Asymptotics for Tests on Mean Profiles, Additional Information and Dimensionality under Non-normality

Solomon W. Harrar*

Abstract

We consider the comparison of mean vectors for k groups when k is large and sample size per group is fixed. The asymptotic null and non-null distributions of the normal theory Likelihood Ratio, Lawley-Hotelling and Bartlett-Nanda-Pillai statistics are derived under general conditions. We extend the results to tests on the profiles of the mean vectors, tests for additional information (provided by a sub-vector of the responses over and beyond the remaining sub-vector of responses in separating the groups) and tests on the dimension of the hyperplane formed by the mean vectors. Our techniques are based on perturbation expansions and limit theorems applied to independent but non-identically distributed sequences of quadratic forms in random matrices. In all these four MANOVA problems, the asymptotic null and non-null distributions are normal. Both the null and non-null distributions are asymptotically invariant to non-normality when the group sample sizes are equal. In the unbalanced case, a slight modification of the test statistics will lead to asymptotically robust tests. Based on the robustness results, some approaches for finite approximation are introduced. The numerical results provide strong support for the asymptotic results and finiteness approximations.

AMS 2000 subject classification: Primary 62H10; Secondary 62H15

Key words and phrases: Asymptotics, Dimensionality, Distribution of Eigenvalues, MANOVA, Perturbation Expansion, Tests for Additional Information, Robustness

1 Introduction

1.1 Model and Background

Let $\mathbf{y}_{i1}, \mathbf{y}_{i2}, \ldots, \mathbf{y}_{in_i}$ be n_i identically and independently distributed p-dimensional vectors of observations from population (in the treatment group) Π_i, $i = 1, \ldots, k$. Assume that $\mathbf{y}_{ij}$ can be modeled as

$$\mathbf{y}_{ij} = \boldsymbol{\mu}_i + \Sigma^{1/2}\boldsymbol{\varepsilon}_{ij}, \quad j = 1, \ldots, n_i \quad \text{and} \quad i = 1, \ldots, k \tag{1.1}$$

*Department of Mathematical Sciences, University of Montana, Missoula, MT 59812, Email:harrar@mso.umt.edu

where Σ is a $p \times p$ positive definite matrix of unknown constants and the $\boldsymbol{\varepsilon}_{ij}$'s are identically and independently distributed with mean $\mathbf{0}$ and variance I_p. Clearly, model (1.1) implies that $E(\mathbf{y}_{ij}) = \boldsymbol{\mu}_i$ and $\mathrm{var}(\mathbf{y}_{ij}) = \Sigma$.

In this paper we are concerned with various tests on the mean vectors $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_k$. More precisely, we first consider the problem of testing equality of the mean vectors. Second, we consider testing problems that arise when observations are collected at successive but fixed time points for all subjects in the study groups. In these testing problems interest lies in analyzing the mean profiles in the different groups. The analysis based on these tests is called *Profile Analysis*. The third testing problem we consider deals with questions on equality of mean vectors given that the groups are known to be the same in terms of some other concomitant variables. This testing problem is referred to as *Testing for Additional Information*. The last testing problem we consider is closely related to the first testing problem. If the null hypothesis of common mean vectors is rejected then we would like to know the dimension of the hyperplane containing the mean vectors. This dimension is also the number of discriminant functions needed to separate the groups. The dimension is determined by conducting a sequence of tests known as *Tests of Dimensionality*. For a detailed description of the above four testing problems see Rao (1973, Section 8c) and Rencher (2002, Chapters 5 and 6).

For each of these testing problems, we obtain the asymptotic null and non-null distributions of some common multivariate tests statistics without assuming normality. The statistics are known as the *Likelihood Ratio*, *Lawley-Hotelling* and *Bartlett-Nanda-Pillai* statistics. The tests based on these statistics are known to have some desirable properties under normality. It is also known that none of these tests uniformly most powerful than the other two under normality. Indeed for large sample size (large n_i's), the three tests are equivalent up to the order $O(N^{-1})$, where $N = \sum_{i=1}^{k} n_i$, for all the MANOVA problems discussed in the previous paragraph regardless of the normality assumption . Moreover, except for the tests of dimensionality, the other tests are robust (up to the order $O(N^{-1})$) to non-normality both in terms of size and power (see Seo, Kanda and Fujikoshi, 1995, Fujikoshi, 2002, Gupta, Xu and Fujikoshi, 2006 and Maruyama, 2007).

An interesting feature of this paper is the asymptotic framework in which we let the number of populations (treatment groups) to go to infinity keeping the sample sizes per populations (treatment groups) fixed. This asymptotic framework is very important in the analysis of DNA gene expression data. For example, Storey et al. (2005) proposes a model for assessing temporal changes in differential expressions of genes in control and endotoxin treated individuals. One interesting aspect of this experiment is the comparison of the temporal differential expression profiles across the genes (large in number, say k) for each of the control and treatment groups. There has been some advancement in this asymptotic framework in the parametric ANOVA setup (see Akritas and Arnold, 2000, Bathke, 2002, Akritas and Papadatos, 2004 and Gupta, Harrar and Fujikoshi, 2006, 2007 and Harrar and Gupta, 2007). The extension to the multivariate situation is particularly interesting and non-trivial mainly for two reasons. First, there is a multitude of possible test statistics in the multivariate situation, and, secondly, some of the multivariate problems do not have a univariate counter part. For example, profile analysis, testing for additional information and testing for dimensionality do not have univariate counter parts.

1.2 Preliminaries

Two matrices that are central in this paper are the hypothesis and error sum of squares and cross products, denoted by H and E, respectively. These matrices are given by

$$H = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{\mathbf{y}}_{i.} - \bar{\mathbf{y}}_{..})(\bar{\mathbf{y}}_{i.} - \bar{\mathbf{y}}_{..})' \quad \text{and} \quad E = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})' \tag{1.2}$$

where

$$\bar{\mathbf{y}}_{i.} = n_i^{-1}\sum_{j=1}^{n_i}\mathbf{y}_{ij}, \quad \bar{\mathbf{y}}_{..} = N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mathbf{y}_{ij}, \quad \text{and} \quad N = \sum_{i=1}^{k} n_i. \tag{1.3}$$

We will need the following assumptions for the development of the results of the paper. We will also need others assumptions peculiar to the testing problem considered in the later sections of the paper. We shall present those as we need them.

A1 : $N = \bar{n}k$ where $\bar{n}$ does not depend on k.

A2 : $\displaystyle\sum_{i=1}^{k} n_i^{-1} = \underline{n}k$ where $\underline{n}$ does not depend on k.

A3 : $E(\boldsymbol{\varepsilon}_{11}'\boldsymbol{\varepsilon}_{11})^{2+\delta} < \infty$ for some $\delta > 0$.

A4 : $\displaystyle\sup_{k\geq 1}\frac{1}{k}\sum_{i=1}^{k} n_i^{3+\delta} < \infty$ for some $\delta > 0$.

Assumptions $A1$ and $A2$ say that for large k the averages of the sample sizes and inverses of the sample sizes grow with k in a linear fashion. Since the n_i's are fixed, these assumptions are not very restrictive. For example if $n_1 = n_2 = \ldots = n_k$ then these assumptions are automatically satisfied. In fact we can make these assumptions even milder with out affecting the validly of our results by simply requiring

$$(1/k)\sum_{i=1}^{k} n_i = O(1) \quad \text{and} \quad (1/k)\sum_{i=1}^{k} n_i = O(1)$$

as $k \to \infty$.

The asymptotic distributions of the test statistics will be shown to be normal. At the center of the derivations is the application of Lindebeg-Feller's Central Limit Theorem to independently (not necessarily identically) distributed quadratic forms in random matrices. The assumptions $A3$ and $A4$ guarantee the validity of the Lindeberg's condition for the application in this paper. Nevertheless, when $n_1 = n_2 = \ldots = n_k$ these assumptions will not be needed except for the test of dimensionality even in that case only A_3 is needed (see Section 5).

In our presentation we use the following notations. First, $\mathbf{0}$ will denote the vector $(0, \ldots, 0)'$, the dimension will be clear from the context. Second, $\mathbf{1}_n$ denotes an n-dimensional vector $(1, \ldots, 1)'$ consisting of ones. Third, I_n is the identity matrix of order n, $J_n = \mathbf{1}_n\mathbf{1}_n'$ and $P_n = I_n - n^{-1}J_n$. We write

$\mathrm{diag}(a_1, a_2, \ldots, a_n)$ to refer to the diagonal matrix whose diagonal entries are $a_1, a_2, \ldots, a_n$. The symbol "$\xrightarrow{\mathcal{L}}$" stands as an abbreviation for "converges in distribution to". Further, we will use the Kronecker (or direct) product $A \otimes B$, of matrices A and B, 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 the definition and properties of the commutation matrix.

Some of the proofs in this paper appeal to limit theorems for sums of independent quadratic forms in random marines. These limit theorem require computation of the first two moments of the quadratic forms. The following Lemma will aid in these computations.

Lemma 1.1. *Suppose $\mathcal{E} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_n)$ where $\boldsymbol{\varepsilon}_i$ is a $p \times 1$ vector. Assume $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_n$ independently and identically distributed with mean $\mathbf{0}$ and variance Σ. Let $A = (a_{ij})$ be an $n \times n$ symmetric matrix. Then,*

i. $E(\mathcal{E}A\mathcal{E}') = \mathrm{tr}(A)\Sigma$ *and*

ii. $\mathrm{var}(\mathcal{E}A\mathcal{E}') = \mathrm{tr}(A^2)(I_{p^2} + K_{p,p})(\Sigma \otimes \Sigma) + \mathrm{tr}(\mathrm{Diag}(A)^2)K_4(\boldsymbol{\varepsilon}_1)$

where $K_4(\boldsymbol{\varepsilon}_1) = \mathrm{E}(\mathrm{vec}(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1')\mathrm{vec}(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1')') - (I_{p^2} + K_{p,p})(\Sigma \otimes \Sigma) - \mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)'$ *and* $\mathrm{Diag}(A) = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn})$.

The following Corollary is an immediate consequence of Lemma 1.1.

Corollary 1.1. *Let the notations and assumptions be as in Lemma 1.1. Then,*

$$\mathrm{var}\left(\mathrm{tr}(\mathcal{E}A\mathcal{E}')\right) = 2\mathrm{tr}(A^2)\mathrm{tr}(\Sigma^2) + \mathrm{tr}(\mathrm{Diag}(A)^2)(\mu_4 - 2\mathrm{tr}(\Sigma^2) - (\mathrm{tr}\Sigma)^2)$$

where $\mu_4 = \sum_{a,b}^{p} E(\varepsilon_{1a}^2\varepsilon_{1b}^2)$.

Note that when all the diagonal entries of A are zeros, the second terms in the right hand side of the variances vanish. The proofs of Lemma 1.1 and Corollary 1.1 are found in Bathke and Harrar (2006).

The rest of the paper is organized as follows. In section 2, the asymptotic distributions of the three test statistics for testing equality of mean vectors are obtained. The results are extended, in Sections 3-5, when the statistics are applied for testing hypotheses on the mean profiles, testing for additional information and testing a sequence of hypotheses to determine the dimension of the hyperplane containing the mean vectors, respectively. Finite sample approximations are discussed in Section 6. In Section 7, the performance of the asymptotic results and the finite sample approximations are investigated via simulation studies. Section 8 contains some concluding remarks.

2 Tests for Equality of Mean Vectors

2.1 Hypotheses and Test Statistics

Consider the model 1.1 and testing the hypothesis

$$\mathrm{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_k \quad \text{versus} \quad \mathrm{H}_1 : \text{ Not } \mathrm{H}_0.$$

Under the assumption ε_{ij} has normal distibution, the three commonly used tests are based on the Likelihood Ratio, Lawley-Hotelling and Bartlett-Nanda-Pilai crteria. These criteria, denoted by $\mathrm{T_{LR}}$, $\mathrm{T_{LH}}$ and $\mathrm{T_{BNP}}$, respectively, are defined as follows.

$$\mathrm{T_{LR}} = -\log(|E|/|H+E|), \quad \mathrm{T_{LH}} = \mathrm{tr}HE^{-1} \quad \text{and} \quad \mathrm{T_{BNP}} = \mathrm{tr}H(H+E)^{-1}. \tag{2.1}$$

The exact null and non-null distributions of these statistics take complicated forms except in a few special cases. Tabulations have been provided in some cases (see Schatzoff, 1966 , Pillai and Gupta, 1969, and Gupta, 1971). On the other hand, these statistics are known to have asymptotic chi-square distribution as sample size tends to infinity. There are also satisfactory asymptotic expansions as functions of chi–square variables, see for example Anderson (1984). Asymptotics in the large dimension framework have been recently considered by some researchers (see, for example, Mudholkar and Trivedi, 1980, Tonda and Fujikoshi, 2004, Srivastava and Fujikoshi, 2006). For a review of the works in this direction see Fujikoshi (2004).

Under non–normality, the null distributions of these statistics are known to converge to chi-square distribution if $N/n_i = O(1)$ as the n_i's tend to infinity. There are also few recent works on asymptotic expansions of the distributions of these statistics under non-normality (Fujikoshi, 2002).

The focus of this section is the asymptotic null and non-null distribution of the test statistics (2.1) when k tends to infinity but the n_i's are fixed (small). In this asymptotic framework the condition $N/n_i = O(1)$ does not hold and, therefore, the usual large sample asymptotic χ^2 results do not hold. Under normality, Fujikoshi (1975) derived asymptotic expansion for the null and non–null distributions of the three statistics in the general linear hypothesis context when both the hypothesis and error degrees of freedom are large. His asymptotic framework, stated in the one way layout context, is equivalent to k and N tending to infinity (assuming $n_i \geq 2$ fixed) at the same rate such that $k/N \to c \in (0, 1/2]$, . Gupta, Harrar and Fujikoshi (2007) generalized Fujikoshi's work in the one-way layout case by replacing the normality assumption with some moment conditions. Namely, they assumed existence of fourth order moments of the errors to derive the null and non-null distributions. In addition, they required the third order moments of the errors to be zero in the non-null case. Another generalization to the multivariate variance components model was done by Gupta, Harrar and Fujikoshi (2006). This section presents a generalization of Gupta, Harrar and Fujikoshi (2007) by dropping the third moment condition in the non-null case and dropping the forth order moment condition in the balance null case. The work in this section can also be viewed as a multivariate generalization of Akritas and Arnold (2000), Bathke (2002) and Akritas and Papadatos (2004). Akritas and Arnold (2000) derive the asymptotic distribution of the ANOVA F-statistic in one-way and two-way layout models and Bathke (2002) considers the balanced multi-factor case. In contrast, Akritas and Papadatos (2004) consider the one-way layout in the heteroscedastic case. Our techniques differ from those used in all three works; notably, and they facilitate a unified approach to tackling the testing problems considered in the later sections of this paper.

2.2 Asymptotic Distributions

Denote the non-centrality parameter matrix by

$$\Omega_k = \sum_{i=1}^{k} n_i \Sigma^{-1/2}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\Sigma^{-1/2}, \quad \text{where} \quad \bar{\boldsymbol{\mu}} = \sum_{i=1}^{k}(n_i/N)\boldsymbol{\mu}_i$$

and assume

A5 : $\Omega_k = \sqrt{k}\Theta$ for some $p \times p$ matrix Θ which does not depend on k.

