\mathbf{Y}_{ijk}$ be independent p-dimensional random vectors with mean vector $\boldsymbol{\mu}_{ij}$ and covariance matrix Σ_{ij} for $i = 1, \ldots, a$, $j = 1, \ldots, b$, and $k = 1, \ldots, n_{ij}$. Consider the model $\boldsymbol{\mu}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij}$, where $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_j$, and $\boldsymbol{\gamma}_{ij}$ are unknown constants corresponding to the effects due to the two factors A and B and their interaction AB. We assume the identifiability constraints $\sum_i \boldsymbol{\alpha}_i = \sum_j \boldsymbol{\beta}_j = \mathbf{0}$, $\sum_j \boldsymbol{\gamma}_{ij} = \mathbf{0}$ for $i = 1, \ldots, a$, and $\sum_i \boldsymbol{\gamma}_{ij} = \mathbf{0}$ for $j = 1, \ldots, b$.

The hypotheses of interest for this treatment design are as follows.

(a) $\mathcal{H}_0^{(A)} : \boldsymbol{\alpha}_i = \mathbf{0}$ for $i = 1, 2, \ldots, a$—which means no main effects of levels of factor A,
(b) $\mathcal{H}_0^{(A|B)} : \boldsymbol{\alpha}_i + \boldsymbol{\gamma}_{ij} = \mathbf{0}$ for $i = 1, 2, \ldots, a$ and $j = 1, 2, \ldots, b$—which means no simple effects of levels of factor A,
(c) $\mathcal{H}_0^{(B)} : \boldsymbol{\beta}_j = \mathbf{0}$ for $j = 1, 2, \ldots, b$—which means no main effects of levels of factor B,
(d) $\mathcal{H}_0^{(B|A)} : \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij} = \mathbf{0}$ for $i = 1, 2, \ldots, a$ and $j = 1, 2, \ldots, b$—which means no simple effects of levels of factor B, and
(e) $\mathcal{H}_0^{(AB)} : \boldsymbol{\gamma}_{ij} = \mathbf{0}$ for $i = 1, 2, \ldots, a$ and $j = 1, 2, \ldots, b$—which means no interaction effects of levels of factor A and levels of factor B.

The asymptotic setup is that the number of levels of one of the factors, say A, is large but the sample size and the number of levels of the other factor remain fixed. In this asymptotic situation, the problems of testing main and simple effects of factor B seemingly fall in the usual large n asymptotic framework. However, the techniques and the results will be new and nontrivial due to the modification of sum of squares and cross products matrices to be introduced next.

Define,

$$
\begin{aligned}
H^{(A)} &= \frac{1}{a-1} \sum_{i=1}^{a} \sum_{j=1}^{b} (\tilde{\mathbf{Y}}_{i..} - \tilde{\mathbf{Y}}_{...})(\tilde{\mathbf{Y}}_{i..} - \tilde{\mathbf{Y}}_{...})', \\
H^{(A|B)} &= \frac{1}{(a-1)b} \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{\mathbf{Y}}_{ij.} - \tilde{\mathbf{Y}}_{.j.})(\bar{\mathbf{Y}}_{ij.} - \tilde{\mathbf{Y}}_{.j.})', \\
H^{(B)} &= \frac{1}{b-1} \sum_{i=1}^{a} \sum_{j=1}^{b} (\tilde{\mathbf{Y}}_{.j.} - \tilde{\mathbf{Y}}_{...})(\tilde{\mathbf{Y}}_{.j.} - \tilde{\mathbf{Y}}_{...})', \qquad (1) \\
H^{(B|A)} &= \frac{1}{a(b-1)} \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{\mathbf{Y}}_{ij.} - \tilde{\mathbf{Y}}_{i..})(\bar{\mathbf{Y}}_{ij.} - \tilde{\mathbf{Y}}_{i..})', \\
H^{(AB)} &= \frac{1}{(a-1)(b-1)} \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{\mathbf{Y}}_{ij.} - \tilde{\mathbf{Y}}_{i..} - \tilde{\mathbf{Y}}_{.j.} + \tilde{\mathbf{Y}}_{...})(\bar{\mathbf{Y}}_{ij.} - \tilde{\mathbf{Y}}_{i..} - \tilde{\mathbf{Y}}_{.j.} + \tilde{\mathbf{Y}}_{...})',
\end{aligned}
$$

and

$$G = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\sum_{k=1}^{n_{ij}}(\mathbf{Y}_{ijk}-\bar{\mathbf{Y}}_{ij.})(\mathbf{Y}_{ijk}-\bar{\mathbf{Y}}_{ij.})' = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}}S_{ij},$$

where $\bar{\mathbf{Y}}_{ij.} = \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}}\mathbf{Y}_{ijk}$, $\tilde{\mathbf{Y}}_{i..} = \frac{1}{b}\sum_{j=1}^{b}\bar{\mathbf{Y}}_{ij.}$, $\tilde{\mathbf{Y}}_{.j.} = \frac{1}{a}\sum_{i=1}^{a}\bar{\mathbf{Y}}_{ij.}$, $\tilde{\mathbf{Y}}_{...} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{\mathbf{Y}}_{ij.}$, and $S_{ij} = \frac{1}{(n_{ij}-1)}\sum_{k=1}^{n_{ij}}(\mathbf{Y}_{ijk}-\bar{\mathbf{Y}}_{ij.})(\mathbf{Y}_{ijk}-\bar{\mathbf{Y}}_{ij.})'$.

The mean vectors $\boldsymbol{\mu}_{ij}$ and covariance matrices Σ_{ij} differ from cell to cell. Thus, it does not make sense to use the weighting scheme used in homoscedastic MANOVA when combining the data in the process of forming the hypothesis mean squares and cross products matrices. More specifically, let $\tilde{\boldsymbol{\mu}}_{p\times ab} = (\boldsymbol{\mu}_{11},\ldots,\boldsymbol{\mu}_{1b},\boldsymbol{\mu}_{21},\ldots,\boldsymbol{\mu}_{ab})$. The hypothesis of no interaction effects can be written as $\mathcal{H}_0^{(AB)} : \tilde{\boldsymbol{\mu}}(\mathbf{P}_a\otimes\mathbf{P}_b) = \mathbf{0}$. Define $\bar{\mathbf{Y}}_{.} = (\bar{\mathbf{Y}}_{11.},\ldots,\bar{\mathbf{Y}}_{1b.},\bar{\mathbf{Y}}_{21.},\ldots,\bar{\mathbf{Y}}_{ab.})$. It is reasonable to consider $\mathbf{Q}^{(AB)} = \bar{\mathbf{Y}}_{.}(\mathbf{P}_a\otimes\mathbf{P}_b)\bar{\mathbf{Y}}_{.}'$ as a hypothesis sum of squares and cross products matrix for $\mathcal{H}_0^{(AB)}$. Then, with the aid of Eqs. (9), one can see that $\mathbf{Q}^{(AB)} = (a-1)(b-1)H^{(AB)}$. This idea is similar to the quadratic forms in Brunner et al. (1997) for their ANOVA-type statistics in the univariate setting. Analogous arguments can be used to justify the other sum of squares and cross products matrices.

As an estimate of the within variability, we take the average of the estimators of the variances of the cell mean vectors. These sum of squares and cross products matrices correspond to the type III sum of squares and cross products matrices in unweighted means analysis of multivariate linear models. In the univariate case these types of mean squares have been used, e.g., by Wang and Akritas (2004, 2006).

2.2 Test statistics

Before we define the test statistics, we show that the sum of squares and cross products matrices given in the previous section have the same expected values under the null hypotheses. First, define Σ by $\Sigma = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}}\Sigma_{ij}$. From here on, we assume that $\Sigma = O(1)$.

It is shown in Appendix A that for each of $H^{(AB)}$, $H^{(A|B)}$, $H^{(A)}$, $H^{(B)}$, and $H^{(B|A)}$, the expected value under the corresponding null hypothesis is equal to Σ. That is,

$$\text{under } \mathcal{H}_0^{(\psi)},\quad E\left(H^{(\psi)}\right) = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}}\Sigma_{ij} = \Sigma \tag{2}$$

where ψ can be each of the effects under consideration: AB, $A|B$, A, B or $B|A$.

In view of (2) and the fact that $E(G) = \Sigma$, we can compare the hypothesis and error matrices to obtain a meaningful test statistic. In multivariate analysis of variance (MANOVA), there is a multitude of test statistics to choose from. None of these statistics perform uniformly better than the others in the whole parameter space. For this manuscript, we consider the four most commonly used test statistics, namely the

Dempster, LR, LH, and BNP criteria. In what follows, we present the test statistics for $\mathcal{H}_0^{(\psi)}$. Here also, ψ can be each of the effects under consideration: AB, $A|B$, A, B, or $B|A$.

(a) Dempster's ANOVA Type criterion: $T_{\mathrm{D}}^{(\psi)} = \mathrm{tr}(H^{(\psi)})/\mathrm{tr}(G)$.
(b) Wilks' Lambda (Likelihood Ratio) criterion: $T_{\mathrm{LR}}^{(\psi)} = -\log(|G|/|H^{(\psi)} + G|)$.
(c) The Lawley–Hotelling criterion: $T_{\mathrm{LH}}^{(\psi)} = \mathrm{tr}(H^{(\psi)}G^{-1})$.
(d) The Bartlett–Nanda–Pillai criterion: $T_{\mathrm{BNP}}^{(\psi)} = \mathrm{tr}\left(H^{(\psi)}(H^{(\psi)} + G)^{-1}\right)$.

2.3 Asymptotic distributions under the null hypotheses

We have seen that the matrix G is an unbiased estimator of Σ. The following theorem asserts that the difference $G - \Sigma$ is asymptotically ($a \to \infty$) negligible.

Theorem 1 *Assume that the n_{ij} are bounded and that $\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}^{-2}(n_{ij} - 1)^{-1}\Sigma_{ij} \otimes \Sigma_{ij} = o(a^2)$ and $\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}^{-3}K_4(\mathbf{Y}_{ij1}) = o(a^2)$ as $a \to \infty$. Then $G - \Sigma \xrightarrow{p} 0$ as $a \to \infty$.*

The conditions in Theorem 1 basically require the weighted sum of the fourth and second order moment do not grow at a rate faster than a^2. Apparently, these conditions become rather mild (in terms of a cell sample sizes) if we assume that the fourth and second order moments of the data in the different treatment groups are uniformly bounded.

In the remainder of this section we obtain the asymptotic null distributions of the four test statistics for testing the main, simple, and interaction effects. Since the results for testing $\mathcal{H}^{(AB)}$, $\mathcal{H}^{(A)}$, $\mathcal{H}^{(A|B)}$ and $\mathcal{H}^{(B|A)}$ are similar in form and their derivations proceed along the same lines, we group them under the same heading in the following subsection.

2.3.1 Testing $\mathcal{H}^{(AB)}$, $\mathcal{H}^{(A)}$, $\mathcal{H}^{(A|B)}$ and $\mathcal{H}^{(B|A)}$

We know from Theorem 1 that, under two technical assumptions, $G - \Sigma = o_p(1)$ as $a \to \infty$ and it is established in Theorem 2 below that $\sqrt{a}(H^{(\psi)} - G)\Omega = O_p(1)$ as $a \to \infty$ and for any matrix of constants Ω. Then the following expansions can be easily verified.

$$T_D^{(\psi)} = 1 + \frac{1}{\sqrt{a}}\left[\sqrt{a}\ \mathrm{tr}(H^{(\psi)} - G)\cdot\frac{1}{\mathrm{tr}(\Sigma)}\right] + o_p(a^{-1/2}),$$
$$T_{\mathrm{LR}}^{(\psi)} = \log|I_p + H^{(\psi)}G^{-1}| = p\log 2 + \frac{1}{2\sqrt{a}}\left[\sqrt{a}\,\mathrm{tr}(H^{(\psi)} - G)\Sigma^{-1}\right] + o_p(a^{-1/2}),$$
$$T_{\mathrm{LH}}^{(\psi)} = p + \frac{1}{\sqrt{a}}\left[\sqrt{a}\,\mathrm{tr}(H^{(\psi)} - G)\Sigma^{-1}\right] + o_p(a^{-1/2}) \tag{3}$$

and

$$T_{\text{BNP}}^{(\psi)} = \operatorname{tr}\left(H^{(\psi)}G^{-1}(I_p + H^{(\psi)}G^{-1})^{-1}\right)$$
$$= \frac{p}{2} + \frac{1}{4\sqrt{a}}\left[\sqrt{a}\operatorname{tr}(H^{(\psi)} - G)\Sigma^{-1}\right] + o_p(a^{-1/2}).$$

Considering the expansions in (3), one can see that all four test statistics, scaled and centered suitably, may be expressed as

$$\sqrt{a}\big(\ell T_{\mathcal{G}}^{(\psi)} - h\big) = \sqrt{a}\operatorname{tr}\big(H^{(\psi)} - G\big)\Omega + o_p(1), \tag{4}$$

where $\ell = 1, 2, 1, 4$, $h = 1, 2p\log 2, p, 2p$ and $\Omega = (1/\operatorname{tr}\Sigma)\mathrm{I}_p, \Sigma^{-1}, \Sigma^{-1}, \Sigma^{-1}$ for $\mathcal{G} = \mathrm{D}, \mathrm{LR}, \mathrm{LH}, \mathrm{BNP}$, respectively.

