NATIONAL UNIVERSITY OF SINGAPORE

DEPARTMENT OF MATHEMATICS

MA4199 HONOURS PROJECT IN MATHEMATICS

Approximate Solutions to the Multivariate Behrens-Fisher Problem

AY 2018/2019 Final Year Project

March 2019

Name: Kevin Christian Wibisono

Matriculation Number: A0144886N

Supervisor: Associate Professor Zhang Jin-Ting

Co-Supervisor: Assistant Professor Pang Chin How, Jeffrey

Acknowledgements

To begin with, I would like to express my sincerest gratitude to Associate Professor Zhang Jin-Ting from the Department of Statistics and Applied Probability, who has supervised and guided me throughout the entire period of this project. Had he not been as patient and supportive as he was, punctual completion of this project would certainly have been impossible. It is quite rare to find a supervisor that constantly strives to encourage and facilitate his students to do well, and I am very fortunate to have done my project under Prof Zhang.

I am also indebted to Assistant Professor Pang Chin How, Jeffrey for his willingness to become my co-supervisor in this project. His suggestions for improving my presentation slides and thesis are invaluable.

I would also like to thank my parents for their continuous moral and emotional support even though we are separated by distance. In particular, I am grateful to my mother for coming to Singapore to look after me when I was hospitalised amidst the completion of this project.

For my friends, words cannot express how much you have positively impacted my life. I am thankful for a lot of things: the consolation you all provided when I fell down, the quality discussions we often had, the precious moments we went through together, and many more. To my senior, Anthony Prayugo, thank you for answering my project-related questions and sending me your LaTeX templates.

To end this, I want to thank the Department of Mathematics and the Department of Statistics and Applied Probability for providing me with this valuable opportunity to conduct an undergraduate-level statistical research under the guidance of your professors.

Summary

In statistics, we are often interested in comparing the means of multiple populations. Commonly used mean comparison methods invoke the assumption of homogeneous variance-covariance matrices. When this assumption is relaxed, this problem becomes increasingly complex and exact solutions are often computationally intractable.

This project deals with approximate solutions to the multivariate Behrens-Fisher (MBF) problem. The MBF problem is the problem of testing equality of means of two normally distributed populations with unequal variance-covariance matrices. As previously explained, the difficulty of this problem is attributed to this variance constraint. Assuming equality of variances, it is easy to construct a natural Wald-type test statistic which exactly follows the well-known Hotelling's T^2 distribution under the null hypothesis. On the other hand, the exact distribution of the test statistic under the condition of unequal variances is infeasible to compute. Numerous researchers, e.g. James (1954) and Johansen (1980), have attempted to formulate approximate solutions to this problem.

The main paper used for this project is Yanagihara and Yuan's (2005) *Three Approximate Solutions to the Multivariate Behrens-Fisher Problem.* The authors developed 3 (three) methods of approximate solutions to the MBF problem. Their main method involved approximating the null distribution of the natural test statistic with an F distribution, while the other two incorporated the Bartlett (1937) and modified Bartlett correction (Fujikoshi, 2000). In addition, they also compared the Type I errors (size) of their methods with 5 (five) other methods by means of Monte-Carlo simulations.

The first section of this thesis starts with a detailed explanation about the univariate and multivariate Behrens-Fisher problems, followed by a brief literature review about numerous approximate solutions developed by researchers. The nonsingular invariance, affine invariance and independence of different labelling schemes

properties desirable for an approximate solution are also examined in this section. Moreover, we also discuss some theoretical knowledge of the Wishart and Hotelling's T^2 distributions, which will be alluded to in the subsequent sections.

The second section details Yanagihara and Yuan's (2005) paper. The authors' approaches and approximation methods are comprehensively presented. A lot of results in the paper were stated without any proofs, and this section includes proofs and derivations of most of the results. In addition, we also discuss the performance of each method assessed through simulations performed by the authors as well as additional insights obtained from our more exhaustive simulation studies.

The third section mainly talks about the general linear hypothesis testing (GLHT) problem in heteroscedastic one-way MANOVA, which is a natural generalisation of the MBF problem. We first discuss the setting of the problem, and summarise Zhang's (2012) approach of modifying Krishnamoorthy and Yu's (2004) method to obtain an approximate solution to the GLHT problem. We then introduce a modification of Yanagihara and Yuan's (2005) F approximation method based on Zhang's (2012) idea. Both modified methods will be shown to be non-singular, affine invariant and independent of different labelling schemes, and reduce to the original methods in the context of the MBF problem. Moreover, the simulation results comparing both methods are also presented and explored. We also discuss an alternative method to deal with the case of high-dimensional multivariate normal distributions, in which the existing methods perform very badly.

The last section attempts to evaluate the performance of the methods derived in the previous section by using real-life data. The five-dimensional Egyptian Skull data, which contains measurement of male Egyptian skulls from 5 (five) different time periods, is used. These two methods are compared with the powerful yet inefficient parametric bootstrap (PB) method introduced by Krishnamoorthy and Lu (2010) in terms of the p-value.

Contents

1 Introduction

1.1 The Univariate Behrens-Fisher Problem

Let $\mathcal{N}(\mu, \sigma^2)$ denote the univariate normal distribution with mean μ and variance σ^2. Suppose that $X_1, X_2, \cdots, X_{n_1}$ is an independent and identically distributed (i.i.d.) sample drawn from $\mathcal{N}(\mu_1, \sigma_1^2)$, and $Y_1, Y_2, \cdots, Y_{n_2}$ is an i.i.d. sample drawn from $\mathcal{N}(\mu_2, \sigma_2^2)$. We are interested in testing the null hypothesis $H_0 : \mu_1 = \mu_2$ versus the two-sided alternative hypothesis $H_1 : \mu_1 \neq \mu_2$.

Consider the case where $\sigma_1^2 = \sigma_2^2 = \sigma^2$, i.e. the two populations have the same variance. In this setting, we are actually testing whether the two populations have the same distribution. For the sample drawn from the first normal population, let $\overline{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$ denote the sample mean and $s_1^2 = (n_1 - 1) \sum_{i=1}^{n_1} (X_i - \overline{X})^2$ denote the sample variance. Similarly, $\overline{Y}$ and s_2^2 respectively denote the sample mean and sample variance of the sample drawn from the second normal population.

It is well-known that under H_0, the test statistic

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{1}$$

exactly follows the $t_{n_1+n_2-2}$ distribution. In order to perform a hypothesis testing with a specified significance level α, it suffices to check whether $|t| > t_{n_1+n_2-2}(\alpha/2)$, where $t_k(\beta)$ denotes the critical value for a t distribution with k degrees of freedom when the right-tail probability is β. We reject H_0 if the above inequality holds and accept H_0 otherwise.

Now, we consider the other case where $\sigma_1^2 \neq \sigma_2^2$. This problem of testing the equality of means of two normally distributed populations when the variances are not equal is referred to as the univariate Behrens-Fisher problem. As mentioned in the Summary, the condition of unequal variances imposed is what makes the Behrens-Fisher problem particularly difficult. Indeed, although exact solutions to this problem have been investigated, they are difficult to compute (Yanagihara and Yuan, 2005) and thus not of our interest. Researchers have been more interested in studying approximate solutions to the Behrens-Fisher problem.

A well-known approximate solution to the univariate Behrens-Fisher problem was developed by Welch (1938), which asserts that the test statistic

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{2}$$

approximately follows the t_k distribution, where

$$k = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}}. \tag{3}$$

1.2 The Multivariate Behrens-Fisher (MBF) Problem

The multivariate Behrens-Fisher (MBF) problem is a natural generalisation of the univariate counterpart. Instead of considering two univariate normal populations, we now consider two multivariate normal populations. Let $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the p-variate normal distribution with mean vector μ and variance-covariance matrix $\boldsymbol{\Sigma}$. Suppose that $\boldsymbol{y_{11}}$, $\boldsymbol{y_{21}}$, $\cdots$, $\boldsymbol{y_{n_1 1}}$ is an i.i.d. sample drawn from $\mathcal{N}_p(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})$, and $\boldsymbol{y_{12}}$, $\boldsymbol{y_{22}}$, $\cdots$, $\boldsymbol{y_{n_2 2}}$ is an i.i.d. sample drawn from $\mathcal{N}_p(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2})$. For $j \in \{1, 2\}$, let $\overline{\boldsymbol{y}}_{\boldsymbol{j}} = \frac{1}{n_j}\sum_{i=1}^{n_j} \boldsymbol{y_{ij}}$ denote the sample mean and $\boldsymbol{S_j} = \frac{1}{n_j-1}\sum_{i=1}^{n_j}(\boldsymbol{y_{ij}} - \overline{\boldsymbol{y}}_{\boldsymbol{j}})(\boldsymbol{y_{ij}} - \overline{\boldsymbol{y}}_{\boldsymbol{j}})'$ denote the sample variance. Same as the univariate case, we are interested in testing the null hypothesis $H_0 : \boldsymbol{\mu_1} = \boldsymbol{\mu_2}$ versus the two-sided alternative hypothesis $H_1 : \boldsymbol{\mu_1} \neq \boldsymbol{\mu_2}$.

When $\boldsymbol{\Sigma_1} = \boldsymbol{\Sigma_2}$, it is known that under H_0,

$$\frac{n_1 n_2}{n_1 + n_2}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}})'\Big(\frac{(n_1 - 1)\boldsymbol{S_1} + (n_2 - 1)\boldsymbol{S_2}}{n_1 + n_2 - 2}\Big)^{-1}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}) \tag{4}$$

exactly follows the $\mathcal{T}^2(p, n_1 + n_2 - 2)$ distribution, i.e. the Hotelling's T^2 distribution with p and n_1+n_2-2 degrees of freedom. Hotelling's T^2 distribution is actually just a scaled F distribution. We will explain this distribution more thoroughly in the subsequent section. A hypothesis testing can be performed using a similar technique as for the univariate case.

When $\boldsymbol{\Sigma_1} \neq \boldsymbol{\Sigma_2}$, this problem is known as the MBF problem. In this case, a natural Wald-type statistic for testing H_0 is

$$T = (\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}})'\Big(\frac{\boldsymbol{S_1}}{n_1} + \frac{\boldsymbol{S_2}}{n_2}\Big)^{-1}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}). \tag{5}$$

It is worth mentioning that when $n_1 = n_2$, the statistic in (4) reduces to the statistic in (5). Yanagihara and Yuan (2005) noted that when n_1 and n_2 approach infinity, T converges to χ^2_p. Predictably, an approximate solution based on this convergence performs very badly when one of the sample sizes is small.

1.3 Approximate Solutions to the MBF Problem and Their Desirable Properties

Nel et al. (1990) managed to obtain the exact null distribution of T; however, it is "computatioinally intractable" (Yanagihara and Yuan, 2005). Instead of studying exact solutions, it is therefore more plausible to develop approximate solutions with superior size and power. Numerous researches (e.g. James (1954), Johansen (1980), Yao (1965), Nel and van der Merwe (1986)) have been devoted to finding approximate solutions to the MBF problem.

According to Zhang (2012), there are 3 (three) properties which are desirable for an approximate solution to the MBF problem, namely:

- **Affine invariance**

  Ideally, a solution to the MBF problem must give the same null distribution and value of the test statistic if the samples $\boldsymbol{y_{11}}$, $\boldsymbol{y_{21}}$, $\cdots$, $\boldsymbol{y_{n_1 1}}$ and $\boldsymbol{y_{22}}$, $\cdots$, $\boldsymbol{y_{n_2 2}}$ are re-centred or rescaled (Zhang, 2012). In particular, transforming each of the $\boldsymbol{y_{ij}}$'s to $\tilde{\boldsymbol{y}}_{\boldsymbol{ij}} = \boldsymbol{B}\boldsymbol{y_{ij}} + \boldsymbol{b}$ ($i \in \{1, 2, \cdots, n_j\}$, $j \in \{1, 2\}$), where $\boldsymbol{B}$ is any invertible constant matrix with p columns and $\boldsymbol{b}$ any constant vector of length p, should not affect the hypothesis testing.

- **Nonsingular invariance**

  Note that the null hypothesis $H_0 : \boldsymbol{\mu_1} = \boldsymbol{\mu_2}$ is equivalent to $H_A : \boldsymbol{A}\boldsymbol{\mu_1} = \boldsymbol{A}\boldsymbol{\mu_2}$ for any invertible matrix $\boldsymbol{A}$ with p columns. A desirable method should therefore be nonsingular invariant, i.e. independent of the choice of $\boldsymbol{A}$.

- **Independence of different labelling schemes**

  Recall that we have an i.i.d. sample $\boldsymbol{y_{11}}$, $\boldsymbol{y_{21}}$, $\cdots$, $\boldsymbol{y_{n_1 1}}$drawn from $\mathcal{N}_p(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})$, and an i.i.d. sample $\boldsymbol{y_{12}}$, $\boldsymbol{y_{22}}$, $\cdots$, $\boldsymbol{y_{n_2 2}}$ drawn from $\mathcal{N}_p(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2})$. We would clearly expect testing $\boldsymbol{\mu_1} = \boldsymbol{\mu_2}$ to give the same result as testing $\boldsymbol{\mu_2} = \boldsymbol{\mu_1}$.

Yanagihara and Yuan (2005) stated that the statistic obtained by Nel and van der Merwe (1986) was not affine invariant. Krishnamoorthy and Yu (2004) then modified the statistic to become invariant.

1.4 The Wishart and Hotelling's T^2 Distribution

Throughout this thesis, we will use the Wishart and Hotelling's T^2 distributions in several instances so as to derive some approximate solutions to the MBF problem. This section is devoted to explaining the definition and useful properties of both distributions. The results below are cited from Jung's (2013) and Hanson's (2014) lecture notes.

Definition 1. *Let $\boldsymbol{X_1}, \boldsymbol{X_2}, \cdots, \boldsymbol{X_n}$ be an i.i.d. sample drawn from $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, and suppose $\boldsymbol{X} = [\boldsymbol{X_1}, \boldsymbol{X_2}, \cdots, \boldsymbol{X_n}]$. Then $\boldsymbol{X}\boldsymbol{X}'$ is said to follow $\mathcal{W}_p(n, \boldsymbol{\Sigma})$, i.e. the Wishart distribution with n degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$.*

Proposition 1.1.

1. *Let $\boldsymbol{M} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$ and $\boldsymbol{B}$ be a $p \times m$ real matrix. Then $\boldsymbol{B}'\boldsymbol{M}\boldsymbol{B} \sim \mathcal{W}_m(n, \boldsymbol{B}'\boldsymbol{\Sigma}\boldsymbol{B})$.*

2. *Let* $\boldsymbol{M} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$. *Then* $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{M}\boldsymbol{\Sigma}^{-1/2} \sim \mathcal{W}_p(n, \boldsymbol{I_p})$.

3. *Let* $\boldsymbol{M_i} \sim \mathcal{W}_p(n_i, \boldsymbol{\Sigma})$ $(i = 1, 2, \cdots k)$ *be independent. Then* $\sum_{i=1}^{k} \boldsymbol{M_i} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$, *where* $n = n_1 + n_2 + \cdots + n_k$.

4. *Let* $\boldsymbol{M} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$. *Then* $\mathrm{E}[\boldsymbol{M}] = n\boldsymbol{\Sigma}$.

5. *Let* $\boldsymbol{M} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$. *Then* $\mathrm{V}[\boldsymbol{M}] = n(\mathrm{tr}(\boldsymbol{\Sigma}^2) + (\mathrm{tr}(\boldsymbol{\Sigma}))^2)$. *Here,* $\mathrm{V}[\boldsymbol{M}]$ *denotes the sum of the variances of all the entries of* $\boldsymbol{M}$.

6. *Let* $\boldsymbol{X_1}, \boldsymbol{X_2}, \cdots, \boldsymbol{X_n}$ *be an i.i.d. sample from* $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and* $\boldsymbol{S}$ *be the sample variance. Then* $(n-1)\boldsymbol{S} \sim \mathcal{W}_p(n-1, \boldsymbol{\Sigma})$.

Definition 2. *Let* $\boldsymbol{d} \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{I_p})$ *and* $\boldsymbol{M} \sim \mathcal{W}_p(m, \boldsymbol{I_p})$ *be independent. Then* $m\boldsymbol{d}'\boldsymbol{M}^{-1}\boldsymbol{d}$ *is said to follow* $\mathcal{T}^2(p, m)$, *i.e. the Hotelling's* T^2 *distribution with* p *and* m *degrees of freedom.*

Proposition 1.2. *Let* $\boldsymbol{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and* $\boldsymbol{M} \sim \mathcal{W}_p(m, \boldsymbol{\Sigma})$ *be independent. Then* $m(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{M}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \sim \mathcal{T}^2(p, m)$.

Proof. Note that $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}) \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{I_p})$ and $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{M}\boldsymbol{\Sigma}^{-1/2} \sim \mathcal{W}_p(m, \boldsymbol{I_p})$ by Proposition 1.1. Using Definition 2 with the substitutions $\boldsymbol{d} := \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})$ and $\boldsymbol{M} := \boldsymbol{\Sigma}^{-1/2}\boldsymbol{M}\boldsymbol{\Sigma}^{-1/2}$, the result immediately follows. □

Now, let $\boldsymbol{X_1}, \boldsymbol{X_2}, \cdots, \boldsymbol{X_n}$ be an i.i.d. sample from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Denote the sample mean and sample variance by $\overline{\boldsymbol{X}}$ and $\boldsymbol{S}$, respectively. We have the following proposition:

Proposition 1.3. *Following the above notation, we have* $n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \mathcal{T}^2(p, n-1)$.

Proof. Note that $\sqrt{n}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $(n-1)\boldsymbol{S} \sim \mathcal{W}_p(n-1, \boldsymbol{\Sigma})$. By using Proposition 1.2 and the well-known fact that $\overline{\boldsymbol{X}}$ and $\boldsymbol{S}$ are independent, the result immediately follows. □

We end this section with the following important proposition that connects Hotelling's T^2 distribution with the F distribution. A proof of this proposition can be found in Hanson's (2014) lecture notes.