Substituting (1.1) in H and E given in (1.2) and expanding the quadratic forms yields

$$\begin{aligned}\frac{1}{k}\Sigma^{-1/2}H\Sigma^{-1/2} = {} & \frac{1}{k}\sum_{i=1}^{k} n_i \bar{\boldsymbol{\varepsilon}}_{i.}\bar{\boldsymbol{\varepsilon}}_{i.}' - \frac{N}{k}\bar{\boldsymbol{\varepsilon}}_{..}\bar{\boldsymbol{\varepsilon}}_{..}' + \frac{1}{k}\sum_{i=1}^{k} n_i(\bar{\boldsymbol{\varepsilon}}_{i.} - \bar{\boldsymbol{\varepsilon}}_{..})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\Sigma^{-1/2} \\ & + \frac{1}{k}\sum_{i=1}^{k} n_i \Sigma^{-1/2}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\bar{\boldsymbol{\varepsilon}}_{i.} - \bar{\boldsymbol{\varepsilon}}_{..})' + \frac{1}{k}\sum_{i=1}^{k} n_i \Sigma^{-1/2}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\Sigma^{-1/2}\end{aligned} \tag{2.2}$$

and

$$\frac{1}{k}\Sigma^{-1/2}E\Sigma^{-1/2} = \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_{i.})(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_{i.})'. \tag{2.3}$$

In the next proposition we assert that some of the terms of $(1/k)\Sigma^{-1/2}H\Sigma^{-1/2}$ are of order $o_P(k^{-1/2})$ and, hence, are negligible up to the order $k^{-1/2}$.

Proposition 2.1. *As $k \to \infty$,*

i. $\frac{N}{k}\bar{\boldsymbol{\varepsilon}}_{..}\bar{\boldsymbol{\varepsilon}}_{..}' = o_P(\frac{1}{\sqrt{k}})$ *and*

ii. $\frac{1}{k}\sum_{i=1}^{k} n_i(\bar{\boldsymbol{\varepsilon}}_{i.} - \bar{\boldsymbol{\varepsilon}}_{..})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\Sigma^{-1/2} = o_P(\frac{1}{\sqrt{k}})$.

Proof. For the sake of brevity, we will prove part (ii) in detail, and sketch the proof of part (i). It is obvious that

$$\frac{1}{k}\sum_{i=1}^{k} n_i(\bar{\boldsymbol{\varepsilon}}_{i.} - \bar{\boldsymbol{\varepsilon}}_{..})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\Sigma^{-1/2} = \frac{1}{k}\sum_{i=1}^{k} n_i\bar{\boldsymbol{\varepsilon}}_{i.}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\Sigma^{-1/2}.$$

Since $E(\boldsymbol{\varepsilon}_{ij}) = 0$ and $\mathrm{var}(\boldsymbol{\varepsilon}_{ij}) = I_p$, it follows that

$$E\left\{\frac{1}{\sqrt{k}}\sum_{i=1}^{k} n_i\bar{\boldsymbol{\varepsilon}}_{i.}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\Sigma^{-1/2}\right\} = 0$$

and

$$\mathrm{var}\left\{\frac{1}{\sqrt{k}}\sum_{i=1}^{k} n_i\bar{\boldsymbol{\varepsilon}}_{i.}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\Sigma^{-1/2}\right\} = \frac{1}{\sqrt{k}}\Theta = O(\frac{1}{\sqrt{k}}).$$

Therefore,

$$\frac{1}{k}\sum_{i=1}^{k} n_i(\bar{\boldsymbol{\varepsilon}}_{i.} - \bar{\boldsymbol{\varepsilon}}_{..})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\Sigma^{-1/2} = \frac{1}{\sqrt{k}}o_P(1) = o_P(\frac{1}{\sqrt{k}}).$$

The proof for part (i) proceeds similarly using $E\{(N/\sqrt{k})\bar{\boldsymbol{\varepsilon}}_{..}\bar{\boldsymbol{\varepsilon}}'_{..}\} = (1/\sqrt{k})I_p$ and $\mathrm{var}\{(N/\sqrt{k})\bar{\boldsymbol{\varepsilon}}_{..}\bar{\boldsymbol{\varepsilon}}'_{..}\} = o(1)$. This last fact can be shown with the aid of Lemma 1.1. □

Let $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k)$ where $\mathcal{E}_i = (\boldsymbol{\varepsilon}_{i1}, \boldsymbol{\varepsilon}_{i2}, \ldots, \boldsymbol{\varepsilon}_{in_i})$. It is easy to verify $n_i\bar{\boldsymbol{\varepsilon}}_{i.}\bar{\boldsymbol{\varepsilon}}'_{i.} = \mathcal{E}_i(1/n_i)J_{n_i}\mathcal{E}'_i$ and $\sum_{j=1}^{n_i}(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_{i.})(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_{i.})' = \mathcal{E}_iP_{n_i}\mathcal{E}'_i$. Equations (2.2) and (2.3) together with Proposition 2.1 and assumption A4 imply

$$\frac{1}{k}\Sigma^{-1/2}H\Sigma^{-1/2} = \frac{1}{k}\sum_{i=1}^{k}\mathcal{E}_i[\frac{1}{n_i}J_{n_i}]\mathcal{E}'_i + \frac{1}{\sqrt{k}}\Theta + o_P(\frac{1}{\sqrt{k}}) \quad \text{and} \tag{2.4}$$

$$\frac{1}{k}\Sigma^{-1/2}E\Sigma^{-1/2} = \frac{1}{k}\sum_{i=1}^{k}\mathcal{E}_iP_{n_i}\mathcal{E}'_i. \tag{2.5}$$

Define

$$U_k = \frac{1}{\sqrt{k}}\sum_{i=1}^{k}[\mathcal{E}_i(\frac{1}{n_i}J_{n_i})\mathcal{E}'_i - I_p] \quad \text{and} \quad V_k = \frac{1}{\sqrt{k}}\sum_{i=1}^{k}[\mathcal{E}_i(I_{n_i} - \frac{1}{n_i}J_{n_i})\mathcal{E}'_i - (\bar{n}-1)I_p]$$

and write

$$\frac{1}{k}\Sigma^{-1/2}H\Sigma^{-1/2} = I_p + \frac{1}{\sqrt{k}}\Theta + \frac{1}{\sqrt{k}}U_k + o_P(\frac{1}{\sqrt{k}}) \quad \text{and} \quad \frac{1}{k}\Sigma^{-1/2}E\Sigma^{-1/2} = (\bar{n}-1)I_p + \frac{1}{\sqrt{k}}V_k. \tag{2.6}$$

We can see from the second equation of (2.6) that $(1/(N-k))E$ converges in probability to Σ. Now let

$$Z_k = (\frac{1}{k}\Sigma^{-1/2}E\Sigma^{-1/2})^{-1/2}(\frac{1}{k}\Sigma^{-1/2}H\Sigma^{-1/2})(\frac{1}{k}\Sigma^{-1/2}E\Sigma^{-1/2})^{-1/2}. \tag{2.7}$$

The test statistics in (2.1) can be expressed in terms of Z_k as

$$\mathrm{T_{LR}} = \log|I_p + Z_k|, \quad \mathrm{T_{LH}} = \mathrm{tr}Z_k \quad \text{and} \quad \mathrm{T_{BNP}} = \mathrm{tr}Z_k(I_p + Z_k)^{-1}. \tag{2.8}$$

Substituting (2.6) in (2.7) we obtain,

$$Z_k = \left((\bar{n}-1)I_p + \frac{1}{\sqrt{k}}V_k\right)^{-1/2}(I_p + \frac{1}{\sqrt{k}}\Theta + \frac{1}{\sqrt{k}}U_k + o_P(\frac{1}{\sqrt{k}}))\left((\bar{n}-1)I_p + \frac{1}{\sqrt{k}}V_k\right)^{-1/2}.$$

Note that, under the assumptions $A1$–$A3$, both U_k and V_k converge in distribution to a normal random matrix as $k \to \infty$. Thus, $U_k = O_P(1)$ and $V_k = O_P(1)$ as $k \to \infty$. Now we can expand

$$\left((\bar{n}-1)I_p + \frac{1}{\sqrt{k}}V_k\right)^{-1/2} = (\bar{n}-1)^{-1/2}\left(I_p - \frac{1}{2}\frac{1}{\sqrt{k}(\bar{n}-1)}V_k + o_P(\frac{1}{\sqrt{k}})\right). \tag{2.9}$$

This expansion and some algebra leads to

$$(\bar{n}-1)Z_k = I_p + \frac{1}{\sqrt{k}}\Theta + \frac{1}{\sqrt{k}}(U_k - \frac{1}{(\bar{n}-1)}V_k) + o_P(\frac{1}{\sqrt{k}}). \tag{2.10}$$

The expansion formulae

$$-\log|I_p - tA| = \sum_{j=1}^{l}\frac{t^j}{j}\mathrm{tr}A^j + O(t^{l+1}) \quad \text{and} \quad (I_p - tA)^{-1} = \sum_{j=1}^{l} t^j A^j + O(t^{l+1}) \tag{2.11}$$

hold for small $t \in (0,1)$ and any $p \times p$ matrix A such that $A = O(1)$ as $k \to \infty$. Substituting (2.10) in (2.8) and applying the expansion formulae (2.11) keeping in mind $U_k = O_P(1)$ and $V_k = O_P(1)$, we see that the three test statistics, centered and scaled appropriately, are asymptotically equivalent to $\mathrm{tr}(U_k - \frac{1}{(\bar{n}-1)}V_k) + \mathrm{tr}\Theta$. More precisely,

$$\sqrt{k}(\ell T_{\mathrm{G}} - h) = \mathrm{tr}(U_k - \frac{1}{(\bar{n}-1)}V_k) + \mathrm{tr}\Theta + o_P(1), \tag{2.12}$$

where for Likelihood Ratio criteria, $\mathrm{G} = \mathrm{LR}$, $\ell = \bar{n}$ and $h = \bar{n}p\log(\bar{n}/(\bar{n}-1))$. Similarly for the Lawley-Hotelling criterion $\mathrm{G} = \mathrm{LH}$, $\ell = \bar{n}-1$ and $h = p$, and for the Bartlett-Nanda-Pillai criterion $\mathrm{G} = \mathrm{BNP}$, $\ell = \bar{n}^2/(\bar{n}-1)$ and $h = \bar{n}p/(\bar{n}-1)$.

Theorem 2.1. *Under assumptions* A1–A5*,*

$$\sqrt{k}(\ell T_{\mathrm{G}} - h) \xrightarrow{\mathcal{L}} N\left(\mathrm{tr}\Theta, \frac{2\bar{n}p}{\bar{n}-1} + \frac{\bar{n}(\bar{n}\underline{n}-1)}{(\bar{n}-1)^2}\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11})\right)$$

as $k \to \infty$ *where* $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}) = (\mu_4(\boldsymbol{\varepsilon}_{11}) - 2p - p^2)$ *and* $\mu_4(\boldsymbol{\varepsilon}_{11}) = \sum_{a,b}^{p} E(\varepsilon_{11a}^2\varepsilon_{11b}^2)$.

Proof. In view of (2.12), it suffices to show that

$$\mathrm{tr}(U_k - \frac{1}{(\bar{n}-1)}V_k) \xrightarrow{\mathcal{L}} N\left(0, \frac{2\bar{n}p}{\bar{n}-1} + \frac{\bar{n}(\bar{n}\underline{n}-1)}{(\bar{n}-1)^2}\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11})\right).$$

To that end notice that,

$$\mathrm{tr}(U_k - \frac{1}{(\bar{n}-1)}V_k) = \frac{1}{\sqrt{k}}\sum_{i=1}^{k}\mathrm{tr}\{\mathcal{E}_i\frac{1}{\bar{n}-1}(\frac{\bar{n}}{n_i}J_{n_i} - I_{n_i})\mathcal{E}_i'\}, \tag{2.13}$$

and that,

$$\lim_{k\to\infty} E\left[\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\mathrm{tr}\{\mathcal{E}_i\frac{1}{\bar{n}-1}(\frac{\bar{n}}{n_i}J_{n_i} - I_{n_i})\mathcal{E}_i'\}\right] = 0 \quad \text{and}$$

$$\lim_{k\to\infty}\mathrm{var}(\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\mathrm{tr}\{\mathcal{E}_i\frac{1}{\bar{n}-1}(\frac{\bar{n}}{n_i}J_{n_i} - I_{n_i})\mathcal{E}_i'\}) = \frac{2\bar{n}p}{\bar{n}-1} + \frac{\bar{n}(\bar{n}\underline{n}-1)}{(\bar{n}-1)^2}\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11})$$

where the second equality follows from Corollary 1.1.

The right hand side of (2.13) will have asymptotic normal distribution if we establish that the Lindeberg's condition holds. Indeed, under the assumption $A3$, the more stronger Liaponov's condition holds. To see this, first observe that

$$\mathrm{tr}\{\mathcal{E}_i\frac{1}{\bar{n}-1}(\frac{\bar{n}}{n_i}J_{n_i}-I_{n_i})\mathcal{E}_i'\}=\mathrm{vec}(\mathcal{E}_i')'\left(I_p\otimes\frac{1}{\bar{n}-1}(\frac{\bar{n}}{n_i}J_{n_i}-I_{n_i})\right)\mathrm{vec}(\mathcal{E}_i').$$

Then by inequality (2.3.10) of Rao and Kleffe (1988), with $m=1$ in their notation,

$$\begin{aligned}
&E\left|\mathrm{tr}\{\mathcal{E}_i\frac{1}{\bar{n}-1}(\frac{\bar{n}}{n_i}J_{n_i}-I_{n_i})\mathcal{E}_i'\}-E\left[\mathrm{tr}\{\mathcal{E}_i\frac{1}{\bar{n}-1}(\frac{\bar{n}}{n_i}J_{n_i}-I_{n_i})\mathcal{E}_i'\}\right]\right|^{2+\delta}\\
&\quad\le\xi_\delta\mathrm{tr}\left\{\frac{1}{\bar{n}-1}(\frac{\bar{n}}{n_i}J_{n_i}-I_{n_i})\right\}^2E\{\mathrm{vec}(\mathcal{E}_i')'\mathrm{vec}(\mathcal{E}_i)\}^{2+\delta}\\
&\quad=\xi_\delta\frac{1}{(\bar{n}-1)^2}(\bar{n}^2-2\bar{n}+n_i)E\{\mathrm{tr}(\mathcal{E}_i'\mathcal{E}_i)\}^{2+\delta}
\end{aligned}$$

where $\xi_\delta<\infty$ depends only on δ. Now appealing to Minkowski's inequality, see for example inequality (2.6.5) in Rao and Kleffe (1988),

$$E\{\mathrm{tr}(\mathcal{E}_i'\mathcal{E}_i)\}^{2+\delta}=E(\sum_{j=1}^{n_i}\boldsymbol{\varepsilon}_{ij}'\boldsymbol{\varepsilon}_{ij})^{2+\delta}\le\left\{\sum_{j=1}^{n_i}(E(\boldsymbol{\varepsilon}_{ij}'\boldsymbol{\varepsilon}_{ij})^{2+\delta})^{1/(2+\delta)}\right\}^{2+\delta}=n_i^{2+\delta}E(\boldsymbol{\varepsilon}_{11}'\boldsymbol{\varepsilon}_{11})^{2+\delta}.$$

Finally it is a straightforward matter to verify that

$$\begin{aligned}
&\sum_{i=1}^{k}E\left|\mathrm{tr}\{\mathcal{E}_i\frac{1}{\bar{n}-1}(\frac{\bar{n}}{n_i}J_{n_i}-I_{n_i})\mathcal{E}_i'\}-E\left[\mathrm{tr}\{\mathcal{E}_i\frac{1}{\bar{n}-1}(\frac{\bar{n}}{n_i}J_{n_i}-I_{n_i})\mathcal{E}_i'\}\right]\right|^{2+\delta}\\
&\qquad=o\left(\left\{\sum_{i=1}^{k}\mathrm{var}(\mathrm{tr}\{\mathcal{E}_i\frac{1}{\bar{n}-1}(\frac{\bar{n}}{n_i}J_{n_i}-I_{n_i})\mathcal{E}_i'\})\right\}^{1+\delta/2}\right)
\end{aligned}$$

as $k\to\infty$. □

From Theorem 2.1 we see that the three tests are equivalent in terms their power when k is large. In general the null ($\Theta=0$) as well as the non-null distributions depend on $\kappa_4^{(1)}(\boldsymbol{\varepsilon})$ which is a measure of multivariate kurtosis (Mardia, 1970). Therefore, the tests are not, in general, asymptotically robust to non-normality. The effect of non-normality depends on the magnitudes of $\bar{n}\underline{n}-1$ and $\kappa_4^{(1)}(\boldsymbol{\varepsilon})$.