In light of the expression (4), the null distributions of the four test statistics can be derived in a unified manner by obtaining the null distribution of $\sqrt{a}\operatorname{tr}(H^{(\psi)} - G)\Omega$ for any fixed matrix Ω. The null distribution of the latter quantity is given in Theorem 2. The theorem needs the following technical assumptions to hold.

Assumption 1 For some $\delta > 0$, $E|(\mathbf{Y}_{ij1} - \boldsymbol{\mu}_{ij}^{(\psi)})'\Sigma_{ij}^{-1}(\mathbf{Y}_{ij1} - \boldsymbol{\mu}_{ij}^{(\psi)})|^{2+\delta} < \infty$ where $\mathbf{Y}_{ijk}$ and $\boldsymbol{\mu}_{ij}^{(\psi)}$ are as defined in Theorem 2.

Assumption 2 For some $\delta > 0$,

$$\lim_{a\to\infty}\frac{1}{a}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}^{1+\delta/2}(n_{ij}-1)^{1+\delta/2}}\operatorname{tr}(\Omega\Sigma_{ij})^{2+\delta} < \infty$$

and

$$\lim_{a\to\infty}\frac{1}{a}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\frac{1}{n_{ij}^{1+\delta/2}n_{ij'}^{1+\delta/2}}\operatorname{tr}(\Omega\Sigma_{ij}\Omega\Sigma_{ij'})^{1+\delta/2} < \infty.$$

As one can readily understand, Assumptions 1 and 2 are needed for the application of Liaponauv's Central Limit Theorem in proving the following theorem. Before stating the theorem, though, we introduce a notation which greatly facilitates a succinct presentation. Let

$$\boldsymbol{\mu}_{ij}^{(\psi)} = \begin{cases} \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j & \text{if } \psi = AB \\ \boldsymbol{\mu} + \boldsymbol{\beta}_j & \text{if } \psi = A \text{ or } A|B\,. \\ \boldsymbol{\mu} + \boldsymbol{\alpha}_i & \text{if } \psi = B \text{ or } B|A \end{cases} \tag{5}$$

Theorem 2 *Let $\psi = AB, A, A|B$ or $B|A$. Suppose that under the hypothesis $\mathcal{H}_0^{(\psi)}$, $\mathbf{Y}_{ijk}$ are independently distributed with mean vector $\boldsymbol{\mu}_{ij}^{(\psi)}$ and covariance matrix Σ_{ij} for $i = 1, \ldots, a$, $j = 1, \ldots, b$ and $k = 1, \ldots, n_{ij}$. Then, under the assumptions 1 and 2, $\sqrt{a}\operatorname{tr}(H^{(\psi)} - G)\Omega \xrightarrow{\mathcal{L}} N\big(0, \tau_\psi^2(\Omega)\big)$ as $a \to \infty$ and n_{ij} and b bounded, where*

$$\tau_{\psi}^{2}(\Omega)=\begin{cases}\frac{2}{b}\left\{v_1(\Omega)+\frac{v_2(\Omega)}{(b-1)^2}\right\} & \textit{when} \quad \psi=AB\\ \frac{2}{b}\{v_1(\Omega)+v_2(\Omega)\} & \textit{when} \quad \psi=A\\ \frac{2}{b}v_1(\Omega) & \textit{when} \quad \psi=A|B\\ \frac{2}{b^2}\left\{v_1(\Omega)+\frac{v_2(\Omega)}{(b-1)^2}\right\} & \textit{when} \quad \psi=B|A\end{cases}.$$

Here, $v_1(\Omega)=\lim_{a\to\infty}\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{\operatorname{tr}(\Omega\Sigma_{ij})^2}{n_{ij}(n_{ij}-1)}$ *and* $v_2=\lim_{a\to\infty}\frac{1}{ab}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\frac{\operatorname{tr}(\Omega\Sigma_{ij}\Omega\Sigma_{ij'})}{n_{ij}n_{ij'}}$*, assuming the limits exist.*

Under the assumptions and notations of Theorem 2, the asymptotic distribution of Dempster's ANOVA type criterion can be obtained by setting $\Omega=(1/\operatorname{tr}\Sigma)I_p$. For the other three criteria, we set $\Omega=\Sigma^{-1}$ to get the asymptotic null distributions.

Needless to say, the asymptotic null distributions of T_{LR}, T_{LH} and T_{BNP}, scaled and centered as in (4), are the same up to the order $O(a^{-1/2})$. A comparison of the asymptotic variances in Theorem 2 reveals that the test statistic for the interaction effect has smaller variance than that of the main effect. Also we see from the asymptotic variances in Theorem 2 that the test statistic for the simple effect of A has smaller variance compared to that of either the interaction or main effects. In addition, it is apparent from the theorem that the sizes of the four tests are asymptotically robust.

2.3.2 Consistent estimator of $\tau_{\psi}^{2}(\Omega)$

In practice, we need a consistent estimator of $\tau_{\psi}^{2}(\Omega)$ for $\Omega=(1/\operatorname{tr}\Sigma)I_p$ and $\Omega=\Sigma^{-1}$ to apply Theorem 2. Since $G-\Sigma=o_p(1)$ as $a\to\infty$, we only need to find a consistent estimator of $\tau_{\psi}^{2}(\Omega)$ assuming Ω is a known constant matrix. The following Theorem provides such an estimator. A similar estimator has been used in a univariate setting in Wang and Akritas (2009).

Theorem 3 *Let the model and assumptions be as in Theorem 2. Further assume the eighth order moments of* $\mathbf{Y}_{ijk}$ *exist and define*

$$\begin{aligned}\widehat{\Psi}_{ij}(\Omega)=\frac{1}{4c_{ij}}\sum_{(k_1,k_2,k_3,k_4)\in\mathcal{K}}^{n_{ij}}&\Omega(\mathbf{Y}_{ijk_1}-\mathbf{Y}_{ijk_2})(\mathbf{Y}_{ijk_1}-\mathbf{Y}_{ijk_2})'\\ &\times\Omega(\mathbf{Y}_{ijk_3}-\mathbf{Y}_{ijk_4})(\mathbf{Y}_{ijk_3}-\mathbf{Y}_{ijk_4})',\end{aligned}$$

where $\mathcal{K}$ *is the set of all quadruples* $\kappa=(k_1,k_2,k_3,k_4)$ *where no element in* κ *is equal to any other element in* κ*, and* $c_{ij}=n_{ij}(n_{ij}-1)(n_{ij}-2)(n_{ij}-3)$*. Then, as* $a\to\infty$*,*

$$\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\operatorname{tr}(\widehat{\Psi}_{ij}(\Omega))-\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\operatorname{tr}(\Omega\Sigma_{ij})^2=o_p(1)$$

and

$$\frac{1}{ab}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\frac{1}{n_{ij}n_{ij'}}\mathrm{tr}(\Omega S_{ij}\Omega S_{ij'})-\frac{1}{ab}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\frac{1}{n_{ij}n_{ij'}}\mathrm{tr}(\Omega\Sigma_{ij}\Omega\Sigma_{ij'})=o_p(1).$$

In some cases it may be appropriate to assume that $\Sigma_{ij} = \Sigma_j$ (cf. Bathke 2004). Under this assumption, $\frac{1}{a}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}}\Sigma_{ij} = \sum_{j=1}^{b}\underline{n}_{.j}\Sigma_j$, where $\underline{n}_{.j} = \frac{1}{a}\sum_{j=1}^{a}\frac{1}{n_{ij}}$. Thus, assuming $\underline{n}_{.j} = O(1)$ as $a \to \infty$, a consistent estimator of Σ_j and, therefore, of Σ can be obtained by pooling the estimates from each level of factor A. This estimator will obviously be consistent as $a \to \infty$ and uses all information in the samples.

2.3.3 Testing for the main effects of factor B

For testing the main effects of factor B, the asymptotic framework $a \to \infty$ is conceptually similar to the usual large replication size asymptotics, and therefore, we expect a chi-squared type asymptotic null distribution. In the proof of Theorem 4 (see Appendix B), we will see that $(b-1)\mathrm{tr}(H^{(B)}G^{-1}) = O_p(1)$. Thus, the expansion $(b-1)T_{\mathcal{G}}^{(B)} = (b-1)\mathrm{tr}(H^{(B)}G^{-1}) + O_p(\frac{1}{a})$ holds for $\mathcal{G} = $ LR and BNP. Therefore, in light of the fact that $G - \Sigma = o_p(1)$ as $a \to \infty$, the asymptotic null distribution of the four test statistics can be derived in a single stroke if we have the asymptotic null distribution of $(b-1)$ $\mathrm{tr}(H^{(B)}\Omega)$ for any constant matrix Ω. Before we present the the result, we state an assumption needed for the application of Liaponauv's Central Limit Theorem.

Assumption 3 For some $\delta > 0$, $E\|\mathbf{Y}_{ij1} - \boldsymbol{\mu}_{ij}^{(B)}\|^{2+\delta} < \infty$ where $\mathbf{Y}_{ijk}$ and $\boldsymbol{\mu}_{ij}^{(B)}$ are as defined in Theorem 4.

Theorem 4 *Suppose $\mathbf{Y}_{ijk}$ are independently distributed with mean vector $\boldsymbol{\mu}_{ij}^{(B)} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i$ and covariance matrix Σ_{ij} for $i = 1, \ldots, a$, $j = 1, \ldots, b$ and $k = 1, \ldots, n_{ij}$. Then, under Assumption 3 and the assumption $(1/a)\sum_{i=1}^{a} n_{ij}^{-1}\Sigma_{ij} = O(1)$ as $a \to \infty$,*

$$(b-1)\mathrm{tr}H^{(B)}\Omega \xrightarrow{\mathcal{L}} \sum_{k=1}^{(b-1)p}\lambda_k\chi_{1,k}^2 \tag{6}$$

where n_{ij} and b are bounded and λ_k is the kth largest eigenvalue of Λ defined by $\Lambda = (P_b \otimes \Omega^{1/2})(\frac{1}{a}\sum_{i=1}^{a}(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\Sigma_{ij}))(P_b \otimes \Omega^{1/2})$. Here, $\chi_{1,k}^2$, $k = 1, 2, \ldots, (b-1)p$, stands for independent chi-square random variables each with one degree of freedom.

The application of Theorem 4 requires a reasonable approximation to the right hand side of (6) and a consistent estimator of Λ or at least of its $p(b-1)$ nonzero eigenvalues $\lambda_1, \ldots, \lambda_{(b-1)p}$. A well known approximation to the distribution of a linear combination of independent chi–squared random variables is a constant multiple of a chi-square distribution, say $g\chi_f^2$, where g and f are determined by the method of moments. The moment matching principle applied here leads to $f = \frac{\mathrm{tr}(\Lambda)^2}{\mathrm{tr}(\Lambda^2)}$ and $g = \frac{\mathrm{tr}(\Lambda^2)}{\mathrm{tr}(\Lambda)}$. Now consistent estimators $\hat{f}$ and $\hat{g}$ can be obtained by replacing Λ with its consistent estimator,

$$\hat{\Lambda} = (P_b \otimes \hat{\Omega}^{1/2}) \left(\frac{1}{a} \sum_{i=1}^{a} \left(\bigoplus_{j=1}^{b} \frac{1}{n_{ij}} S_{ij} \right) \right) \left(P_b \otimes \hat{\Omega}^{1/2} \right)$$

where $\hat{\Omega}$ is obtained from Ω by replacing Σ with G. The consistency of G for Σ is established in Theorem 1. The same argument as in the proof of Theorem 1 can be used to establish

$$\frac{1}{a} \sum_{i=1}^{a} \left(\bigoplus_{j=1}^{b} \frac{1}{n_{ij}} S_{ij} \right) - \frac{1}{a} \sum_{i=1}^{a} \left(\bigoplus_{j=1}^{b} \frac{1}{n_{ij}} \Sigma_{ij} \right) = o_p(1) \quad \text{as } a \to \infty.$$

On the other hand, $\Lambda = P_b \otimes I_p$ for the Wilks Lambda, Lawley–Hotelling and Bartlett–Nanda–Pillai criteria when the quantity $(1/a) \sum_{i=1}^{a} n_{ij}^{-1} \Sigma_{ij}$ does not depend on the index j. In this case Λ is idempotent and therefore $(b-1) T_{\mathrm{G}}^{(B)} \xrightarrow{\mathcal{L}} \chi^2_{(b-1)p}$ for $\mathrm{G} = \mathrm{LR}, \mathrm{LH}, \mathrm{BNP}$. In the case of $T_{\mathrm{D}}^{(B)}$, $\Lambda = P_b \otimes \Sigma$. Then,

$$T_{\mathrm{D}}^{(B)} \xrightarrow{\mathcal{L}} \sum_{k=1}^{p} \frac{\sigma_k}{\mathrm{tr}\,\Sigma} \chi^2_{(b-1),k} \tag{7}$$

where $\sigma_1, \ldots, \sigma_p$ are the eigenvalues of Σ. Here again applying the moment matching principle to the right hand side of (7), we can use the approximation $g\chi^2_f$ where consistent estimators of f and g are given by $\mathrm{tr}(G)^2/\mathrm{tr}(G^2)$ and $\mathrm{tr}(G^2)/\mathrm{tr}(G)^2$, respectively.