Proposition 1.4. *Let* $X \sim \mathcal{T}^2(p, m)$. *Then* $\frac{m-p+1}{mp} X \sim \mathcal{F}_{p,m-p+1}$. *In other words, Hotelling's* T^2 *distribution is just a scaled* F *distribution.*

2 Yanagihara and Yuan's (2005) Paper

2.1 The Outline of Yanagihara and Yuan's (2005) Paper

In 2005, Yanagihara and Yuan published a research paper titled *Three Approximate Solutions to the Multivariate Behrens-Fisher Problem.* As the title suggests, the authors developed 3 (three) approximate solutions to the MBF problem. Their main solution involved approximating the natural Wald-type test statistic with an F distribution with appropriate degrees of freedom, while their other two solutions utilised the well-known Bartlett correction (Bartlett, 1937) and modified Bartlett correction (Fujikoshi, 2000). The next subsection provides a comprehensive study of Yanagihara and Yuan's (2005) main method.

In addition, the authors compared the Type I errors, i.e. the probability of rejecting a true null hypothesis, of their methods with 5 (five) other methods by means of Monte-Carlo simulations. They also measured a quantity called the average absolute discrepancy (AAD), which is the average of the differences between the nominal and empirical (Type I error) sizes. An approximate solution with a small AAD is more desirable. The authors concluded that their main method is better than the other methods in terms of the AAD. The subsection after the next subsection details these simulation studies.

The 8 (eight) methods considered in their research paper can be summarised as follows:

- The simple chi-square approximation method, which approximates T with a chi-square distribution with p degrees of freedom.

- Yanagihara and Yuan's (2005) main method, which approximates $T_F = \frac{n-2-\hat{\theta}_1}{(n-2)p}T$ with an F distribution with p and $\hat{\upsilon}$ degrees of freedom, where

$$\hat{\theta}_1 = \frac{p\hat{\psi}_1 + (p-2)\hat{\psi}_2}{p(p+2)}, \quad \hat{\theta}_2 = \frac{\hat{\psi}_1 + 2\hat{\psi}_2}{p(p+2)} \text{ and } \hat{\upsilon} = \frac{(n-2-\hat{\theta}_1)^2}{(n-2)\hat{\theta}_2 - \hat{\theta}_1}.$$

- Yanagihara and Yuan's (2005) second method, which approximates

$$T_B = \Big(1 - \frac{\hat{\psi}_1 + \hat{\psi}_2}{p(n-2)}\Big)T$$

with a chi-square distribution with p degrees of freedom.

- Yanagihara and Yuan's (2005) third method, which approximates

$$T_{MB} = \frac{(p+2)(2p(n-2) - \hat{\psi}_1)}{2(\hat{\psi}_1 + 2\hat{\psi}_2)} \log\Big(1 + \frac{T(\hat{\psi}_1 + 2\hat{\psi}_2)}{(n-2)p(p+2)}\Big)$$

with a chi-square distribution with p degrees of freedom.

- James' (1954) second-order approximation, which approximates the critical point of the null distribution of T with

$$c_\alpha^{(p)}\Big(1 + \frac{(p+2)\hat{\psi}_1 + (\hat{\psi}_1 + 2\hat{\psi}_2)c_\alpha^{(p)}}{2p(p+2)(n-2)}\Big),$$

where $c_\alpha^{(p)}$ is the upper α percentage point of a chi-square distribution with p degrees of freedom.

- Yao's (1965) method, which approximates $T_{FY} = \frac{T}{\hat{v}_Y p}$ with an F distribution with p and $\hat{v}_Y - p + 1$ degrees of freedom, where

$$\hat{v}_Y = \frac{n^2(n_1-1)(n_2-1)(\overline{\boldsymbol{y}}_{\boldsymbol{d}}'\overline{\boldsymbol{S}}^{-1}\overline{\boldsymbol{y}}_{\boldsymbol{d}})^2}{n_2^2(n_2-1)(\overline{\boldsymbol{y}}_{\boldsymbol{d}}'\overline{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{1}}\overline{\boldsymbol{S}}^{-1}\overline{\boldsymbol{y}}_{\boldsymbol{d}})^2 + n_1^2(n_1-1)(\overline{\boldsymbol{y}}_{\boldsymbol{d}}'\overline{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{2}}\overline{\boldsymbol{S}}^{-1}\overline{\boldsymbol{y}}_{\boldsymbol{d}})^2}$$

with $\overline{\boldsymbol{y}}_{\boldsymbol{d}} = \overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}$ and $\overline{\boldsymbol{S}} = \frac{n_2}{n}\boldsymbol{S}_{\boldsymbol{1}} + \frac{n_1}{n}\boldsymbol{S}_{\boldsymbol{2}}$.

- Johansen's (1980) method, which approximates $T_{FJ} = \frac{T}{\hat{\phi}_J}$ with an F distribution with p and $\hat{v}_J$ degrees of freedom, where

$$\hat{\phi}_J = p + \frac{(p-1)(\hat{\psi}_1 + \hat{\psi}_2)}{(p+2)(n-2)} \text{ and } \hat{v}_J = \frac{2p(p+2)(n-2)}{3(\hat{\psi}_1 + \hat{\psi}_2)}.$$

- Modified Nel and Van der Merwe's Method (Krishnamoorthy and Yu, 2004), which approximates $T_{FM} = \frac{(\hat{v}_M - p + 1)T}{\hat{v}_M p}$ with an F distribution with p and $\hat{v}_M - p + 1$ degrees of freedom, where

$$\hat{v}_M = \frac{p(p+1)(n-2)}{\hat{\psi}_1 + \hat{\psi}_2}.$$

Here, T denotes the test statistic mentioned in (5), p denotes the dimension of the multivariate normal distributions, n_1 and n_2 denote the sample sizes and $n = n_1 + n_2$, $\overline{\boldsymbol{y}}_{\boldsymbol{1}}$ and $\overline{\boldsymbol{y}}_{\boldsymbol{2}}$ denote the sample means, $\overline{\boldsymbol{S}}_{\boldsymbol{1}}$ and $\overline{\boldsymbol{S}}_{\boldsymbol{2}}$ denote the sample variances,

$$\hat{\psi}_1 = \frac{n_2^2(n-2)}{n^2(n_1-1)}(\mathrm{tr}(\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}))^2 + \frac{n_1^2(n-2)}{n^2(n_2-1)}(\mathrm{tr}(\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}))^2 \text{ and} \tag{6}$$

$$\hat{\psi}_2 = \frac{n_2^2(n-2)}{n^2(n_1-1)}\mathrm{tr}(\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}) + \frac{n_1^2(n-2)}{n^2(n_2-1)}\mathrm{tr}(\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}). \tag{7}$$

2.2 Yanagihara and Yuan's (2005) Main Method

Let us recall the setting of the MBF problem as mentioned in Section 1.2. Suppose that $\boldsymbol{y_{11}}, \boldsymbol{y_{21}}, \cdots, \boldsymbol{y_{n_1 1}}$ is an i.i.d. sample drawn from $\mathcal{N}_p(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})$, and $\boldsymbol{y_{12}}, \boldsymbol{y_{22}}, \cdots, \boldsymbol{y_{n_2 2}}$ is an i.i.d. sample drawn from $\mathcal{N}_p(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2})$. For $j \in \{1, 2\}$, let $\overline{\boldsymbol{y}}_{\boldsymbol{j}} = \frac{1}{n_j}\sum_{i=1}^{n_j} \boldsymbol{y_{ij}}$ denote the sample mean and $\boldsymbol{S_j} = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(\boldsymbol{y_{ij}} - \overline{\boldsymbol{y}}_{\boldsymbol{j}})(\boldsymbol{y_{ij}} - \overline{\boldsymbol{y}}_{\boldsymbol{j}})'$ denote the sample variance. We are interested in testing the null hypothesis $H_0 : \boldsymbol{\mu_1} = \boldsymbol{\mu_2}$ versus the two-sided alternative hypothesis $H_1 : \boldsymbol{\mu_1} \neq \boldsymbol{\mu_2}$ assuming that $\boldsymbol{\Sigma_1} \neq \boldsymbol{\Sigma_2}$. A natural Wald-type statistic for testing H_0 is, as mentioned in (5),

$$T = (\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}})'\Big(\frac{\boldsymbol{S_1}}{n_1} + \frac{\boldsymbol{S_2}}{n_2}\Big)^{-1}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}).$$

To begin with, let $\overline{\boldsymbol{\Sigma}} = \frac{n_2}{n}\boldsymbol{\Sigma_1} + \frac{n_1}{n}\boldsymbol{\Sigma_2}$, where $n = n_1 + n_2$. Since $\boldsymbol{\Sigma_1}$ and $\boldsymbol{\Sigma_2}$ are symmetric, it follows that $\overline{\boldsymbol{\Sigma}}$ is also symmetric. Moreover, let

$$\boldsymbol{z} = \sqrt{\frac{n_1 n_2}{n}}\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}) \text{ and } \boldsymbol{W} = \overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\Big(\frac{n_2}{n}\boldsymbol{S}_{\boldsymbol{1}} + \frac{n_1}{n}\boldsymbol{S}_{\boldsymbol{2}}\Big)\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}. \tag{8}$$

This implies

$$\begin{aligned}
\boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z} &= \sqrt{\frac{n_1 n_2}{n}}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}})'\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\overline{\boldsymbol{\Sigma}}^{\frac{1}{2}}\Big(\frac{n_2}{n}\boldsymbol{S}_{\boldsymbol{1}} + \frac{n_1}{n}\boldsymbol{S}_{\boldsymbol{2}}\Big)^{-1}\overline{\boldsymbol{\Sigma}}^{\frac{1}{2}}\sqrt{\frac{n_1 n_2}{n}}\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}) \\
&= \frac{n_1 n_2}{n}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}})'\Big(\frac{n_2}{n}\boldsymbol{S}_{\boldsymbol{1}} + \frac{n_1}{n}\boldsymbol{S}_{\boldsymbol{2}}\Big)^{-1}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}) \\
&= (\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}})'\Big(\frac{\boldsymbol{S}_{\boldsymbol{1}}}{n_1} + \frac{\boldsymbol{S}_{\boldsymbol{2}}}{n_2}\Big)^{-1}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}) \\
&= T.
\end{aligned}$$

Next, we write $T = \boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z} = \frac{\boldsymbol{z}'\boldsymbol{z}}{U}$, where $U = \frac{\boldsymbol{z}'\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z}}$. Notice that under H_0,

$$\boldsymbol{z}'\boldsymbol{z} = (\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}})'\Big(\frac{\boldsymbol{\Sigma}_{\boldsymbol{1}}}{n_1} + \frac{\boldsymbol{\Sigma}_{\boldsymbol{2}}}{n_2}\Big)^{-1}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}) \sim \chi^2_p \tag{9}$$

since $\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}} \sim N_p(\boldsymbol{0}, \frac{\boldsymbol{\Sigma}_{\boldsymbol{1}}}{n_1} + \frac{\boldsymbol{\Sigma}_{\boldsymbol{2}}}{n_2})$. Moreover, it can be shown that $\boldsymbol{z}'\boldsymbol{z}$ and U are mutually independent (Fang et al., 1990). Our aim is to approximate the null distribution of T with a constant multiple of an F distribution.

Since $\boldsymbol{z}'\boldsymbol{z} \sim \chi^2_p$, in order to achieve our ultimate goal it is natural to assume

$$U \approx \frac{\chi^2_v}{\phi}, \tag{10}$$

which means

$$T = \frac{\boldsymbol{z}'\boldsymbol{z}}{U} \approx \frac{\chi^2_p}{\chi^2_v/\phi} = \frac{\phi p}{v}\frac{\chi^2_p/p}{\chi^2_v/v} \text{ or}$$

$$\frac{v}{\phi p}T \overset{a}{\sim} F_{p,v}, \tag{11}$$

where $\overset{a}{\sim}$ means "approximately follows".

Now, using the well-known fact about the mean and variance of a chi-square distribution, we obtain from (10) that

$$\mathrm{E}[U] \approx \frac{v}{\phi} \text{ and } \mathrm{E}[U^2] \approx \frac{v(v+2)}{\phi^2}. \tag{12}$$

In order to approximate the constant v and ϕ, we calculate the first and second moments of U asymptotically by means of the identity $U = \frac{\boldsymbol{z}'\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z}}$ and match them with (12). First, observe that

$$\begin{aligned}
\boldsymbol{W}^{-1} &= \overline{\boldsymbol{\Sigma}}^{\frac{1}{2}}\Big(\frac{n_2}{n}\boldsymbol{S_1} + \frac{n_1}{n}\boldsymbol{S_2}\Big)^{-1}\overline{\boldsymbol{\Sigma}}^{\frac{1}{2}} \\
&= \Big(\frac{n_2}{n}\boldsymbol{\Sigma_1} + \frac{n_1}{n}\boldsymbol{\Sigma_2}\Big)^{\frac{1}{2}}\Big(\frac{n_2}{n}\boldsymbol{S_1} + \frac{n_1}{n}\boldsymbol{S_2}\Big)^{-1}\Big(\frac{n_2}{n}\boldsymbol{\Sigma_1} + \frac{n_1}{n}\boldsymbol{\Sigma_2}\Big)^{\frac{1}{2}} \\
&= \boldsymbol{I_p} - \frac{1}{\sqrt{n-2}}\overline{\boldsymbol{V}} + \frac{1}{n-2}\overline{\boldsymbol{V}}^2 + \mathrm{O}_p((n-2)^{-\frac{3}{2}}),
\end{aligned}$$

where

$$\overline{\boldsymbol{V}} = \sqrt{n-2}\,\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\Big(\frac{n_2}{n}(\boldsymbol{S_1} - \boldsymbol{\Sigma_1}) + \frac{n_1}{n}(\boldsymbol{S_2} - \boldsymbol{\Sigma_2})\Big)\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}. \tag{13}$$

Since $U = \frac{\boldsymbol{z}'\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z}}$, it follows that

$$U = 1 + \frac{1}{\sqrt{n-2}}\Big(\frac{\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}\Big) + \frac{1}{n-2}\Big(\frac{(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2}{(\boldsymbol{z}'\boldsymbol{z})^2} - \frac{\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}\Big) + \mathrm{O}_p((n-2)^{-\frac{3}{2}}) \tag{14}$$

and

$$U^2 = 1 + \frac{2}{\sqrt{n-2}}\Big(\frac{\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}\Big) + \frac{1}{n-2}\Big(\frac{3(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2}{(\boldsymbol{z}'\boldsymbol{z})^2} - \frac{2\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}\Big) + \mathrm{O}_p((n-2)^{-\frac{3}{2}}). \tag{15}$$

$\overline{\boldsymbol{V}}$ and $\boldsymbol{z}$ are independent, and so are $\frac{\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}$ and $\boldsymbol{z}'\boldsymbol{z}$ as well as $\frac{\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}$ and $\boldsymbol{z}'\boldsymbol{z}$. A proof of this result can be found in Fang et al.'s (1990) *Symmetric Multivariate and Related Distributions.* Taking expectations on both sides of (14) and (15) and using the established independence conditions, we easily obtain

$$\mathrm{E}[U] \approx 1 + \frac{1}{\sqrt{n-2}}\frac{\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}]}{\mathrm{E}[\boldsymbol{z}'\boldsymbol{z}]} + \frac{1}{n-2}\Big(\frac{\mathrm{E}[(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2]}{\mathrm{E}[(\boldsymbol{z}'\boldsymbol{z})^2]} - \frac{\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}]}{\mathrm{E}[\boldsymbol{z}'\boldsymbol{z}]}\Big) \text{ and} \tag{16}$$

$$\mathrm{E}[U^2] \approx 1 + \frac{2}{\sqrt{n-2}} \frac{\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}]}{\mathrm{E}[\boldsymbol{z}'\boldsymbol{z}]} + \frac{1}{n-2}\Big(3\frac{\mathrm{E}[(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2]}{\mathrm{E}[(\boldsymbol{z}'\boldsymbol{z})^2]} - 2\frac{\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}]}{\mathrm{E}[\boldsymbol{z}'\boldsymbol{z}]}\Big). \tag{17}$$

Recall that we already have $\boldsymbol{z}'\boldsymbol{z} \sim \chi^2_p$, so that $\mathrm{E}[\boldsymbol{z}'\boldsymbol{z}] = p$ and $\mathrm{E}[(\boldsymbol{z}'\boldsymbol{z})^2] = p(p+2)$. Hence, it remains for us to calculate the following 3 (three) quantities: $\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}]$, $\mathrm{E}[(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2]$ and $\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}]$.

The first quantity is actually very simple to compute. We just need to apply the law of total expectation and use the facts that $\overline{\boldsymbol{V}}$ and $\boldsymbol{z}$ are independent and $\mathrm{E}[\overline{\boldsymbol{V}}] = \boldsymbol{0}$ (since $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$ are both unbiased estimators for $\boldsymbol{\Sigma_1}$ and $\boldsymbol{\Sigma_2}$). These give us $\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}] = 0$. In order to compute the other two quantities, we need the following proposition:

Proposition 2.1. *Let* $\boldsymbol{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *For any symmetric* $p \times p$ *constant matrix* $\boldsymbol{A}$, *we have*

1. $\mathrm{E}[\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}] = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$ *and*

2. $\mathrm{Var}[\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}] = 2\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\mu}$.