Let us now examine the effect of p on the upper quantiles of the test statistics if we assumed normality when normality did not hold. Let z_α be the upper $\alpha\in(0,1)$ quantile of a standard normal variate. The difference in the quantiles under non-normality and normality when the quantiles are based on limiting distribution in Theorem 2.1 is given by

$$\begin{aligned}
&z_\alpha\sqrt{\frac{2\bar{n}p}{\bar{n}-1}+\frac{\bar{n}(\bar{n}\underline{n}-1)}{(\bar{n}-1)^2}\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11})}-z_\alpha\sqrt{\frac{2\bar{n}p}{\bar{n}-1}}\\
&\qquad=z_\alpha p\sqrt{\frac{2\bar{n}}{\bar{n}-1}}\left[\left(\frac{2(\bar{n}\underline{n}-1)}{(\bar{n}-1)}\left(\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11})/p^2\right)+\frac{1}{p}\right)^{1/2}-\frac{1}{\sqrt{p}}\right]\\
&\qquad=z_\alpha p\sqrt{\frac{2\bar{n}}{\bar{n}-1}}\left[\sqrt{\frac{2(\bar{n}\underline{n}-1)}{(\bar{n}-1)}\left(\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11})/p^2\right)}-\frac{1}{\sqrt{p}}+O(\frac{1}{p})\right]
\end{aligned}$$

which holds for large p and under the assumption that $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}) = O(p^2)$. The second equality follows from a Taylor series expansion applied to the square root term in the square bracket of the preceding line. The assumption $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}) = O(p^2)$ is tenable, for example, for elliptically contoured populations for which $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11})$ is proportional to $p(p+2)$. Therefore, the effect of non-normality gets worse as p increases.

On the other hand when $n_1 = \ldots = n_k$, the second term in the variances vanish. Thus all the three tests are null as well as non-null robust. Moreover, in this equal sample size case none of the assumptions A1–A4 are needed.

To apply Theorem 2.1 in the unbalanced case one needs a consistent ($k \to \infty$) estimator of $\kappa_4^{(1)}$. One can use the family of estimators proposed by Yanagihara (2006). These estimators are unbiased under normality and less biased than Mardia's (1970) estimator under non-normality. The family is indexed by a tuning parameter λ and its expression when $\lambda = 0$ is

$$\hat{\kappa}_4^{(1)} = \frac{N+1}{N(N-1)} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left\{ (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{..})' \hat{\Sigma}^{-1} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{..}) \right\}^2 - p(p+2) \tag{2.14}$$

where $\hat{\Sigma} = N^{-1} \sum_{=1}^{k} \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{..})(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{..})'$.

2.3 Robustness in the Unbalanced Case

We have seen in Theorem 2.1 that the three multivariate test statistics are not robust in the unbalanced case. However, it turns out that a slight modification to E will make the tests asymptotically robust. Define,

$$\tilde{E} = \sum_{i=1}^{k} \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})' = \sum_{i=1}^{k} \Sigma^{1/2} \mathcal{E}_i \left[\frac{1}{n_i - 1} P_{n_i} \right] \mathcal{E}_i' \Sigma^{1/2}.$$

Accordingly, we modify the test statistics as,

$$\tilde{\mathrm{T}}_{\mathrm{LR}} = -\log(|\tilde{E}|/|H + \tilde{E}|), \quad \tilde{\mathrm{T}}_{\mathrm{LH}} = \mathrm{tr} H \tilde{E}^{-1} \quad \text{and} \quad \tilde{\mathrm{T}}_{\mathrm{BNP}} = \mathrm{tr} H (H + \tilde{E})^{-1}. \tag{2.15}$$

Comparing $\tilde{E}$ with E in (2.5), we see that the only difference between the two is that in the case of $\tilde{E}$ the sum of squares and cross products within each group is divided by $n_i - 1$. It may be noted that as an estimator of Σ, $E/(N-k)$ might be better than $(1/k)\tilde{E}$ under normality. However, as we will see later under non-normality this estimation efficiency is traded off for robustness. In the balanced case T_{LH} and $\tilde{T}_{\mathrm{LH}}$ are equivalent whereas T_{LR} and T_{BNP} differ slightly from $\tilde{T}_{\mathrm{LR}}$ and $\tilde{T}_{\mathrm{BNP}}$, respectively.

It can, now, easily be seen that

$$\frac{1}{k} \Sigma^{-1/2} \tilde{E} \Sigma^{-1/2} = I_p + \frac{1}{\sqrt{k}} \tilde{V}_k \quad \text{where} \quad \tilde{V}_k = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} [\mathcal{E}_i (\frac{1}{n_i - 1} P_{n_i}) \mathcal{E}_i' - I_p].$$

Defining $\tilde{Z}_k$ the same way as in (2.7) but in terms of $\tilde{E}$ instead of E, and carrying out similar algebra as that which led to (2.10), we obtain

$$\tilde{Z}_k = I_p + \frac{1}{\sqrt{k}}\Theta + \frac{1}{\sqrt{k}}(U_k - \tilde{V}_k) + o_P(\frac{1}{\sqrt{k}}). \tag{2.16}$$

The test statistics, first expressed in terms of $\tilde{Z}_k$ and then expansion formulae (2.11) applied to them, can be summarized as

$$\sqrt{k}(\tilde{\ell}\tilde{\mathrm{T}}_{\mathrm{G}} - \tilde{h}) = \mathrm{tr}(U_k - \tilde{V}_k) + \mathrm{tr}\Theta + o_P(1).$$

where for Likelihood Ratio criteria $\mathrm{G} = \mathrm{LR}$, $\tilde{\ell} = 2$ and $\tilde{h} = 2p\log 2$; for Lawley-Hotelling criteria $\mathrm{G} = \mathrm{LH}$, $\tilde{\ell} = 1$ and $\tilde{h} = p$; and for Bartlett-Nanda-Pillai criteria $\mathrm{G} = \mathrm{BNP}$, $\tilde{\ell} = 4$ and $\tilde{h} = 2p$.

Theorem 2.2. *Under the assumptions* A1–A5*,*

$$\sqrt{k}(\tilde{\ell}\tilde{T}_{\mathrm{G}} - \tilde{h}) \xrightarrow{\mathcal{L}} N(\mathrm{tr}\Theta, 2p\tau_0^2)$$

as $k \to \infty$ *assuming the limit* $\tau_0^2 = \lim_{k\to\infty} \sum_{i=1}^{k} k^{-1} n_i (n_i - 1)^{-1}$ *exists.*

Proof. As in the proof of Theorem 2.1 here also it suffices to show that

$$\mathrm{tr}(U_k - V_k) = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \mathrm{tr}\{\mathcal{E}_i \frac{1}{n_i - 1}(J_{n_i} - I_{n_i})\mathcal{E}_i'\} \xrightarrow{\mathcal{L}} N(0, 2p\tau_0^2).$$

Since the diagonal entries of $(J_{n_i} - I_{n_i})$ are all 0, the variance of $\mathrm{tr}\{\mathcal{E}_i \frac{1}{n_i-1}(J_{n_i} - I_{n_i})\mathcal{E}_i'\}$ will not depend on the fourth cumulant (see Corollary 1.1). The remaining part of the proof goes exactly along the same lines as that of Theorem 2.1. □

It is clear from Theorem 2.2 that $\tilde{T}_{\mathrm{LR}}$, $\tilde{T}_{\mathrm{LH}}$ and $\tilde{T}_{\mathrm{BNP}}$ are asymptotically robust to non-normality under the null as well as non-null hypotheses. Moreover, they are also asymptotically equivalent.

2.4 Power Comparison

Let $\beta_k^{T_G}(\Omega_k)$ and $\beta_k^{\tilde{T}_G}(\Omega_k)$ be the powers of size α tests based on T_G and $\tilde{T}_G$, respectively. Applying Theorems 2.1 and 2.2, it follows that

$$\lim_{k\to\infty} \left[\beta_k^{T_G}(\Omega_k) - \beta_k^{\tilde{T}_G}(\Omega_k)\right] = \Phi(z_\alpha - \frac{\mathrm{tr}(\Theta)}{\tilde{\sigma}_G}) - \Phi(z_\alpha - \frac{\mathrm{tr}(\Theta)}{\sigma_G})$$

under the local alternatives defined by A5 where $\Phi(\cdot)$ is the CDF of a standard normal variate and, σ_{G}^2 and $\tilde{\sigma}_{\mathrm{G}}^2$ are the asymptotic variances of $\sqrt{k}(\ell T_{\mathrm{G}} - h)$ and $\sqrt{k}(\tilde{\ell}\tilde{T}_{\mathrm{G}} - \tilde{h})$, respectively. Since $\mathrm{tr}(\Theta) \geq 0$ (because Θ is positive semi-definite), T_{G} is more (less) powerful than $\tilde{T}_{\mathrm{G}}$ for large k if and only if $\tilde{\sigma}_{\mathrm{G}}^2 > (<)\sigma_{\mathrm{G}}^2$. Now,

$$\sigma_{\mathrm{G}}^2 - \tilde{\sigma}_{\mathrm{G}}^2 \geq 0 \iff \kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}) \geq \frac{2(\bar{n} - 1)}{(\bar{n}\underline{n} - 1)}(\frac{\tau_0^2}{\bar{n}/(\bar{n} - 1)} - 1)p. \tag{2.17}$$

Notice that the function $g(x) = x/(x-1)$ is convex for $x > 1$. Thus, it follows from Jensen's Inequality (for $n_i > 1$) that,

$$\tau_0^2 = \lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} \frac{n_i}{n_i - 1} \geq \lim_{k\to\infty} \frac{\sum_{i=1}^{k} n_i/k}{\sum_{i=1}^{k} n_i/k - 1} = \frac{\bar{n}}{\bar{n}-1}.$$

Furthermore, by applying the arithmetic-harmonic mean inequality $\bar{n}\underline{n} - 1 \geq 0$. Consequently,

$$\frac{2(\bar{n}-1)}{(\bar{n}\underline{n}-1)}\left(\frac{\tau_0^2}{\bar{n}/(\bar{n}-1)} - 1\right)p \geq 0.$$

Therefore, for distributions with tails lighter than the normal distribution T_{G} is more powerful than $\tilde{T}_{\mathrm{G}}$. Under normality, T_{G} has higher power than $\tilde{T}_{\mathrm{G}}$. Furthermore, for "slightly" heavy tailed distributions, T_G is more powerful than $\tilde{T}_{\mathrm{G}}$ if inequality (2.17) is met. However, for distributions with substantial heaviness in the tails (compared to the the normal distribution), $\tilde{T}_{\mathrm{G}}$'s power excels that of T_{G}. This can happen, for example, if $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11})$ grows with p at a fast rate. For instance, for elliptically contoured populations $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}) = \kappa p(p+2)$ where κ is the kurtosis parameter (see Section 5). For larger value of p or κ, $\tilde{T}_{\mathrm{G}}$ can be more powerful than T_{G}.

3 Profile Analysis

Assume model (1.1) holds. Suppose the p observations Y_{ij} represent p measurements taken on the jth subject in the ith group under each of the p different experimental conditions (time points). The three hypotheses of interest in profile analysis are (i) whether or not the mean profiles for the k groups are parallel (similar) (ii) Given that the k mean profiles are parallel, are they at the same level (coincident)? (iii) Given that the mean profiles are at the same level, are there differences among the experimental conditions or are the k profiles flat?

Let $\boldsymbol{\mu}_i = (\mu_i^{(1)}, \mu_i^{(2)}, \ldots, \mu_i^{(p)})'$ and $B = (\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_k)$. Further, let $C = (I_{k-1}, -\mathbf{1}_{k-1})'$ and M be $m \times p$ known matrix of rank $m \leq p$. Consider the general linear hypothesis regarding B formulated as

$$\mathrm{H}_0 : MBC = 0. \tag{3.1}$$

This testing problem can be equivalently formulated by modifying the model and the hypothesis as

$$\mathbf{y}_{ij}^* = \boldsymbol{\mu}_i^* + \Sigma^{*1/2}\boldsymbol{\varepsilon}_{ij}^*$$

and

$$H_0 : \boldsymbol{\mu}_1^* = \boldsymbol{\mu}_2^* = \ldots = \boldsymbol{\mu}_k^*, \tag{3.2}$$

where $\mathbf{y}_{ij}^* = M\mathbf{y}_{ij}$, $\boldsymbol{\mu}_i^* = M\boldsymbol{\mu}_i$, $\Sigma^* = M\Sigma M'$ and $\boldsymbol{\varepsilon}_{ij}^* = (M\Sigma M')^{-1/2}M\Sigma^{1/2}\boldsymbol{\varepsilon}_{ij}$. Notice that $E(\boldsymbol{\varepsilon}_{ij}^*) = \mathbf{0}$, $\mathrm{var}(\boldsymbol{\varepsilon}_{ij}^*) = I_m$ and, using the identity $\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}(B)$,

$$K_4(\boldsymbol{\varepsilon}_{ij}^*) = \{((M\Sigma M')^{-1/2}M\Sigma^{1/2}) \otimes ((M\Sigma M')^{-1/2}M\Sigma^{1/2})\}K_4(\boldsymbol{\varepsilon}_{ij}) \times \{(\Sigma^{1/2}M'(M\Sigma M')^{-1/2}) \otimes (\Sigma^{1/2}M'(M\Sigma M')^{-1/2})\}. \tag{3.3}$$

Notice also that $\kappa_4^{(1)}(\boldsymbol{\varepsilon}^*)$ can be calculated using the relation,

$$\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{ij}^*) = E\{\boldsymbol{\varepsilon}_{ij}'\Sigma^{1/2}M'(M\Sigma M')^{-1}M\Sigma^{1/2}\boldsymbol{\varepsilon}_{ij}\}^2 - 2m - m^2$$

which can easily be verified by a straightforward noting that $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{ij}^*) = \mathrm{vec}(I_m)'K_4(\boldsymbol{\varepsilon}_{ij}^*)\mathrm{vec}(I_m)$ and using (3.3). Also, it can easily be shown that the non-centrality matrix in this formulation is related to the non-centrality matrix in the original formulation as

$$\Omega_k^* = (M\Sigma M')^{-1/2}M\Sigma^{1/2}\Omega_k\Sigma^{1/2}M'(M\Sigma M')^{-1/2}.$$

Further, assumption A4 implies that $\Omega_k = \sqrt{k}\Theta^*$ where

$$\Theta^* = (M\Sigma M')^{-1/2}M\Sigma^{1/2}\Theta\Sigma^{1/2}M'(M\Sigma M')^{-1/2}.$$

Therefore, the testing problem (3.1) under the model (1.1) reduces to the one discussed in Section 2. That is, the test statistics T_{G} and $\tilde{T}_{\mathrm{G}}$ in (2.1) and (2.15) for $\mathrm{G} = \mathrm{LR}, \mathrm{LH}$ and BNP can be applied on $M\mathbf{y}_{ij}$. We denote these test statistics by T_{G}^* and $\tilde{T}_{\mathrm{G}}^*$ for $\mathrm{G} = \mathrm{LR}, \mathrm{LH}$ and BNP; their distributions are given in Theorems 3.1 and 3.2 below.