2.4 Asymptotic distributions under local alternatives

In this section, we give the asymptotic distributions under some reasonable local alternatives. Here also we present the results for the general cases and the distributions for each of the four test statistics can be obtained by making the necessary substitution for Ω. Here also we introduce a handy notation which allows a concise presentation of the results. Denote

$$\boldsymbol{\mu}_{ij,a}^{\psi} = \boldsymbol{\mu}_{ij}^{\psi} + \begin{cases} \boldsymbol{\gamma}_{ij,a} & \text{if} \quad \psi = AB \\ \boldsymbol{\alpha}_{i,a} & \text{if} \quad \psi = A \\ \boldsymbol{\alpha}_{i,a} + \boldsymbol{\gamma}_{ij,a} & \text{if} \quad \psi = A|B \\ \boldsymbol{\beta}_{j,a} & \text{if} \quad \psi = B \\ \boldsymbol{\beta}_{j,a} + \boldsymbol{\gamma}_{ij,a} & \text{if} \quad \psi = B|A \end{cases},$$

where $\boldsymbol{\mu}_{ij}^{\psi}$ is as defined by (5) and the triangular arrays $\boldsymbol{\alpha}_{i,a}$, $\boldsymbol{\beta}_{j,a}$ and $\boldsymbol{\gamma}_{ij,a}$ satisfy the following constraints for each positive integer a: $\sum_{i=1}^{a} \boldsymbol{\alpha}_{i,a} = \sum_{j=1}^{b} \boldsymbol{\beta}_{j,a} = \mathbf{0}$, $\sum_{j=1}^{b} \boldsymbol{\gamma}_{ij,a} = \mathbf{0}$ for $i = 1, \ldots, a$ and $\sum_{i=1}^{a} \boldsymbol{\gamma}_{ij,a} = \mathbf{0}$ for $j = 1, \ldots, b$. Also let

$$\Theta_{\psi} = \begin{cases} \dfrac{1}{\sqrt{a}} \dfrac{1}{(b-1)} \displaystyle\sum_{i=1}^{a} \sum_{j=1}^{b} \boldsymbol{\gamma}_{ij,a} \boldsymbol{\gamma}'_{ij,a} & \text{if } \psi = AB \\ \dfrac{1}{\sqrt{a}} \displaystyle\sum_{i=1}^{a} \boldsymbol{\alpha}_{i,a} \boldsymbol{\alpha}'_{i,a} & \text{if } \psi = A \\ \dfrac{1}{\sqrt{a}} \dfrac{1}{b} \displaystyle\sum_{i=1}^{a} \sum_{j=1}^{b} (\boldsymbol{\alpha}_{i,a} + \boldsymbol{\gamma}_{ij,a})(\boldsymbol{\alpha}_{i,a} + \boldsymbol{\gamma}_{ij,a})' & \text{if } \psi = A|B \\ \dfrac{1}{\sqrt{a}} \dfrac{1}{(b-1)} \displaystyle\sum_{i=1}^{a} \sum_{j=1}^{b} (\boldsymbol{\beta}_{j,a} + \boldsymbol{\gamma}_{ij,a})(\boldsymbol{\beta}_{j,a} + \boldsymbol{\gamma}_{ij,a})' & \text{if } \psi = B|A \end{cases}.$$

We assume $\Theta_{\psi} = O(1)$. For example, when $\psi = AB$ this assumption will be reasonable if $\boldsymbol{\gamma}_{ij,a} = O(a^{-1/4})$. So the following Theorem states that the four test statistics will be able to detect a sequence of alternative points converging to $\mathbf{0}$ at the rate $a^{-1/4}$.

Theorem 5 *Let $\psi \in \{AB, A, A|B, B|A\}$. Suppose $\mathbf{Y}_{ijk}$ are independently distributed with mean $\boldsymbol{\mu}_{ij,a}^{\psi}$ and covariance Σ_{ij}. Then under the assumptions and notations of Theorem 2, $\sqrt{a}\,\mathrm{tr}(H^{(\psi)} - G)\Omega \xrightarrow{\mathcal{L}} N(\mathrm{tr}\Theta_{\psi}\Omega, \tau_{\psi}^2(\Omega))$ where $\tau_{\psi}^2(\Omega)$ is as defined in Theorem 2.*

Notice that the asymptotic power of the four tests under the local alternative in Theorem 5 is a monotone function of $\mathrm{tr}\ \Theta_{\psi}\Omega/\sqrt{\tau_{\psi}^2(\Omega)}$. The four test statistics will have the same asymptotic power if $\Omega = cI_p$ for any non-negative constant c. That would be the case, for example, when $\Sigma_{ij} = I_p$.

Next we present the asymptotic distribution of the test statistics for the main effects of factor B under local alternatives.

Theorem 6 *Suppose $\mathbf{Y}_{ijk}$ are independently distributed with mean vector $\boldsymbol{\mu}_{ij,a}^{B}$ and covariance matrix Σ_{ij} for $i = 1, \ldots, a$, $j = 1, \ldots, b$ and $k = 1, \ldots, n_{ij}$. Assume $\boldsymbol{\beta}_{j,a} = a^{-1/2}\boldsymbol{\beta}_{j,0}$ where $\boldsymbol{\beta}_{j,0}$ are fixed vectors of constants. Then under Assumption 3 and the assumption that $(1/a)\sum_{i=1}^{a} n_{ij}^{-1}\Sigma_{ij} = O(1)$ as $a \to \infty$,*

$$(b-1)\mathrm{tr} H^{(B)}\Omega \xrightarrow{\mathcal{L}} \sum_{k=1}^{(b-1)p} \lambda_k \chi_{1,k}^2(\mathrm{tr} B P_b B'\Omega)$$

as $a \to \infty$ where $B = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_b)$, Λ is as defined in Theorem 4, λ_i is the kth largest eigenvalue of Λ as $a \to \infty$ and, n_{ij} and b are bounded where $\chi_{1,k}^2(\theta)$ for

$k = 1, 2, \ldots, (b-1)p$ are independent chi-square random variables each with one degree of freedom and non-centrality parameter θ.

2.5 Numerical accuracy of the asymptotic distributions

To assess the quality of the asymptotic null distributions in Theorem 2, a simulation study was conducted by generating data from multivariate normal and multivariate skew normal (Azzalini and Capitanio 1999) distributions with dimension $p = 3$. In each case, three different structures for the covariance matrices were considered. These structures are (i) Σ_{ij} = free of i and j, (ii) $\Sigma_{ij} = (1-\rho_{ij})I_p + \rho_{ij}\mathbf{1}_p\mathbf{1}_p'$ where $\rho_{ij} = (ij)^{1/2}/(1+ij)$ and (iii) $\Sigma_{ij} = ijI_p$. We denote these three structures by Σ_1, Σ_2 and Σ_3, respectively. Note that covariance structure Σ_3 violates the assumptions of Theorem 2.2, and the tests are, therefore, not necessarily expected to perform well in the simulation study. Five values of $a = 10, 20, 35, 50$, and 100 were considered, along with $b = 3$. For the replication sizes, we set $n_{ij} = 4$ for $a-1$ cells, $n_{ij} = 5$ for $(b-1)$ cells and $n_{ij} = 6$ for $(a-1)(b-1)+1$ cells.

In Table 1, results for testing the main effect of A and the interaction effect AB are displayed. Considering the results for the structures Σ_1 and Σ_2, it can be seen from the table that the test statistic T_{BNP} appeared to be conservative for all values of a considered, in particular, when testing the main effect. In this case, the quality of approximation was not quite adequate even for $a = 50$ and 100. On the other hand, T_{D} and T_{LH} appeared to be liberal for testing both the main effect of factor A and the interaction effect. However, the quality of approximation improved quickly and attained the desired size fairly close at $a = 50$ and 100. The test statistic T_{LR} appeared to be the best in terms of achieving the desired size. More so, in particular, for testing the interaction effect. Table 1 also shows that heteroscedasticity and skewness did not adversely affect the size of the tests.

Despite the fact that structure Σ_3 violates the assumptions of Theorems 2, we see that this did not seem to have an effect under normality. When the data contained skewness, however, the numerical accuracy of the sizes was only acceptable for the test of interaction effect, but not for testing the main effects. Nevertheless, the increasing tendency of the sizes as a gets bigger is undesirable. This indicates that the assumptions seem to be needed in practice, and that they are not merely abstract technical conditions. In sum, the simulations give some compelling evidence that the assumptions of the theorem on the covariances are important only for testing the main effect. Even in this case the assumption seems to be required only when there is departure from normality.

Overall, the actual type I error rates were better controlled at the desired level for the interaction effect than the main effect. This is expected because the sampling variance of the test statistic for the interaction effect is smaller than that of the main effect.

3 Finite sample approximations

Let $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ be mutually independently distributed as $N_p(\mathbf{0}, \Sigma_i)$ where $\Sigma_i > 0$. Define $\mathbf{Q} = \mathbf{Y}C\mathbf{Y}'$ where $C = (c_{ij})$ is an $n \times n$ symmetric non-negative

Table 1 Simulated actual sizes (expressed as percentages) for the four test statistics when sampling from multivariate normal and multivariate Skew–Normal distributions under the three heteroscedasticity structures Σ_1, Σ_2, and Σ_3

Cov.	Pop.	Test Stat	$a = 10$		$a = 20$		$a = 35$		$a = 50$		$a = 100$	
			A	AB	A	AB	A	AB	A	AB	A	AB
Σ_1	MN	T_{D}	10.0	8.0	7.3	7.1	7.1	5.5	7.3	7.5	6.3	6.0
		T_{LR}	4.0	4.1	2.7	5.0	3.8	4.9	4.9	4.9	3.1	6.5
		T_{LH}	10.7	9.1	7.6	8.9	7.7	7.2	7.8	7.4	5.5	7.4
		T_{BNP}	1.3	1.2	1.3	2.6	2.1	3.2	3.3	3.6	2.4	5.1
	MSN	T_{D}	9.5	6.7	5.9	8.6	7.1	7.2	6.8	5.9	6.4	6.8
		T_{LR}	2.8	3.4	4.1	6.0	4.9	5.5	4.0	4.3	4.8	6.0
		T_{LH}	10.8	8.5	8.2	8.9	7.8	7.6	6.5	5.6	6.9	7.2
		T_{BNP}	1.3	1.3	1.7	3.9	2.7	3.1	2.7	3.6	3.3	5.0
Σ_2	MN	T_{D}	7.3	8.8	6.8	7.5	5.8	6.3	6.0	4.9	6.0	6.8
		T_{LR}	3.0	5.7	3.8	4.7	3.9	5.5	3.8	4.7	4.4	5.8
		T_{LH}	9.6	11.8	7.6	9.1	7.3	7.8	5.9	6.9	6.4	7.6
		T_{BNP}	0.5	1.4	1.6	2.8	2.1	3.4	2.0	2.6	2.9	4.8
	MSN	T_{D}	7.8	8.1	7.4	7.2	6.0	5.9	8.0	5.6	6.6	4.8
		T_{LR}	3.2	4.8	3.2	5.6	3.8	4.4	5.1	5.0	4.5	5.2
		T_{LH}	9.6	10.9	7.8	8.7	6.3	6.1	7.9	6.7	7.5	6.9
		T_{BNP}	0.6	1.7	1.2	2.5	2.0	3.3	2.7	3.3	3.2	3.6
Σ_3	MN	T_{D}	7.1	8.8	6.4	5.4	6.0	7.1	6.7	5.9	5.4	6.2
		T_{LR}	2.2	5.2	3.1	3.4	3.4	4.8	4.3	4.2	3.8	5.5
		T_{LH}	10.0	12.1	8.6	8.1	7.0	8.6	7.0	7.0	5.9	6.8
		T_{BNP}	0.6	1.4	1.3	1.7	1.6	3.2	2.2	2.3	2.6	3.5
	MSN	T_{D}	27.9	7.8	42.4	7.7	59.1	9.3	75.9	7.3	93.9	6.1
		T_{LR}	11.9	5.0	24.4	5.3	43.1	7.4	60.4	5.4	86.3	5.1
		T_{LH}	32.8	11.5	46	9.6	23.9	10.9	77.5	8.5	94.6	7.2
		T_{BNP}	2.5	1.8	8.4	2.8	23.9	4.3	38.5	3.4	69.9	3.5

Here, $b = 3$, $p = 3$ and $n_{ij} = 4$ for $a-1$ cells, $n_{ij} = 5$ for $(b-1)$ cells and $n_{ij} = 6$ for $(a-1)(b-1)+1$ cells. The desired size of the tests is 0.05
Simulation size is 1,000

definite matrix and $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n)$. We are interested in approximating the distribution of $\mathbf{Q}$ by a p dimensional central Wishart distribution with degrees of freedom f and mean $f\Psi$, denoted by $W_p(f, \Psi)$, where $\Psi > 0$. The quantities f and Ψ are to be approximated by matching the means and the total variances of $\mathbf{Q}$ and $W_p(f, \Psi)$. By the total variance is meant the trace of the variance–covariance matrix.