Proof. For the first part, we note that by definition, $\boldsymbol{\Sigma} = \mathrm{E}[\boldsymbol{y}\boldsymbol{y}'] - \boldsymbol{\mu}\boldsymbol{\mu}'$. We also note that $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$ is a scalar, so that its expectation is the same as the expectation of its trace. We thus have

$$\begin{aligned}
\mathrm{E}[\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}] &= \mathrm{E}[\mathrm{tr}(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y})] \text{ since } \boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} \text{ is a scalar} \\
&= \mathrm{E}[\mathrm{tr}(\boldsymbol{A}\boldsymbol{y}\boldsymbol{y}')] \text{ by the cyclic property of the trace function} \\
&= \mathrm{tr}(\mathrm{E}[\boldsymbol{A}\boldsymbol{y}\boldsymbol{y}']) \text{ by the commutativity of the trace and expectation functions} \\
&= \mathrm{tr}(\boldsymbol{A}\mathrm{E}[\boldsymbol{y}\boldsymbol{y}']) \text{ since } \boldsymbol{A} \text{ is a constant matrix} \\
&= \mathrm{tr}(\boldsymbol{A}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}')) \text{ since } \boldsymbol{\Sigma} = \mathrm{E}[\boldsymbol{y}\boldsymbol{y}'] - \boldsymbol{\mu}\boldsymbol{\mu}' \text{ by the definition of } \boldsymbol{\Sigma} \\
&= \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \mathrm{tr}(\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}) \text{ by the properties of the trace function} \\
&= \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} \text{ since } \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} \text{ is a scalar.}
\end{aligned}$$

For the second part, see Rencher and Schaalje's (2008) *Linear Models in Statistics.*

□

We also have the following corollary:

Corollary 2.1.1. *Let* $\boldsymbol{u} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{I_p})$ *and* $\boldsymbol{A}$ *be a symmetric* $p \times p$ *constant matrix. We have*

1. $\mathrm{E}[\boldsymbol{u}'\boldsymbol{A}\boldsymbol{u}] = \mathrm{tr}(\boldsymbol{A})$ *and*

2. $\mathrm{E}[(\boldsymbol{u}'\boldsymbol{A}\boldsymbol{u})^2] = 2\mathrm{tr}(\boldsymbol{A}^2) + (\mathrm{tr}(\boldsymbol{A}))^2$.

Proof. Simply use Proposition 2.1 with the substitutions $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I_p}$. □

We also need the following proposition taken from Gupta and Nagar's (1999) *Matrix Variate Distributions*:

Proposition 2.2. *If* $\boldsymbol{T} \sim \mathcal{W}_p(n-1, \boldsymbol{\Sigma})$ *and* $\boldsymbol{A}$ *is a* $p \times p$ *constant matrix, the following holds:*

1. $\mathrm{E}[\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}] = (n-1)\boldsymbol{\Sigma}\boldsymbol{A}'\boldsymbol{\Sigma} + (n-1)\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A})\Sigma + (n-1)^2\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}$ *and*

2. $\mathrm{E}[\mathrm{tr}(\boldsymbol{A}\boldsymbol{T})\boldsymbol{T}] = (n-1)\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma} + (n-1)\boldsymbol{\Sigma}\boldsymbol{A}'\boldsymbol{\Sigma} + (n-1)^2\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})\boldsymbol{\Sigma}$.

The following corollary follows from Proposition 2.2:

Corollary 2.2.1. *Let* $(n-1)\boldsymbol{S} \sim \mathcal{W}_p(n-1, \boldsymbol{\Sigma})$*,* $\boldsymbol{V} = \sqrt{n-1}(\boldsymbol{S} - \boldsymbol{\Sigma})$ *and* $\boldsymbol{A}$ *be a* $p \times p$ *symmetric constant matrix. We have*

1. $\mathrm{E}[(\mathrm{tr}(\boldsymbol{A}\boldsymbol{V}))^2] = 2\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma})$ *and*

2. $\mathrm{E}[\mathrm{tr}(\boldsymbol{A}\boldsymbol{V}\boldsymbol{A}\boldsymbol{V})] = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}) + (\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}))^2$.

Proof. To show the first part, see that

$$\begin{aligned}
\mathrm{E}[(\mathrm{tr}(\boldsymbol{A}\boldsymbol{V}))^2] &= (n-1)\mathrm{E}[(\mathrm{tr}(\boldsymbol{A}\boldsymbol{S} - \boldsymbol{A}\boldsymbol{\Sigma}))^2] \\
&= (n-1)\mathrm{E}[(\mathrm{tr}(\boldsymbol{A}\boldsymbol{S}) - \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}))^2] \\
&= (n-1)(\mathrm{E}[(\mathrm{tr}(\boldsymbol{A}\boldsymbol{S}))^2] - 2\mathrm{E}[\mathrm{tr}(\boldsymbol{A}\boldsymbol{S})]\mathrm{E}[\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})] + \mathrm{E}[(\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}))^2]).
\end{aligned}$$

Notice also that since $\boldsymbol{\Sigma}$ is a constant matrix,

$$\mathrm{E}[\mathrm{tr}(\boldsymbol{AS})] = \mathrm{tr}(\mathrm{E}[\boldsymbol{AS}]) = \mathrm{tr}(\boldsymbol{A}\mathrm{E}[\boldsymbol{S}]) = \mathrm{tr}(\boldsymbol{A\Sigma}) = \mathrm{E}[\mathrm{tr}(\boldsymbol{A\Sigma})].$$

Therefore, $\mathrm{E}[(\mathrm{tr}(\boldsymbol{AV}))^2] = (n-1)(\mathrm{E}[(\mathrm{tr}(\boldsymbol{AS}))^2] - (\mathrm{tr}(\boldsymbol{A\Sigma}))^2)$.

Now, see that

$$\mathrm{E}[(\mathrm{tr}(\boldsymbol{AS}))^2] = \mathrm{E}[\mathrm{tr}(\mathrm{tr}(\boldsymbol{AS})\boldsymbol{SA})] = \mathrm{tr}(\mathrm{E}[\mathrm{tr}(\boldsymbol{AS})\boldsymbol{SA}]) = \mathrm{tr}(\mathrm{E}[\mathrm{tr}(\boldsymbol{AS})\boldsymbol{S}]\boldsymbol{A}).$$

By Proposition 2.2(2),

$$\mathrm{E}[\mathrm{tr}(\boldsymbol{AS})\boldsymbol{S}] = \frac{\boldsymbol{\Sigma A \Sigma}}{n-1} + \frac{\boldsymbol{\Sigma A' \Sigma}}{n-1} + \mathrm{tr}(\boldsymbol{A\Sigma})\boldsymbol{\Sigma} = \frac{2\boldsymbol{\Sigma A \Sigma}}{n-1} + \mathrm{tr}(\boldsymbol{A\Sigma})\boldsymbol{\Sigma}$$

since $\boldsymbol{A}$ is symmetric and $(n-1)\boldsymbol{S} \sim \mathcal{W}_p(n-1, \boldsymbol{\Sigma})$. Hence,

$$\mathrm{E}[(\mathrm{tr}(\boldsymbol{AS}))^2] = \mathrm{tr}\Big(\frac{2\boldsymbol{\Sigma A \Sigma A}}{n-1} + \mathrm{tr}(\boldsymbol{A\Sigma})\boldsymbol{\Sigma A}\Big) = \frac{2}{n-1}\mathrm{tr}(\boldsymbol{\Sigma A \Sigma A}) + (\mathrm{tr}(\boldsymbol{A\Sigma}))^2.$$

This, together with the last result in the previous paragraph, clearly prove the first part of Corollary 2.2.1.

To show the second part, see that

$$\begin{aligned}
\mathrm{E}[\mathrm{tr}(\boldsymbol{AVAV})] &= (n-1)\mathrm{E}[\mathrm{tr}(\boldsymbol{A}(\boldsymbol{S}-\boldsymbol{\Sigma})\boldsymbol{A}(\boldsymbol{S}-\boldsymbol{\Sigma}))] \\
&= (n-1)\mathrm{E}[\mathrm{tr}(\boldsymbol{ASAS}) - \mathrm{tr}(\boldsymbol{ASA\Sigma}) - \mathrm{tr}(\boldsymbol{A\Sigma AS}) + \mathrm{tr}(\boldsymbol{A\Sigma A\Sigma})] \\
&= (n-1)(\mathrm{E}[\mathrm{tr}(\boldsymbol{ASAS})] - \mathrm{tr}(\boldsymbol{A\Sigma A\Sigma}))
\end{aligned}$$

since

$$\mathrm{E}[\mathrm{tr}(\boldsymbol{ASA\Sigma})] = \mathrm{E}[\mathrm{tr}(\boldsymbol{A\Sigma AS})] = \mathrm{tr}(\mathrm{E}[\boldsymbol{A\Sigma AS}]) = \mathrm{tr}(\boldsymbol{A\Sigma A}\mathrm{E}[\boldsymbol{S}]) = \mathrm{tr}(\boldsymbol{A\Sigma A\Sigma}).$$

By Proposition 2.2(1),

$$\begin{aligned}\mathrm{E}[\mathrm{tr}(\boldsymbol{ASAS})] &= \mathrm{tr}(\mathrm{E}[\boldsymbol{ASAS}]) = \mathrm{tr}(\boldsymbol{A}\mathrm{E}[\boldsymbol{SAS}]) \\ &= \mathrm{tr}\Big(\frac{\boldsymbol{A\Sigma A\Sigma}}{n-1} + \frac{\boldsymbol{A}\mathrm{tr}(\boldsymbol{\Sigma A})\boldsymbol{\Sigma}}{n-1} + \boldsymbol{A\Sigma A\Sigma}\Big) \\ &= \frac{n}{n-1}\mathrm{tr}(\boldsymbol{A\Sigma A\Sigma}) + \frac{(\mathrm{tr}(\boldsymbol{A\Sigma}))^2}{n-1}\end{aligned}$$

since $\boldsymbol{A}$ is symmetric and $(n-1)\boldsymbol{S} \sim \mathcal{W}_p(n-1, \boldsymbol{\Sigma})$. Therefore,

$$\mathrm{E}[\mathrm{tr}(\boldsymbol{AVAV})] = (n-1)\Big(\frac{1}{n-1}\mathrm{tr}(\boldsymbol{A\Sigma A\Sigma}) + \frac{(\mathrm{tr}(\boldsymbol{A\Sigma}))^2}{n-1}\Big) = \mathrm{tr}(\boldsymbol{A\Sigma A\Sigma}) + (\mathrm{tr}(\boldsymbol{A\Sigma}))^2,$$

completing the proof. □

We now present Proposition 2.3, which pertains to the value of the quantities $\mathrm{E}[(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2]$ and $\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}]$.

Proposition 2.3. *Let*

$$\psi_1 = \frac{n_2^2(n-2)}{n^2(n_1-1)}[\mathrm{tr}(\boldsymbol{\Sigma_1}\overline{\boldsymbol{\Sigma}}^{-1})]^2 + \frac{n_1^2(n-2)}{n^2(n_2-1)}[\mathrm{tr}(\boldsymbol{\Sigma}_2\overline{\boldsymbol{\Sigma}}^{-1})]^2 \ \textit{and}$$

$$\psi_2 = \frac{n_2^2(n-2)}{n^2(n_1-1)}\mathrm{tr}(\boldsymbol{\Sigma_1}\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_1}\overline{\boldsymbol{\Sigma}}^{-1}) + \frac{n_1^2(n-2)}{n^2(n_2-1)}\mathrm{tr}(\boldsymbol{\Sigma_2}\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_2}\overline{\boldsymbol{\Sigma}}^{-1}).$$

Then, under H_0, $\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}] = \psi_1 + \psi_2$ *and* $\mathrm{E}[(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2] = 2\psi_1 + 4\psi_2$.

Proof. Since $\overline{\boldsymbol{y}}_{\boldsymbol{1}} \sim \mathcal{N}_p(\boldsymbol{\mu_1}, \frac{\boldsymbol{\Sigma_1}}{n_1})$ and $\overline{\boldsymbol{y}}_{\boldsymbol{2}} \sim \mathcal{N}_p(\boldsymbol{\mu_2}, \frac{\boldsymbol{\Sigma_2}}{n_2})$ and they are independent, $\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}} \sim \mathcal{N}_p(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}, \frac{\boldsymbol{\Sigma_1}}{n_1} + \frac{\boldsymbol{\Sigma_2}}{n_2})$. This means under H_0, $\boldsymbol{z} = \sqrt{\frac{n_1 n_2}{n}}\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}) \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{I_p})$. Since $\boldsymbol{z}$ and $\overline{\boldsymbol{V}}$ are independent, using Corollary 2.1.1(1) and the law of total expectation, we obtain $\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}] = \mathrm{E}[\mathrm{tr}(\overline{\boldsymbol{V}}^2)]$. Moreover, applying Corollary 2.2.1(2) with the substitutions $\boldsymbol{S} := \boldsymbol{S_1}$, $n := n_1$, $\boldsymbol{\Sigma} := \boldsymbol{\Sigma_1}$ and $\boldsymbol{A} := \overline{\boldsymbol{\Sigma}}^{-1}$, we have

$$\mathrm{E}[\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{S_1} - \boldsymbol{\Sigma_1})\overline{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{S_1} - \boldsymbol{\Sigma_1}))] = \frac{\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_1}\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_1}) + (\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_1}))^2}{n_1 - 1}.$$

Similarly,

$$\mathrm{E}[\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{S_2}-\boldsymbol{\Sigma_2})\overline{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{S_2}-\boldsymbol{\Sigma_2}))] = \frac{\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_2}\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_2}) + (\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_2}))^2}{n_2-1}.$$

Now, by the definition of $\overline{\boldsymbol{V}}$ in (13), we have $\mathrm{E}[\mathrm{tr}(\overline{\boldsymbol{V}}^2)] = (n-2)(P+Q+R+S)$, where

$$P = \mathrm{E}\big[\mathrm{tr}\big(\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_2}{n}(\boldsymbol{S_1}-\boldsymbol{\Sigma_1})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_2}{n}(\boldsymbol{S_1}-\boldsymbol{\Sigma_1})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\big)\big],$$

$$Q = \mathrm{E}\big[\mathrm{tr}\big(\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_1}{n}(\boldsymbol{S_2}-\boldsymbol{\Sigma_2})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_1}{n}(\boldsymbol{S_2}-\boldsymbol{\Sigma_2})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\big)\big],$$

$$R = \mathrm{E}\big[\mathrm{tr}\big(\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_2}{n}(\boldsymbol{S_1}-\boldsymbol{\Sigma_1})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_1}{n}(\boldsymbol{S_2}-\boldsymbol{\Sigma_2})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\big)\big] \text{ and}$$

$$S = \mathrm{E}\big[\mathrm{tr}\big(\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_1}{n}(\boldsymbol{S_2}-\boldsymbol{\Sigma_2})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_2}{n}(\boldsymbol{S_1}-\boldsymbol{\Sigma_1})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\big)\big].$$

Using the results from the previous paragraph and the cyclic property of the trace function, we have

$$P = \frac{n_2^2}{n^2(n_1-1)}[\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_1}\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_1}) + (\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_1}))^2].$$

Similarly,

$$Q = \frac{n_1^2}{n^2(n_2-1)}[\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_2}\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_2}) + (\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_2}))^2].$$

Notice that $R = S = 0$ since $\boldsymbol{S_1}-\boldsymbol{\Sigma_1}$ and $\boldsymbol{S_2}-\boldsymbol{\Sigma_2}$ are independent and $\mathrm{E}[\boldsymbol{S_i}] = \boldsymbol{\Sigma_i}$ for $i \in \{1,2\}$. Therefore, $\mathrm{E}[\mathrm{tr}(\overline{\boldsymbol{V}}^2)] = (n-2)(P+Q) = \psi_1+\psi_2$ using the fact that $\mathrm{tr}(\boldsymbol{Z}) = \mathrm{tr}(\boldsymbol{Z}')$ for every square matrix $\boldsymbol{Z}$. This proves the first part of the proposition.