Theorem 3.1. *Under assumptions* A1–A4*,*

$$\sqrt{k}(\ell^* T_{\mathrm{G}}^* - h^*) \xrightarrow{\mathcal{L}} N\left(\mathrm{tr}\Theta^*, \frac{2\bar{n}m}{\bar{n}-1} + \frac{\bar{n}(\bar{n}\underline{n}-1)}{(\bar{n}-1)^2}\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}^*)\right)$$

where for Likelihood Ratio criterion $\ell^* = \bar{n}$ and $h^* = \bar{n}m\log(\bar{n}/(\bar{n}-1))$; for Lawley-Hotelling criterion $\ell^* = \bar{n}-1$ and $h^* = m$; and for Bartlett-Nanda-Pillai criterion $\ell^* = \bar{n}^2/(\bar{n}-1)$ and $h^* = \bar{n}m/(\bar{n}-1)$.

It is clear from Theorem 3.1 that in the balanced case both the null and non-null distributions are free from $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}^*)$. One also notes that in the balanced case, the asymptotic null distributions depend on M only through its rank m.

In the unbalanced case one can use a consistent estimator of $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}^*)$ (see equation (2.14)) given by,

$$\hat{\kappa}_4^{(1)}(\boldsymbol{\varepsilon}_{11}^*) = \frac{N+1}{N(N-1)}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left\{(\mathbf{y}_{ij}-\bar{\mathbf{y}}_{..})'M'\hat{\Sigma}_M^{-1}M(\mathbf{y}_{ij}-\bar{\mathbf{y}}_{..})\right\}^2 - m(m+2)$$

where $\hat{\Sigma}_M = N^{-1}\sum_{=1}^{k}\sum_{j=1}^{n_i}M(\mathbf{y}_{ij}-\bar{\mathbf{y}}_{..})(\mathbf{y}_{ij}-\bar{\mathbf{y}}_{..})'M'$ to apply the result of Theorem 3.1. One can also choose to use the modified statistics $\tilde{T}_{\mathrm{G}}^*$ whose distributions are given in Theorem 3.2 below.

Theorem 3.2. *Under the assumptions* A1–A5 *and assuming the limit $\tau_0^2 = \lim_{k\to\infty} k^{-1}\sum_{i=1}^{k} n_i(n_i-1)^{-1}$ exists,*

$$\sqrt{k}(\tilde{\ell}^*\tilde{T}_{\mathrm{G}}^* - \tilde{h}^*) \xrightarrow{\mathcal{L}} N(\mathrm{tr}\Theta^*, 2m\tau_0^2)$$

where for Likelihood Ratio criterion $\tilde{\ell}^* = 2$ and $\tilde{h}^* = 2m\log 2$; for Lawley-Hotelling criterion $\tilde{\ell}^* = 1$ and $\tilde{h}^* = m$; and for Bartlett-Nanda-Pillai criterion $\tilde{\ell}^* = 4$ and $\tilde{h}^* = 2m$.

Let L be a $(p-1)\times p$ matrix such that $L\mathbf{1}_p = \mathbf{0}$. The three hypotheses in profile analysis can be expressed in terms of the group mean vectors $\boldsymbol{\mu}_i$ as (see Rencher, 2002),

$$\mathrm{H}_{0\mathrm{P}} : L\boldsymbol{\mu}_1 = L\boldsymbol{\mu}_2 = \cdots = L\boldsymbol{\mu}_k$$

$$\mathrm{H}_{0\mathrm{C}} : \mathbf{1}_p'\boldsymbol{\mu}_1 = \mathbf{1}_p'\boldsymbol{\mu}_2 = \cdots = \mathbf{1}_p'\boldsymbol{\mu}_k \quad \text{and}$$

$$\mathrm{H}_{0\mathrm{F}} : L\bar{\boldsymbol{\mu}} = \mathbf{0},$$

respectively. Now, since $\mathrm{H}_{0\mathrm{P}}$ and $\mathrm{H}_{0\mathrm{C}}$ are special cases of the hypothesis (3.1), their asymptotic results can be accommodated in Theorems 3.1 and 3.2 by choosing $M = L$ and $M = \mathbf{1}_p$, respectively.

For testing the hypothesis $\mathrm{H}_{0\mathrm{P}}$, it is clear that the test statistics T^*_{G}, for $\mathrm{G} = \mathrm{LR}, \mathrm{LH}, \mathrm{BNP}$, depend on the data only through the eigenvalues of $LHL'(LEL')^{-1}$ which are the same as the nonzero eigenvalues of $HL'(LEL')^{-1}L$. Let L_1 and L_2 be two $(p-1)\times p$ matrices of rank $p-1$ whose rows are orthogonal to $\mathbf{1}_p$. Then we must have that $L_1 = BL_2$ for some non-singular matrix B. Consequently, $HL_1'(L_1EL_1')^{-1}L_1 = HL_2'(L_2EL_2')^{-1}L_2$. Therefore the distributions of T_{G}'s are invariant to the choice of L. Similarly, the distributions of $\tilde{T}^*_{\mathrm{G}}$ are invariant to the choice of L.

A reduction in expression can be obtained in the case of testing $\mathrm{H}_{0\mathrm{C}}$. Indeed in this case the three test statistics are equivalent because $M\mathbf{y}_{ij}$ is univariate. Hence it suffices to consider the Loweley-Hotelling criterion. Let $\boldsymbol{\alpha}_i = (\alpha_i^{(1)}, \alpha_i^{(2)}, \cdots, \alpha_i^{(p)})' = \boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}$, $\Sigma = (\sigma_{lj})$ and $\theta = \lim_{k\to\infty}\left(\sum_{i=1}^{k}(\sum_{j=1}^{p}\alpha_i^{(j)})^2 \big/ (\sqrt{k}\sum_{l,j}^{p}\sigma_{lj})\right)$ and $\varepsilon_{11} = \mathbf{1}_p'\Sigma^{1/2}\boldsymbol{\varepsilon}_{11}$.

Corollary 3.1. *Under the assumptions $A1$–$A5$, the asymptotic distribution of $\sqrt{k}((\bar{n}-1)T^*_{\mathrm{LH}}-1)$ when $M = \mathbf{1}_p'$ is $N(\theta, \sigma^2)$ where*

$$\sigma^2 = \frac{2\bar{n}}{\bar{n}-1} + \frac{\bar{n}(\bar{n}\underline{n}-1)}{(\bar{n}-1)^2}(\xi_4/(\sum_{l,j}^{p}\sigma_{lj})^2 - 3)$$

and ξ_4 is the fourth moment of $\mathbf{1}_p'\Sigma^{1/2}\boldsymbol{\varepsilon}_{11}$.

Corollary 3.2. *Under the assumptions $A1$–$A5$ and assuming $\tau_0^2 = \lim_{k\to\infty} k^{-1}\sum_{i=1}^{k} n_i(n_i-1)^{-1}$ exists, the asymptotic distribution of $\sqrt{k}(\tilde{T}^*_{\mathrm{LH}}-1)$ when $M = \mathbf{1}_p'$ is $N(\theta, 2\tau_0^2)$.*

Given that the profiles are parallel and coincident (i.e $\mathrm{H}_{0\mathrm{P}}$ and $\mathrm{H}_{0\mathrm{C}}$ are true), the Hotelling T^2 test statistic

$$T^2 = N(L\bar{\mathbf{y}}_{..})'(\frac{1}{N-k}LEL')^{-1}(L\bar{\mathbf{y}}_{..})$$

is used for testing $\mathrm{H}_{0\mathrm{F}}$. Assume $L\bar{\boldsymbol{\mu}} = N^{-1/2}\boldsymbol{\delta}$ for some fixed $\boldsymbol{\delta} \in \mathbb{R}^p$. Then, because $k \to \infty$ implies $N \to \infty$, $\sqrt{N}L\bar{\mathbf{y}}_{..} \xrightarrow{\mathcal{L}} N_{p-1}(\boldsymbol{\delta}, L\Sigma L')$ and $(N-k)^{-1}LEL' \xrightarrow{P} L\Sigma L'$, it holds that $T^2 \xrightarrow{\mathcal{L}} \chi^2_{(p-1)}(\boldsymbol{\delta}'\boldsymbol{\delta})$. Further, $T^2 \xrightarrow{\mathcal{L}} \chi^2_{(p-1)}$ under the null hypothesis because $\boldsymbol{\delta} = \mathbf{0}$ in that case. The statistic T^2 is invariant to non-normality as $k \to \infty$ under the null as well as non-null cases.

4 Tests for Additional Information

Assume model (1.1) holds. Let us partition $\mathbf{y}_{ij}$ as $(\mathbf{y}_{ij}^{(1)\prime}, \mathbf{y}_{ij}^{(2)\prime})'$ where $\mathbf{y}_{ij}^{(1)}$ is $q \times 1$ and $\mathbf{y}_{ij}^{(2)}$ is a $(p-q) \times 1$, respectively, vectors. We would like to determine whether the observations $\mathbf{y}_{ij}^{(2)}$ make a significant contribution towards the test of $\mathrm{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_k$ above and beyond the $\mathbf{y}_{ij}^{(1)}$'s. Such a testing problem is important in particular if it is costly to observe the $\mathbf{y}_{ij}^{(2)}$'s. For example, in comparing k different treatments for a certain type of cancer, it would be of great interest to known if DNA gene expression information provides any additional information in separating the treatments over and above what is provided by phenotypic information because DNA information is very expensive to collect.

Partition $\boldsymbol{\mu}_i$, Σ, H and E accordingly as follows

$$\boldsymbol{\mu}_i = \begin{pmatrix} \boldsymbol{\mu}_i^{(1)} \\ \boldsymbol{\mu}_i^{(2)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \quad \text{and} \quad E = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}.$$

The hypothesis of additional information can be formulated as

$$\mathrm{H}_{0\mathrm{A}} : \boldsymbol{\mu}_1^{(1)} = \boldsymbol{\mu}_2^{(1)} = \ldots = \boldsymbol{\mu}_k^{(1)} \quad \text{versus} \quad \mathrm{H}_{1\mathrm{A}} : \text{Not } \mathrm{H}_{0\mathrm{A}}$$

given that $\boldsymbol{\mu}_1^{(2)} = \boldsymbol{\mu}_2^{(2)} = \ldots = \boldsymbol{\mu}_k^{(2)}$.

Let $\Sigma = AA'$ denote the Cholesky decomposition of Σ, i.e A is a lower triangular matrix with positive diagonal elements, and partition A in a similar way. Notice that $A_{12} = 0$ and A_{11} and A_{22} are lower diagonal matrices of dimension $q \times q$ and $(p-q) \times (p-q)$, respectively. Then model (1.1) can be written as, $\mathbf{y}_{ij}^{(1)} = \boldsymbol{\mu}_i^{(1)} + A_{11}\boldsymbol{\varepsilon}_{ij}^{(1)}$ and $\mathbf{y}_{ij}^{(2)} = \boldsymbol{\mu}_i^{(2)} + \Xi^{1/2}\boldsymbol{\epsilon}_{ij}$ where $\Xi = (A_{12}A_{12}' + A_{22}A_{22}')$ and $\boldsymbol{\epsilon}_{ij} = \Xi^{-1/2}(A_{21}\boldsymbol{\varepsilon}_{ij}^{(1)} + A_{22}\boldsymbol{\varepsilon}_{ij}^{(2)})$. Obviously, $E(\boldsymbol{\epsilon}_{ij}) = 0$ and $\mathrm{var}(\boldsymbol{\epsilon}_{ij}) = I_{p-q}$. Moreover, $\Omega_{11,k} = \sum_{i=1}^{k} n_i A_{11}^{-1}(\boldsymbol{\mu}_i^{(1)} - \bar{\boldsymbol{\mu}}^{(1)})(\boldsymbol{\mu}_i^{(1)} - \bar{\boldsymbol{\mu}}^{(1)})' A_{11}'^{-1}$ and $\Omega_{11,k} = \sqrt{k}\Theta_{11}$ where $\Omega_{11,k}$ and Θ_{11} are the $q \times q$ sub-matrices in the top-left of Ω_k and Θ, respectively.

Let $T = H + E$ and define $H_{11.2} = H_{11} - H_{12}H_{22}^{-1}H_{21}$. Also define $E_{11.2}$ and $T_{11.2}$ similarly. Under normality the likelihood ratio criterion (Fujikoshi, 1981) is given by

$$T_{\mathrm{LR}}^{(11.2)} = -\log \frac{|E_{11.2}|}{|T_{11.2}|}.$$

In a manner analogous to (2.1), Gupta, Xu and Fujikoshi (2006) defined the Lawley-Hotelling and Bartlett-Nanda-Pillai statistics as

$$T_{\mathrm{LH}}^{(11.2)} = \mathrm{tr}(T_{11.2} - E_{11.2})E_{11.2}^{-1} \quad \text{and} \quad T_{\mathrm{BNP}}^{(11.2)} = \mathrm{tr}(T_{11.2} - E_{11.2})T_{11.2}^{-1}.$$

Fujikoshi (1981) explains how to obtain the asymptotic expansion for the null as well as non-null distributions of $T_{\mathrm{LR}}^{(11.2)}$ using existing results for multivariate linear models found, for example, in Anderson (1984, pp. 327-8). We note that the same methods can be applied to obtain the asymptotic expansions of the null and non-null distributions of $T_{\mathrm{LH}}^{(11.2)}$ and $T_{\mathrm{BNP}}^{(11.2)}$ under normality. Gupta, Xu and Fujikoshi (2006) derived the asymptotic expansion for the null distributions of these statistics under

non-normality. In both Fujikoshi (1981) and Gupta, Xu and Fujikoshi (2006), the asymptotic framework required all group sample sizes tend to infinity at the same rate but k remained fixed. In this section we derive the asymptotic null as well as non-null distributions of the three statistics in the large k asymptotic framework under non-normality.

From (2.6) it follows that

$$\sqrt{k}(\frac{1}{k}H - \frac{1}{N-k}E) = \Sigma^{1/2}(U_k - \frac{1}{N/k-1}V_k)\Sigma^{1/2} + \Sigma^{1/2}\Theta\Sigma^{1/2} + O_P(\frac{1}{\sqrt{k}}).$$

Because $(U_k - (N/k-1)^{-1}V_k) = O_P(1)$ it follows that $\sqrt{k}(k^{-1}H - (N-k)^{-1}E) = O_P(1)$ which also implies that $\sqrt{k}(k^{-1}H_{ij} - (N-k)^{-1}E_{ij}) = O_P(1)$ for $i, j = 1, 2$. Moreover, $(N-k)^{-1}E = \Sigma + o_P(1)$ implies $(N-k)^{-1}E_{ij} = \Sigma_{ij} + o_P(1)$ for $i, j = 1, 2$. Now we are ready to prove the following Proposition.