For a random matrix $\mathbf{W}$ distributed as $W_p(f, \Psi)$, the mean and variance are $E(\mathbf{W}) = f\Psi$ and $\mathrm{Var}(\mathbf{W}) = f(I_{p^2} + K_{p,p})(\Psi \otimes \Psi)$. On the other hand, the mean and variance of $\mathbf{Q}$ can be obtained using Lemma 1 noting that, under normality, $K_4(\mathbf{Y}_i) = 0$. Now setting the means and the total variances of $\mathbf{Q}$ and $\mathbf{W}$ equal,

$$f\Psi = \sum_{i=1}^{n} c_{ii}\Sigma_i \quad \text{and}$$

$$\operatorname{tr}\left(f(I_{p^2}+K_{p,p})(\Psi\otimes\Psi)\right) = \operatorname{tr}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}^2 (I_{p^2}+K_{p,p})(\Sigma_i\otimes\Sigma_j)\right).$$

This leads to

$$f = \frac{\operatorname{tr}\left(\sum_{i=1}^n c_{ii}\Sigma_i\right)^2}{\sum_{i=1}^n\sum_{j=1}^n c_{ij}^2 \operatorname{tr}(\Sigma_i\Sigma_j)} \quad \text{and} \quad \Psi = \frac{\sum_{i=1}^n\sum_{j=1}^n c_{ij}^2 \operatorname{tr}(\Sigma_i\Sigma_j)}{\operatorname{tr}\left(\sum_{i=1}^n c_{ii}\Sigma_i\right)^2}\sum_{i=1}^{n} c_{ii}\Sigma_i.$$

It must be noted that one can also consider matching other functions of the eigenvalues of the variances of $\mathbf{W}$ and $\mathbf{Q}$ such as the determinant to get different approximations.

Assuming normality, we propose approximating the distributions of $H^{(A)}$, $H^{(B)}$, $H^{(A|B)}$, $H^{(B|A)}$, $H^{(AB)}$ and G by Wishart distributions with the respective degrees of freedom denoted by f_A, f_B, $f_{A|B}$, $f_{B|A}$, f_{AB} and f_G. The rationale behind this approximation is that it is shown in Sect. 2 that the asymptotic distributions of the test statistics do not depend on the distribution of the data. Therefore, small sample size approximations such as those based on asymptotic expansions may give good approximations for moderate sample sizes under non-normality as well. The degrees of freedoms for the approximating Wishart distributions of the sum of squares and cross products matrices are obtained, after lengthy algebra, as:

$$\begin{aligned}
f_A &= \operatorname{tr}(\Sigma^2)(ab)^2\\
&\quad\times\left(\sum_{i=1}^{a}\sum_{j,j'=1}^{b}\frac{1}{n_{ij}n_{ij'}}\operatorname{tr}(\Sigma_{ij}\Sigma_{ij'})+\frac{1}{(a-1)^2}\sum_{i\neq i'=1}^{a}\sum_{j,j'=1}^{b}\frac{1}{n_{ij}n_{i'j'}}\operatorname{tr}(\Sigma_{ij}\Sigma_{i'j'})\right)^{-1},\\
f_{A|B} &= \operatorname{tr}(\Sigma^2)(ab)^2\\
&\quad\times\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}^2}\operatorname{tr}(\Sigma_{ij}^2)+\frac{1}{(a-1)^2}\sum_{i\neq i'=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}n_{i'j}}\operatorname{tr}(\Sigma_{ij}\Sigma_{i'j})\right)^{-1},\\
f_B &= \operatorname{tr}(\Sigma^2)(ab)^2\\
&\quad\times\left(\sum_{i,i'=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}n_{i'j}}\operatorname{tr}(\Sigma_{ij}\Sigma_{i'j})+\frac{1}{(b-1)^2}\sum_{i,i'=1}^{a}\sum_{j\neq j'=1}^{b}\frac{1}{n_{ij}n_{i'j'}}\operatorname{tr}(\Sigma_{ij}\Sigma_{i'j'})\right)^{-1},\\
f_{B|A} &= \operatorname{tr}(\Sigma^2)(ab)^2\\
&\quad\times\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}^2}\operatorname{tr}(\Sigma_{ij}^2)+\frac{1}{(b-1)^2}\sum_{i=1}^{a}\sum_{j\neq j'=1}^{b}\frac{1}{n_{ij}n_{ij'}}\operatorname{tr}(\Sigma_{ij}\Sigma_{ij'})\right)^{-1},
\end{aligned}$$

$$
\begin{aligned}
f_{AB} &= \operatorname{tr}(\Sigma^2)(ab)^2 \\
&\times \left(\sum_{i=1}^{a}\sum_{j=1}^{b} \frac{1}{n_{ij}^2}\operatorname{tr}(\Sigma_{ij}^2) + \frac{1}{(a-1)^2} \sum_{i\neq i'=1}^{a}\sum_{j=1}^{b} \frac{1}{n_{ij}n_{i'j}}\operatorname{tr}(\Sigma_{ij}\Sigma_{i'j}) + \frac{1}{(b-1)^2} \right.\\
&\times \sum_{i=1}^{a}\sum_{j\neq j'=1}^{b} \frac{1}{n_{ij}n_{ij'}}\operatorname{tr}(\Sigma_{ij}\Sigma_{ij'}) + \frac{1}{(a-1)^2(b-1)^2} \\
&\left. \times \sum_{i\neq i'=1}^{a}\sum_{j\neq j'=1}^{b} \frac{1}{n_{ij}n_{i'j'}}\operatorname{tr}(\Sigma_{ij}\Sigma_{i'j'}) \right)^{-1}
\end{aligned}
$$

and

$$
f_G = \operatorname{tr}(\Sigma^2)(ab)^2 \left(\sum_{i=1}^{a}\sum_{j=1}^{b} \frac{1}{n_{ij}^2(n_{ij}-1)}\operatorname{tr}(\Sigma_{ij}^2) \right)^{-1}
$$

where $\Sigma = (ab)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}^{-1}\Sigma_{ij}$. In practice, Σ_{ij} and Σ can be replace by their unbiased estimators. It may be noted that f_ψ diverges as $a \to \infty$ for each $\psi \in \{A, A|B, B, B|A, AB, G\}$.

In the remainder of this section, we show the details of the proposed approximations when applied to the test statistics for testing the main effects of factor A. The details for the others can be obtained along the same lines. Also to facilitate ease of presentation, we will use the following notations. Let $\mathbf{B}$ and $\mathbf{W}$ be independent random matrices distributed as $\mathbf{B} \sim W_p(n_h, \Sigma)$ and $\mathbf{W} \sim W_p(n_e, \Sigma)$. Based on $\mathbf{B}$ and $\mathbf{W}$, define $T_1 = -\log|\mathbf{W}(\mathbf{B}+\mathbf{W})^{-1}|$, $T_2 = \operatorname{tr}(\mathbf{B}\mathbf{W}^{-1})$ and $T_3 = \operatorname{tr}\{\mathbf{B}(\mathbf{B}+\mathbf{W})^{-1}\}$.

Since $H^{(A)}$ depends on the data only through the cell means $\bar{\mathbf{Y}}_{ij}$ and G depends on the data only through S_{ij}, we see that $f_A H^{(A)}$ and $f_G G$ are independent. Also according to the approximation discussed above, $f_A H^{(A)}$ and $f_G G$ are approximately distributed as $W_p(f_A, \Sigma)$ and $W_p(f_G, \Sigma)$, respectively.

3.1 χ^2-Based asymptotic expansion

It is well known in the multivariate literature (see, for example, Anderson 2003) that the upper α quantile of the distributions of mT_i can be expanded as

$$
\chi^2_{\nu,\alpha} + \frac{1}{n_e}\left(\left\{ \frac{(p-n_h+1)}{2} - d \right\} \chi^2_{\nu,\alpha} + \frac{a(p+n_h+1)}{2(\nu+2)}\chi^4_{\nu,\alpha} \right) + O\left(\frac{1}{n_e^2}\right)
$$

where $m = n_e(1+d/n_e), 0 < \alpha < 1, \nu = p \cdot n_h$ and $\chi^2_{\nu,\alpha}$ is the upper α quantile of the chi-square distribution with ν degrees of freedom. The values taken by d are $-(p-n_h+1)/2, 0$ and $n_h - 1$ for $i = 1, 2, 3$, respectively, and those taken by a are $0, 1, -1$ for $i = 1, 2, 3$, respectively. Notice that this expansions slightly differs from those given in Anderson (2003) because we are giving the modified versions after Bartlett's corrections (accounting for the multiplying factor $1 + d/n_e$). These formulae will reduce to those given in Anderson (2003) when $d = 0$. The sign of the second term in equation (22) on page 332 of Anderson (2003) should be corrected to minus.

This approximation can be applied to the test statistics for the main effects of factor A by redefining $H^{(A)}$ as $f_A H^{(A)}$ and G as $f_G G$ and setting $n_h = f_A$ and $n_e = f_G$. In that case, T_1, T_2 and T_3 coincide with $T_{\mathrm{LR}}^{(A)}$, $T_{\mathrm{LH}}^{(A)}$ and $T_{\mathrm{BNP}}^{(A)}$, respectively.

3.2 Normal-based asymptotic expansion

In a similar asymptotic framework as in this paper, Fujikoshi (1975) obtained asymptotic expansions for the distributions of centered and scaled versions of T_1, T_2 and T_3. More precisely, Fujikoshi's asymptotic framework is that $n_h = nh$, $n_e = ne$, $h > 0$, $e > 0$ and $h + e = 1$. The asymptotic expansion is in the order of n meaning that both n_h and n_e tend to infinity at the same rate. Accordingly, the Cornish-Fisher expansion for the upper α quantile of $\sqrt{(m/\tau^2)}(T_i - l)$ is

$$z_\alpha + \frac{1}{\sqrt{m}}\{a_1 h_1(z_\alpha) + a_3 h_3(z_\alpha)\} - \frac{1}{m}\left\{b_2 h_2(z_\alpha) + b_4 h_4(z_\alpha) + b_6 h_6(z_\alpha)\right.$$
$$\left. + z_\alpha(a_1 + a_3 h_3(z_\alpha))\left(\frac{1}{2}a_1 + a_3\left[\frac{1}{2}h_3(z_\alpha) - 2\right]\right)\right\} + O\left(m^{-3/2}\right) \tag{8}$$

where z_α denotes the upper α-quantile of a standard normal variate and the functions $h_1, \ldots, h_6$ are the first six Hermite polynomials defined as $h_1(x) = 1$, $h_2(x) = -x$, $h_3(x) = x^2 - 1$, $h_4(x) = -x^3 + 3x$, $h_5(x) = x^4 - 6x^2 + 3$ and $h_6(x) = -x^5 + 10x^3 - 15x$. The values taken by the coefficients $m, \tau, l, a_1, a_3, b_2, b_4$ and b_6 depend on i. For $i = 1$, $m = \{(1+e)n - (p+1)\}/2$, $\tau^2 = 2ph(\mu e)^{-1}$, $l = -p\log e$, $a_1 = \tau^{-1}p(p+1)h(2\mu e)^{-1}$, $a_3 = 2\tau^{-3}ph(1+e)(\mu e)^{-2}/3$, $b_2 = (1/2)\tau^{-2}p(p+1)h(\mu e)^{-1}\{[p(p+1)+4(1+e)](\mu e)^{-1}/4 - 1\}$, $b_4 = \tau^{-4}ph\{p(p+1)(1+e)h + 2(1+e+e^2)\}(\mu e)^{-3}/3$ and $b_6 = (1/2)a_3^2$ where $\mu = 2(1+e)^{-1}$. For $i = 2$, $m = ne$, $\tau^2 = 2phe^{-2}$, $l = phe^{-1}$, $a_1 = \tau^{-1}p(p+1)he^{-1}$, $a_3 = 4\tau^{-3}ph(2-e)(3e^3)^{-1}$, $b_2 = \tau^{-2}p(p+1)[(1/2)(p^2+p+8)h^2 + 3he]e^{-2}$, $b_4 = 2\tau^{-4}ph\{(2/3)p(p+1)h(2-e) + e^2 - 5e + 5\}e^{-4}$ and $b_6 = (1/2)a_3^2$. For $i = 3$, $m = n$, $\tau^2 = 2phe$, $l = ph$, $a_1 = 0$, $a_3 = (4/3)\tau^{-3}phe(e-h)$, $b_2 = -\tau^{-2}phe(p+1)$, $b_4 = 2\tau^{-4}phe(e^2 + h^2 - 3he)$ and $b_6 = (1/2)a_3^2$.