In order to prove the second part, notice that Corollary 2.1.1(2) and the law of total expectation give us

$$\mathrm{E}[(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2] = 2\mathrm{E}[\mathrm{tr}(\overline{\boldsymbol{V}}^2)] + \mathrm{E}[(\mathrm{tr}(\overline{\boldsymbol{V}}))^2] = 2(\psi_1+\psi_2) + \mathrm{E}[(\mathrm{tr}(\overline{\boldsymbol{V}}))^2].$$

Therefore, it suffices to show that $\mathrm{E}[(\mathrm{tr}(\overline{\boldsymbol{V}}))^2] = 2\psi_2$. By Corollary 2.2.1(1) and the same substitutions as above, we have

$$\mathrm{E}[(\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\sqrt{n_1-1}(\boldsymbol{S_1}-\boldsymbol{\Sigma_1})))^2] = 2\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_1}\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_1}) \text{ and}$$

$$\mathrm{E}[(\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\sqrt{n_2-1}(\boldsymbol{S_2}-\boldsymbol{\Sigma_2})))^2] = 2\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_2}\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_2}).$$

By the definition of $\overline{\boldsymbol{V}}$ in (13), we have $\mathrm{E}[(\mathrm{tr}(\overline{\boldsymbol{V}}))^2] = (n-2)(T+U+2V)$, where

$$T = \mathrm{E}\big[\big(\mathrm{tr}\big(\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_2}{n}(\boldsymbol{S_1}-\boldsymbol{\Sigma_1})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\big)\big)^2\big],$$

$$U = \mathrm{E}\big[\big(\mathrm{tr}\big(\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_1}{n}(\boldsymbol{S_2}-\boldsymbol{\Sigma_2})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\big)\big)^2\big] \text{ and}$$

$$V = \mathrm{E}\big[\mathrm{tr}\big(\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_2}{n}(\boldsymbol{S_1}-\boldsymbol{\Sigma_1})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\big)\mathrm{tr}\big(\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\frac{n_1}{n}(\boldsymbol{S_2}-\boldsymbol{\Sigma_2})\overline{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\big)\big].$$

See that $V = 0$ as $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$ are independent and $E(\boldsymbol{S_i}) = \boldsymbol{\Sigma_i}$ for $i \in \{1,2\}$. Moreover, using the results from the previous paragraph and the cyclic property of the trace function, we obtain

$$T = \frac{2n_2^2}{n^2(n_1-1)}\mathrm{tr}(\boldsymbol{\Sigma_1}\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_1}\overline{\boldsymbol{\Sigma}}^{-1}) \text{ and } U = \frac{2n_1^2}{n^2(n_2-1)}\mathrm{tr}(\boldsymbol{\Sigma_2}\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma_2}\overline{\boldsymbol{\Sigma}}^{-1}).$$

Combined with the fact that $\mathrm{tr}(\boldsymbol{Z}) = \mathrm{tr}(\boldsymbol{Z}')$ for every square matrix $\boldsymbol{Z}$, we have proven the second part of the proposition. We are done. □

Using Proposition 2.3, (16) and (17) become

$$\mathrm{E}[U] \approx 1 - \frac{\theta_1}{n-2} \text{ and} \tag{18}$$

$$\mathrm{E}[U^2] \approx 1 - \frac{2}{n-2}(\theta_1-\theta_2), \tag{19}$$

where

$$\theta_1 = \frac{p\psi_1 + (p-2)\psi_2}{p(p+2)} \text{ and } \theta_2 = \frac{\psi_1 + 2\psi_2}{p(p+2)}. \tag{20}$$

By equating (18) and (19) with (12), we obtain

$$v = \frac{(n-2-\theta_1)^2}{(n-2)\theta_2 - \frac{\theta_1^2}{2}} \text{ and } \phi = \frac{v(n-2)}{n-2-\theta_1}. \quad (21)$$

In order to make (11) exact when $\Sigma_1 = \Sigma_2$ and $n_1 = n_2$, we make a slight adjustment to v. It now becomes

$$v = \frac{(n-2-\theta_1)^2}{(n-2)\theta_2 - \theta_1}. \quad (22)$$

An F statistic is obtained by using consistent estimates of v and ϕ. We let

$$\hat{\psi}_1 = \frac{n_2^2(n-2)}{n^2(n_1-1)}(\mathrm{tr}(\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}))^2 + \frac{n_1^2(n-2)}{n^2(n_2-1)}(\mathrm{tr}(\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}))^2 \text{ and} \quad (23)$$

$$\hat{\psi}_2 = \frac{n_2^2(n-2)}{n^2(n_1-1)}\mathrm{tr}(\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}) + \frac{n_1^2(n-2)}{n^2(n_2-1)}\mathrm{tr}(\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}), \quad (24)$$

where $\overline{\boldsymbol{S}} = \frac{n_2}{n}\boldsymbol{S_1} + \frac{n_1}{n}\boldsymbol{S_2}$. The F statistic is

$$T_F = \frac{n-2-\hat{\theta}_1}{(n-2)p}T \overset{a}{\sim} F_{p,\hat{v}}, \quad (25)$$

where $\hat{\theta}_1 = \frac{p\hat{\psi}_1+(p-2)\hat{\psi}_2}{p(p+2)}$, $\hat{\theta}_2 = \frac{\hat{\psi}_1+2\hat{\psi}_2}{p(p+2)}$ and $\hat{v} = \frac{(n-2-\hat{\theta}_1)^2}{(n-2)\hat{\theta}_2-\hat{\theta}_1}$. This result concludes this subsection.

2.3 Simulation Studies

This subsection summarises the simulation studies conducted by Yanagihara and Yuan (2005) in their research paper and provides additional insights relating to the performance of the methods. As mentioned in Section 2.1, the authors compared the 3 (three) methods they developed with 5 (five) other methods. The comparison metric used is the Type I error (empirical size), i.e. the probability of rejecting a true null hypothesis. For example, the Type I error of Yanagihara and Yuan's (2005) main method (as mentioned in (25)) is $\mathrm{P}(T_F > u_\alpha(p, \hat{v}))$, where $u_\alpha(p, \hat{v})$ is the upper α critical value for an F distribution with p and $\hat{v}$ degrees of freedom.

The authors chose the nominal size $\alpha \in \{0.1, 0.05, 0.01\}$ and performed simulations on numerous combinations of p, n_1, n_2, $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma_2}$. For each combination, the Type I error is estimated by averaging over 30,000 replications. The authors assumed (WLOG, as shown by Yao (1965)) that $\mathbf{\Sigma_1} = \text{diag}(\lambda_1, \cdots, \lambda_p)$, where $0 < \lambda_1 \leq \cdots \leq \lambda_p < 1$ and $\mathbf{\Sigma_2} = \boldsymbol{I_p} - \mathbf{\Sigma_1}$. The quantity called the average absolute discrepancy (AAD) was introduced. This measures the average of the differences between the nominal and empirical sizes over a combination of conditions. The figure below shows one part of the reproduced simulation results. Here, the nominal size is fixed

alpha = 0.1											
p	n1	n2	λ	T	T_F	T_B	T_MB	T2	T_FY	T_FJ	T_FM
2	10	10	(1)	0.1712	0.0994	0.1061	0.1039	0.1141	0.1042	0.1039	0.1053
			(2)	0.15	0.0962	0.0998	0.0949	0.1076	0.0978	0.102	0.0973
			(3)	0.1472	0.0973	0.1014	0.0968	0.1057	0.0959	0.098	0.0932
			(4)	0.1523	0.098	0.0988	0.0981	0.1053	0.0897	0.1022	0.099
			(5)	0.1472	0.0928	0.1015	0.0921	0.1047	0.0984	0.0717	0.0953
		20	(1)	0.1312	0.0976	0.1029	0.0995	0.1056	0.0995	0.1013	0.1007
			(2)	0.1323	0.099	0.0995	0.0955	0.1017	0.0995	0.0997	0.0996
			(3)	0.1388	0.0992	0.1016	0.1006	0.1044	0.1009	0.099	0.0995
			(4)	0.1432	0.0987	0.1016	0.1008	0.1061	0.0969	0.1008	0.0995
			(5)	0.1445	0.1004	0.0995	0.1015	0.1041	0.1051	0.1012	0.1012
			(6)	0.1785	0.1024	0.1069	0.1024	0.116	0.106	0.1093	0.105
8	20	20	(i)	0.3618	0.0986	0.1338	0.1138	0.1748	0.131	0.1432	0.1056
			(ii)	0.2571	0.0965	0.1153	0.1035	0.1297	0.0945	0.1128	0.0987
			(iii)	0.2553	0.0951	0.1129	0.096	0.1256	0.0916	0.1096	0.0957
			(iv)	0.3159	0.1033	0.1328	0.1096	0.153	0.1121	0.1334	0.1084
			(v)	0.2805	0.099	0.1212	0.1082	0.1366	0.1015	0.1192	0.1018
			(vi)	0.251	0.0973	0.1153	0.1014	0.1303	0.0955	0.1113	0.0959
		40	(i)	0.2004	0.0976	0.1115	0.1008	0.113	0.1002	0.1068	0.1018
			(ii)	0.2329	0.1013	0.1207	0.1049	0.1238	0.1024	0.1111	0.1034
			(iii)	0.2375	0.1014	0.12	0.1023	0.1215	0.117	0.1159	0.1016
			(iv)	0.1952	0.0973	0.1135	0.0978	0.1128	0.096	0.1052	0.0984
			(v)	0.2001	0.1009	0.1166	0.0989	0.1109	0.0993	0.1018	0.093
			(vi)	0.2359	0.0997	0.121	0.1052	0.1236	0.1108	0.1151	0.1044
			(vii)	0.4009	0.0912	0.1398	0.1144	0.1867	0.1368	0.1563	0.1061
AAD				0.1109	0.0025	0.0125	0.0043	0.0216	0.0072	0.0123	0.0033
AAD from paper				0.1113	0.0025	0.0132	0.0041	0.0277	0.0074	0.0122	0.0033

to be 0.1. T, T_F, T_B, T_{MB}, T_2, T_{FY}, T_{FJ} and T_{FM} refer to the 8 (eight) test considered mentioned in the same order as in Section 2.1. (1), (2), (3), (4), (5) and (6) refer to c(0.1, 0.1), c(0.2, 0.5), c(0.2, 0.7), c(0.1, 0.9), c(0.5, 0.5) and c(0.9, 0.9) respectively;

while (i), (ii), (iii), (iv), (v), (vi) and (vii) refer to $c(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)$, $c(0.1, 0.1, 0.1, 0.5, 0.5, 0.9, 0.9, 0.9)$, $c(0.5, 0.5, 0.5, 0.5, 0.5, 0, 5, 0.5, 0.5)$, $c(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.9)$,$c(0.1, 0.1, 0.1, 0.1, 0.5, 0.5, 0.5, 0.5)$, $c(0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.9)$ and $c(0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9)$ respectively. The black-coloured numbers refer to the Type I errors, and the red-coloured numbers refer to the AADs for all 8 (eight) methods. It can be seen that the AADs of our simulation match those of the authors' (the blue-coloured numbers).

It is clear that a small AAD is preferred since it means that on average, the difference between the empirical size and the fixed nominal size is small. From the figure above, we see that Yanagihara and Yuan's (2005) main method gives the smallest AAD, followed by Krishnamoorthy and Yu's (2004) method. The exact conclusion is obtained when the nominal size is fixed to be 0.05 or 0.01.

After doing more thorough simulations, it is found that T_F performs very badly when $\mathbf{\Sigma_1}$ is small, $\mathbf{\Sigma_2}$ is large and $\frac{n_1}{n_2}$ is large, while the performance of T_{FM} is more stable across all conditions. Some examples are given below.

- When $\alpha = 0.1$, $n_1 = 100$, $n_2 = 10$, $\mathbf{\Sigma_1} = \text{diag}(\text{rep}(0.01, 8))$ and $\mathbf{\Sigma_2} = \text{diag}(\text{rep}(0.99, 8))$, the empirical size of T_F is 0.022 and that of T_{FM} is 0.139.

- When $\alpha = 0.1$, $n_1 = 70$, $n_2 = 7$, $\mathbf{\Sigma_1} = \text{diag}(\text{rep}(0.05, 4))$ and $\mathbf{\Sigma_2} = \text{diag}(\text{rep}(0.95, 4))$, the empirical size of T_F is 0.048 and that of T_{FM} is 0.111.

These results show that perhaps T_{FM} performs better than T_F in general.

Apart from comparing Type I errors, it is also useful to compare the powers of the methods. The power of a test is defined as the probability of correctly rejecting a false null hypothesis, which is one minus the Type II error probability. Even though that the simple chi-square approximation method gives the best power, it is not preferred as it results in a large AAD and does not take the difference of the covariance matrices into account. Comparing the powers of Yanagihara and Yuan's (2005) main method and Krishnamoorthy and Yu's (2004) method does not give any valuable insights other than that their power graphs nearly coincide.

3 A More General Case of the MBF Problem

3.1 The General Linear Hypothesis Testing (GLHT) Problem in Heteroscedastic One-Way MANOVA

Recall that for the MBF problem, we have the following formulation:

- $\boldsymbol{y_{11}}, \boldsymbol{y_{21}}, \cdots, \boldsymbol{y_{n_1 1}}$ is an i.i.d. sample drawn from $\mathcal{N}_p(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})$.
- $\boldsymbol{y_{12}}, \boldsymbol{y_{22}}, \cdots, \boldsymbol{y_{n_2 2}}$ is an i.i.d. sample drawn from $\mathcal{N}_p(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2})$.
- $\boldsymbol{\Sigma_1} \neq \boldsymbol{\Sigma_2}$.
- The null hypothesis is $H_0 : \boldsymbol{\mu_1} = \boldsymbol{\mu_2}$.

We can generalise the MBF problem by including more than 2 (two) i.i.d. samples and allowing the null hypothesis to be of the form $\boldsymbol{C\mu} = \boldsymbol{c}$, where $\boldsymbol{\mu} = [\boldsymbol{\mu_1'}, \boldsymbol{\mu_2'}, \cdots, \boldsymbol{\mu_k'}]'$. Consider the following problem formulation, which is known as the general linear hypothesis testing (GLHT) problem in heteroscedastic one-way MANOVA:

- $\boldsymbol{y_{11}}, \boldsymbol{y_{21}}, \cdots, \boldsymbol{y_{n_1 1}}$ is an i.i.d. sample drawn from $\mathcal{N}_p(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})$.
- $\boldsymbol{y_{12}}, \boldsymbol{y_{22}}, \cdots, \boldsymbol{y_{n_2 2}}$ is an i.i.d. sample drawn from $\mathcal{N}_p(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2})$.
- $\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$
- $\boldsymbol{y_{1k}}, \boldsymbol{y_{2k}}, \cdots, \boldsymbol{y_{n_k k}}$ is an i.i.d. sample drawn from $\mathcal{N}_p(\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})$.
- $\boldsymbol{\Sigma_i} \neq \boldsymbol{\Sigma_j}$ for every $i, j \in \{1, 2, \cdots, k\}$ where $i \neq j$.
- The null hypothesis is $H_0 : \boldsymbol{C\mu} = \boldsymbol{c}$, where $\boldsymbol{\mu} = [\boldsymbol{\mu_1'}, \boldsymbol{\mu_2'}, \cdots, \boldsymbol{\mu_k'}]'$ is a vector of length kp, $\boldsymbol{C}$ is a $q \times kp$ matrix with rank q and $\boldsymbol{c}$ is a vector of length q.

It can be easily seen that the MBF problem is just a special case of the general setting described previously. In particular, setting $k := 2$, $\boldsymbol{c} := \boldsymbol{0_p}$ and $\boldsymbol{C} := [1\ -1] \otimes \boldsymbol{I_p}$ give us the MBF problem. Here, $\otimes$ denotes the Kronecker product.

3.2 Zhang's (2012) Generalisation of Krishnamoorthy and Yu's (2004) Method for the GLHT Problem

In his paper titled *An Approximate Hotelling T^2-Test for Heteroscedastic One-Way MANOVA*, Zhang (2012) generalised Krishnamoorthy and Yu's (2004) method to obtain an approximate solution to the GLHT problem. This subsection explains the idea behind Zhang's (2012) method.

For each $l \in \{1, 2, \cdots, k\}$, let $\hat{\boldsymbol{\mu}}_{\boldsymbol{l}}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{l}}$ be sample mean and sample variance, respectively. We also let $\hat{\boldsymbol{\mu}} = [\hat{\boldsymbol{\mu}}'_{\boldsymbol{1}}, \hat{\boldsymbol{\mu}}'_{\boldsymbol{2}}, \cdots, \hat{\boldsymbol{\mu}}'_{\boldsymbol{k}}]'$. Then $\hat{\boldsymbol{\mu}} \sim \mathcal{N}_{kp}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \text{diag}(\frac{\boldsymbol{\Sigma_1}}{n_1}, \frac{\boldsymbol{\Sigma_2}}{n_2}, \cdots, \frac{\boldsymbol{\Sigma_k}}{n_k})$ is a $kp \times kp$ matrix. This means $\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c} \sim \mathcal{N}_q(\boldsymbol{C}\boldsymbol{\mu} - \boldsymbol{c}, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')$, suggesting the following Wald-type test statistic:

$$T = (\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})'(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c}), \tag{26}$$

where $\hat{\boldsymbol{\Sigma}} = \text{diag}(\frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{1}}}{n_1}, \frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{2}}}{n_2}, \cdots, \frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}}{n_k})$.

We write $\boldsymbol{z} = (\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-\frac{1}{2}}(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})$ and $\boldsymbol{W} = (\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-\frac{1}{2}}(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-\frac{1}{2}}$. It can be easily seen that under the null hypothesis, $\boldsymbol{z}'\boldsymbol{z} \sim \chi^2_q$. Notice also that $T = \boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z}$ and $\boldsymbol{z} \sim \mathcal{N}_q(\boldsymbol{\mu_z}, \boldsymbol{I_q})$, where $\boldsymbol{\mu_z} = (\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-\frac{1}{2}}(\boldsymbol{C}\boldsymbol{\mu} - \boldsymbol{c})$.

The main idea of this method is to express $\boldsymbol{W}$ as a Wishart mixture, i.e. a linear combination of several independent Wishart random matrices, then approximate $\boldsymbol{W}$ with a single Wishart random matrix $\boldsymbol{R} \sim \mathcal{W}_q(d, \boldsymbol{\Omega})$. Since there are 2 (two) unknown parameters, we need to establish 2 (two) equations in order to uniquely determine the value of the unknown parameters. In this case, we match the expectation and total variation of $\boldsymbol{W}$ and $\boldsymbol{R}$. The total variation $\text{V}[\boldsymbol{X}]$ of a random matrix $\boldsymbol{X}$ is defined as the sum of the variances of all the entries of $\boldsymbol{X}$.

Recall that from Proposition 1.1, we have that if $\boldsymbol{Y} \sim \mathcal{W}_p(n, \boldsymbol{V})$, then $\text{E}[\boldsymbol{Y}] = n\boldsymbol{V}$ and $\text{V}[\boldsymbol{Y}] = n(\text{tr}(\boldsymbol{V}^2) + (\text{tr}(\boldsymbol{V}))^2)$. In order to derive this method, we first express $\boldsymbol{C}$ as $[\boldsymbol{C_1}, \boldsymbol{C_2}, \cdots, \boldsymbol{C_k}]$, where $\boldsymbol{C_i}$ is a $q \times p$ matrix for every $i \in \{1, 2, \cdots, k\}$. We also set $\boldsymbol{H_l} = (\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-\frac{1}{2}}\boldsymbol{C_l}$ for every $l \in \{1, 2, \cdots, k\}$. This means $\boldsymbol{H} =$

$(\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}\boldsymbol{C} = [\boldsymbol{H_1}, \boldsymbol{H_2}, \cdots, \boldsymbol{H_k}]$.