Proposition 4.1. *As $k \to \infty$,*

$$\sqrt{k}\left((T_{11.2} - E_{11.2})E_{11.2}^{-1} - H_{11}E_{11}^{-1}\right) = o_P(1).$$

Proof. Notice that

$$\frac{1}{k}(T_{11.2} - E_{11.2}) = \frac{1}{k}H_{11} - \frac{1}{k}(T_{12}T_{22}^{-1}T_{21} - E_{12}E_{22}^{-1}E_{21}) \text{ and } T_{12}T_{22}^{-1}T_{21} = (\frac{1}{\sqrt{k}}T_{12})(\frac{1}{k}T_{22})^{-1}(\frac{1}{\sqrt{k}}T_{21}).$$

Then,

$$\frac{1}{k}T_{22} = \frac{N/k}{N-k}E_{22} + \frac{1}{\sqrt{k}}\left\{\sqrt{k}(\frac{1}{k}H_{22} - \frac{1}{N-k}E_{22})\right\} = \bar{n}\Sigma_{22} + o_P(1)$$

and, hence, $\{(1/k)T_{22}\}^{-1} = O_P(1)$. On the other hand,

$$\frac{1}{\sqrt{k}}T_{12} = \frac{N/\sqrt{k}}{N-k}E_{12} + \sqrt{k}(\frac{1}{k}H_{12} - \frac{1}{N-k}E_{12}) = O_P(1) + O_P(1) = O_P(1),$$

and, hence, $k^{-1/2}T_{21} = k^{-1/2}T_{12}' = O_P(1)$. Therefore, $T_{12}T_{22}^{-1}T_{21} = O_P(1)$. Similar arguments can be used to establish $E_{12}E_{22}^{-1}E_{21} = O_P(1)$. Consequently, by noting that $(1/k)E_{11} = O_P(1)$, we obtain

$$\frac{1}{k}E_{11.2} = \frac{1}{k}E_{11} - \frac{1}{k}E_{12}E_{22}^{-1}E_{21} = \frac{1}{k}E_{11} + O_P(\frac{1}{k})$$

and $k^{-1}(T_{11.2} - E_{11.2}) = o_P(k^{-1/2})$.

Thus,

$$(\frac{1}{k}E_{11.2})^{-1} = (\frac{1}{k}E_{11})^{-1} + O_P(\frac{1}{k}).$$

Finally,

$$\begin{aligned}\sqrt{k}\left((T_{11.2} - E_{11.2})E_{11.2}^{-1} - H_{11}E_{11}^{-1}\right) &= \sqrt{k}\left(\frac{1}{k}(T_{11.2} - E_{11.2})(\frac{1}{k}E_{11.2})^{-1} - \frac{1}{k}H_{11}(\frac{1}{k}E_{11})^{-1}\right) \\ &= o_P(1).\end{aligned}$$

□

The upshot of Proposition 4.1 is that $\sqrt{k}\{(T_{11.2} - E_{11.2})E_{11.2}^{-1} - I_q\}$ and $\sqrt{k}(H_{11}E_{11}^{-1} - I_q)$ have the same asymptotic distribution. By virtue of the fact that $\mathbf{y}_{ij}^{(1)} = \boldsymbol{\mu}_i^{(1)} + A_{11}\boldsymbol{\varepsilon}_{ij}^{(1)}$ and Theorem 2.1 the following Theorem is true.

Theorem 4.1. *Under assumptions* A1–A4*,*

$$\sqrt{k}(\ell T_{\mathrm{G}}^{(11.2)} - h) \xrightarrow{\mathcal{L}} N\left(\mathrm{tr}\Theta_{11}, \frac{2\bar{n}q}{\bar{n}-1} + \frac{\bar{n}(\bar{n}\underline{n}-1)}{(\bar{n}-1)^2}\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}^{(1)})\right)$$

as $k \to \infty$ *where* $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}^{(1)}) = (\mu_4(\boldsymbol{\varepsilon}_{11}^{(1)}) - 2q - q^2)$ *and* $\mu_4(\boldsymbol{\varepsilon}_{11}^{(1)}) = \sum_{a,b}^{q} E(\varepsilon_{11a}^2\varepsilon_{11b}^2)$.

In the above Theorem, the values assumed by G, h and ℓ are as follows. For the Likelihood Ratio criterion $\mathrm{G} = \mathrm{LR}$ $\ell = \bar{n}$ and $h = \bar{n}q\log(\bar{n}/(\bar{n}-1))$; for the Lawley-Hotelling criterion $\mathrm{G} = \mathrm{LH}$, $\ell = \bar{n} - 1$ and $h = p$; and for the Bartlett-Nanda-Pillai criterion $\mathrm{G} = \mathrm{BNP}$, $\ell = \bar{n}^2/(\bar{n}-1)$ and $h = \bar{n}q/(\bar{n}-1)$.

We can easily obtain a consistent estimator for $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}^{(1)})$ from (2.14). Furthermore, robust tests can be obtained by defining $\tilde{T} = H + \tilde{E}$ and the three tests $\tilde{T}_{\mathrm{G}}^{(11.2)}$ for $\mathrm{G} = \mathrm{LR}, \mathrm{LH}, \mathrm{BNP}$ defined in terms $\tilde{T}$ and $\tilde{E}$ in the obvious way. Similar asymptotic equivalence result as in Proposition 4.1 holds in this case as well.

Proposition 4.2. *As* $k \to \infty$*,*

$$\sqrt{k}\left((\tilde{T}_{11.2} - \tilde{E}_{11.2})\tilde{E}_{11.2}^{-1} - H_{11}\tilde{E}_{11}^{-1}\right) = o_P(1).$$

Finally the asymptotic distribution of $\tilde{T}_{\mathrm{G}}^{(11.2)}$ is given in the following Theorem.

Theorem 4.2. *Under the assumptions* A1–A5*,*

$$\sqrt{k}(\tilde{\ell}\tilde{T}_{\mathrm{G}}^{(11.2)} - \tilde{h}) \xrightarrow{\mathcal{L}} N(\mathrm{tr}\Theta_{11}, 2q\tau_0^2)$$

as $k \to \infty$ *assuming* $\tau_0^2 = \lim_{k\to\infty}\sum_{i=1}^{k} k^{-1}n_i(n_i-1)^{-1}$ *exists where for Likelihood Ratio criterion* $\mathrm{G} = \mathrm{LR}$*,* $\tilde{\ell} = 2$ *and* $\tilde{h} = 2q\log 2$*; for Lawley-Hotelling criterion* $\mathrm{G} = \mathrm{LH}$*,* $\tilde{\ell} = 1$ *and* $\tilde{h} = q$*; and for Bartlett-Nanda-Pillai criterion* $\mathrm{G} = \mathrm{BNP}$*,* $\tilde{\ell} = 4$ *and* $\tilde{h} = 2q$*.*

5 Tests of Dimensionality

Consider model (1.1) and assume,

$$\mathrm{A6} : \boldsymbol{\varepsilon}_{11} \sim (-2\phi'(0))^{-1/2}\mathrm{EL_p}(0, I_p, \phi)$$

and

$$\mathrm{A7} : \Omega_k = k\Theta.$$

The notation $\mathrm{EL_p}(0, I_p, \phi)$ stands for the p-variate spherical distribution with the characteristic function generator $\phi : [0, \infty) \to [0, \infty)$ (see Fang and Zhang, 1990). It can be shown that (see Gang, 1990) $E(\boldsymbol{\varepsilon}_{11}) = \mathbf{0}$, $\mathrm{var}(\boldsymbol{\varepsilon}_{11}) = I_p$ and $K_4(\boldsymbol{\varepsilon}_{11}) = \kappa_4(I_{p^2} + K_{p,p} + \mathrm{vec}(I_p)\mathrm{vec}(I_p)')$ where κ_4 is known as the kurtosis parameter defined by $\kappa_4 = \phi''(0)/(\phi'(0))^2 - 1$ where ϕ' and ϕ'' are the first and second derivatives of ϕ, respectively.

Suppose the hypothesis H_0 in Section 2 is rejected. We would like to know the dimension of the hyperplane that contains the mean vectors $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, ..., $\boldsymbol{\mu}_k$. This number, sometimes referred to as the *dimensionality*, is also the number of discriminant functions needed to separate the k populations. Since k is presumed to be large in this paper, we assume $k > p$.

Various methods are available for estimating the dimensionality (Backhouse and McKay, 1982). The most common ones involve a sequence of Hypothesis tests. Let $\omega_1 \geq \omega_2 \geq \ldots \geq \omega_p \geq 0$ be the eigenvalues of Ω_k. The dimension of the hyperplane is the largest r for which the hypothesis

$$H_{0r} : \omega_r > \omega_{r+1} = \ldots = \omega_p = 0$$

is not rejected where the tests are conducted sequentially for $r = 1, 2, \ldots, p-1$. Let $d_1 > d_2 > \ldots > d_p > 0$ be the eigenvalues of HE^{-1}. The three common normal theory test statistics for testing H_{0r} are the likelihood ratio ($T_{\mathrm{LR}}^{(r)}$), Lawley-Hotteling ($T_{\mathrm{LH}}^{(r)}$) and Bartlett-Nanda-Pillai statistics ($T_{\mathrm{BNP}}^{(r)}$) defined by

$$T_{\mathrm{LR}}^{(r)} = -\sum_{\alpha=r+1}^{p} \log(1 + d_\alpha), \quad T_{\mathrm{LH}}^{(r)} = \sum_{\alpha=r+1}^{p} d_\alpha \quad \text{and} \quad T_{\mathrm{BNP}}^{(r)} = \sum_{\alpha=r+1}^{p} d_\alpha/(1 + d_\alpha), \tag{5.1}$$

respectively.

These tests have asymptotic χ^2 distributions under normality when all the sample sizes n_i tend to infinity at the same rate. The small sample distributions of the test statistics depend on the nuisance parameters $\omega_1, \omega_2, \ldots, \omega_r$. Schott (1984) obtained optimal upper bound for the null distributions of $T_{\mathrm{LR}}^{(r)}$ and $T_{\mathrm{LH}}^{(r)}$ which can be used to obtain critical values. Seo, Kanda and Fujikoshi (1995) obtained non-normality correction factors for the test statistics under the assumption $A6$ by matching the moments of the null-distributions with that of the normal theory limiting chi-square distribution up to the order $O(N^{-1})$. The optimal bounds of Schott (1984) have been extended by Yoshida, Imai and Sato (2002, 2004) to the situation where the samples are coming from a matrix elliptical distribution (see Fang and Zhang, 1990, Chapter 5). However, this matrix distribution does not allow to model independent sampling except in the special case of normality. In this paper, we derive the large k asymptotic distributions of the three statistics under independent sampling from elliptically contoured distributions.

Calculations similar to those of Section 2 yield

$$\frac{1}{k}\Sigma^{-1/2}H\Sigma^{-1/2} = I_p + \Theta + \frac{1}{\sqrt{k}}U_k + o_P(\frac{1}{\sqrt{k}}) \quad \text{and} \quad \frac{1}{k}\Sigma^{-1/2}E\Sigma^{-1/2} = (\bar{n} - 1)I_p + \frac{1}{\sqrt{k}}V_k. \tag{5.2}$$

Notice that the eigenvalues of HE^{-1} are the same as those of

$$Z_k = (\frac{1}{k}\Sigma^{-1/2}E\Sigma^{-1/2})^{-1/2}(\frac{1}{k}\Sigma^{-1/2}H\Sigma^{-1/2})(\frac{1}{k}\Sigma^{-1/2}E\Sigma^{-1/2})^{-1/2}.$$

Substituting (5.2), expanding the resulting equation in the same way as (2.9) and performing some algebraic manipulation leads to

$$(\bar{n}-1)Z_k = I_p + \Theta + \frac{1}{\sqrt{k}}\left\{(U_k - \frac{1}{(\bar{n}-1)}V_k) - \frac{1}{2}\frac{1}{(\bar{n}-1)}(\Theta V_k + V_k\Theta)\right\} + o_P(\frac{1}{\sqrt{k}}). \tag{5.3}$$

The distribution of ε_{ij}'s is invariant under orthogonal transformation. Therefore, the distributions of U_k and V_k are invariant to pre- or post- multiplication by an orthogonal matrix. Hence, without loss of generality we can assume Ω_k and Θ are diagonal matrices. In particular, we can set $\Theta = \mathrm{diag}\{\theta_1, \theta_2, \ldots, \theta_p\}$. It is for this purpose that we mainly need assumption $A6$ in this paper.

The rest of the derivations in this section require the following Lemma on perturbation expansion for eigenvalues.

Lemma 5.1 (Fujikoshi, 1977)**.** *Suppose the $p \times p$ symmetric random matrix Z can be expressed as,*

$$Z = \Delta + \frac{1}{\sqrt{m}}V$$

where $\Delta = \mathrm{diag}(\delta_1, \delta_2, \ldots, \delta_p)$ and V is a $p \times p$ random matrix such that $V = O_P(1)$ as $m \to \infty$. Assume $\delta_1 = \delta_2 = \ldots = \delta_{q_1} = \lambda_1$, $\delta_{q_1+1} = \delta_{q_1+2} = \ldots = \delta_{q_2} = \lambda_2$, $\cdots$, $\delta_{p-q_t+1} = \delta_{p-q_t+2} = \ldots = \delta_p = \lambda_t$ where $\lambda_1 > \lambda_2 > \ldots > \lambda_t > 0$ and $q_1, q_2, \ldots, q_t$ are positive integers such that $q_1 + q_2 + \cdots + q_t = p$. Then the $(q_1 + q_2 + \ldots + q_\alpha + j)$th largest eigenvalue of Z can be expressed as the jth largest eigenvalue of

$$W = \lambda_\alpha I_{q_\alpha} + \frac{1}{\sqrt{m}}V_{\alpha\alpha} + \frac{1}{m}\sum_{\alpha\neq\beta}\lambda_{\alpha\beta}V_{\alpha\beta}V_{\beta\alpha} + O(m^{-3/2})$$

for any $j \in \{1, 2, \ldots, q_\alpha\}$ where $\lambda_{\alpha\beta} = (\lambda_\alpha - \lambda_\beta)^{-1}$ and the matrices $V_{\alpha\beta}$ are the (α, β)th block of V partitioned into $q_1, q_2, \ldots, q_t$ rows and columns.

In Theorem 5.1 below we present the asymptotic distribution of the eigenvalues of HE^{-1} in the case where the eigenvalues of Θ are simple. The case with multiplicity of the eigenvalues is far more complicated.