To use Fujikoshi's approximation to our case, we first redefine $H^{(A)}$ as $f_A H^{(A)}$ and G as $f_G G$. For h, e and n, we take $h = \frac{f_A}{f_A + f_G}$, $e = \frac{f_G}{f_A + f_G}$, and $n = f_A + f_G$. Finally, note that T_1, T_2 and T_3 coincide with $T_{\mathrm{LR}}^{(A)}$, $T_{\mathrm{LH}}^{(A)}$ and $T_{\mathrm{BNP}}^{(A)}$, respectively.

3.3 Numerical accuracy of the small sample approximations

We investigated the accuracy of the small sample approximations proposed for the null distributions of the test statistics T_{LR}, T_{LH} and T_{BNP} for testing the main effects of A and interaction effects AB. Table 2 displays simulation results for the approximations based on the χ^2-based asymptotic expansion (AEC) and normal-based asymptotic

Table 2 Simulated actual sizes (expressed as percentages) for T_{LR}, T_{LH} and T_{BNP} tests based on Normal and Chi-Square asymptotic expansion approximations

TS	Σ_1						Σ_2					
	MV Normal			MV Skew Normal			MV Normal			MV Skew Normal		
	T_{LR}	T_{LH}	T_{BNP}	T_{LR}	T_{LH}	T_{BNP}	T_{LR}	T_{LH}	T_{BNP}	T_{LR}	T_{LH}	T_{BNP}
$a = 6$												
A												
AEC	5.4	5.6	6.1	5.5	5.8	5.9	4.5	4.3	5.1	5.6	6.5	6.1
AEN	4.9	5.8	5.5	4.9	5.9	5.2	3.8	4.7	4.6	5.0	6.5	5.5
AB												
AEC	5.1	5.1	6.8	4.8	4.9	7.0	4.6	4.8	6.6	5.1	5.3	7.3
AEN	4.6	5.1	4.8	4.3	4.9	4.8	3.6	4.8	4.2	4.1	5.3	4.4
$a = 10$												
A												
AEC	5.4	5.2	6.0	5.5	5.4	6.1	4.3	4.1	4.6	4.6	4.6	5.1
AEN	5.3	5.2	5.3	4.9	5.4	5.4	3.9	4.2	4.2	4.4	4.6	4.6
AB												
AEC	4.4	4.2	6.3	4.8	4.1	6.5	5.9	6.0	9.8	4.8	4.9	7.8
AEN	3.6	4.2	4.3	3.6	4.2	3.9	5.6	6.0	5.1	4.4	4.9	4.6
$a = 20$												
A												
AEC	4.1	4.4	4.8	4.8	4.7	5.2	4.4	4.5	5.1	4.2	4.0	4.7
AEN	3.8	4.4	3.9	4.6	4.7	4.8	4.2	4.5	4.3	4.0	4.0	4.1
AB												
AEC	5.8	5.2	9.9	6.8	6.2	10.1	6.3	5.5	10.1	6.8	6.5	9.9
AEN	5.1	5.2	5.1	6.1	6.3	6.0	4.9	5.5	5.3	5.6	6.5	5.6
$a = 35$												
A												
AEC	4.8	4.7	5.4	6.1	6.2	6.8	4.9	4.6	5.6	4.6	4.5	5.2
AEN	4.5	4.7	4.7	6.0	6.2	6.0	4.4	4.6	4.7	4.3	4.5	4.2
AB												
AEC	5.8	4.8	11.1	6.1	5.1	11.7	6.6	5.7	11.1	5.2	4.5	10.4
AEN	4.8	4.8	4.8	5.1	5.1	5.2	5.8	5.7	5.9	4.4	4.5	4.2
$a = 50$												
A												
AEC	5.8	5.9	6.9	4.8	4.6	5.8	4.9	4.5	5.6	6.4	6.1	7.0
AEN	5.7	5.9	5.6	4.4	4.6	4.3	4.6	4.6	4.6	6.1	6.1	6.0

Table 2 continued

TS	Σ_1						Σ_2					
	MV Normal			MV Skew Normal			MV Normal			MV Skew Normal		
	T_{LR}	T_{LH}	T_{BNP}	T_{LR}	T_{LH}	T_{BNP}	T_{LR}	T_{LH}	T_{BNP}	T_{LR}	T_{LH}	T_{BNP}
AB												
AEC	6.9	5.1	13.6	5.4	4.5	10.3	6.6	4.5	12.0	6.0	4.6	11.2
AEN	5.0	5.2	5.2	4.5	4.5	4.5	4.5	4.6	4.7	4.8	4.6	4.7

Data are generated from multivariate normal (MN Normal) and multivariate Skew–Normal (MV Skew Normal) distributions under the heteroscedasticity structures Σ_1 and Σ_2. Here, $b = 3$, $p = 3$ and $n_{ij} = 4$ for $a - 1$ cells, $n_{ij} = 5$ for $(b - 1)$ cells and 6 for $b - 1$ cells and $n_{ij} = 6$ for $(a - 1)(b - 1) + 1$. The desired size of the tests is set at 0.05

Simulation size is 1,000

expansion (AEN). For the sake of brevity, we consider only the heteroscedastic structures Σ_1 and Σ_2.

The results shown in Table 2 demonstrate that both the chi-square and normal based small sample approximations maintained the desired size ($\alpha = 0.05$) fairly accurately for T_{LR} and T_{LH}. On the other hand, we notice that the quality of the chi-square based approximation for T_{BNP} when testing the interaction effect deteriorated as a increased. So the chi-square approximation for this test can not be recommended.

4 Real data example: randomization for a smoking cessation trial

The Greek Health Project (*NIH R01 CA107191*) was intended to assess the efficacy of a motivational interviewing (a form of counseling) versus an attention matched control on smoking quit rate. The subjects for the research are students in the Greek houses (fraternities and sororities) of the University of Missouri-Colombia. To avoid a contamination effect, the researchers decided to use a cluster-randomized design (individual fraternity or sorority chapters taken as clusters). That is, a whole chapter is assigned to either the treatment or control arm. Prior to the assignment of the chapters to the treatment and control arm, it was necessary to know whether nicotine dependence of the subjects depended on the chapter they came from, in order to avoid unintentional selection bias. Another important variable believed to be highly associated with nicotine dependence, in particular in the context of college students, is depression. The level of depression for each subject was determined as low or high based on the Center for Epidemiologic Studies Depression Scale (Kohout et al. 1993). Three well known scales of nicotine dependence are the Fagerström Test for Nicotine Dependence (Heatherton et al. 1991), Hooked on Nicotine Checklist (Wellman et al. 2005) and Minnesota Tobacco Withdrawal Scale (Hatsukami et al. 1984). In the analysis that follows, we shorthand these variables as FTSD, HOOKNCTN and WDRSYPTM, respectively. For this data set we have two factors (Chapter with twenty levels and Depression with two levels) and three dependent variables (FTSD, HOOKNCTN and WDRSYPTM).

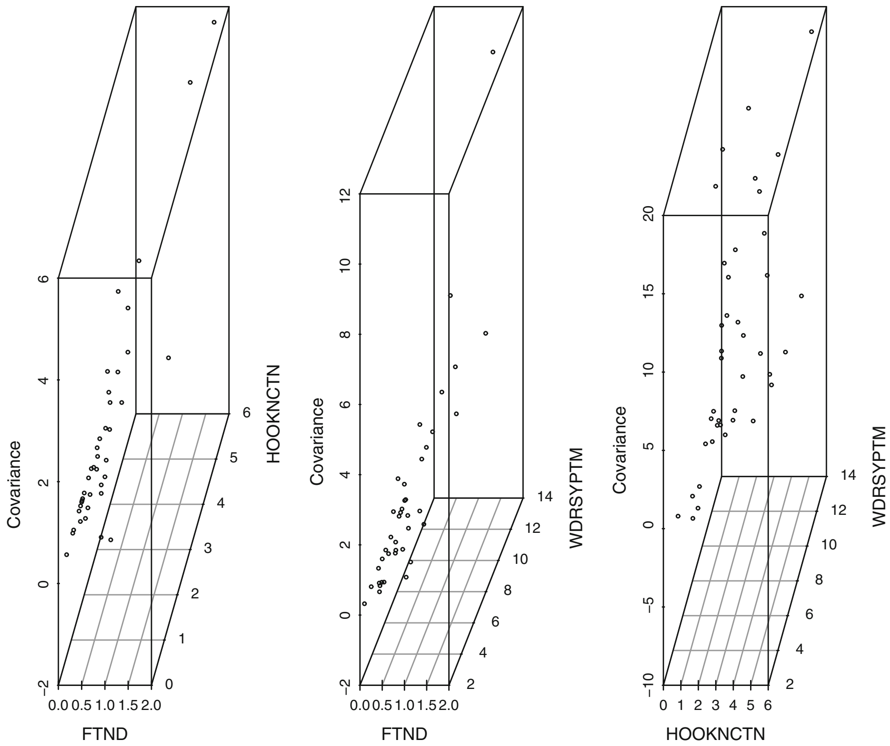

Fig. 1 Three-dimensional scatter plot of the cell covariance between two variables against cell means separately for each pair of the dependent variables

In Fig. 1, the sample covariances (for each pair of the dependent variables separately) against the sample means of the variables are plotted to assess if the assumption of constant covariance matrix is reasonable. It is clear from the three plots that cell covariances change with the cell means indicating the violation of the constant covariance assumption. A univariate Levene test for homogeneity of variance (Levene 1960) for FTSD, HOOKNCTN and WDRSYPTM resulted in p-values $< 0.001, 0.011$ and <0.001, respectively, all of which leading to the rejection of the homogeneity of variance assumption. Given the linear dependence between the means and covariances discerned in the plots of Fig. 1, one might think that the heteroscedasticity may be resolved after a log transformation of each of the dependent variables. That does not appear to be the case in view of Levene's test p values $< 0.001, 0.242$ and 0.009 for the log of FTSD, HOOKNCTN and WDRSYPTM, respectively. Also, a plot of the covariances against the means for the log-transformed data (not shown here in the interest of space) indicates that the transformation does not seem to remove all of the heteroscedasticity.

Having strong evidence against homogeneity of variance–covariance matrices, we analyzed the data using the standard homoscedastic MANOVA procedures as well as the heteroscedastic MANOVA procedure of this paper to investigate the effect of heteroscedasticity on these methods. We conducted a homoscedastic MANOVA on

Table 3 p values $\times 100$ for testing the effect of Chapter and Chapter $\times$ Depression interaction

	Chapter				Chapter $\times$ Depression			
Approximation	T_D	T_{LR}	T_{LH}	T_{BNP}	T_D	T_{LR}	T_{LH}	T_{BNP}
Heteroscedastic MANOVA								
χ^2–Asym. Exp.		36.3	35.3	41.1		46.5	46.9	51.3
Normal-Asym. Exp.		37.1	35.2	38.3		47	35.2	38.3
Large-a Asym.	41.3	47.7	28.5	66	49.6	55.1	41.3	67.6
Data	T_D	T_{LR}	T_{LH}	T_{BNP}	T_D	T_{LR}	T_{LH}	T_{BNP}
Homoscedastic MANOVA								
Original		1.7	1.5	1.9		8.0	7.5	8.5
log-Transformed Data		5	4.6	5.4		15.2	15.2	15.1

both the original as well as the log-transformed observations. The results for the main effects of Chapter and the interaction between Chapter and Depression are displayed in Table 3. In the heteroscedastic MANOVA analysis, we computed p-values based on the two small sample approximations of Sect. 3 for the likelihood ratio, Lawley–Hotelling and Barlett–Nanda–Pillai Criteria and the large a asymptotic null distributions of Sect. 2 for the three statistics as well as Dempster's ANOVA-type statistic.

It is clear from Table 3 that if we analyzed the data assuming a constant covariance matrix and utilized a standard homoscedastic MANOVA, we would conclude that nicotine dependence varied by chapters. Also the evidence for Chapter $\times$ Depression interactions would not be negligible at the level of significance 0.1, though insignificant at the 0.05 level. On the other hand, the heteroscedastic MANOVA results agree with each other and do not show evidence of nicotine dependence variation by Chapter nor presence of Chapter $\times$ Depression interaction. The log transformation which has somewhat reduced, but not fully removed, the heteroscedasticity led to borderline insignificant results for the Chapter effects. This gives a clear indication that in this data set, heteroscedasticity was disguised as Chapter effects, which do not actually seem to be present. Since heteroscedasticity was indicated by the original as well as the transformed data, the heteroscedastic MANOVA results appear more trustworthy, and we conclude that a selection bias due to randomizing chapters is not supported by the data.