Now, see that $\boldsymbol{W} = \boldsymbol{H\hat{\Sigma}H'} = \sum_{l=1}^{k} \boldsymbol{W_l}$, where $\boldsymbol{W_l} = n_l^{-1}\boldsymbol{H_l\hat{\Sigma}_l H_l'}$ for every $l \in \{1, 2, \cdots, k\}$. Recall that $\boldsymbol{\hat{\Sigma}_l} \sim W_p(n_l - 1, \frac{\boldsymbol{\Sigma_l}}{n_l-1})$ for every $l \in \{1, 2, \cdots, k\}$, which implies $\boldsymbol{W_l} \sim W_q(n_l - 1, \frac{\boldsymbol{\Omega_l}}{n_l-1})$. Here, $\boldsymbol{\Omega_l} = \mathrm{E}[\boldsymbol{W_l}] = n_l^{-1}\boldsymbol{H_l\Sigma_l H_l'}$. Also, we have $\mathrm{V}[\boldsymbol{W_l}] = \frac{\mathrm{tr}(\boldsymbol{\Omega_l^2}) + (\mathrm{tr}(\boldsymbol{\Omega_l}))^2}{n_l-1}$.

Since the $\boldsymbol{W_l}$'s are independent, we obtain

$$\mathrm{E}[\boldsymbol{W}] = \sum_{l=1}^{k} \boldsymbol{\Omega_l} = \boldsymbol{H\Sigma H'} = \boldsymbol{I_q} \text{ and } \mathrm{V}[\boldsymbol{W}] = \sum_{l=1}^{k} \frac{\mathrm{tr}(\boldsymbol{\Omega_l^2}) + (\mathrm{tr}(\boldsymbol{\Omega_l}))^2}{n_l - 1}. \tag{27}$$

By equating (27) with the expectation and total variation of $\boldsymbol{R} \sim \mathcal{W}_q(d, \boldsymbol{\Omega})$, we easily obtain

$$\boldsymbol{\Omega} = \frac{\boldsymbol{I_q}}{d} \text{ and } d = \frac{q(q+1)}{\sum_{l=1}^{k} \frac{\mathrm{tr}(\boldsymbol{\Omega_l^2}) + (\mathrm{tr}(\boldsymbol{\Omega_l}))^2}{n_l-1}}. \tag{28}$$

We are almost done. Recall that we have obtained the approximation $\boldsymbol{W} \overset{a}{\sim} \mathcal{W}_q(d, \boldsymbol{\Omega})$. From Proposition 1.2, we derive that under H_0, $T = \boldsymbol{z'W}^{-1}\boldsymbol{z} \overset{a}{\sim} \mathcal{T}^2(q, d)$. Using Proposition 1.4, equivalently we have

$$\frac{d - q + 1}{qd} T \overset{a}{\sim} \mathcal{F}_{q, d-q+1}. \tag{29}$$

In reality, the $\boldsymbol{\Omega_l}$'s are unknown; they are replaced by their estimators

$$\boldsymbol{\hat{\Omega}_l} = n_l^{-1}(\boldsymbol{C\hat{\Sigma}C'})^{-\frac{1}{2}}\boldsymbol{C_l\hat{\Sigma}_l C_l'}(\boldsymbol{C\hat{\Sigma}C'})^{-\frac{1}{2}}. \tag{30}$$

The test statistic is therefore

$$T_{FM} = \frac{\hat{d} - q + 1}{q\hat{d}} T \overset{a}{\sim} \mathcal{F}_{q, \hat{d}-q+1}, \tag{31}$$

where $\hat{d}$ is obtained from d by replacing each occurrence of $\boldsymbol{\Omega_l}$ with $\boldsymbol{\hat{\Omega}_l}$.

3.3 A Generalisation of Yanagihara and Yuan's (2005) Main Method for the GLHT Problem

In this subsection, we apply Zhang's idea to generalise Yanagihara and Yuan's (2005) main method to deal with the GLHT problem.

For each $l \in \{1, 2, \cdots, k\}$, let $\hat{\boldsymbol{\mu}}_{\boldsymbol{l}}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{l}}$ be sample mean and sample variance, respectively. We also let $\hat{\boldsymbol{\mu}} = [\hat{\boldsymbol{\mu}}'_{\boldsymbol{1}}, \hat{\boldsymbol{\mu}}'_{\boldsymbol{2}}, \cdots, \hat{\boldsymbol{\mu}}'_{\boldsymbol{k}}]'$. Then $\hat{\boldsymbol{\mu}} \sim \mathcal{N}_{kp}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \text{diag}(\frac{\boldsymbol{\Sigma_1}}{n_1}, \frac{\boldsymbol{\Sigma_2}}{n_2}, \cdots, \frac{\boldsymbol{\Sigma_k}}{n_k})$ is a $kp \times kp$ matrix. This means $\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c} \sim \mathcal{N}_q(\boldsymbol{C}\boldsymbol{\mu} - \boldsymbol{c}, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')$, suggesting the following Wald-type test statistic:

$$T = (\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})'(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c}), \tag{32}$$

where $\hat{\boldsymbol{\Sigma}} = \text{diag}(\frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{1}}}{n_1}, \frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{2}}}{n_2}, \cdots, \frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}}{n_k})$.

We write $\boldsymbol{z} = (\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-\frac{1}{2}}(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})$ and $\boldsymbol{W} = (\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-\frac{1}{2}}(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-\frac{1}{2}}$. It can be easily seen that under the null hypothesis, $\boldsymbol{z}'\boldsymbol{z} \sim \chi^2_q$. Now, notice that $T = \boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z} = \frac{\boldsymbol{z}'\boldsymbol{z}}{U}$. Here, $U = \frac{\boldsymbol{z}'\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z}}$.

Since $\boldsymbol{z}'\boldsymbol{z} \sim \chi^2_p$, similar to Yanagihara and Yuan's (2005) original method for the MBF problem, it is natural to assume $U \approx \frac{\chi^2_v}{\phi}$, which means $\frac{v}{\phi p} T \overset{a}{\sim} F_{p,v}$. In order to approximate the constant v and ϕ, we calculate the first and second moments of U asymptotically by means of the identity $U = \frac{\boldsymbol{z}'\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z}}$ and match them with the facts that $\text{E}[U] \approx \frac{v}{\phi}$ and $\text{E}[U^2] \approx \frac{v(v+2)}{\phi^2}$.

It can be shown that $\boldsymbol{W}^{-1} = \boldsymbol{I_p} - \frac{1}{\sqrt{N}}\overline{\boldsymbol{V}} + \frac{1}{N}\overline{\boldsymbol{V}}^2 + \text{O}_p(N^{-\frac{3}{2}})$, where

$$\overline{\boldsymbol{V}} = \sqrt{N}[(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-\frac{1}{2}}(\boldsymbol{C}[\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}]\boldsymbol{C}')(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-\frac{1}{2}}] \tag{33}$$

and $N = n_1 + n_2 + \cdots + n_k - k$. Similar to the original method explained in Section 2.2, we have

$$U = 1 + \frac{1}{\sqrt{N}}\Big(\frac{\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}\Big) + \frac{1}{N}\Big(\frac{(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2}{(\boldsymbol{z}'\boldsymbol{z})^2} - \frac{\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}\Big) + \text{O}_p(N^{-\frac{3}{2}}) \text{ and} \tag{34}$$

$$U^2 = 1 + \frac{2}{\sqrt{N}}\Big(\frac{\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}\Big) + \frac{1}{N}\Big(\frac{3(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2}{(\boldsymbol{z}'\boldsymbol{z})^2} - \frac{2\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}\Big) + \mathrm{O}_p(N^{-\frac{3}{2}}). \tag{35}$$

We also have that $\overline{\boldsymbol{V}}$ and $\boldsymbol{z}$ are independent, and so are $\frac{\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}$ and $\boldsymbol{z}'\boldsymbol{z}$ as well as $\frac{\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}}$ and $\boldsymbol{z}'\boldsymbol{z}$ (Fang et al., 1990). Taking expectations on both sides of (34) and (35) and using the established independence conditions, we easily obtain

$$\mathrm{E}[U] \approx 1 + \frac{1}{\sqrt{n-2}}\frac{\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}]}{\mathrm{E}[\boldsymbol{z}'\boldsymbol{z}]} + \frac{1}{n-2}\Big(\frac{\mathrm{E}[(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2]}{\mathrm{E}[(\boldsymbol{z}'\boldsymbol{z})^2]} - \frac{\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}]}{\mathrm{E}[\boldsymbol{z}'\boldsymbol{z}]}\Big) \text{ and} \tag{36}$$

$$\mathrm{E}[U^2] \approx 1 + \frac{2}{\sqrt{n-2}}\frac{\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}]}{\mathrm{E}[\boldsymbol{z}'\boldsymbol{z}]} + \frac{1}{n-2}\Big(3\frac{\mathrm{E}[(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2]}{\mathrm{E}[(\boldsymbol{z}'\boldsymbol{z})^2]} - 2\frac{\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}]}{\mathrm{E}[\boldsymbol{z}'\boldsymbol{z}]}\Big). \tag{37}$$

Recall that we already have $\boldsymbol{z}'\boldsymbol{z} \sim \chi^2_p$, so that $\mathrm{E}[\boldsymbol{z}'\boldsymbol{z}] = p$ and $\mathrm{E}[(\boldsymbol{z}'\boldsymbol{z})^2] = p(p+2)$. Using the same reasoning as mentioned in the derivation of the original method, we have $\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z}] = 0$. Hence, it remains for us to calculate $\mathrm{E}[(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2]$ and $\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}]$. We need the following propositions:

Proposition 3.1. *Suppose that* $\boldsymbol{C} = [\boldsymbol{C_1}, \boldsymbol{C_2}, \cdots, \boldsymbol{C_k}]$ *where* $\boldsymbol{C_i}$ *is a* $q \times p$ *matrix for every* $i \in \{1, 2, \cdots, k\}$. *Then, under* H_0,

$$\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}] = N\sum_{i=1}^{k}\frac{\mathrm{tr}(((\boldsymbol{C\Sigma C'})^{-1}\boldsymbol{C_i}\boldsymbol{\Sigma_i}\boldsymbol{C_i'})^2) + (\mathrm{tr}((\boldsymbol{C\Sigma C'})^{-1}\boldsymbol{C_i}\boldsymbol{\Sigma_i}\boldsymbol{C_i'}))^2}{n_i^2(n_i-1)}.$$

Proof. Notice that under H_0, $\boldsymbol{z} \sim \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{I_q})$. Moreover, $\overline{\boldsymbol{V}}$ is symmetric. Using Corollary 2.1.1(1), the law of total expectation and the definition of $\overline{\boldsymbol{V}}$ as mentioned in (33), we have

$$\begin{aligned}\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}] &= \mathrm{E}[\mathrm{tr}(\overline{\boldsymbol{V}}^2)] \\ &= N\,\mathrm{E}[\mathrm{tr}((\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}(\boldsymbol{C}[\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}]\boldsymbol{C'})(\boldsymbol{C\Sigma C'})^{-1}(\boldsymbol{C}[\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}]\boldsymbol{C'})(\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}))].\end{aligned}$$

See also that $\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} = \boldsymbol{a_1} + \boldsymbol{a_2} + \cdots + \boldsymbol{a_k}$, where $\boldsymbol{a_i} = \mathrm{diag}(0, \cdots, 0, \frac{\hat{\Sigma}_i - \Sigma_i}{n_i}, 0, \cdots, 0)$

for every $i \in \{1, 2, \cdots, k\}$. Therefore,

$$\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}] = N \sum_{i,j\in\{1,2,\cdots,k\}} \mathrm{E}[\mathrm{tr}((\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}(\boldsymbol{Ca_iC'})(\boldsymbol{C\Sigma C'})^{-1}(\boldsymbol{Ca_jC'})(\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}))].$$

Observe that when $i \neq j$,

$$\mathrm{E}[\mathrm{tr}((\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}(\boldsymbol{Ca_iC'})(\boldsymbol{C\Sigma C'})^{-1}(\boldsymbol{Ca_jC'})(\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}))] = 0$$

since a_i and a_j are independent and $\mathrm{E}[\boldsymbol{a_i}] = \mathrm{E}[\boldsymbol{a_j}] = \boldsymbol{0}$. Hence,

$$\mathrm{E}[\boldsymbol{z}'\overline{\boldsymbol{V}}^2\boldsymbol{z}] = N \sum_{i=1}^{k} \mathrm{E}[\mathrm{tr}((\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}(\boldsymbol{Ca_iC'})(\boldsymbol{C\Sigma C'})^{-1}(\boldsymbol{Ca_iC'})(\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}))].$$

Now, observe that $\boldsymbol{Ca_iC'} = \boldsymbol{C_i}\frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}}{n_i}\boldsymbol{C_i'} - \boldsymbol{C_i}\frac{\boldsymbol{\Sigma_i}}{n_i}\boldsymbol{C_i'}$ for every $i \in \{1, 2, \cdots, k\}$. It can be easily shown that

$$\frac{(n_i - 1)\boldsymbol{C_i}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\boldsymbol{C_i'}}{n_i} \sim \mathcal{W}_q\Big(n_i - 1, \frac{\boldsymbol{C_i\Sigma_iC_i'}}{n_i}\Big).$$

Using Corollary 2.2.1(2) by plugging $\boldsymbol{S} := \frac{\boldsymbol{C_i}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\boldsymbol{C_i'}}{n_i}$, $n := n_i$, $\boldsymbol{\Sigma} := \frac{\boldsymbol{C_i\Sigma_iC_i'}}{n_i}$ and $\boldsymbol{A} := (\boldsymbol{C\Sigma C'})^{-1}$ and the fact that $\mathrm{tr}(\boldsymbol{PQ}) = \mathrm{tr}(\boldsymbol{QP})$ for any square matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ of the same size, we obtain the desired result. □

Proposition 3.2. *Suppose that* $\boldsymbol{C} = [\boldsymbol{C_1}, \boldsymbol{C_2}, \cdots, \boldsymbol{C_k}]$ *where* $\boldsymbol{C_i}$ *is a* $q \times p$ *matrix for every* $i \in \{1, 2, \cdots, k\}$. *Then, under* H_0,

$$\mathrm{E}[(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2] = N \sum_{i=1}^{k} \frac{4\mathrm{tr}(((\boldsymbol{C\Sigma C'})^{-1}\boldsymbol{C_i\Sigma_iC_i'})^2) + 2(\mathrm{tr}((\boldsymbol{C\Sigma C'})^{-1}\boldsymbol{C_i\Sigma_iC_i'}))^2}{n_i^2(n_i - 1)}.$$

Proof. Note that using Corollary 2.1.1(2) and the law of total expectation, we have $\mathrm{E}[(\boldsymbol{z}'\overline{\boldsymbol{V}}\boldsymbol{z})^2] = 2\mathrm{E}[\mathrm{tr}(\overline{\boldsymbol{V}}^2)] + \mathrm{E}[(\mathrm{tr}(\overline{\boldsymbol{V}}))^2]$. Hence, we only need to prove that

$$\mathrm{E}[(\mathrm{tr}(\overline{\boldsymbol{V}}))^2] = N \sum_{i=1}^{k} \frac{2\mathrm{tr}(((\boldsymbol{C\Sigma C'})^{-1}\boldsymbol{C_i\Sigma_iC_i'})^2)}{n_i^2(n_i - 1)}.$$

See that

$$\mathrm{E}[(\mathrm{tr}(\overline{\boldsymbol{V}}))^2] = N\sum \mathrm{E}[\mathrm{tr}((\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}\boldsymbol{Ca_iC'}(\boldsymbol{C\Sigma C'})^{-\frac{1}{2}})\mathrm{tr}((\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}\boldsymbol{Ca_jC'}(\boldsymbol{C\Sigma C'})^{-\frac{1}{2}})],$$

where the sum is taken over $i, j \in \{1, 2, \cdots, k\}$. Using the same argument as in the previous proposition, we obtain that

$$\mathrm{E}[\mathrm{tr}((\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}\boldsymbol{Ca_iC'}(\boldsymbol{C\Sigma C'})^{-\frac{1}{2}})\mathrm{tr}((\boldsymbol{C\Sigma C'})^{-\frac{1}{2}}\boldsymbol{Ca_jC'}(\boldsymbol{C\Sigma C'})^{-\frac{1}{2}})] = 0$$

when $i \neq j$. Furthermore, by means of Corollary 2.2.1(1) (with the same substitutions as above) and the fact that $\mathrm{tr}(\boldsymbol{PQ}) = \mathrm{tr}(\boldsymbol{QP})$ for any square matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ of the same size, we are done. □

For simplicity, we write

$$\psi_1 = N\sum_{i=1}^{k}\frac{(\mathrm{tr}((\boldsymbol{C\Sigma C'})^{-1}\boldsymbol{C_i\Sigma_i C_i'}))^2}{n_i^2(n_i-1)} \text{ and } \psi_2 = N\sum_{i=1}^{k}\frac{\mathrm{tr}(((\boldsymbol{C\Sigma C'})^{-1}\boldsymbol{C_i\Sigma_i C_i'})^2)}{n_i^2(n_i-1)}. \tag{38}$$

Using Proposition 3.1, Proposition 3.2 and the fact that $\boldsymbol{z'z} \sim \chi_q^2$, we easily obtain

$$\mathrm{E}\Big[\frac{\boldsymbol{z'}\overline{\boldsymbol{V}}^2\boldsymbol{z}}{\boldsymbol{z'z}}\Big] = \frac{\psi_1+\psi_2}{q} \text{ and } \mathrm{E}\Big[\frac{(\boldsymbol{z'}\overline{\boldsymbol{V}}\boldsymbol{z})^2}{(\boldsymbol{z'z})^2}\Big] = \frac{2\psi_1+4\psi_2}{q(q+2)}. \tag{39}$$

We may use the same method as in Section 2.2 in order to determine the corresponding F statistic. First, we let

$$\hat{\psi}_1 = N\sum_{i=1}^{k}\frac{(\mathrm{tr}((\boldsymbol{C\hat{\Sigma} C'})^{-1}\boldsymbol{C_i\hat{\Sigma}_i C_i'}))^2}{n_i^2(n_i-1)} \text{ and } \hat{\psi}_2 = N\sum_{i=1}^{k}\frac{\mathrm{tr}(((\boldsymbol{C\hat{\Sigma} C'})^{-1}\boldsymbol{C_i\hat{\Sigma}_i C_i'})^2)}{n_i^2(n_i-1)}. \tag{40}$$

The F statistic can be found to be

$$T_F = \frac{N-\hat{\theta}_1}{Nq}T \overset{a}{\sim} \mathcal{F}_{q,\hat{v}}, \tag{41}$$

where $\hat{\theta}_1 = \frac{q\hat{\psi}_1+(q-2)\hat{\psi}_2}{q(q+2)}$, $\hat{\theta}_2 = \frac{\hat{\psi}_1+2\hat{\psi}_2}{q(q+2)}$ and $\hat{v} = \frac{(N-\hat{\theta}_1)^2}{N\hat{\theta}_2-\hat{\theta}_1}$.