Theorem 5.1. *Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$ where $\theta_1 > \theta_2 > \ldots > \theta_p > 0$ be the nonzero eigenvalues values of Θ. Let $\boldsymbol{\ell} = (l_1, l_2, \ldots, l_p)'$ where $l_1 > l_2 > \ldots > l_p$ are the eigenvalues of $((\bar{n}-1)Z - I_p)$. Then, under the assumptions $A1$–$A4$, $A6$ and $A7$,*

$$\sqrt{k}(\boldsymbol{\ell} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N_p(\mathbf{0}, \Gamma)$$

where $\Gamma = (\gamma_{\alpha\beta})$ for $\alpha, \beta = 1, 2, \ldots, p$,

$$\gamma_{\alpha\alpha} = 2\left(1 + (1+\theta_\alpha)^2\frac{1-\underline{n}}{(\bar{n}-1)^2} + (1+\theta_\alpha)\frac{\bar{n}-2+\underline{n}}{\bar{n}-1}\right) + 3\kappa_4\left(1 + (1+\theta_\alpha)^2\frac{\bar{n}-2+\underline{n}}{(\bar{n}-1)^2} + (1+\theta_\alpha)\frac{1-\underline{n}}{\bar{n}-1}\right),$$

and

$$\gamma_{\alpha\beta} = \kappa_4\left(\underline{n} - (2+\theta_\alpha+\theta_\beta) + (1+\theta_\alpha)(1+\theta_\beta)\frac{\bar{n}-2+\underline{n}}{(\bar{n}-1)^2}\right) \quad \textit{for } \alpha \neq \beta.$$

Proof. In view of Lemma 5.1, we get the expansion for the αth largest eigenvalue of $((\bar{n}-1)Z - I_p)$ as

$$l_\alpha = \theta_\alpha + \frac{1}{\sqrt{k}}(u_{\alpha\alpha} - \frac{1+\theta_\alpha}{\bar{n}-1}v_{\alpha\alpha}) + o_P(\frac{1}{\sqrt{k}})$$

where $u_{\alpha\alpha}$ and $v_{\alpha\alpha}$ are the (α,α)th entries of U_k and V_k, respectively. In terms of the error vectors $\boldsymbol{\varepsilon}_{ij}$'s, we have

$$l_\alpha = \theta_\alpha + \frac{1}{\sqrt{k}}Q^{(\alpha)} + o_P(\frac{1}{\sqrt{k}})$$

where

$$Q^{(\alpha)} = \frac{1}{\sqrt{k}}\sum_{i=1}^{k} Q_i^{(\alpha)}, \quad Q_i^{(\alpha)} = \mathcal{E}_i^{(\alpha)}(\frac{1}{n_i}J_{n_i} - \frac{1+\theta_\alpha}{\bar{n}-1}P_{n_i})\mathcal{E}_i^{(\alpha)\prime} - \theta_\alpha$$

and $\mathcal{E}_i^{(\alpha)}$ is the αth row of $\mathcal{E}_i$. Put $\mathbf{Q} = (Q^{(1)}, Q^{(2)}, \ldots, Q^{(p)})'$. Direct but lengthy calculations show that,

$$E(Q_i^{(\alpha)}) = \mathrm{tr}(\frac{1}{n_i}J_{n_i} - \frac{1+\theta_\alpha}{\bar{n}-1}P_{n_i}) - \theta_\alpha = 1 - \frac{1+\theta_\alpha}{\bar{n}-1}(n_i-1) - \theta_\alpha \tag{5.4}$$

$$\begin{aligned}\mathrm{Var}(Q_i^{(\alpha)}) = 2\left(1 + (1+\theta_\alpha)^2\frac{n_i-1}{n_i(\bar{n}-1)^2} + (1+\theta_\alpha)\frac{(n_i-1)^2}{n_i(\bar{n}-1)}\right)\\ + 3\kappa_4\left(1 + (1+\theta_\alpha)^2\frac{(n_i-1)^2}{n_i(\bar{n}-1)^2} + (1+\theta_\alpha)\frac{n_i-1}{n_i(\bar{n}-1)}\right)\end{aligned} \tag{5.5}$$

and

$$\mathrm{Cov}(Q_i^{(\alpha)}, Q_i^{(\beta)}) = \kappa_4\left(\frac{1}{n_i} - (2+\theta_\alpha+\theta_\beta)\frac{n_i-1}{\bar{n}-1} + (1+\theta_\alpha)(1+\theta_\beta)\frac{(n_i-1)^2}{n_i(\bar{n}-1)^2}\right). \tag{5.6}$$

Along the same lines as in the proof of Theorem 2.1, we can verify Lyapunov's condition for $\mathbf{t}'\mathbf{Q}$ for $\mathbf{t} \in \mathbb{R}^p$. Then, averaging each of (5.4), (5.5) and (5.6) over i, and taking the the limit as k tends to infinity of each establish that $\mathbf{Q} \xrightarrow{\mathcal{L}} N_p(\mathbf{0}, \Gamma)$. Since $\mathbf{Q}$ and $\sqrt{k}(\boldsymbol{\ell} - \boldsymbol{\theta})$ are asymptotically equivalent the theorem is proved. □

Now consider the case $\theta_r > \theta_{r+1} = \ldots = \theta_p = 0$. However, the r largest eigenvalues Θ do not necessarily have to be simple. Partition U_k and V_k as

$$U_k = \begin{pmatrix} U_{11} & U_{12} \\ U_{12}' & U_{22} \end{pmatrix} \quad \text{and} \quad V_k = \begin{pmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{pmatrix}$$

where the sub matrices in the top left corners are of dimension $r \times r$. From (5.3) and the perturbation expansion in Lemma 5.1, the $p-r$ smallest eigenvalues $(\bar{n}-1)Z_k$ are the same as that of

$$(\bar{n}-1)W_k^{(r)} = I_{p-r} + \frac{1}{\sqrt{k}}(U_{22} - \frac{1}{(\bar{n}-1)}V_{22}) + o_P(\frac{1}{\sqrt{k}}).$$

Therefore, the three test statistics can equivalently be expressed as

$$\mathrm{T}_{\mathrm{LR}}^{(r)} = \log|I_p + W_k^{(r)}|, \quad \mathrm{T}_{\mathrm{LH}}^{(r)} = \mathrm{tr}W_k^{(r)} \quad \text{and} \quad \mathrm{T}_{\mathrm{BNP}}^{(r)} = \mathrm{tr}W_k^{(r)}(I_p + W_k^{(r)})^{-1}. \tag{5.7}$$

Hence, we get the following Theorem as a direct application of Theorem 2.1.

Theorem 5.2. *Under assumptions* A1–A4, A6, A7 *and the null hypotheses* H_{0r}

$$\sqrt{k}(\ell T_{\mathrm{G}}^{(r)} - h) \xrightarrow{\mathcal{L}} N\left(0, \frac{2\bar{n}(p-r)}{\bar{n}-1} + \frac{\bar{n}(\bar{n}\underline{n}-1)}{(\bar{n}-1)^2}(\frac{\phi''(0)}{\phi'(0)^2} - 1)\{(p-r)^2 + 2(p-r)\}\right)$$

as $k \to \infty$ *where for* $\mathrm{G} = \mathrm{LR}$, $\ell = \bar{n}$ *and* $h = \bar{n}(p-r)\log(\bar{n}(\bar{n}-1)^{-1})$*; for* $\mathrm{G} = \mathrm{LH}$, $\ell = \bar{n}-1$ *and* $h = p-r$*; and for* $\mathrm{G} = \mathrm{BNP}$, $\ell = \bar{n}^2(\bar{n}-1)^{-1}$ *and* $h = \bar{n}(p-r)(\bar{n}-1)^{-1}$.

Here also we can proceed along the same lines as in Sections 2–4 to obtain robust tests in the unbalanced case. The asymptotic null distributions of the test statistics $\tilde{T}_{\mathrm{G}}^{(r)}$, defined in the usual way, are summarized in the following Theorem.

Theorem 5.3. *Under assumptions* A1–A7 *and the null hypotheses* H_{0r}

$$\sqrt{k}(\ell T_{\mathrm{G}}^{(r)} - h) \xrightarrow{\mathcal{L}} N(0, 2(p-r)\tau_0^2)$$

as $k \to \infty$ *assuming* $\tau_0^2 = \lim_{k\to\infty} \sum_{i=1}^{k} k^{-1} n_i(n_i-1)^{-1}$ *exists where for Likelihood Ratio criterion* $\mathrm{G} = \mathrm{LR}$, $\tilde{\ell} = 2$ *and* $\tilde{h} = 2(p-r)\log 2$*; for Lawley-Hotelling criterion* $\mathrm{G} = \mathrm{LH}$, $\tilde{\ell} = 1$ *and* $\tilde{h} = p-r$*; and for Bartlett-Nanda-Pillai criterion* $\mathrm{G} = \mathrm{BNP}$, $\tilde{\ell} = 4$ *and* $\tilde{h} = 2(p-r)$.

6 Finite Sample Approximations

In this section we provide some reasonable finite k approximations for the quantiles of the test statistics in the four MANOVA problems discussed in the previous sections. The main rationale for the approximations is the fact that the null distributions derived in this paper are, for the most part, asymptotically invariant to non-normality. That is, the critical values obtained under normality are expected to be asymptotically correct under non-normality as well. As a result the tests based on asymptotic expansions for the quantiles derived under normality and large k asymptotic framework are expected to give tests whose actual sizes are fairly close to the desired size under non-normality as well.

The finite performance of this type of approximation was evaluated by Bathke and Harrar (2007) for T_{LH} and T_{BNP} statistics in the nonparametric rank transformation context. However, how to apply them in the context of the testing problems considered in this paper has never been addressed. Moreover, their performance is not known in the situation where a sequence of tests are carried out as in the profile analysis and tests of dimensionality. Further, it is not quite clear how to apply them for the robust test statistics in the unbalanced case, i.e $\tilde{T}_{\mathrm{G}}$, $\tilde{T}_{\mathrm{G}}^{*}$, $\tilde{T}_{\mathrm{G}}^{(11.2)}$ and $\tilde{T}_{\mathrm{G}}^{(r)}$ for $\mathrm{G} = \mathrm{LR}, \mathrm{LH}$ and BNP.

6.1 Tests for Equality of Mean Vectors

Fujikoshi(1975) obtained asymptotic expansions for the distributions of centered and scaled versions of $T_1 = -\log|W(B+W)^{-1}|$, $T_2 = \mathrm{tr}(BW^{-1})$ and $T_3 = \mathrm{tr}\{B(B+W)^{-1}\}$ where B and W are independent random matrices distributed as $B \sim W_p(n_h, \Sigma, \Omega)$ and $W \sim W_p(n_e, \Sigma)$ such that $n_h = nh$,

$n_e = ne$ and $h > 0$, $e > 0$ and $h + e = 1$. The notation $W_p(n_h, \Sigma, \Omega)$ denotes the noncentral Wishart distribution with degrees of freedom n_h, mean $n_h\Sigma$ and non-centrality matrix Ω, and $W_p(n_e, \Sigma) = W_p(n_e, \Sigma, 0)$ which is a cental Wishart distribution. The asymptotic expansion is of the order of n meaning that when both n_h and n_e tend to infinity at the same rate. The Cornish-Fisher expansion for the upper α quantile of $\sqrt{(m/\tau^2)}(T_i - l)$ when $\Omega = 0$ is

$$z_\alpha + \frac{1}{\sqrt{m}}\{a_1 h_1(z_\alpha) + a_3 h_3(z_\alpha)\} - \frac{1}{m}\{b_2 h_2(z_\alpha) + b_4 h_4(z_\alpha) + b_6 h_6(z_\alpha) + z_\alpha(a_1 + a_3 h_3(z_\alpha))(\frac{1}{2}a_1 + a_3[\frac{1}{2}h_3(z_\alpha) - 2])\} + O(m^{-3/2}) \quad (6.1)$$

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, τ, 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 $b_6 = (1/2)a_3^2$.

Let the notations be as in section 2. Suppose $\mathbf{y}_{ij} \sim N_p(\boldsymbol{\mu}_i, \Sigma)$ where $N_p(\mu_i, \Sigma)$ denotes a p-variate normal distribution with mean $\boldsymbol{\mu}_i$ and variance-covariance matrix Σ. Then it is known that $H \sim W_p(k-1, \Sigma, \Omega_k)$, $E \sim W_p(N-k, \Sigma)$ and, H and E are independent. Hence, when $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_k$ (i.e $\Omega_k = 0$), the approximation (6.1) can be applied by taking $h = (k-1)/(N-1)$, $e = (N-k)/(N-1)$ and $n = N - 1$. In this case T_1, T_2 and T_3 coincide with T_{LR}, T_{LH} and T_{BNP}, respectively.

In the unbalanced case, it can easily be shown that H and $\tilde{E}$ are independent. However, $\tilde{E}$ does not have Wishart distribution. Therefore, we can not apply (6.1) directly. We propose the following solution. Let us first approximate the distribution $\tilde{E}$ by a constant multiple of Wishart distribution, i.e $\tilde{E} \overset{\text{approx}}{\sim} c \cdot W_p(\nu, \Sigma)$ where the constants c and ν, determined by matching the first two moments, are given by $c = \sum_{i=1}^k (k(n_i - 1))^{-1}$ and $\nu = k^2(\sum_{i=1}^k (n_i - 1)^{-1})^{-1}$. Then we have $(\sum_{i=1}^k (n_i - 1)^{-1})^{-1}\tilde{E} \overset{\text{approx}}{\sim} W_p((\sum_{i=1}^k (n_i - 1)^{-1})^{-1}, \Sigma)$. Now (6.1) may be applied by setting $h = (k-1)/(k - 1 + k^2(\sum_{i=1}^k (n_i - 1)^{-1})^{-1})$, $e = k^2(\sum_{i=1}^k (n_i - 1)^{-1})^{-1}/(k - 1 + k^2(\sum_{i=1}^k (n_i - 1)^{-1})^{-1})$ and $n = k - 1 + k^2(\sum_{i=1}^k (n_i - 1)^{-1})^{-1}$. Here also T_1, T_2 and T_3 coincide with $\tilde{T}_{\mathrm{LR}}$, $\tilde{T}_{\mathrm{LH}}$ and $\tilde{T}_{\mathrm{BNP}}$, respectively.

6.2 Profile Analysis

We use the same notations as in section 3. For testing $H_{0\mathrm{P}}$ and $H_{0\mathrm{F}}$ we note that $MHM' \sim W_m(k-1, M\Sigma M', \Omega_k^*)$, $MEM' \sim W_m(N-k, M\Sigma M', 0)$ and, MHM' and MEM' are independent under normality (i.e. $\mathbf{y}_{ij} \sim N_p(\boldsymbol{\mu}_i, \Sigma)$). Therefor the approximation (6.1) can be applied to T_{G}^* for $\mathrm{G} = \mathrm{LR}, \mathrm{LH}$ and BNP by setting $p = m$, $h = (k-1)/(N-1)$, $e = (N-k)/(N-1)$ and $n = N-1$. In the

unbalanced case, the same manipulation as in section 6.1 leads to a finite approximation to the quatiles of $\tilde{T}^*_{\mathrm{G}}$; $\mathrm{G} = \mathrm{LR}, \mathrm{LH}$ and BNP.

In the case testing $H_{0\mathrm{C}}$, it is known that $(N-k)(k-1)^{-1}T^*_{LH} \sim F(k-1, N-k)$ where $F(k-1, N-k)$ denotes the F-distribution with degrees of freedom $k-1$ and $N-k$. Hence the quantiles from the $F(k-1, N-k)$ distribution can be used as an alternative finite approximation in the balanced case. In the unbalanced case, based on arguments similar to those in 6.1 we have $(k-1)^{-1}\tilde{T}^*_{\mathrm{G}} \overset{\text{approx}}{\sim} F(k-1, (\sum_{i=1}^{k}(n_i-1)^{-1})^{-1})$ as a finite sample approximation.

In regards to testing $H_{0\mathrm{F}}$, it is known that $(N-k-p+2)(N-k)^{-1}(p-1)^{-1}T^2 \sim F(p-1, N-k)$. Moreover, we have seen in section 3 that the large k asymptotic reduces to the large N asymptotics. Therefore a quantile obtained from $F(p-1, N-k)$ provides a finite sample approximation.

6.3 Tests for Additional Information

It is proved in Fujikoshi (1981) that if $\mathbf{y}_{ij} \sim N(\boldsymbol{\mu}_i, \Sigma)$ and $H_{0\mathrm{A}}$ is true then $E_{11.2} \sim W_q(N-k-p+q, \Sigma_{11.2})$, $(T_{11.2} - E_{11.2}) \sim W_q(k-1, \Sigma_{11.2})$ and, $E_{11.2}$ and $(T_{11.2} - E_{11.2})$ are independent where $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Therefore, the approximation (6.1) can be applied by replacing p with q and setting $h = (k-1)/(N-1-p+q)$, $e = (N-k-p+q)/(N-1-p+q)$ and $n = N-1-p+q$. In this case $T^{(11.2)}_{\mathrm{LR}}$, $T^{(11.2)}_{\mathrm{LH}}$ and $T^{(11.2)}_{\mathrm{BNP}}$ coincide with T_1, T_2 and T_3, respectively. Here again the same manipulation as in section 6.1 leads to a finite approximation to the quantiles of $\tilde{T}^{(11.2)}_{\mathrm{G}}$; $\mathrm{G} = \mathrm{LR}, \mathrm{LH}$ and BNP in the unbalanced case.