5 Discussion and conclusion

In this paper, we considered a heteroscedastic MANOVA model. The usual MANOVA sum of squares and cross products matrices were modified to adapt to heteroscedasticity. Although these sum of squares and cross products matrices were motivated from suitability for testing the corresponding hypotheses, as it happens they can neatly be expressed as matrix quadratic forms. Defining the four standard multivariate test statistics based on the new sum of squares and cross products matrices, their asymptotic distributions were derived in a unified manner both under the null as well as reasonable local alternatives.

Adjustment factors to the sum of squares and cross products matrices were derived so that their first two moments match those of respective Wishart distributions. With these factors, the asymptotic expansion approximations for the Likelihood ratio, Lawley–Hotelling and Bartlett–Nanda–Pillai statistics under normality and homoscedasticity were shown, in a numerical study, to perform quite well even when both non-normality and heteroscedasticity were present.

The asymptotic framework considered here is such that the number of levels of one of the factors tends to infinity. A real data example has been provided to illustrate the practicality of the asymptotic framework and the application of the methods.

The quadratic form expressions for the sum of squares and cross products matrices vividly suggest a formal extension to the multi-factor case. Although similar asymptotic results appear to come through, the technical details seem cumbersome unless handy notations and techniques are introduced. We plan to explore this in a future work.

The results of this paper may also serve as a theoretical tool for obtaining the corresponding results in a fully nonparametric setting. The test statistics computed on componentwise ranks are typically asymptotically equivalent to the same statistics based on so-called Asymptotic Rank Transforms (e.g. see Harrar and Bathke 2008). In the two-way layout case, the asymptotic rank transforms would still be heteroscedastic even under the null hypotheses. However, the asymptotic equivalence alluded to above needs to be established and consistent estimators of the asymptotic variances need to be found. In the interest of space, we chose to delegate this problem to a separate treatise.

Appendix A: Moments of sum of squares and cross products matrices

The following lemma is handy in calculating the first two moments of matrix quadratic forms. Its proof is given in Bathke and Harrar (2008).

Lemma 1 *Suppose* $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ *is a* $p \times n$ *random matrix whose columns* $\mathbf{Y}_i, i = 1, \ldots, n$, *are independently distributed with mean* $\mathbf{0}$ *and covariance* Σ_i. *Let* $A = (a_{ij})$ *and* $B = (b_{ij})$ *be* $n \times n$ *symmetric matrices. Then,* $E(\mathbf{Y}A\mathbf{Y}') = \sum_{i=1}^n a_{ii}\Sigma_i$ *and* $\mathrm{Cov}\left(\mathrm{vec}(\mathbf{Y}A\mathbf{Y}'), \mathrm{vec}(\mathbf{Y}B\mathbf{Y}')\right) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}b_{ij}(I_{p^2} + K_{p,p})(\Sigma_i \otimes \Sigma_j) + \sum_{i=1}^n a_{ii}b_{ii}K_4(\mathbf{Y}_i)$, *where* $K_4(\mathbf{Y}_i) = \mathrm{E}\left(\mathrm{vec}(\mathbf{Y}_i\mathbf{Y}_i')\mathrm{vec}(\mathbf{Y}_i\mathbf{Y}_i')'\right) - (I_{p^2} + K_{p,p})(\Sigma_i \otimes \Sigma_i) - \mathrm{vec}(\Sigma_i)\mathrm{vec}(\Sigma_i)'$.

Let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_a), \mathbf{Y}_i = (\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{ib})$ and $\mathbf{Y}_{ij} = (\mathbf{Y}_{ij1}, \ldots, \mathbf{Y}_{ijn_{ij}})$. The sum of squares and cross products matrices $H^{(A)}, H^{(A|B)}, H^{(B|A)}, H^{(AB)}$ and G can be written as $\mathbf{Y}C_A\mathbf{Y}', \mathbf{Y}C_{A|B}\mathbf{Y}'$, $H^{(B|A)} = \mathbf{Y}C_{B|A}\mathbf{Y}', \mathbf{Y}C_{AB}\mathbf{Y}'$ and $\mathbf{Y}C_G\mathbf{Y}'$, respectively, where

$$C_A = \frac{1}{a-1}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)\left(P_a \otimes \frac{1}{b}J_b\right)\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right),$$

$$C_{A|B} = \frac{1}{(a-1)b}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)(P_a \otimes I_b)\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right),$$

$$C_{B|A} = \frac{1}{a(b-1)} \left(\bigoplus_{i=1}^{a} \bigoplus_{j=1}^{b} \frac{1}{n_{ij}} \mathbf{1}_{n_{ij}} \right) (I_a \otimes P_b) \left(\bigoplus_{i=1}^{a} \bigoplus_{j=1}^{b} \frac{1}{n_{ij}} \mathbf{1}'_{n_{ij}} \right),$$

$$C_{AB} = \frac{1}{(a-1)(b-1)} \left(\bigoplus_{i=1}^{a} \bigoplus_{j=1}^{b} \frac{1}{n_{ij}} \mathbf{1}_{n_{ij}} \right) (P_a \otimes P_b) \left(\bigoplus_{i=1}^{a} \bigoplus_{j=1}^{b} \frac{1}{n_{ij}} \mathbf{1}'_{n_{ij}} \right)$$

and

$$C_G = \frac{1}{ab} \bigoplus_{i=1}^{a} \bigoplus_{j=1}^{b} \frac{1}{n_{ij}(n_{ij}-1)} P_{n_{ij}}. \tag{9}$$

It is straightforward that $E(G) = \Sigma$ both under the null and alternative hypotheses. The expected value of the random matrix $H^{(A)}$ can be expressed as $E(\mathbf{Y}C_A\mathbf{Y}') = E(\mathbf{Y}C_A^{(1)}\mathbf{Y}) - E(\mathbf{Y}C_A^{(2)}\mathbf{Y})$, where

$$C_A^{(1)} = \frac{1}{b(a-1)} \left(\bigoplus_{i=1}^{a} \bigoplus_{j=1}^{b} \frac{1}{n_{ij}} \mathbf{1}_{n_{ij}} \right) (I_a \otimes J_b) \left(\bigoplus_{i=1}^{a} \bigoplus_{j=1}^{b} \frac{1}{n_{ij}} \mathbf{1}'_{n_{ij}} \right)$$

and

$$C_A^{(2)} = \frac{1}{ab(a-1)} \left(\bigoplus_{i=1}^{a} \bigoplus_{j=1}^{b} \frac{1}{n_{ij}} \mathbf{1}_{n_{ij}} \right) J_{ab} \left(\bigoplus_{i=1}^{a} \bigoplus_{j=1}^{b} \frac{1}{n_{ij}} \mathbf{1}'_{n_{ij}} \right).$$

Each of the expected values on the right hand side, under the null hypothesis $H^{(A)}$, can be computed with the aid of Lemma 1 to get

$$\begin{aligned} E(\mathbf{Y}C_A\mathbf{Y}') &= \frac{1}{b(a-1)} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{1}{n_{ij}} \Sigma_{ij} - \frac{1}{ab(a-1)} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{1}{n_{ij}} \Sigma_{ij} \\ &= \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{1}{n_{ij}} \Sigma_{ij} = \Sigma. \end{aligned}$$

The calculations for $E(\mathbf{Y}C_{A|B}\mathbf{Y}')$, $E(\mathbf{Y}C_{B|A}\mathbf{Y}')$ and $E(\mathbf{Y}C_{AB}\mathbf{Y}')$, under their respective null hypotheses, are similar.

Appendix B: Asymptotic distributions

Proof of Theorem 1 In view of Lemma 1,

$$
\begin{aligned}
&\operatorname{var}\left(\frac{1}{a}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}}\operatorname{vec}(S_{ij})\right)\\
&\quad=\frac{1}{a^2}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}^2}\frac{1}{(n_{ij}-1)^2}\operatorname{var}\left(\operatorname{vec}\left(\mathbf{Y}_{ij}\left(I_{n_{ij}}-\frac{1}{n_{ij}}J_{n_{ij}}\right)\mathbf{Y}_{ij}'\right)\right)\\
&\quad=(I_{p^2}+K_{p,p})\left(\frac{1}{a^2}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}^2(n_{ij}-1)}(\Sigma_{ij}\otimes\Sigma_{ij})\right)\\
&\qquad+\frac{1}{a^2}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}^3}K_4(\mathbf{Y}_{ij1})
\end{aligned}
$$

which becomes asymptotically negligible provided $\sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}^{-2}(n_{ij}-1)^{-1}(\Sigma_{ij}\otimes\Sigma_{ij})=o(a^2)$ and $\sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}^{-3}K_4(\mathbf{Y}_{ij1})=o(a^2)$ as $a\to\infty$. $\square$

Proof of Theorem 2 It is evident from (1) that $H^{(\psi)}$ and G are invariant to the location translation $\mathbf{Y}_{ijk}-\boldsymbol{\mu}_{ij}^{(\psi)}$. Hence we can, without loss of generality, take $\boldsymbol{\mu}_{ij}^{(\psi)}=\mathbf{0}$ when considering the test for $\mathcal{H}_0^{(\psi)}$.

Let us write $\sqrt{a}(H^{(\psi)}-G)=\sqrt{a}\mathbf{Y}(C_\psi-C_G)\mathbf{Y}'$. One can then, after lengthy algebra, express

$$
\begin{aligned}
C_A-C_G&=\frac{1}{ab}\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}}-I_{n_{ij}})\\
&\quad+\frac{1}{ab}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)(I_a\otimes(J_b-I_b))\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}'\right)\\
&\quad-\frac{1}{a(a-1)b}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)((J_a-I_a)\otimes J_b)\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}'\right),\\
C_{A|B}-C_G&=\frac{1}{ab}\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}}-I_{n_{ij}})\\
&\quad+\frac{1}{a(a-1)b}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)((J_a-I_a)\otimes I_b)\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}'\right),
\end{aligned}
$$

$$C_{B|A} - C_G = \frac{1}{a}\bigoplus_{i=1}^{a}\left[\frac{1}{(b-1)}\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)P_b\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right.$$
$$\left.-\frac{1}{b}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}P_{n_{ij}}\right],$$

and

$$C_{AB} - C_G = \frac{1}{ab}\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}} - I_{n_{ij}})$$
$$-\frac{1}{ab(b-1)}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)(I_a\otimes(J_b - I_b))\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)$$
$$-\frac{1}{a(a-1)(b-1)}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)((J_a - I_a)\otimes P_b)\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right).$$

Then in view of Lemma 2,

$$\sqrt{a}\cdot\mathrm{tr}\left\{\mathbf{Y}(C_A - C_G)\mathbf{Y}'\Omega\right\}$$
$$\doteq \sqrt{a}\cdot\mathrm{tr}\left\{\mathbf{Y}\left[\frac{1}{ab}\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}} - I_{n_{ij}})\right.\right.$$
$$\left.\left.+\frac{1}{ab}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)(I_a\otimes(J_b - I_b))\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right]\mathbf{Y}'\Omega\right\}$$
$$= \frac{1}{\sqrt{a}}\sum_{i=1}^{a}\frac{1}{b}\mathrm{tr}\left\{\mathbf{Y}_i\left[\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}} - I_{n_{ij}})\right.\right.$$
$$\left.\left.+\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)(J_b - I_b)\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right]\mathbf{Y}'_i\Omega\right\} = \frac{1}{\sqrt{a}}\sum_{i=1}^{a}Z_i^{(A)},$$

$$\sqrt{a}\cdot\mathrm{tr}\left\{\mathbf{Y}(C_{A|B} - C_G)\mathbf{Y}'\Omega\right\}$$
$$\doteq \sqrt{a}\cdot\mathrm{tr}\left\{\mathbf{Y}\left[\frac{1}{ab}\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}} - I_{n_{ij}})\right]\mathbf{Y}'\Omega\right\}$$
$$= \frac{1}{\sqrt{a}}\sum_{i=1}^{a}\frac{1}{b}\mathrm{tr}\left\{\mathbf{Y}_i\left[\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}} - I_{n_{ij}})\right]\mathbf{Y}'_i\Omega\right\} = \frac{1}{\sqrt{a}}\sum_{i=1}^{a}Z_i^{(A|B)},$$

$$
\begin{aligned}
&\sqrt{a}\cdot \mathrm{tr}\left\{\mathbf{Y}(C_{B|A}-C_G)\mathbf{Y}'\Omega\right\}\\
&=\frac{1}{\sqrt{a}}\sum_{i=1}^{a}\mathrm{tr}\left\{\mathbf{Y}_i\left[\frac{1}{(b-1)}\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)P_b\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right.\right.\\
&\quad\left.\left.-\frac{1}{b}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}P_{n_{ij}}\right]\mathbf{Y}_i\Omega\right\}=\frac{1}{\sqrt{a}}\sum_{i=1}^{a}Z_i^{(B|A)},
\end{aligned}
$$