3.4 A Proof of the Invariance Properties of the Generalised Methods

As mentioned in Section 1.3, a desirable approximate solution should satisfy 3 (three) properties, namely affine invariance, nonsingular invariance and independence of different labelling schemes. In this subsection, we will show that the generalised methods described in the previous subsections indeed possess the aforementioned properties. Note that the definitions of the properties presented in this subsection are different from those stated in Section 1.3 as we are now dealing with the generalised MBF problem.

We start by proving the following proposition:

Proposition 3.3. *The generalised Krishnamoorthy and Yu's (2004) method is affine invariant, nonsingular invariant and independent of different labelling schemes of the mean vectors.*

Proof. From (31), it is easy to see that in order to prove the invariance and independence conditions, it suffices to prove that those conditions hold for T and $\hat{d}$. Recall that in order to show that an approximate solution is affine invariant, we need to transform each point $\boldsymbol{y_{ij}}$ in each sample into $\tilde{\boldsymbol{y}}_{\boldsymbol{ij}} = \boldsymbol{B}\boldsymbol{y_{ij}} + \boldsymbol{b}$ where $\boldsymbol{B}$ is any invertible constant matrix with p columns and $\boldsymbol{b}$ any constant vector of length p. For every $l \in \{1, 2, \cdots, k\}$, denote by $\tilde{\boldsymbol{\mu}}_{\boldsymbol{l}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{l}}$ the mean and variance of $\tilde{\boldsymbol{y}}_{\boldsymbol{ij}}$. This implies $\tilde{\boldsymbol{\mu}}_{\boldsymbol{l}} = \boldsymbol{B}\boldsymbol{\mu_l} + \boldsymbol{b}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{l}} = \boldsymbol{B}\Sigma_l\boldsymbol{B}'$, which means $\boldsymbol{\mu_l} = \boldsymbol{B}^{-1}(\tilde{\boldsymbol{\mu}}_{\boldsymbol{l}} - \boldsymbol{b})$.

We define $\tilde{\boldsymbol{\mu}} = [\tilde{\boldsymbol{\mu}}'_{\boldsymbol{1}}, \tilde{\boldsymbol{\mu}}'_{\boldsymbol{2}}, \cdots, \tilde{\boldsymbol{\mu}}'_{\boldsymbol{k}}]'$ and $\tilde{\boldsymbol{\Sigma}} = \text{diag}(\frac{\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{1}}}{n_1}, \frac{\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{2}}}{n_2}, \cdots, \frac{\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{k}}}{n_k})$. It can be shown that $\boldsymbol{\mu} = \tilde{\boldsymbol{B}}^{-1}(\tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{b}})$ and $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{B}}\boldsymbol{\Sigma}\tilde{\boldsymbol{B}}'$, where $\tilde{\boldsymbol{B}} = \boldsymbol{I_k} \otimes \boldsymbol{B}$ and $\tilde{\boldsymbol{b}} = \boldsymbol{I_k} \otimes \boldsymbol{b}$. Now, we can rewrite the null hypothesis as $\tilde{H}_0 : \tilde{\boldsymbol{C}}\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{c}}$ where $\tilde{\boldsymbol{C}} = \boldsymbol{C}\tilde{\boldsymbol{B}}^{-1}$ and $\tilde{\boldsymbol{c}} = \boldsymbol{C}\tilde{\boldsymbol{B}}^{-1}\tilde{\boldsymbol{b}} + \boldsymbol{c}$.

Let $\hat{\tilde{\boldsymbol{\mu}}}_{\boldsymbol{l}}$ and $\hat{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{l}}$ be the unbiased estimators for $\tilde{\boldsymbol{\mu}}_{\boldsymbol{l}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{l}}$. Observe that $\hat{\tilde{\boldsymbol{\mu}}}_{\boldsymbol{l}} =$

$\tilde{\boldsymbol{B}}\hat{\boldsymbol{\mu}}_l + \tilde{\boldsymbol{b}}$ and $\hat{\tilde{\boldsymbol{\Sigma}}}_l = \tilde{\boldsymbol{B}}\hat{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{B}}'$. It follows that $\hat{\tilde{\boldsymbol{\mu}}} = \tilde{\boldsymbol{B}}\hat{\boldsymbol{\mu}} + \tilde{\boldsymbol{b}}$ and $\hat{\tilde{\boldsymbol{\Sigma}}} = \tilde{\boldsymbol{B}}\hat{\boldsymbol{\Sigma}}\tilde{\boldsymbol{B}}'$. Now, see that $\tilde{\boldsymbol{C}}\hat{\tilde{\boldsymbol{\mu}}} - \tilde{\boldsymbol{c}} = \boldsymbol{C}\tilde{\boldsymbol{B}}^{-1}(\tilde{\boldsymbol{B}}\hat{\boldsymbol{\mu}} + \tilde{\boldsymbol{b}}) - (\boldsymbol{C}\tilde{\boldsymbol{B}}^{-1}\tilde{\boldsymbol{b}} + \boldsymbol{c}) = \boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c}$ and $\tilde{\boldsymbol{C}}\hat{\tilde{\boldsymbol{\Sigma}}}\tilde{\boldsymbol{C}}' = (\boldsymbol{C}\tilde{\boldsymbol{B}}^{-1})\tilde{\boldsymbol{B}}\hat{\boldsymbol{\Sigma}}\tilde{\boldsymbol{B}}'(\boldsymbol{C}\tilde{\boldsymbol{B}}^{-1})' = \boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}'$. This shows the affine invariance of $T = (\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})'(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})$.

Now, we aim to show the affine invariance of $\hat{d}$. Since

$$\hat{d} = \frac{q(q+1)}{\sum_{l=1}^{k} \frac{\text{tr}(\hat{\boldsymbol{\Omega}}_l^2) + (\text{tr}(\hat{\boldsymbol{\Omega}}_l))^2}{n_l - 1}},$$

it remains for us to show the affine invariance of $\text{tr}(\hat{\boldsymbol{\Omega}}_l^2)$ and $\text{tr}(\hat{\boldsymbol{\Omega}}_l)$, where $\hat{\boldsymbol{\Omega}}_l = n_l^{-1}(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-\frac{1}{2}}\boldsymbol{C}_l\hat{\boldsymbol{\Sigma}}_l\boldsymbol{C}_l'(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-\frac{1}{2}}$. We let $\boldsymbol{G}_l = n_l^{-1}\boldsymbol{C}_l\hat{\boldsymbol{\Sigma}}_l\boldsymbol{C}_l'$ and $\boldsymbol{G} = \boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}'$. Notice that $\boldsymbol{G} = \sum_{l=1}^{k}\boldsymbol{G}_l$ and $\hat{\boldsymbol{\Omega}}_l = \boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{G}_l\boldsymbol{G}^{-\frac{1}{2}}$. Recall that we have established the affine invariance property of $\boldsymbol{G}$. Hence, it remains for us to show that $\boldsymbol{G}_l$ is also affine invariant. Indeed, this is clear since $\tilde{\boldsymbol{C}}_l = \boldsymbol{C}_l\tilde{\boldsymbol{B}}^{-1}$ and $\hat{\tilde{\boldsymbol{\Sigma}}}_l = \tilde{\boldsymbol{B}}\hat{\boldsymbol{\Sigma}}_l\tilde{\boldsymbol{B}}'$. Thus, we have shown that the Krishnamoorthy and Yu's (2004) method is affine invariant.

Next, we wish to prove the nonsingular invariance of this generalised method. Note that the nonsingular invariance condition described in Section 1.3 is meant for the MBF problem. For the generalised setting, we need to show that under for any invertible constant matrix with q columns $\boldsymbol{P}$, the null hypothesis $\tilde{H}_0 : \tilde{\boldsymbol{C}}\boldsymbol{\mu} = \tilde{\boldsymbol{c}}$ where $\tilde{\boldsymbol{C}} = \boldsymbol{P}\boldsymbol{C}$ and $\tilde{\boldsymbol{c}} = \boldsymbol{P}\boldsymbol{c}$ is equivalent to H_0.

Note that we have $\tilde{\boldsymbol{C}}\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{c}} = \boldsymbol{P}(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})$ and $\tilde{\boldsymbol{C}}\hat{\Sigma}\tilde{\boldsymbol{C}}' = \boldsymbol{P}(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')\boldsymbol{P}'$, which clearly show the nonsingular invariance of T. It remains for us to show the nonsingular invariance of $\hat{d}$. As above, it is sufficient to establish the nonsingular invariance of $\text{tr}(\hat{\boldsymbol{\Omega}}_l^2)$ and $\text{tr}(\hat{\boldsymbol{\Omega}}_l)$. We note that $\tilde{\boldsymbol{C}}_l = \boldsymbol{P}\boldsymbol{C}_l$. This implies $\tilde{\boldsymbol{G}}_l = \boldsymbol{P}\boldsymbol{G}_l\boldsymbol{P}'$ and $\tilde{\boldsymbol{G}} = \sum_{l=1}^{k}\boldsymbol{G}_l = \boldsymbol{P}\boldsymbol{G}\boldsymbol{P}'$. From here, we have $\text{tr}(\hat{\tilde{\boldsymbol{\Omega}}}_l) = \text{tr}(\tilde{\boldsymbol{G}}^{-\frac{1}{2}}\tilde{\boldsymbol{G}}_l\tilde{\boldsymbol{G}}^{-\frac{1}{2}}) = \text{tr}(\tilde{\boldsymbol{G}}_l\tilde{\boldsymbol{G}}^{-1}) = \text{tr}(\boldsymbol{P}\boldsymbol{G}_l\boldsymbol{G}^{-1}\boldsymbol{P}^{-1}) = \text{tr}(\boldsymbol{G}_l\boldsymbol{G}^{-1}) = \text{tr}(\boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{G}_l\boldsymbol{G}^{-\frac{1}{2}}) = \text{tr}(\hat{\boldsymbol{\Omega}}_l)$. Analogously, we obtain that $\text{tr}(\hat{\boldsymbol{\Omega}}_l^2)$ is also nonsingular invariant.

Lastly, we need to establish that the generalised Krishnamoorthy and Yu's (2004) method is independent of different labelling schemes of the mean vectors. To see this, we let $\{l_1, l_2, \cdots, l_k\}$ be any permutation of $\{1, 2, \cdots, k\}$. It is easy to see that $\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}' = \sum_{l=1}^{k} n_l^{-1}\boldsymbol{C_l}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{l}}\boldsymbol{C}_{\boldsymbol{l}}' = \sum_{u=1}^{k} n_{l_u}^{-1}\boldsymbol{C}_{\boldsymbol{l_u}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{l_u}}\boldsymbol{C}_{\boldsymbol{l_u}}'$ is independent of different labelling schemes of the mean vectors. Similarly, $\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c}$ can be shown to be independent of different labelling schemes of the mean vectors, and so does T. In order to show the same holds for $\hat{d}$, we note that

$$\sum_{l=1}^{k} \frac{\mathrm{tr}(\hat{\boldsymbol{\Omega}}_{\boldsymbol{l}}^2) + (\mathrm{tr}(\hat{\boldsymbol{\Omega}}_{\boldsymbol{l}}))^2}{n_l - 1} = \sum_{l=1}^{k} \frac{\mathrm{tr}((\boldsymbol{G_l}\boldsymbol{G}^{-1})^2) + (\mathrm{tr}(\boldsymbol{G_l}\boldsymbol{G}^{-1}))^2}{n_l - 1}$$
$$= \sum_{u=1}^{k} \frac{\mathrm{tr}((\boldsymbol{G}_{\boldsymbol{l_u}}\boldsymbol{G}^{-1})^2) + (\mathrm{tr}(\boldsymbol{G}_{\boldsymbol{l_u}}\boldsymbol{G}^{-1}))^2}{n_{l_u} - 1}.$$

□

Now, we will prove that the generalised Yanagihara and Yuan's (2005) method also satisfies the invariance and independence conditions.

Proposition 3.4. *The generalised Yanagihara and Yuan's (2005) method is affine invariant, nonsingular invariant and independent of different labelling schemes of the mean vectors.*

Proof. From (41), it is easy to see that in order to prove the invariance and independence conditions, it suffices to prove that those conditions hold for $\hat{\psi}_1$ and $\hat{\psi}_2$. We do not need to provide a proof that the conditions hold for T as it has already been mentioned in the proof of Proposition 3.3.

Following the notations used in the proof of Proposition 3.3, we have that

$$\hat{\psi}_1 = N \sum_{i=1}^{k} \frac{(\mathrm{tr}(\boldsymbol{G_i}\boldsymbol{G}^{-1}))^2}{n_i - 1} \text{ and } \hat{\psi}_2 = N \sum_{i=1}^{k} \frac{\mathrm{tr}((\boldsymbol{G_i}\boldsymbol{G}^{-1})^2)}{n_i - 1}.$$

Since $\boldsymbol{G}$ and $\boldsymbol{G_l}$ (for every $l \in \{1, 2, \cdots, k\}$) have been shown to be affine invariant, we can immediately conclude that $\hat{\psi}_1$ and $\hat{\psi}_2$ are affine invariant.

Now, recall that we have established the nonsingular invariance of $\mathrm{tr}(\boldsymbol{G_l}\boldsymbol{G}^{-1})$ and $\mathrm{tr}((\boldsymbol{G_l}\boldsymbol{G}^{-1})^2)$, which clearly imply that $\hat{\psi}_1$ and $\hat{\psi}_2$ are nonsingular invariant.

Lastly, note that

$$\hat{\psi}_1 = N \sum_{i=1}^{k} \frac{(\operatorname{tr}(\boldsymbol{G_i}\boldsymbol{G}^{-1}))^2}{n_i - 1} = N \sum_{u=1}^{k} \frac{(\operatorname{tr}(\boldsymbol{G_{i_u}}\boldsymbol{G}^{-1}))^2}{n_{i_u} - 1}$$

for any permutation $\{i_1, i_2, \cdots, i_k\}$ of $\{1, 2, \cdots, k\}$, which shows that $\hat{\psi}_1$ is independent of different labelling schemes of the mean vectors. Similarly, $\hat{\psi}_2$ is also independent of different labelling schemes of the mean vectors; and we are done. □

3.5 A Proof of the Equivalence of the Generalised Methods to the Original Methods in the Context of the MBF Problem

Recall that the MBF problem is just a special case of the GLHT problem described in Section 3.1. In particular,the GLHT problem becomes the MBF problem if we set $k := 2$, $\boldsymbol{c} := \boldsymbol{0_p}$ and $\boldsymbol{C} := [1\ -1] \otimes \boldsymbol{I_p}$. Our aim is to show that setting those values will reduce both generalised methods to the original methods.

We consider the following propositions:

Proposition 3.5. *The generalised Krishnamoorthy and Yu's (2004) method reduces to the Krishnamoorthy and Yu's (2004) method when $k := 2$, $\boldsymbol{c} := \boldsymbol{0_p}$ and $\boldsymbol{C} := [1\ -1] \otimes \boldsymbol{I_p}$.*

Proof. According to (31), the test statistic for the generalised Krishnamoorthy and Yu's (2004) method is

$$T_{FM} = \frac{\hat{d} - q + 1}{q\hat{d}} T \overset{a}{\sim} \mathcal{F}_{q, \hat{d}-q+1},$$

where

$$\hat{d} = \frac{q(q+1)}{\sum_{l=1}^{k} \frac{\operatorname{tr}(\hat{\boldsymbol{\Omega}}_{\boldsymbol{l}}^2) + (\operatorname{tr}(\hat{\boldsymbol{\Omega}}_{\boldsymbol{l}}))^2}{n_l - 1}}$$

with $\hat{\boldsymbol{\Omega}}_{\boldsymbol{l}} = n_l^{-1}(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-\frac{1}{2}}\boldsymbol{C_l}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{l}}\boldsymbol{C}_{\boldsymbol{l}}'(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-\frac{1}{2}}$.

Notice that in the context of the MBF problem, $N := n - 2$ and $q := p$. Comparing this statistic with the statistic on the top of Page 9 suggests that we only need to establish the following:

- The test statistic

$$T = (\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})'(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})$$

  is the same as

$$T = (\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}})'\Big(\frac{\boldsymbol{S_1}}{n_1} + \frac{\boldsymbol{S_2}}{n_2}\Big)^{-1}(\overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}).$$

- The quantity

$$\hat{d} = \frac{q(q+1)}{\sum_{l=1}^{k} \frac{\operatorname{tr}(\hat{\boldsymbol{\Omega}}_{\boldsymbol{l}}^2) + (\operatorname{tr}(\hat{\boldsymbol{\Omega}}_{\boldsymbol{l}}))^2}{n_l - 1}}$$

  is the same as the quantity

$$\hat{v}_M = \frac{p(p+1)(n-2)}{\hat{\psi}_1 + \hat{\psi}_2}.$$

First, note that $\boldsymbol{C} = [\boldsymbol{I_p} \ \ -\boldsymbol{I_p}]$, and $\hat{\boldsymbol{\mu}} = [\overline{\boldsymbol{y}}_{\boldsymbol{1}} \ \overline{\boldsymbol{y}}_{\boldsymbol{2}}]'$ and $\boldsymbol{c} := \boldsymbol{0_p}$, so that $\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c} = \overline{\boldsymbol{y}}_{\boldsymbol{1}} - \overline{\boldsymbol{y}}_{\boldsymbol{2}}$. Moreover, we observe that $\hat{\boldsymbol{\Sigma}} = \operatorname{diag}\big(\frac{\boldsymbol{S_1}}{n_1}, \frac{\boldsymbol{S_2}}{n_2}\big)$. This means $(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1} = \big(\frac{\boldsymbol{S_1}}{n_1} + \frac{\boldsymbol{S_2}}{n_2}\big)^{-1}$. This shows the first statement.