6.4 Tests of Dimensionality

Under the same assumptions as in Fujikoshi(1975) and the assumption $\omega_j = n\theta_j$ for $j = 1, 2, \ldots, r$ where θ_j is a constant which does not depend on n, Isogai(1977) derived the asymptotic expansion for the distribution of $T^{(r)}_{\mathrm{G}}$. Accordingly the cornish-fisher expansion for $\sqrt{(m/\tau^2)}(T^{(r)}_{\mathrm{G}} - l)$ when $\phi_{r+1} = \ldots = \phi_p = 0$ is given by,

$$z_\alpha + \frac{1}{\sqrt{m}}\{a_1 h_1(z_\alpha) + a_3 h_3(z_\alpha)\} \tag{6.2}$$

where the values of m, τ^2, l, a_1 and a_2 depend on G. The functions $h_1(.)$ and $h_2(.)$ are defined as in section 6.1. For $\mathrm{G} = \mathrm{LR}$, $m = (n-1) - (1/2)(p+r)$, $\tau^2 = (1+e)(p-r)h/e$, $l = -(p-r)\log e$, $a_1 = \tau^{-1}(p-r)(1+e)\{(h-e)p - r + 1 - h\sum_{j=1}^{r}\theta_j^{-1}\}/(4e)$ and $a_3 = \tau^{-3}((1+e)/e)^2(p-r)(1+e)h/6$. For $\mathrm{G} = \mathrm{LH}$, $m = ne$, $\tau^2 = 2(p-r)h/e^2$, $l = (p-r)h/e$, $a_1 = \tau^{-1}(p-r)\{2(p+1)h - 2r - h\sum_{j=1}^{r}\theta_j^{-1}\}/(2e)$ and $a_3 = 4\tau^{-3}(p-r)(2-e)h/(3e^3)$. For $\mathrm{G} = \mathrm{BNP}$, $m = n$, $\tau^2 = 2e(p-r)eh$, $l = (p-r)h$, $a_1 = \tau^{-1}(p-r)(2rh - 2r - h\sum_{j=1}^{r}\theta_j^{-1})/2$ and $a_3 = 4\tau^{-3}(p-4)(e-h)eh/3$.

The asymptotic expansion (6.2) can be applied for the asymptotic framework of Section 5 by setting $h = (k-1)/(N-1)$, $e = (N-k)/(N-1)$ and $n = N-1$. In the unbalanced case also we can make similar arguments to obtain finite sample approximation for $\tilde{T}^{(r)}_{\mathrm{G}}$ by applying (6.2) with h, e and n exactly as in the last paragraph of Section 6.1.

7 Simulation Study

In this section we report simulation results for the null distributions of the tests of equality of mean vectors and dimensionality. The simulation was carried out by sampling from the multivariate normal $[MN]$, the multivariate t with 8 degrees of freedom $[MT]$ and the contaminated multivariate normal $0.9\phi_p(\mathbf{0}, I_p) + 0.1\phi(\mathbf{0}, 9I_p)$ $[CN]$ distributions. (The notations indicated in [] will be used for later reference.) In each case data was generated taking $p = 3$. Previously it was established that the effect of non-normality depends on the product $\bar{n}\underline{n}$. To investigate the effect of non-normality in the unbalanced case, we consider two patterns for sample sample sizes. They are (i) half of the sample sizes are set to 3 and the remaining half are set to 5 ($\bar{n}\underline{n} = 1.067$) and (ii) four fifth of the sample sizes are set to 3 and a fifth set to 15 ($\bar{n}\underline{n} = 1.5120$). We refer to these two patterns as the more balanced and less balanced patterns, respectively. The effect of non-normality is expected to be higher with the less balanced pattern. We assess the accuracy of the following approximations for the test statistics T_{G} and $\tilde{T}_{\mathrm{G}}$ for $\mathrm{G} = \mathrm{LR}, \mathrm{LH}, \mathrm{BNP}$.

1. Percentiles of T_{G} calculated based on Theorem 2.1 with the true value of $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11})$ $[T_G(\kappa_4^{(1)})]$,
2. Percentiles of T_{G} calculated based on Theorem 2.1 using the estimator of $\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11})$ given in (2.14) $[T_G(\hat{\kappa}_4^{(1)})]$,
3. Percentiles of $\tilde{T}_{\mathrm{G}}$ calculated based on Theorem 2.2 $[\tilde{T}_{\mathrm{G}}]$,
4. Percentiles of T_{G} calculated by (6.1) $[T_{\mathrm{G}}(F75)]$,
5. Percentiles of $\tilde{T}_{\mathrm{G}}$ calculated by (6.1) and the manipulation discussed in Section 6.1 $[\tilde{T}_{\mathrm{G}}(F75)]$ and
6. Percentiles of T_{G} obtained using the large n_i asymptotic expansions found in Anderson (1984, pp. 327-8, eqs. (20)-(22)) $[T_{\mathrm{G}}(A84)]$.

In Table 1, the result are presented for the more balanced sample size. It is clear from this table that the approximation based on Theorem 2.1 performs very well for likelihood ratio statistic. Further, it is also seen that the results with the actual and estimated $\kappa_4^{(1)}$ closely agree. It is quite clear that the approximations based on Fujikosi's (1975) work quite well for $\tilde{T}_{\mathrm{G}}$ for all values of G and all values of k considered. Since the sample size pattern is more balanced, the effect of non-normality is expected to be less. As a result we expect Fujikoshi (1975) approximation to work well for T_{G} as well. The simulation confirms our expectation. A striking observation is that the approximation based on the large n_i asymptotic expansions does a good job for the Lawley-Hotelling criteria T_{LH}. However, the behavior of this approximation is unclear for the other two statistics. Simulations for larger values of p (not reported here) show that the quality of approximations deteriorate slowly with increasing values of p. However, non-normality appear to have a negligible effect.

In the less balanced case (Table 2), we observe basically the same pattern as the more balanced case under normality. As expected, the actual sized based on the true value of $\kappa_4(1)$ and its consistent

Table 1: Simulated actual 5% sizes ($\times 100$) for the more balanced sample size pattern ($n_1 = \ldots = n_{k/2} = 3$ and $n_{k/2+1} = \ldots = n_k = 5$). Simulation size is $10,000$.

Population	k	G	$T_{\mathrm{G}}(\kappa_4^{(1)})$	$T_{\mathrm{G}}(\hat{\kappa}_4^{(1)})$	$\tilde{T}_{\mathrm{G}}$	$T_{\mathrm{G}}(F75)$	$\tilde{T}_{\mathrm{G}}(F75)$	$T_{\mathrm{G}}(A84)$
MN	20	LR	5.2	5.0	3.7	4.5	4.2	6.5
$\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}) = 0$		LH	8.4	8.1	8.8	5.4	5.4	5.4
		BNP	3.1	3.0	1.0	5.1	5.0	1.1
	30	LR	5.7	5.8	4.2	4.9	4.7	7.7
		LH	8.0	8.1	8.1	5.6	5.5	5.6
		BNP	3.7	3.8	1.8	5.4	5.3	1.6
	40	LR	5.2	4.8	4.1	4.7	4.6	7.9
		LH	7.2	6.7	7.6	5.2	5.2	5.1
		BNP	3.7	3.4	1.9	5.1	4.9	1.3
	50	LR	5.1	5.6	4.2	4.6	4.6	8.5
		LH	6.9	7.5	7.1	5.1	5.0	5.0
		BNP	3.6	4.0	2.3	4.9	4.9	1.5
MT	20	LR	5.5	4.2	3.8	4.9	4.3	7.1
$\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}) = 7.5$		LH	8.7	6.9	9.0	5.9	5.7	5.9
		BNP	3.2	2.2	1.2	5.4	5.0	1.3
	30	LR	5.1	5.1	3.7	4.7	4.4	7.6
		LH	7.6	7.6	8.0	5.4	5.2	5.4
		BNP	3.4	3.4	1.5	5.2	4.8	1.4
	40	LR	5.2	3.2	4.1	4.9	4.7	8.0
		LH	7.1	4.7	7.5	5.2	5.1	5.1
		BNP	3.7	2.3	2.2	5.2	5.1	1.6
	50	LR	5.1	3.6	3.7	4.9	4.2	8.5
		LH	6.8	5.2	7.1	5.2	4.6	5.1
		BNP	3.5	2.4	2.0	5.1	4.6	1.4
CN	20	LR	4.8	5.0	3.6	4.6	4.3	6.3
$\kappa_4^{(1)}(\boldsymbol{\varepsilon}_{11}) = 26.7$		LH	7.6	7.8	9.0	5.4	5.3	5.5
		BNP	2.7	2.9	1.3	4.9	4.7	1.2
	30	LR	4.3	4.8	3.7	4.3	4.2	7.3
		LH	6.6	7.2	7.9	5.1	5.0	5.0
		BNP	2.7	3.0	1.6	4.8	4.8	1.2
	40	LR	4.3	4.8	3.5	4.6	4.1	8.3
		LH	6.4	6.8	7.5	5.1	4.7	5.1
		BNP	3.0	3.1	1.9	4.8	4.6	1.4
	50	LR	4.6	4.7	4.3	4.8	4.8	8.2
		LH	6.2	6.4	7.2	5.1	5.2	5.0
		BNP	3.6	3.7	2.4	5.1	5.1	1.6

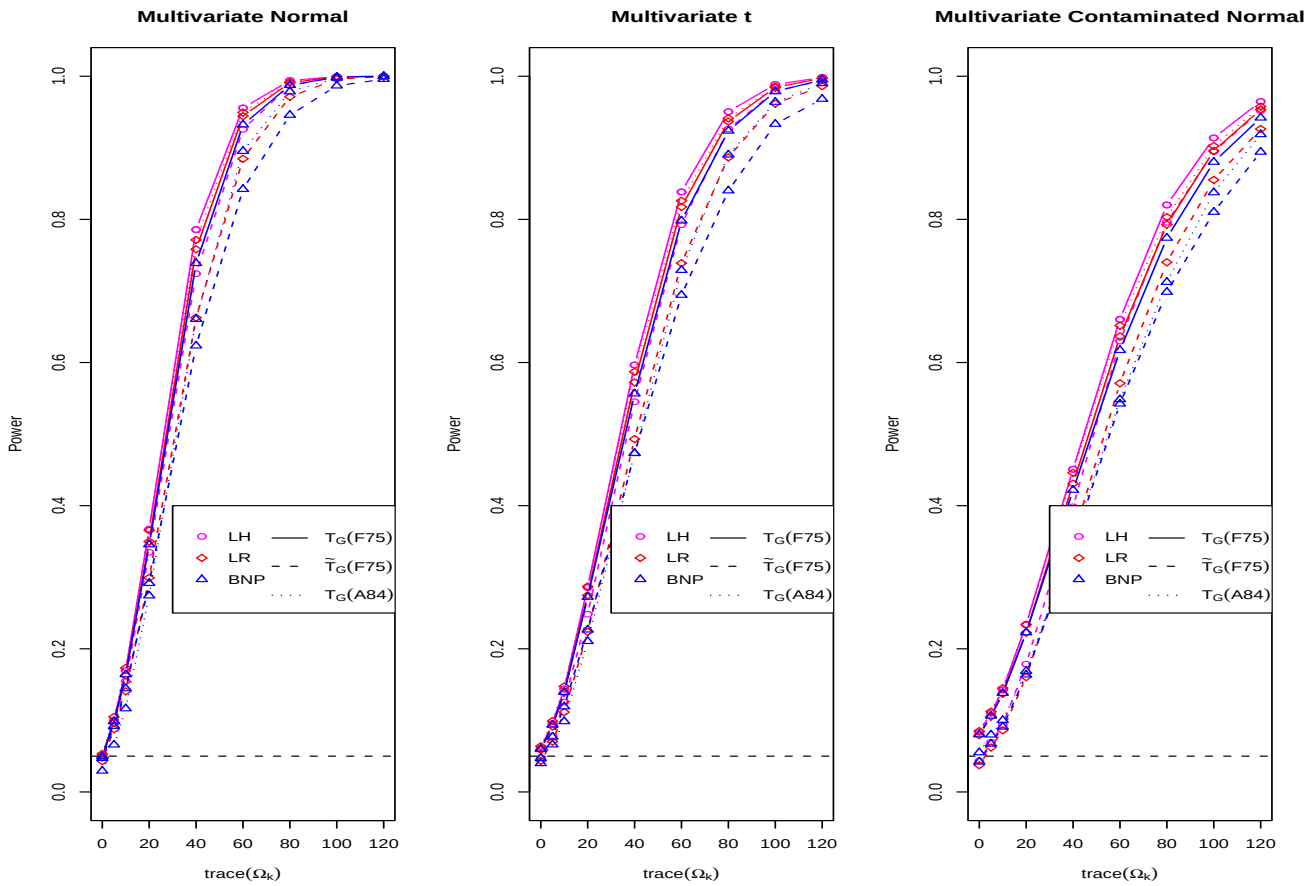
Table 2: Simulated actual 5% sizes ($\times 100$) for the less balanced sample size pattern ($n_1 = \ldots = n_{k/5} = 15$ and $n_{k/5+1} = \ldots = n_k = 3$). Simulation size is $10,000$.

	k	G	$T_{\mathrm{G}}(\kappa_4^{(1)})$	$T_{\mathrm{G}}(\hat{\kappa}_4^{(1)})$	$\tilde{T}_{\mathrm{G}}$	$T_{\mathrm{G}}(F75)$	$\tilde{T}_{\mathrm{G}}(F75)$	$T_{\mathrm{G}}(A84)$
MN	20	LR	4.8	5.0	4.0	4.7	4.2	5.8
		LH	6.8	7.2	9.6	5.5	5.6	5.5
		BNP	3.3	3.5	1.3	5.0	4.9	2.0
	30	LR	4.7	4.8	4.3	4.7	4.6	6.1
		LH	6.4	6.4	8.5	5.0	5.3	5.0
		BNP	3.5	3.5	1.7	5.1	5.2	2.0
	40	LR	4.7	4.5	4.3	4.7	4.7	6.2
		LH	6.0	5.7	8.1	5.0	5.3	4.9
		BNP	3.6	3.4	2.0	4.9	5.1	1.9
	50	LR	4.6	4.5	4.1	4.6	4.4	6.3
		LH	5.8	5.7	7.5	4.9	4.9	4.8
		BNP	3.7	3.7	2.2	4.7	4.7	2.1
MT(8)	20	LR	4.6	4.7	3.7	5.5	4.0	6.9
		LH	6.6	6.7	9.4	6.2	5.5	6.2
		BNP	2.9	3.0	1.2	5.8	4.7	2.3
	30	LR	5.0	5.3	3.8	5.9	4.0	7.3
		LH	6.4	6.8	8.7	6.3	5.1	6.3
		BNP	3.5	3.8	1.6	6.2	4.6	2.5
	40	LR	4.5	4.7	4.2	5.6	4.4	7.3
		LH	5.9	6.2	7.7	6.0	5.0	6.0
		BNP	3.5	3.6	1.9	5.6	4.9	2.6
	50	LR	5.0	5.1	4.2	6.1	4.5	8.0
		LH	6.2	6.4	7.7	6.5	5.0	6.5
		BNP	3.9	4.0	2.3	6.3	4.9	2.8
$CN(0.1, 3)$	20	LR	4.5	5.9	3.1	7.4	3.2	7.5
		LH	6.5	8.2	7.9	8.4	4.3	8.4
		BNP	3.0	4.1	0.9	7.6	4.0	3.7
	30	LR	4.6	6.0	3.4	7.9	3.9	7.7
		LH	6.1	7.8	7.4	8.4	4.7	8.4
		BNP	3.3	4.5	1.3	8.1	4.3	4.0
	40	LR	4.5	5.1	3.5	7.8	3.7	7.4
		LH	5.8	6.5	6.9	8.5	4.2	8.5
		BNP	3.4	3.9	1.7	7.8	4.1	4.1
	50	LR	4.7	4.7	3.9	8.2	4.3	7.7
		LH	5.9	5.8	6.9	8.8	4.7	8.8
		BNP	3.7	3.6	2.1	8.3	4.6	4.4

estimator $\hat{\kappa}_4(1)$ agree very well. Moreover, Fujikoshi's (1975) approximation is seen to be better for the modified statistics (robust statistics), and more so for Lawley-Hotelling and Bartlett-Nanda-Pillai statistics. This is somewhat expected because from theory we saw that the tests are not asymptotically robust for large k if the sample sizes are unbalanced. The approximation from the large n_i asymptotic expansion is not as good as it was for the more balanced case. Furthermore, unlike the other approximations the behavior of this approximation is unclear as k gets bigger even for Lawley-Hotelling criteria for which this approximation showed good performance in the more balanced case.