and

$$
\begin{aligned}
&\sqrt{a}\cdot \mathrm{tr}\left\{\mathbf{Y}(C_{AB}-C_G)\mathbf{Y}'\Omega\right\}\\
&\doteqdot\sqrt{a}\cdot\mathrm{tr}\left\{\mathbf{Y}\left[\frac{1}{ab}\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}}-I_{n_{ij}})-\frac{1}{ab(b-1)}\right.\right.\\
&\quad\left.\left.\times\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)(I_a\otimes(J_b-I_b))\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right]\mathbf{Y}'\Omega\right\}\\
&=\frac{1}{\sqrt{a}}\sum_{i=1}^{a}\frac{1}{b}\mathrm{tr}\left\{\mathbf{Y}_i\left[\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}}-I_{n_{ij}})\right.\right.\\
&\quad\left.\left.-\frac{1}{(b-1)}\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)(J_b-I_b)\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right]\mathbf{Y}'_i\Omega\right\}\\
&=\frac{1}{\sqrt{a}}\sum_{i=1}^{a}Z_i^{(AB)}
\end{aligned}
$$

where $V_a \doteqdot W_a$ means $V_a - W_a \xrightarrow{p} 0$ as $a\to\infty$. It can be verified that $E(Z_i^{(A)})=0$ and,

$$
\begin{aligned}
\mathrm{Var}(Z_i^{(A)})&=\frac{1}{b^2}\mathrm{var}\left(\mathrm{vec}(\Omega)'\mathrm{vec}\left(\mathbf{Y}_i\left[\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}}-I_{n_{ij}})\right.\right.\right.\\
&\quad\left.\left.\left.-\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)(J_b-I_b)\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right]\mathbf{Y}'_i\right)\right)\\
&=\frac{1}{b^2}\mathrm{vec}(\Omega)'\mathrm{var}\left(\mathrm{vec}\left(\mathbf{Y}_i\left[\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}}-I_{n_{ij}})\right.\right.\right.\\
&\quad\left.\left.\left.-\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)(J_b-I_b)\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right]\mathbf{Y}'_i\right)\right)\mathrm{vec}(\Omega).
\end{aligned}
\tag{10}
$$

Applying Lemma 1

$$\mathrm{var}\left(\mathrm{vec}\left(\mathbf{Y}_i\left[\bigoplus_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}(J_{n_{ij}}-I_{n_{ij}})-\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)(J_b-I_b)\left(\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right]\mathbf{Y}'_i\right)\right)$$
$$=(I_{p^2}+K_{p,p})\left(\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\Sigma_{ij}\otimes\Sigma_{ij}+\sum_{j\neq j'}^{b}\frac{1}{n_{ij}n_{ij'}}\Sigma_{ij}\otimes\Sigma_{ij'}\right)$$

Putting this in (10),

$$\mathrm{Var}(Z_i^{(A)})=\frac{2}{b^2}\left(\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\mathrm{tr}(\Omega\Sigma_{ij})^2+\sum_{j\neq j'}^{b}\frac{1}{n_{ij}n_{ij'}}\mathrm{tr}(\Omega\Sigma_{ij}\Omega\Sigma_{ij'})\right). \tag{11}$$

Furthermore,

$$\lim_{a\to\infty}\frac{1}{a}\sum_{i=1}^{a}\mathrm{var}(Z_i^{(A)})=\frac{2}{b}\left[\lim_{a\to\infty}\left(\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\mathrm{tr}(\Omega\Sigma_{ij})^2\right)\right.$$
$$\left.+\lim_{a\to\infty}\left(\frac{1}{ab}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\frac{1}{n_{ij}n_{ij'}}\mathrm{tr}(\Omega\Sigma_{ij}\Omega\Sigma_{ij'})\right)\right].$$

Notice that $Z_i^{(A)}$'s are independently but not identically distributed. In the following we will show that Liaponauv's conditions hold. Applying Lemma 3 by setting $s=1+\delta/2$,

$$\lim_{a\to\infty}\frac{\sum_{i=1}^{a}E|Z_i^{(A)}-E(Z_i^{(A)})|^{2+\delta}}{\left(\sqrt{\sum_{i=1}^{a}\mathrm{var}(Z_i^{(A)})}\right)^{2+\delta}}$$
$$\leq c_\delta E(\mathbf{Y}'_{ij1}\Sigma_{ij}^{-1}\mathbf{Y}_{ij1})^{2+\delta}\frac{\lim_{a\to\infty}\left(\frac{1}{a^{1+\delta/2}}\sum_{i=1}^{a}(\mathrm{var}(Z_i^{(A)}))^{1+\delta/2}\right)}{\lim_{a\to\infty}\left(\frac{1}{a}\sum_{i=1}^{a}\mathrm{var}(Z_i^{(A)})\right)^{1+\delta/2}}=0$$

provided $\lim_{a\to\infty}(1/a)\sum_{i=1}^{a}(\mathrm{var}(Z_i^{(A)}))^{1+\delta/2}<\infty$ where c_δ does not depend on a. But from (11),

$$(\mathrm{var}(Z_i^{(A)}))^{1+\delta/2}$$
$$=\frac{2^{1+\delta/2}}{b^{2+\delta}}\left(\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\mathrm{tr}(\Omega\Sigma_{ij})^2+\sum_{j\neq j'}^{b}\frac{1}{n_{ij}n_{ij'}}\mathrm{tr}(\Omega\Sigma_{ij}\Omega\Sigma_{ij'})\right)^{1+\delta/2}$$

$$\leq \frac{2^{1+\delta/2}}{b^{2+\delta}}(bp)^{1+\delta/2}\sum_{j=1}^{b}\frac{1}{n_{ij}^{1+\delta/2}(n_{ij}-1)^{1+\delta/2}}\mathrm{tr}(\Omega\Sigma_{ij})^{2+\delta}$$
$$+\frac{2^{1+\delta/2}}{b^{2+\delta}}(pb(b-1))^{1+\delta/2}\sum_{j\neq j'}^{b}\frac{1}{n_{ij}^{1+\delta/2}n_{ij'}^{1+\delta/2}}\mathrm{tr}(\Omega\Sigma_{ij}\Omega\Sigma_{ij'})^{1+\delta/2}.$$

The last inequality follows from $|\sum_{j=1}^{b} a_j|^s \leq b^s \sum_{j=1}^{b}|a_j|^s$ which holds true for any $s > 0$. This later inequality also implies $(\mathrm{tr}(A))^s \leq p^s \sum_{i=1}^{p}\lambda_i^s = p^s\mathrm{tr}(A)^s$ for a non-negative definite matrix A with eigenvalues $\lambda_1, \ldots, \lambda_p$. Thus, $\lim_{a\to\infty}(1/a)\sum_{i=1}^{a}(\mathrm{var}(Z_i^{(A)}))^{1+\delta/2} < \infty$ as long as Assumptions 1 and 2 hold.

Likewise, $E(Z_i^{(A|B)}) = E(Z_i^{(B|A)}) = E(Z_i^{(AB)}) = 0$,

$$\lim_{a\to\infty}\frac{1}{a}\sum_{i=1}^{a}\mathrm{var}(Z_i^{(A|B)}) = \frac{2}{b}\lim_{a\to\infty}\left(\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\mathrm{tr}(\Omega\Sigma_{ij})^2\right),$$

$$\lim_{a\to\infty}\frac{1}{a}\sum_{i=1}^{a}\mathrm{var}(Z_i^{(B|A)}) = \frac{2}{b^2(b-1)^2}\left[(b-1)^2\lim_{a\to\infty}\left(\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\mathrm{tr}(\Omega\Sigma_{ij})^2\right)\right.$$
$$\left.+\lim_{a\to\infty}\left(\frac{1}{ab}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\frac{1}{n_{ij}n_{ij'}}\mathrm{tr}(\Omega\Sigma_{ij}\Omega\Sigma_{ij'})\right)\right],$$

$$\lim_{a\to\infty}\frac{1}{a}\sum_{i=1}^{a}\mathrm{var}(Z_i^{(AB)}) = \frac{2}{b(b-1)^2}\left[(b-1)^2\lim_{a\to\infty}\left(\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\mathrm{tr}(\Omega\Sigma_{ij})^2\right)\right.$$
$$\left.+\lim_{a\to\infty}\left(\frac{1}{ab}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\frac{1}{n_{ij}n_{ij'}}\mathrm{tr}(\Omega\Sigma_{ij}\Omega\Sigma_{ij'})\right)\right]$$

and Liaponauv's condition can be verified in a similar manner. □

Lemma 2 *As $a \to \infty$,*

$$Q_1 = \sqrt{a}\mathbf{Y}\left[\frac{1}{a(a-1)b}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)((J_a - I_a)\otimes J_b)\right.$$
$$\left.\times\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right]\mathbf{Y}' = o_p(1),$$

$$Q_2 = \sqrt{a}\mathbf{Y}\left[\frac{1}{a(a-1)b}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)((J_a - I_a)\otimes I_b)\right.$$
$$\left.\times\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right]\mathbf{Y}' = o_p(1)$$

and

$$Q_3 = \sqrt{a}\mathbf{Y}\left[\frac{1}{a(a-1)(b-1)}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)((J_a - I_a)\otimes P_b)\right.$$
$$\left.\times\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}'_{n_{ij}}\right)\right]\mathbf{Y}' = o_p(1).$$

Proof Write $Q_i = \mathbf{Y}M_i\mathbf{Y}'$ where M_i are defined in the obvious way. Since the diagonal elements of M_is are zero, we can easily verify by applying Lemma 1 that $E(Q_i) = 0$ and, as $a \to \infty$, $\mathrm{var}(Q_i) = o(1)$ for $i = 1, 2, 3$. □

The following Lemma extends an inequality in Rao and Kleffe (1989, pp 39) to a matrix quadratic forms.

Lemma 3 *Let $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_m)$ such that $\mathcal{E}_i = \Xi_i\mathbf{U}_i$ where $\mathbf{U}_i$ are independent p dimensional random vectors with mean $E(\mathbf{U}_i) = 0$ and covariance matrix I_p, and Ξ_i is a matrix of constants for $i = 1, 2, \ldots, m$. Let D be an $m \times m$ symmetric matrix. If the diagonal entries of D are all zero, then $E|\mathrm{tr}\mathcal{E}D\mathcal{E}' - E(\mathrm{tr}\mathcal{E}D\mathcal{E}')|^{2s} \le 2^{5s}C(2s)B(2s)\{\max_{i=1}^{m} E(\mathbf{U}_i'\mathbf{U}_i)^{2s}\}\{\mathrm{var}(\mathrm{tr}\mathcal{E}D\mathcal{E}')\}^s$ for $s > 1$ where $C(2s) = 2^{s-1}(2\pi)^{-1/2}\Gamma(s+1/2)$, $B(2s) = (18(2s)^{3/2}(2s-1)^{-1/2})^{2s}$ and $\Gamma(\cdot)$ is the Gamma function.*

Proof First note that when the diagonal entries of D are zeros from Lemma 1 we see that $\mathrm{var}(\mathrm{tr}\mathcal{E}D\mathcal{E}') = 2\mathrm{tr}\{(D\otimes I_p)(\bigoplus_{i=1}^{m}\Sigma_i)\}^2$ where $\Sigma_i = \Xi_i\Xi_i'$. On the other hand, note that $\mathrm{tr}(\mathcal{E}D\mathcal{E}') = \mathrm{vec}(\mathcal{E}')'(D\otimes I_p)\mathrm{vec}(\mathcal{E}')$ and that $\mathrm{var}(\mathrm{vec}(\mathcal{E})) = \bigoplus_{i=1}^{m}\Sigma_i$. Now applying inequality (2.3.10) in Rao and Kleffe (1989) yields the desired result. □

Proof of Theorem 3 Noting that $\hat{\Psi}_{ij}(\Omega)$ is a U-statistic (cf. Serfling 1980), it is straight forward to show that

$$E\left[\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\mathrm{tr}(\widehat{\Psi}_{ij}(\Omega)) - \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}(n_{ij}-1)}\mathrm{tr}(\Omega\Sigma_{ij})^2\right] = 0$$

and

$$E\left[\frac{1}{ab}\sum_{i=1}^{a}\sum_{j\ne j'}^{b}\frac{1}{n_{ij}n_{ij'}}\mathrm{tr}(\Omega S_{ij}\Omega S_{ij'}) - \frac{1}{ab}\sum_{i=1}^{a}\sum_{j\ne j'}^{b}\frac{1}{n_{ij}n_{ij'}}\mathrm{tr}(\Omega\Sigma_{ij}\Omega\Sigma_{ij'})\right] = 0.$$

Then it suffices to show that the variances tend to 0 as $a \to \infty$. This can be achieved by a repetitive application of Hölder's inequality to the second moments. □

Proof of Theorem 4 Here also we can set $\mu_{ij}^{(B)} = \mathbf{0}$ without loss of generality. As for the other sum of squares and cross products matrices, express $H^{(B)}$ as quadratic form

$$(b-1)H^{(B)} = \mathbf{Y}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)\left(\frac{1}{a}J_a\otimes P_b\right)$$
$$\times\left(\frac{1}{a}J_a\otimes P_b\right)\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}'\right)\mathbf{Y}'.$$

Noting that $a^{-1}(J_a\otimes P_b)$ is an idempotent matrix,

$$(b-1)H^{(B)} = \left(\mathbf{Y}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)\left(\frac{1}{a}J_a\otimes P_b\right)\right)$$
$$\times\left(\mathbf{Y}\left(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}}\right)\left(\frac{1}{a}J_a\otimes P_b\right)\right)'.$$

Now, $\mathbf{Y}(\bigoplus_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\mathbf{1}_{n_{ij}})(\frac{1}{a}J_a\otimes P_b) = \frac{1}{a}\sum_{i=1}^{a}(\bar{\mathbf{Y}}_{i1.},\ldots,\bar{\mathbf{Y}}_{ib.})P_b\mathbf{1}_a' = \bar{\mathbf{Z}}_a P_b\mathbf{1}_a'$, where $\bar{\mathbf{Z}}_a = (1/a)\sum_{i=1}^{a}\mathbf{Z}_i$ and $\mathbf{Z}_i = (\bar{\mathbf{Y}}_{i1.},\ldots,\bar{\mathbf{Y}}_{ib.})$. It can easily be seen that $\mathrm{var}(\mathrm{vec}(\sqrt{a}\bar{\mathbf{Z}}_a)) = \frac{1}{a}\sum_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\Sigma_{ij}$.