For the second statement, note that $(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1} = \big(\frac{\boldsymbol{S_1}}{n_1} + \frac{\boldsymbol{S_2}}{n_2}\big)^{-1} = \frac{n_1 n_2}{n}\overline{\boldsymbol{S}}^{-1}$ and $\boldsymbol{C_i}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\boldsymbol{C}_{\boldsymbol{i}}' = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}} = \boldsymbol{S_i}$ for $i \in \{1, 2\}$. Thus,

$$\operatorname{tr}(\hat{\boldsymbol{\Omega}}_{\boldsymbol{l}}) = \frac{1}{n_l}\operatorname{tr}((\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}\boldsymbol{C_l}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{l}}\boldsymbol{C}_{\boldsymbol{l}}') = \frac{n_1 n_2}{n n_l}\operatorname{tr}(\boldsymbol{S_i}\overline{\boldsymbol{S}}^{-1})$$

using the fact that $\operatorname{tr}(\boldsymbol{PQ}) = \operatorname{tr}(\boldsymbol{QP})$ for any square matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ of the same size. Similarly,

$$\operatorname{tr}(\hat{\boldsymbol{\Omega}}_{\boldsymbol{l}}^2) = \frac{1}{n_l^2}\operatorname{tr}(((\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}\boldsymbol{C_l}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{l}}\boldsymbol{C}_{\boldsymbol{l}}')^2) = \frac{n_1^2 n_2^2}{n^2 n_l^2}\operatorname{tr}(\boldsymbol{S_i}\overline{\boldsymbol{S}}^{-1}\boldsymbol{S_i}\overline{\boldsymbol{S}}^{-1}).$$

Now, since $p = q$ and $k = 2$, it is now easy to show that $\hat{d} = \hat{v}_M$ using the above identities. Indeed,

$$\begin{aligned}
\hat{d} &= \frac{q(q+1)}{\sum_{l=1}^{k} \frac{\mathrm{tr}(\hat{\boldsymbol{\Omega}}_{\boldsymbol{l}}^2)+(\mathrm{tr}(\hat{\boldsymbol{\Omega}}_{\boldsymbol{l}}))^2}{n_l-1}} \\
&= \frac{p(p+1)}{\frac{n_2^2}{n^2(n_1-1)}\big(\mathrm{tr}(\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}) + \mathrm{tr}(\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1})\big) + \frac{n_1^2}{n^2(n_2-1)}\big(\mathrm{tr}(\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}) + \mathrm{tr}(\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1})\big)} \\
&= \frac{p(p+1)(n-2)}{\hat{\psi}_1 + \hat{\psi}_2} \\
&= \hat{v}_M .
\end{aligned}$$

The proof is now complete. □

Proposition 3.6. *The generalised Yanagihara and Yuan's (2005) method reduces to the Yanagihara and Yuan's (2005) method when* $k := 2$, $\boldsymbol{c} := \boldsymbol{0_p}$ *and* $\boldsymbol{C} := [1\ \ -1] \otimes \boldsymbol{I_p}$.

Proof. According to (41), the F statistic for the generalised Yanagihara and Yuan's (2005) method is

$$T_F = \frac{N - \hat{\theta}_1}{Nq} T \overset{a}{\sim} \mathcal{F}_{q,\hat{v}},$$

where $\hat{\theta}_1 = \frac{q\hat{\psi}_1+(q-2)\hat{\psi}_2}{q(q+2)}$, $\hat{\theta}_2 = \frac{\hat{\psi}_1+2\hat{\psi}_2}{q(q+2)}$ and $\hat{v} = \frac{(N-\hat{\theta}_1)^2}{N\hat{\theta}_2-\hat{\theta}_1}$ with

$$\hat{\psi}_1 = N \sum_{i=1}^{k} \frac{(\mathrm{tr}((\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}\boldsymbol{C_i}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\boldsymbol{C}_{\boldsymbol{i}}'))^2}{n_i^2(n_i-1)} \text{ and } \hat{\psi}_2 = N \sum_{i=1}^{k} \frac{\mathrm{tr}(((\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}\boldsymbol{C_i}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\boldsymbol{C}_{\boldsymbol{i}}')^2)}{n_i^2(n_i-1)}.$$

Notice that in the context of the MBF problem, $N := n - 2$ and $q := p$. Comparing this statistic with (25) suggests that we only need to establish the following:

- The test statistic

$$T = (\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})'(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})$$

is the same as

$$T = (\overline{\boldsymbol{y}}_1 - \overline{\boldsymbol{y}}_2)'\Big(\frac{\boldsymbol{S_1}}{n_1} + \frac{\boldsymbol{S_2}}{n_2}\Big)^{-1}(\overline{\boldsymbol{y}}_1 - \overline{\boldsymbol{y}}_2).$$

- The quantities

$$\hat{\psi}_1 = N\sum_{i=1}^{k}\frac{(\mathrm{tr}((\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}\boldsymbol{C_i}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\boldsymbol{C_i'}))^2}{n_i^2(n_i-1)} \text{ and } \hat{\psi}_2 = N\sum_{i=1}^{k}\frac{\mathrm{tr}(((\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}\boldsymbol{C_i}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\boldsymbol{C_i'})^2)}{n_i^2(n_i-1)}$$

are the same as the quantities

$$\hat{\psi}_1 = \frac{n_2^2(n-2)}{n^2(n_1-1)}(\mathrm{tr}(\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}))^2 + \frac{n_1^2(n-2)}{n^2(n_2-1)}(\mathrm{tr}(\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}))^2 \text{ and}$$

$$\hat{\psi}_2 = \frac{n_2^2(n-2)}{n^2(n_1-1)}\mathrm{tr}(\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}) + \frac{n_1^2(n-2)}{n^2(n_2-1)}\mathrm{tr}(\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}).$$

Note that the first statement has been shown before. In order to show that the second statement holds, recall that we have established $(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1} = \big(\frac{\boldsymbol{S_1}}{n_1} + \frac{\boldsymbol{S_2}}{n_2}\big)^{-1} = \frac{n_1 n_2}{n}\overline{\boldsymbol{S}}^{-1}$ and $\boldsymbol{C_i}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\boldsymbol{C_i'} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}} = \boldsymbol{S_i}$ for $i \in \{1,2\}$. This implies

$$N\sum_{i=1}^{k}\frac{(\mathrm{tr}((\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}\boldsymbol{C_i}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\boldsymbol{C_i'}))^2}{n_i^2(n_i-1)} = (n-2)\sum_{i=1}^{2}\frac{\frac{n_1^2 n_2^2}{n^2}(\mathrm{tr}(\overline{\boldsymbol{S}}^{-1}\boldsymbol{S_i}))^2}{n_i^2(n_i-1)}$$

$$= \frac{n_2^2(n-2)}{n^2(n_1-1)}(\mathrm{tr}(\boldsymbol{S_1}\overline{\boldsymbol{S}}^{-1}))^2 + \frac{n_1^2(n-2)}{n^2(n_2-1)}(\mathrm{tr}(\boldsymbol{S_2}\overline{\boldsymbol{S}}^{-1}))^2$$

using the fact that $\mathrm{tr}(\boldsymbol{PQ}) = \mathrm{tr}(\boldsymbol{QP})$ for any square matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ of the same size. This shows that both the $\hat{\psi}_1$'s above are exactly the same. Analogously, we can show that both the $\hat{\psi}_2$'s are also exactly the same. Therefore, the proof is complete. □

3.6 Simulation Studies

Similar to Section 2.3, we conducted simulation studies in order to compare the performance of the methods described in the previous subsections. We also compare both methods in terms of their Type I errors. Following Zhang (2012), we set $\boldsymbol{\Sigma_1} = \boldsymbol{I_p}$, $\boldsymbol{\Sigma_2} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_p)$ and $\boldsymbol{\Sigma_l}$ $(l \geq 3)$ to be some positive definite matrices, where p, λ_k's and the sample sizes will be specified in the next figure. We also measured the AAD of each method, which indicates the difference between the nominal and empirical sizes. For each combination of parameters, we simulated over 100,000 replications in order to obtain more stable and accurate results.

The figure below shows the simulation results for trivariate one-way MANOVA with 3 (three) samples. Here, the nominal size is fixed to be 0.05. T_F *Gen* refers to the generalised Yanagihara and Yuan's (2005) method, while T_{FM} *Gen* refers to the generalised Krishnamoorthy and Yu's (2004) method (Zhang's method). The black-coloured numbers on the same columns as the methods' names refer to the Type I errors, while the red-coloured numbers refer to the AADs. It can be seen

alpha = 0.05				
Sigma1 = I3; Sigma2 = diag(lambda)				
Sigma3 = (1,rho,rho; rho,1,rho; rho,rho,1)				
lambda	rho	n1, n2, n3	T_F Gen	T_FM Gen
1,1,1	0	7,7,7	0.0309	0.0383
1,1,1	0	10,10,10	0.0406	0.0444
1,1,1	0	15,15,15	0.0464	0.0485
1,1,1	0	7,10,20	0.0486	0.0562
1,1,1	0	10,20,40	0.0517	0.0554
1,1,1	0	20,10,7	0.0474	0.0549
1,1,1	0	40,20,10	0.0523	0.0564
1,5,0.1	0.05	7,7,7	0.033	0.0413
1,5,0.1	0.05	10,10,10	0.043	0.0477
1,5,0.1	0.05	15,15,15	0.0471	0.0493
1,5,0.1	0.05	7,10,20	0.0518	0.0596
1,5,0.1	0.05	10,20,40	0.0566	0.0612
1,5,0.1	0.05	20,10,7	0.0515	0.0591
1,5,0.1	0.05	40,20,10	0.0544	0.0585
1,3,0.1	0.09	7,7,7	0.0323	0.0409
1,3,0.1	0.09	10,10,10	0.0425	0.047
1,3,0.1	0.09	15,15,15	0.0469	0.0496
1,3,0.1	0.09	7,10,20	0.0516	0.0596
1,3,0.1	0.09	10,20,40	0.0552	0.0593
1,3,0.1	0.09	20,10,7	0.0526	0.06
1,3,0.1	0.09	40,20,10	0.0554	0.0597
AAD			0.0059	0.0068

that the performance of the generalised Yanagihara and Yuan's (2005) method is slightly better than that of the generalised Krishnamoorthy and Yu's (2004) method for this particular set of parameters.

The figure below shows the simulation results for five-variate one-way MANOVA with 5 (five) samples. Similar to the above, the nominal size is set to be 0.05. It can be observed that in this case, the generalised Yanagihara and Yuan's (2005) method also has a better performance as compared to the generalised Krishnamoorthy and Yu's (2005) method.

alpha = 0.05						
Sigma1 = I5; Sigma2 = diag(lambda); Sigma3 = diag(eta); Sigma4 = diag(u), Sigma5 = diag(v)						
lambda	eta	u	v	n1,n2,n3,n4,n5	T_F Gen	T_FM Gen
1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	15,15,15,15,15	0.0428	0.0492
1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	25,25,25,25,25	0.0491	0.0519
1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	50,50,50,50,50	0.0498	0.051
1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	20,25,35,40,50	0.0519	0.0546
1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	30,35,40,50,70	0.0493	0.0512
1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	50,40,35,25,20	0.0519	0.0549
1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	1,1,1,1,1	70,50,40,35,30	0.0495	0.0516
12,12,1,24,1	1,0.1,2,24,21	1,3,3,9,10	5,15,15,45,50	15,15,15,15,15	0.0467	0.0542
12,12,1,24,1	1,0.1,2,24,21	1,3,3,9,10	5,15,15,45,50	25,25,25,25,25	0.0495	0.053
12,12,1,24,1	1,0.1,2,24,21	1,3,3,9,10	5,15,15,45,50	50,50,50,50,50	0.0497	0.0511
12,12,1,24,1	1,0.1,2,24,21	1,3,3,9,10	5,15,15,45,50	20,25,35,40,50	0.051	0.0536
12,12,1,24,1	1,0.1,2,24,21	1,3,3,9,10	5,15,15,45,50	30,35,40,50,70	0.0486	0.0503
12,12,1,24,1	1,0.1,2,24,21	1,3,3,9,10	5,15,15,45,50	50,40,35,25,20	0.0556	0.0595
12,12,1,24,1	1,0.1,2,24,21	1,3,3,9,10	5,15,15,45,50	70,50,40,35,30	0.052	0.0544
1,3,9,9,5	5,15,45,45,45	1,3,3,9,30	5,15,15,45,100	15,15,15,15,15	0.0475	0.0558
1,3,9,9,5	5,15,45,45,45	1,3,3,9,30	5,15,15,45,100	25,25,25,25,25	0.0486	0.052
1,3,9,9,5	5,15,45,45,45	1,3,3,9,30	5,15,15,45,100	50,50,50,50,50	0.0496	0.0511
1,3,9,9,5	5,15,45,45,45	1,3,3,9,30	5,15,15,45,100	20,25,35,40,50	0.0504	0.053
1,3,9,9,5	5,15,45,45,45	1,3,3,9,30	5,15,15,45,100	30,35,40,50,70	0.0503	0.0521
1,3,9,9,5	5,15,45,45,45	1,3,3,9,30	5,15,15,45,100	50,40,35,25,20	0.0531	0.0573
1,3,9,9,5	5,15,45,45,45	1,3,3,9,30	5,15,15,45,100	70,50,40,35,30	0.0504	0.053
AAD					0.0017	0.0032

3.7 An Alternative Method for the Case of High-Dimensional Multivariate Normal Distributions

After performing more thorough simulations with different combinations of parameters, it is found that both generalised methods perform very badly when the samples are drawn from high-dimensional multivariate normal distributions. In order to see this more clearly, consider the following results:

- When $\alpha = 0.05$, $n_1 = 10$, $n_2 = 100$, $n_3 = 1000$, $\mathbf{\Sigma_1} = \text{diag}(\text{rep}(10, 3))$, $\mathbf{\Sigma_2} = \text{diag}(\text{rep}(0.05, 3))$ and $\mathbf{\Sigma_3} = \text{diag}(\text{rep}(0.001, 3))$, the empirical size of T_F *Gen* is 0.063 and that of T_{FM} *Gen* is 0.071.

- When $\alpha = 0.05$, $n_1 = 10$, $n_2 = 100$, $n_3 = 1000$, $\mathbf{\Sigma_1} = \text{diag}(\text{rep}(1, 3))$, $\mathbf{\Sigma_2} = \text{diag}(\text{rep}(1, 3))$ and $\mathbf{\Sigma_3} = \text{diag}(\text{rep}(1, 3))$, the empirical size of T_F *Gen* is 0.064 and that of T_{FM} *Gen* is 0.071.

- When $\alpha = 0.05$, $n_1 = 10$, $n_2 = 100$, $n_3 = 1000$, $\mathbf{\Sigma_1} = \text{diag}(\text{rep}(10, 8))$, $\mathbf{\Sigma_2} = \text{diag}(\text{rep}(0.05, 8))$ and $\mathbf{\Sigma_3} = \text{diag}(\text{rep}(0.001, 8))$, the empirical size of T_F *Gen* is 0.320 and that of T_{FM} *Gen* is 0.414.

- When $\alpha = 0.05$, $n_1 = 10$, $n_2 = 100$, $n_3 = 1000$, $\mathbf{\Sigma_1} = \text{diag}(\text{rep}(1, 8))$, $\mathbf{\Sigma_2} = \text{diag}(\text{rep}(1, 8))$ and $\mathbf{\Sigma_3} = \text{diag}(\text{rep}(1, 8))$, the empirical size of T_F *Gen* is 0.309 and that of T_{FM} *Gen* is 0.378.

The results above show that ceteris paribus, increasing the dimension of the multivariate normal populations results in a drastic decline in the performance of both methods. In order to mitigate this, we introduce an alternative method which gives a much better size. However, this alternative method may have a low power so that its usage needs to be controlled.

The idea of this method is to approximate using a constant multiple of a chi-square distribution the following test statistic:

$$T = (\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})'(\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c}), \tag{42}$$

where $\boldsymbol{D} = \text{diag}(\text{rep}(1/n_1, p), \text{rep}(1/n_2, p), \cdots, \text{rep}(1/n_k, p))$. Compare this with the Wald-type test statistic $(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})'(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c})$ used for the generalised methods we discussed earlier.