In Figure 1, the empirical powers of T_{G} and $\tilde{T}_{\mathrm{G}}$, based on the approximation $\tilde{T}_{\mathrm{G}}(F75)$, are graphed against $\mathrm{trace}(\Omega_k)$ for $k = 30$, $p = 3$ and the more balanced sample size pattern. In this figure, the null quantiles for T_{G} are based on the approximations $T_{\mathrm{G}}(F75)$ and $T_{\mathrm{G}}(A84)$ whereas that of $\tilde{T}_{\mathrm{G}}$ ar based on $\tilde{T}_{\mathrm{G}}(F75)$. All the tests are seen to have acceptable powers for all the sampling populations. However, we notice that the power deteriorates as the kurtosis gets bigger. Nevertheless, the difference observed among the individual tests seems to be attributable to the actual test sizes.

Figure 1: Achieved Power for 5% sizes for tests of equality of mean vectors $p = 3$, $k = 30$, and $n_1 = \ldots = n_{15} = 3$ and $n_{16} = \ldots = n_{30} = 5$. Simulation size is $10,000$.



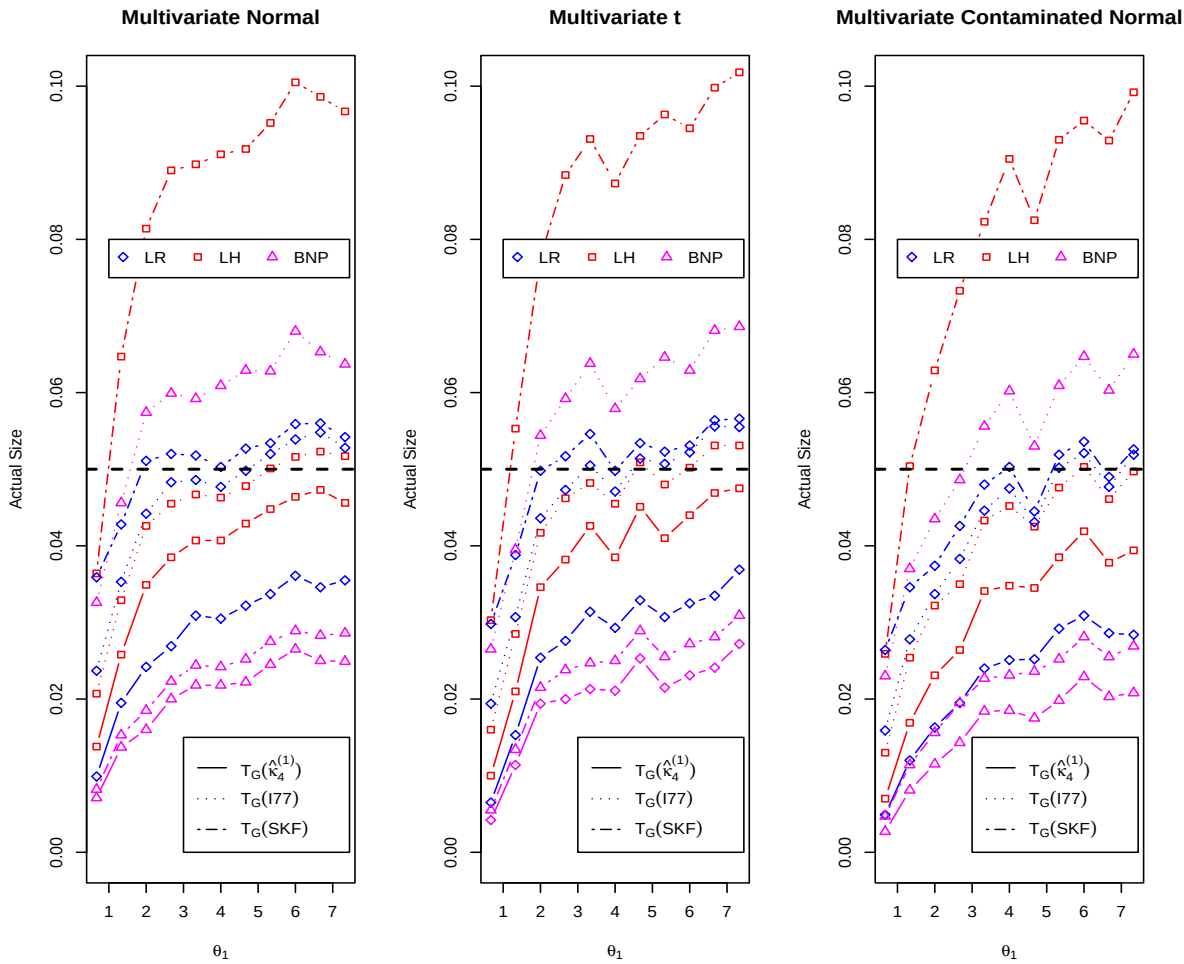
Next we assess the accuracy of the asymptotic results and the finite approximations for tests of dimensionality. We present the results for the more balanced sample size pattern and for $p = 3$, $k = 30$

and $r = 1$. Therefore, the hypothesis under consideration is $H_{01} : \omega_1 > \omega_2 = \omega_3 = 0$. Since we have seen from the theoretical results and from Tables 1 and 2 that the gain in the modified statistics is not much unless $|\bar{n}\underline{n} - 1| >> 0$, we do not present results for the modified statistic from the simulation study.

The approximations considered are:

1. Percentiles obtained from the limiting distribution given in Theorem 5.2 with estimated κ_4 $[T_G(\kappa_4)]$,
2. Percentiles obtained from the large n_i asymptotics with adjustment factor as in Seo, Kanda and Fujikoshi (1995) $[T_G(SKF)]$
   and
3. Percentiles obtained using the asymptotic expansion of Isogai (1977) $[T_G(I77)]$.

Figure 2: Achieved 5% sizes for tests of dimensionality for $p = 3$, $r = 1$, $k = 30$, $n_1 = \ldots = n_{15} = 3$ and $n_{16} = \ldots = n_{30} = 5$. Simulation size is $10,000$.



In Figure 2, the actual 5% sizes are graphed against $\theta_1 = \omega_1/k$. We have placed a horizontal dashed line at the actual size 0.05 for reference. It is clear from this figure that the desired 5% level is best met for the Likelihood Ratio criteria based on the Isogai (1977) and Seo, Kanda and Fujikoshi (1995) approximations and Lawley-Hotelling criteria based on the limiting distribution and Isogai (1977) approximations. The dependence of the actual size on θ_1 subsides quickly. Similar to what we observed in Tables 1 and 2, the approximations for Bartlett-Nanda-Pillai are not quite as accurate.

8 Concluding Remarks

The asymptotic null as well as non-null distributions of multivariate test statistics for some testing problems in one-way layout have been derived under general conditions and in the asymptotic framework that the number of treatment tends to infinity. More specifically, we obtained the asymptotic distributions in the contexts of tests of equality of mean vectors, profile analysis, additional information and dimensionality. The techniques employed allow a unified treatment of all these testing problems.

The asymptotic distributions were found to be normal. The mean of the limiting distributions equal the trace of the non-centrality parameters under normality and the variances depend on a measure of kurtosis. Indeed, the variances do not depend on non-normality if the group sample sizes are equal. Hence, one should be very cautious in applying the normal theory results in the general case. Neglecting non-normality will have a serious consequence even when the number of treatments is large if the sample sizes are unequal. The effect gets worse when p gets large. Therefore, we advise to use the results in this paper which account for the kurtosis in the data if the sample sizes are unbalanced. An alternative to kurtosis adjustment is to use the modified test statistics which were shown to be asymptotically invariant to non-normality. The modification involves only the error sum of squares and cross-products matrix. The modified statistics have power advantage over their unmodified counterparts when data is coming from a distribution with substantially heavy tails. It must also be noted that the results of this paper are also useful under normality. For most of the testing problems, exact distributions are known only in a few special cases. Asymptotic results are mostly available only in the large group sample sizes (replication sizes) asymptotic framework. The numerical results have demonstrated that the behaviors of large sample sizes asymptotic results are unpredictable when the number of treatments gets large.

Given the theoretical robustness of the usual statistics in the balanced case, in particular, and the modified statistics, in general, the numerical evidence confirms the accuracy of previous asymptotic expansion results in the same asymptotic framework but under normality.

Finally, we conclude this section by providing a heuristic argument on the implications of the results of this paper to large p asymptotics under normality. If we denote by $T_G(p, k-1, N-k)$ the exact distribution of T_G under normality, then it is well known that (see Anderson, 1984, Chapter 8)

$$T_G(p, k-1, N-k) = T_G(k-1, p, N-p-1).$$

Therefore the $k \to \infty$ asymptotic results may have some implication to the $p \to \infty$, $N \to \infty$ and $p/N \to c \in (0, 1)$ asymptotics. This idea is at the heart of Tonda and Fujikoshi (2004) who derive large

dimensional ($p \to \infty$) asymptotic expansion of the distribution of the likelihood ratio statistic for the multivariate linear hypothesis.

ACKNOWLEDGEMENT

This research was supported by the Montana Board of Research and Commercialization Project.

References

[1] Anderson, T. W. (1984), An Introduction to Multivariate Analysis (2nd ed.), John Wiley and Sons, Inc., New York.

[2] Akritas, M. A. and 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.

[3] Akritas, M.G. and Papadatos, N. (2004), Heteroscedastic One-way ANOVA and Lack-of-Fit Tests, Journal of the American Statistical Association, 99, 368-382.

[4] Backhouse, A. R. and McKay, R. J. (1982), Tests of Dimensionality in Multivariate Analysis of Variance, Communications in Statistics: Theory and Methods, 11, 1003-1027.

[5] Bathke, A. (2002), ANOVA for Large Number of Treatments, Mathematical Methods of Statistics, 11,118-132.

[6] Bathke, A.C. and Harrar, S.W. (2007), Nonparametric Methods in Multivariate Factorial Designs for Large Number of Factor Levels, Journal of Statistical Planning Inference, doi:10.1016/j.jspi.2006.11.004.

[7] Fang, K. T. and Zhang, Y. T. (1990), Generalized Multivariate Analysis, Springer-Verlag: Berlin.

[8] 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.

[9] Fujikoshi, Y. (1977), An Asymptotic Expansion for the Distributions of the Latent Roots of the Wishart Matrix with Multiple Population Roots, Annals of the Institute of Statistical Mathematics, 29, 379-387.

[10] Fujikoshi, Y. (1981), The Power of the Likelihood Ratio Test for Additional Information in a Multivariate Linear Model, Annals of the Institute of Statistical Mathematics, 279-285.

[11] 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.

[12] Fujikoshi, Y. (2004), Multivariate Analysis for the Case When the Dimension is Large Compared to the Sample Size, Journal of Korean Statistical Society, 33, 1-24.

[13] Gang, L. (1990), Moments of a Random Vector and Its Quadratic Forms, In Statistical Inference in Elliptically Contoured and Related Distributions, ed. Fang, K. T. and Anderson, T. W., Allerton Press Inc., New York, 433-440.

[14] Gupta, A. K. (1971), Noncentral distribution of Wilks' Statistic in MANOVA, The Annals of Mathematical Statistics, 42, 1254- 1261.

[15] Gupta, A. K., Harrar, S. W. and Fujikoshi, Y. (2006), Asymptotics for Testing Hypothesis in Some Multivariate Variance Components Model under Non-normality, Journal of Multivariate Analysis, 97, 148-178.

[16] Gupta, A. K., Harrar, S. W., Fujikoshi, Y. (2007), MANOVA for large hypothesis degrees of freedom under non-normality, Test, doi: 10.1007/s11749-006-0026-6.

[17] Gupta, A. K., Xu, J. and Fujikoshi, Y. (2006), An Asymptotic Expansion of the Distribution of Rao's U-statistic under a General Condition, Journal of Multivariate Analysis, 97, 492-513.

[18] Harrar, S.W. and Gupta, A.K. (2006), Asymptotic Expansion of the Null Distribution of the F-statistic in One-way ANOVA under Non-normality. Annals of the Institute of Statistical Mathematics, doi: 10.1007/s10463-006-0055-7.

[19] Isogai, T. (1977), Asymptotic Expansion for the Distributions of Latent Roots of $S_h S_e^{-1}$ and of Certain Test Statistics in MANOVA, Annals of the Institute of Statistical Mathematics, 29, Part A, 235-246.

[20] Mardia, K. V. (1970), Measures of Multivariate Skewness and Kurtosis with Applications, Biometrika, 57, 519-530.

[21] Maruyama, Y. (2007), Asymptotic Expansions of the Null Distributions of Some Test Statistics for Profile Analysis in General Distributions, Journal of Statistical Planning and Inference, 137, 506-526.

[22] Pillai, K. C. S. and Gupta, A. K. (1969), On Exact Distribution of Wilks' Criterion, Biometrika, 56, 109-118.

[23] Rencher, A.C. (2002), Multivariate Statistical Inference and Applications, 2nd Edition, Wiley:NewYork.

[24] Rao, C. R. (1973), Linear Statistical Inference and Its Applications, 2nd Edition, Wiley: NewYork.

[25] Rao, C. R. and Kleffe, J. (1988), Estimation of Variance Components and Applications, North-Holland: Amsterdam.

[26] Schatzoff, M. (1966), Exact Distribution of Wilks' Likelihood Ratio Criterion, Biometrika, 53, 347-358.

[27] Schott, J. (1984), Optimal Bounds for the Distribution of Some Test Criteria for Tests of Dimensionality, Biometrika, 71, 561-567.

[28] Seo, T., Kanda, T. and Fujikoshi, Y. (1995), The Effects of Nonnormality on Tests of Dimensionality in Cannonical Correlations and MANOVA Models, Journal of Multivariate Analysis, 52, 325-337.

[29] Srivastava, M. S. and Fujikoshi, Y. (2006), Multivariate Analysis of Variance with Fewer Observations than the Dimension, Journal of Multivariate Analysis, 97, 1927-1940.

[30] Storey, J. D., Xiao, W., Leek, J. T., Tompkins, R. G. and Davis, R. W. (2005), Significance Analysis of Time Course Microarray Experiments, Proceedings of the National Academy of Scinces of the United States of America, 102, 12837-12842.

[31] Tonda, T. and Fujikoshi, Y. (2004), Asymptotic Expansion of the Null Distribution of LR Statistic for Multivariate Linear Hypothesis when the Dimension is Large, Communications in Statistics: Theory and Methods, 33, 1205-1220.

[32] Yanagihara, H. (2006), A Family of Estimators for Multivariate Kurtosis in a Nonnormal Linear Regression Model, Journal of Multivariate Analysis, 98, 1-29.

[33] Yoshida, K., Imai, H. and Sato, Y. (2002), New Criteria for tests of Dimensionality under Elliptical Populations, Journal of Japan Statistical Society, 32, 183-192.

[34] Yoshida, K., Imai, H. and Sato, Y. (2004), Upper Limits for Criteria for testsof Dimensionality under Elliptical Populations, Communications in Statistics: Theory and Methods, 33, 2799-2816.