To establish the asymptotic normality of $\bar{\mathbf{Z}}_a$ it suffices to show that $\mathrm{tr}(\mathbf{T}'\bar{\mathbf{Z}}_a)$ has an asymptotic univariate normal distribution for each $\mathbf{T}\in\mathbb{R}^{p\times b}$ (see, for example, Rao 1973). Assuming $(1/a)\sum_{i=1}^{a}n_{ij}^{-1}\Sigma_{ij} = O(1)$ as $a\to\infty$, Liaponauv's condition will be satisfied if $a^{-(1+\delta/2)}\sum_{i=1}^{a}E|\mathrm{tr}(\mathbf{T}'\mathbf{Z}_i)|^{2+\delta}\to 0$ as $a\to\infty$. This latter condition holds if $E\|\mathbf{Y}_{ij1}\|^{2+\delta}<\infty$. Therefore, we have

$$\sqrt{a}\bar{\mathbf{Z}}_a \xrightarrow{\mathcal{L}} N_{p\times b}\left(\mathbf{0}, \frac{1}{a}\sum_{i=1}^{a}\bigoplus_{j=1}^{b}\frac{1}{n_{ij}}\Sigma_{ij}\right). \tag{12}$$

Observe that $(b-1)\mathrm{tr}H^{(B)}\Omega = a\,\mathrm{tr}(\bar{\mathbf{Z}}_a P_b\bar{\mathbf{Z}}_a'\Omega) = a\,\mathrm{tr}((\Omega^{1/2}\bar{\mathbf{Z}}_a P_b)'(\Omega^{1/2}\bar{\mathbf{Z}}_a P_b)) = \mathrm{vec}(\sqrt{a}\Omega^{1/2}\bar{\mathbf{Z}}_a P_b)'\mathrm{vec}(\sqrt{a}\Omega^{1/2}\bar{\mathbf{Z}}_a P_b)$. In view of (12), $\mathrm{vec}(\sqrt{a}\Omega^{1/2}\bar{\mathbf{Z}}_a P_b)\xrightarrow{\mathcal{L}} N_{pb}(\mathbf{0},\Lambda)$. As the rank of Λ is $p(b-1)$, the above pb variate normal distribution is a singular one. At any rate, from the theory of quadratic forms in normal random matrices (see Gupta and Nagar 2000), we have that $aH^{(B)}\Omega = a\bar{\mathbf{Z}}_a P_b\bar{\mathbf{Z}}_a'\Omega = O_p(1)$ and that if $\lambda_1,\lambda_2,\ldots,\lambda_{p(b-1)}$ are the nonzero eigenvalues of Λ then $(b-1)\mathrm{tr}H^{(B)}\Omega\xrightarrow{\mathcal{L}}\sum_{k=1}^{(b-1)p}\lambda_i\chi_{1,k}^2$. □

Proof of Theorem 5 Here we present the proof for $\psi = A$. The proofs for the others follow along the same lines.

Let $H^{(A)}(\mathbf{Y})$ and $G(\mathbf{Y})$ denote $H^{(A)}$ and G matrices, respectively, based on $\mathbf{Y}$. Given that $\sum_{i=1}^{a}\boldsymbol{\alpha}_{i,a} = \sum_{j=1}^{b}\boldsymbol{\beta}_{j,a} = \mathbf{0}$ and $G(\mathbf{Y}) = G(\mathbf{U})$, a direct manipulation reveals that

$$\sqrt{a}\,\mathrm{tr}(H^{(A)}(\mathbf{Y}) - G(\mathbf{Y}))\Omega = \sqrt{a}\,\mathrm{tr}(H^{(A)}(\mathbf{U}) - G(\mathbf{U}))\Omega$$
$$+2\frac{1}{\sqrt{a}}\sum_{i=1}^{a}\boldsymbol{\alpha}'_{i,a}\Omega\tilde{\mathbf{U}}_{i..} + \mathrm{tr}\Theta_A\Omega + O\left(\frac{1}{\sqrt{a}}\right).$$

It is straightforward to show that $a^{-1/2}\sum_{i=1}^{a}\boldsymbol{\alpha}'_{i,a}\Omega\tilde{\mathbf{U}}_{i..}$ converges to $\mathbf{0}$ in probability. Further, the asymptotic distribution of the first term on the right hand side is the same as that given in Theorem 2. The proof is complete by applying Slutsky's Theorem. □

Proof of Theorem 6 In the notations of the proof of Theorem 4, we can see that $\mathrm{vec}(\sqrt{a}\Omega^{1/2}\bar{\mathbf{Z}}_a P_b) \xrightarrow{\mathcal{L}} N_{pb}(\mathrm{vec}(\sqrt{a}\Omega^{1/2}BP_b), \Lambda)$. Then the final result follows from normal theory of quadratic forms. □

Acknowledgments The research of Solomon W. Harrar was supported by the 2007/2008 Small Grant Program of the University of Montana. Part of the manuscript was prepared during a research visit of the second author at the University of Montana, and the authors would like to express their gratitude to both departments for facilitating this visit. The authors are grateful to Dr. Kari J. Harris, principal investigator of the Greek Health Project, for making available the data analyzed in Sect. 4. The authors are also thankful to the editor, the associate editor and the two anonymous referees for their thoughtful suggestions and comments which have led to significant improvements over the original version of the manuscript.

References

Akritas, M. A., Arnold, S. F. (2000). Asymptotics for analysis of variance when the number of levels is large. *Journal of the American Statistical Association, 95*, 212–226.

Anderson, T. W. (2003). *An introduction to multivariate analysis*. New York: Wiley.

Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society B, 61*(3), 579–602.

Bathke, A. C. (2002). ANOVA for large number of treatments. *Mathematical Methods of Statistics, 11*, 118–132.

Bathke, A. C. (2004). The ANOVA F test can still be used in some balanced designs with unequal variances and nonnormal data. *Journal of Statistical Planning and Inference, 126*, 413–422.

Bathke, A. C., Harrar, S. W. (2008). Nonparametric methods in multivariate factorial designs for large number of factor levels. *Journal of Statistical Planning Inference, 138*, 588–610.

Boos, D. D., Brownie, C. (1995). ANOVA and rank tests when number of treatments is large. *Statistics and Probability Letters, 20*, 183–191.

Brunner, E., Dette, H., Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association, 92*, 1494–1502.

Dempster, A. P. (1958). A high dimensional two sample significance test. *Annals of Mathematical Statistics, 29*, 995–1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics, 16*, 41–50.

Fujikoshi, Y. (1975). Asymptotic formulas for the non-null distributions of three statistics for multivariate linear hypothesis. *Annals of the Institute of Statistical Mathematics, 27*, 99–108.

Fujikoshi, Y. (2002). Asymptotic expansions for the distribution of multivariate basic statistics and one-way MANOVA tests under nonnormality. *Journal of Statistical Planning and Inference, 108*, 263–282.

Fujikoshi, Y., Ulyanov, V., Shimizu, R. (2010). *Multivariate statistics: High-dimensional and large-sample approximations*. New York: Wiley.

Gupta, A. K., Harrar, S. W., Fujikoshi, Y. (2006). Asymptotics for testing hypothesis in some multivariate variance components model under non-normality. *Journal of Multivariate Analysis, 97*, 148–178.

Gupta, A. K., Harrar, S. W., Fujikoshi, Y. (2008). MANOVA for large hypothesis degrees of freedom under non-normality. *Test, 17*(1), 120–137.

Gupta, A. K., Nagar, D. K. (2000). *Matrix variate distributions*. Boca Raton: Chapman & Hall/CRC Press.
Harrar, S. W., Bathke, A. C. (2008). Nonparametric methods for unbalanced multivariate data and many factor levels. *Journal of Multivariate Analysis, 99*, 1635–1664.
Harrar, S. W., Gupta, A. K. (2007). Asymptotic expansion for the null distribution of the F-statistic in one-way ANOVA under non-normality. *Annals of the Institute of Statistical Mathematics, 59*(3), 531–556.
Hatsukami, D. K., Hughes, J. R., Pickens, R. W., Svikis, D. (1984). Tobacco withdrawal symptoms: An experimental analysis. *Psychopharmacology, 84*, 231–236.
Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., Fagerström, K. (1991). The Fagerström test for nicotine dependence: A revision of the Fagerström tolerance Questionnaire. *British Journal of Addiction, 86*, 1119–1127.
Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics, 1*, 799–821.
Ito, P. K. (1969). On the effect of heteroscedasticity and non-normality upon some multivariate test procedures. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. 2, pp. 87–120). New York: Academic Press.
Ito, P. K. (1980). Robustness of ANOVA and MANOVA test procedures. In P. R. Krishnaiah (Ed.), *Handbook of statistics* (Vol. 1, pp. 199–236). Amsterdam: North-Holland.
James, G. S. (1954). Test for linear hypothesis in univariate and multivariate analysis when ratios of the population variances are unknown. *Biometrika, 41*, 19–43.
Kakizawa, Y. (2007). A test of equality of mean vectors of several heteroscedastic multivariate populations. *Journal of Japan Statistical Society, 37*, 253–283.
Kakizawa, Y. (2009). Third-order power comparisons for a class of tests for multivariate linear hypothesis under general distributions. *Journal of Multivariate Analysis, 100*, 473–496.
Kakizawa, Y., Iwashita, T. (2008). Hotelling's one-sample and two-sample T^2 tests and the multivariate Behrens–Fisher problem under nonnormality. *Journal of Statistical Planning and Inference, 138*, 3379–3404.
Kohout, F. J., Berkman, L. F., Evans, D. A., Cornoni-Huntley, J. (1993). Two shorter forms of the CES-D depression symptoms index. *Journal of Aging and Health, 5*, 179–193.
Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, H. B. Mann (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278–292). Palo Alto: Stanford University Press.
Magnus, J. R., Neudecker, H. (1979). The commutation matrix: Some properties and applications. *Annals of Statistics, 7*, 381–394.
Muirhead, R. J. (1985). *Aspects of multivariate statistical theory*. New York: Wiley.
Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.
Rao, C. R., Kleffe, J. (1989). *Estimation of variance components and applications*. Amsterdam: Elsevier.
Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.
Siotani, M., Hayakawa, T., Fujikoshi, Y. (1985). *Modern multivariate statistical analysis: A graduate course and handbook*. Syracuse: American Sciences Press.
Wakaki, H., Yanagihara, H., Fujikoshi, Y. (2002). Asymptotic expansions of the null distributions of statistics for multivariate linear hypothesis under non-normality. *Hiroshima Mathematical Journal, 32*, 17–50.
Wang, H., Akritas, M. G. (2004). Rank tests for ANOVA with large number of factor levels. *Journal of Nonparametric Statistics, 16*, 563–589.
Wang, H., Akritas, M. G. (2009). *Asymptotically distribution free tests in heteroscedastic unbalanced high dimensional ANOVA*. Technical Report, II-09-1, Department of Statistics, Kansas State University.
Wang, L., Akritas, M. G. (2006). Two-way heteroscedastic ANOVA when the number of levels is large. *Statistica Sinica, 16*, 1387–1408.
Wellman, R. J., DiFranza, J. R., Savageau, J. A., Godiwala, S., Friedman, K., Hazelton, J. (2005). Measuring adults' loss of autonomy over cigarette use: The hooked on nicotine checklist. *Nicotine and Tobacco, 7*, 157–161.