In Section 3.2, we established that $\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c} \sim \mathcal{N}_q(\boldsymbol{C}\boldsymbol{\mu} - \boldsymbol{c}, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')$, which means that $\boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c} \sim \mathcal{N}_q(\boldsymbol{0}_{\boldsymbol{q}}, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')$ under the null hypothesis. Recall that our aim is to approximate (42) using a constant multiple of a chi-square distribution. In other words, we need to find the appropriate values of β and d such that

$$T \overset{a}{\sim} \beta\chi^2_d. \tag{43}$$

Since there are two unknowns, it is natural to match the expected value and variance of T with those of $\beta\chi^2_d$. First, note that the expected value of $\beta\chi^2_d$ is βd, while the variance is $2\beta^2 d$. In order to calculate the expected value and variance of T, we use Proposition 2.1, which states that if $\boldsymbol{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{A}$ is a symmetric $p \times p$ constant matrix, we have $\mathrm{E}[\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}] = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$ and $\mathrm{Var}[\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}] = 2\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\mu}$. Substituting $\boldsymbol{y} := \boldsymbol{C}\hat{\boldsymbol{\mu}} - \boldsymbol{c}$, $p := q$, $\boldsymbol{\mu} = \boldsymbol{0}_{\boldsymbol{q}}$, $\boldsymbol{\Sigma} := \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}'$ and $\boldsymbol{A} := (\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}')^{-1}$ (it is easy to check that $\boldsymbol{A}$ is symmetric since $\boldsymbol{D}$ is symmetric) gives us

$$\mathrm{E}[T] = \mathrm{tr}((\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}') \text{ and } \mathrm{Var}[T] = 2\mathrm{tr}(((\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^2). \tag{44}$$

By matching the expected values and variances of T and $\beta\chi^2_d$ as well as substituting $\boldsymbol{\Sigma}$ with its unbiased estimator $\hat{\boldsymbol{\Sigma}}$, we obtain that

$$\beta = \frac{\mathrm{tr}(\boldsymbol{X}^2)}{\mathrm{tr}(\boldsymbol{X})} \text{ and } d = \frac{(\mathrm{tr}(\boldsymbol{X}))^2}{\mathrm{tr}(\boldsymbol{X}^2)}, \tag{45}$$

where $\boldsymbol{X} = (\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}')^{-1}\boldsymbol{C}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}'$. Using this method, we obtain the following results:

- When $\alpha = 0.05$, $n_1 = 10$, $n_2 = 100$, $n_3 = 1000$, $\boldsymbol{\Sigma_1} = \mathrm{diag}(\mathrm{rep}(10, 8))$, $\boldsymbol{\Sigma_2} = \mathrm{diag}(\mathrm{rep}(0.05, 8))$ and $\boldsymbol{\Sigma_3} = \mathrm{diag}(\mathrm{rep}(0.001, 8))$, the empirical size of the new method is 0.025 (compare this with 0.320 for T_F *Gen* and 0.414 for T_{FM} *Gen*).

- When $\alpha = 0.05$, $n_1 = 10$, $n_2 = 100$, $n_3 = 1000$, $\boldsymbol{\Sigma_1} = \mathrm{diag}(\mathrm{rep}(1, 8))$, $\boldsymbol{\Sigma_2} =$

diag(rep$(1, 8)$) and $\mathbf{\Sigma_3} =$ diag(rep$(1, 8)$), the empirical size of the new method is 0.034 (compare this with 0.309 for T_F *Gen* and 0.378 for T_{FM} *Gen*).

From these results, it can be seen that this new method gives a much better size for the case when the samples are drawn from high-dimensional normal populations.

4 An Application to the Egyptian Skull Data

4.1 The Egyptian Skull Data

In this subsection, we apply the generalised methods developed in Section 3 to a real dataset known as the Egyptian Skull data. This dataset, which can be obtained from `https://www3.nd.edu/~busiforc/handouts/Data%20and%20Stories/regression/egyptian%20skull%20development/EgyptianSkulls.html`, contains measurements of male Egyptian skulls from 5 (five) different time periods: the early predynastic period (4000 BC), the late predynastic period (3300 BC), the 12th and 13th dynasties (1850 BC), the Ptolemaic period (200 BC) and the Roman period (AD 150). 30 (thirty) skulls are measured from each time period, which means that there are 150 data points in total.

The variables measured for each skull are maximum breadth (X_1), borborygmatic height (X_2), dentoalveolar length (X_3) and nasal height (X_4), all of which are expressed in millimetres. A categorical variable X_5 indicates the time period of the skull. An example of a data point is $(X_1, X_2, X_3, X_4, X_5) = (131, 138, 89, 49, -4000)$.

Following Zhang (2012), we check the significance of the mean vector differences of the first k samples using only the first 10, 20 and 30 observations, for $k \in \{2, 3, 4, 5\}$. In total, we consider $3 \times 4 = 12$ cases. We compare the generalised methods with a method called parametric bootstrap (PB) introduced by Krishnamoorthy and Lu (2010). The PB method has been shown to perform well in numerous conditions, so that it may be used as a benchmark to compare the performance of the two generalised methods we developed previously (Zhang, 2012). Despite the remarkable performance of the PB method (Krishnamoorthy and Lu, 2010), this method is less preferred as it requires a lot of bootstrap replications (Zhang, 2012), which significantly affect the time taken to perform a hypothesis testing.

4.2 A Comparison of the Generalised Methods with the Parametric Bootstrap (PB) Method

The table below compares the p-value obtained from performing hypothesis testings using the generalised Yanagihara and Yuan's (2005) method (abbreviated to T_F *Gen*), generalised Krishnamoorthy and Yu's (2004) method (abbreviated to T_{FM} *Gen*) and the PB method. The mean vectors are ordered chronologically; for example, $\boldsymbol{\mu}_1$ refers to the mean vector corresponding to the skulls from the early predynastic period (4000 BC), and $\boldsymbol{\mu}_5$ refers to the mean vector corresponding to the skulls from the Roman period (AD 150).

Null hypothesis	First 10 observations			First 20 observations			First 30 observations		
	PB	Gen T_F	Gen T_FM	PB	Gen T_F	Gen T_FM	PB	Gen T_F	Gen T_FM
\mu_1 = \mu_2	0.6412	0.6431	0.6448	0.7194	0.7223	0.7227	0.8182	0.8141	0.8142
\mu_1 = \mu_2 = \mu_3	0.6107	0.6179	0.6234	0.2063	0.2083	0.2071	0.0326	0.0306	0.0298
\mu_1 = \mu_2 = \mu_3 = \mu_4	0.105	0.1225	0.1105	0.0225	0.0248	0.0227	0.0002	0.0003	0.0002
\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5	0.0502	0.0669	0.0532	0.0021	0.0032	0.0025	0.0000	0.0000	0.0000

It can be seen that the p-values for all three methods are about the same. However, for the last two null hypotheses, the generalised Krishnamoorthy and Yu's (2004) method gives closer p-values to the benchmark as compared to the generalised Yanagihara and Yuan's (2005) method, which tends to overestimate the p-values. Hence, for this dataset, it seems that the generalised Krishnamoorthy and Yu's (2004) is a better method.

5 Conclusion

In this project, we compare 8 (eight) approximate solutions to the multivariate Behrens-Fisher (MBF) problem. The Monte-Carlo simulation performed shows that Yanagihara and Yuan's (2005) main method (T_F) and Krishnamoorthy and Yu's (2004) method (T_{FM}) are the best in terms of the Type I error. A more thorough simulation reveals that T_{FM} is more stable across different sample sizes and covariance matrices than its counterpart; in the case when $\mathbf{\Sigma_1}$ is small, $\mathbf{\Sigma_2}$ is large and $\frac{n_1}{n_2}$ is large, T_{FM} performs quite well whereas T_F performs very badly.

We also discuss Zhang's (2012) generalisation to Krishnamoorthy and Yu's (2004) method (T_{FM} *Gen*), which is capable of dealing with the general linear hypothesis testing (GLHT) problem in heteroscedastic one-way MANOVA. We then use Zhang's (2012) idea to extend Yanagihara and Yuan's (2005) method (the extension is abbreviated to T_F *Gen*). The Monte-Carlo simulations performed for the trivariate and five-variate cases show that T_F *Gen* is better than T_{FM} *Gen* in terms of the Type I error. However, when both methods are applied to the Egyptian Skull data and compared with the parametric bootstrap (PB) method (Krishnamoorthy and Lu, 2010) as a benchmark, it is found that T_{FM} *Gen* is a better method.

We note that both T_F *Gen* and T_{FM} *Gen* perform very badly when the samples are drawn from high-dimensional multivariate normal distributions. A new method, whose test statistic does not depend on $\hat{\mathbf{\Sigma}}$, is introduced. Although this method gives a much better size as compared to T_F *Gen* and T_{FM} *Gen*, it results in a low power. Further research may therefore focus on developing new approximate solutions to the GLHT problem in heteroscedastic one-way MANOVA which have reasonable size and power even in the case of high-dimensional data.

6 Appendix: R Codes

6.1 Calculating the Empirical Sizes for All Eight Methods (MBF Problem)

```
#import necessary libraries
library(MASS)
library(rlist)

#compute empirical size for all eight methods at once
#(MBF Problem as described in Section 2.3)
tall_es <- function(alpha,p,n1,n2,Sigma1,Sigma2,nrep){
  count <- rep(0,8)
  for(i in seq(nrep)){
    n <- n1+n2
    sample1 <- mvrnorm(n=n1,mu=rep(1,p),Sigma=Sigma1,tol=1e-6)
    sample2 <- mvrnorm(n=n2,mu=rep(1,p),Sigma=Sigma2,tol=1e-6)

    y1bar <- as.matrix(colMeans(sample1))
    y2bar <- as.matrix(colMeans(sample2))
    S1 <- cov(sample1)
    S2 <- cov(sample2)

    Sbar <- (n2/n)*S1+(n1/n)*S2
    T <- t(y1bar-y2bar)%*%solve(S1/n1+S2/n2)%*%(y1bar-y2bar)

    psi1hat <- ((n2**2*(n-2))/(n**2*(n1-1)))*(sum(diag(S1%*%
    solve(Sbar))))**2 + ((n1**2*(n-2))/(n**2*(n2-1)))*
    (sum(diag(S2%*%solve(Sbar))))**2
```

```
psi2hat <- ((n2**2*(n-2))/(n**2*(n1-1)))*(sum(diag(S1%*%
solve(Sbar)%*%S1%*%solve(Sbar))) + ((n1**2*(n-2))/(n**
2*(n2-1)))*(sum(diag(S2%*%solve(Sbar)%*%S2%*%solve(Sbar))))

#Method 2
theta1hat <- (p*psi1hat+(p-2)*psi2hat)/(p*(p+2))
theta2hat <- (psi1hat+2*psi2hat)/(p*(p+2))
vhat <- ((n-2-theta1hat)**2)/((n-2)*theta2hat-theta1hat)
T_F <- ((n-2-theta1hat)/((n-2)*p))*T

#Method 3
c1hat <- (psi1hat+psi2hat)/p
T_B <- (1-c1hat/(n-2))*T

#Method 4
beta1hat <- (p*(p+2))/(psi1hat+2*psi2hat)
beta2hat <- -((p+2)*psi1hat)/(2*(psi1hat+2*psi2hat))
T_MB <- ((n-2)*beta1hat+beta2hat)*log(1+T/((n-2)*beta1hat))

#Method 5
upp <- qchisq(alpha,p,lower.tail=FALSE)
upp_approximated <- upp*(1+((p+2)*psi1hat+(psi1hat+2*psi2hat)*
upp)/(2*p*(p+2)*(n-2)))

#Method 6
ydbar <- y1bar-y2bar
vyhat <- (n**2*(n1-1)*(n2-1)*(t(ydbar)%*%solve(Sbar)%*%ydbar)**
2)/(n2**2*(n2-1)*(t(ydbar)%*%solve(Sbar)%*%S1%*%solve(Sbar)%*%
ydbar)**2+n1**2*(n1-1)*(t(ydbar)%*%solve(Sbar)%*%S2%*%
solve(Sbar)%*%ydbar)**2)
```

```
    T_FY <- ((vyhat-p+1)*T)/(vyhat*p)

    #Method 7
    phijhat <- p+((p-1)*(psi1hat+psi2hat))/((p+2)*(n-2))
    vjhat <- (2*p*(p+2)*(n-2))/(3*(psi1hat+psi2hat))
    T_FJ <- T/phijhat

    #Method 8
    vmhat = (p*(p+1)*(n-2))/(psi1hat+psi2hat)
    T_FM <- ((vmhat-p+1)*T)/(vmhat*p)

    count[1] <- count[1]+as.integer(T>qchisq(alpha,p,lower.tail=FALSE))
    count[2] <- count[2]+as.integer(T_F>qf(alpha,p,vhat,lower.tail=FALSE))
    count[3] <- count[3]+as.integer(T_B>qchisq(alpha,p,lower.tail=FALSE))
    count[4] <- count[4]+as.integer(T_MB>qchisq(alpha,p,lower.tail=FALSE))
    count[5] <- count[5]+as.integer(T>upp_approximated)
    count[6] <- count[6]+as.integer(T_FY>qf(alpha,p,vyhat-p+1,
    lower.tail=FALSE))
    count[7] <- count[7]+as.integer(T_FJ>qf(alpha,p,vjhat,
    lower.tail=FALSE))
    count[8] <- count[8]+as.integer(T_FM>qf(alpha,p,vmhat-p+1,
    lower.tail=FALSE))
  }
  return(round(count/nrep,4))
}
```

6.2 Calculating the Empirical Sizes for All Two Methods (Trivariate One-Way MANOVA)

```
#import necessary libraries
library(MASS)
library(rlist)

#compute empirical size for all two methods at once
#(trivariate one-way MANOVA as described in Section 3.6)
multitall3_es <- function(alpha,p,n1,n2,n3,Sigma1,Sigma2,Sigma3,nrep){
  count <- c(0,2)
  for(i in seq(nrep)){
    n <- n1+n2+n3
    N <- n-3
    q <- 6
    sample1 <- mvrnorm(n=n1,mu=rep(0,3),Sigma=Sigma1,tol=1e-6)
    sample2 <- mvrnorm(n=n2,mu=rep(0,3),Sigma=Sigma2,tol=1e-6)
    sample3 <- mvrnorm(n=n3,mu=rep(0,3),Sigma=Sigma3,tol=1e-6)

    C <- cbind(rbind(diag(3),diag(3)),-1*diag(6))
    C1 <- C[,c(1,2,3)]
    C2 <- C[,c(4,5,6)]
    C3 <- C[,c(7,8,9)]

    muhat1 <- as.matrix(colMeans(sample1))
    muhat2 <- as.matrix(colMeans(sample2))
    muhat3 <- as.matrix(colMeans(sample3))
    muhat <- rbind(muhat1,muhat2,muhat3)

    Sigmahat1 <- cov(sample1)
```

```
Sigmahat2 <- cov(sample2)
Sigmahat3 <- cov(sample3)
zeromatrix <- matrix(0,nrow=3,ncol=3)
Sigmahat <- cbind(rbind(Sigmahat1/n1,zeromatrix,zeromatrix),rbind(
zeromatrix,Sigmahat2/n2,zeromatrix),rbind(zeromatrix,zeromatrix,
Sigmahat3/n3))

T <- t(C%*%muhat)%*%solve(C%*%Sigmahat%*%t(C))%*%(C%*%muhat)

psi1hat <- N*(sum(diag(solve(C%*%Sigmahat%*%t(C))%*%C1%*%
Sigmahat1%*%t(C1)))^2/(n1^2*(n1-1))+sum(diag(solve(C%*%
Sigmahat%*%t(C))%*%C2%*%Sigmahat2%*%t(C2)))^2/(n2^2*(n2-1))+
sum(diag(solve(C%*%Sigmahat%*%t(C))%*%C3%*%Sigmahat3%*%
t(C3)))^2/(n3^2*(n3-1)))

J1 <- solve(C%*%Sigmahat%*%t(C))%*%C1%*%Sigmahat1%*%t(C1)
J2 <- solve(C%*%Sigmahat%*%t(C))%*%C2%*%Sigmahat2%*%t(C2)
J3 <- solve(C%*%Sigmahat%*%t(C))%*%C3%*%Sigmahat3%*%t(C3)
psi2hat <- N*(sum(diag(J1%*%J1))/(n1^2*(n1-1))+sum(diag(J2%*%
J2))/(n2^2*(n2-1))+sum(diag(J3%*%J3))/(n3^2*(n3-1)))

#Method 1
theta1hat <- (q*psi1hat+(q-2)*psi2hat)/(q*(q+2))
theta2hat <- (psi1hat+2*psi2hat)/(q*(q+2))
vhat <- ((N-theta1hat)**2)/((N)*theta2hat-theta1hat)
Omega1hat <- (1/n1) * ((C%*%Sigmahat%*%t(C))%^%(-0.5)) %*%
(C1%*%Sigmahat1%*%t(C1)) %*% ((C%*%Sigmahat%*%t(C))%^%(-0.5))
Omega2hat <- (1/n2) * ((C%*%Sigmahat%*%t(C))%^%(-0.5)) %*%
(C2%*%Sigmahat2%*%t(C2)) %*% ((C%*%Sigmahat%*%t(C))%^%(-0.5))
Omega3hat <- (1/n3) * ((C%*%Sigmahat%*%t(C))%^%(-0.5)) %*%
```

```
    (C3%*%Sigmahat3%*%t(C3)) %*% ((C%*%Sigmahat%*%t(C))%^%(-0.5))
    T_F <- ((N-theta1hat)/((N)*q))*T

    #Method 2
    K1 <- (sum(diag(Omega1hat%*%Omega1hat))+(sum(diag(Omega1hat)))^
    2)/(n1-1)
    K2 <- (sum(diag(Omega2hat%*%Omega2hat))+(sum(diag(Omega2hat)))^
    2)/(n2-1)
    K3 <- (sum(diag(Omega3hat%*%Omega3hat))+(sum(diag(Omega3hat)))^
    2)/(n3-1)
    dhat <- (q*(q+1))/(K1+K2+K3)
    T_FM <- ((dhat-q+1)/(q*dhat))*T

    count[1] <- count[1]+as.integer(T_F>qf(alpha,q,vhat,lower.tail=FALSE))
    count[2] <- count[2]+as.integer(T_FM>qf(alpha,q,dhat-q+1,
    lower.tail=FALSE))
  }
  return(round(count/nrep,4))
}
```

6.3 Calculating the Empirical Sizes for All Two Methods (Five-Variate One-Way MANOVA)

```
#import necessary libraries
library(MASS)
library(rlist)

#compute empirical size for all two methods at once
#(five-variate one-way MANOVA as described in Section 3.6)
multitall5_es <- function(alpha,p,n1,n2,n3,n4,n5,Sigma1,Sigma2,Sigma3,
Sigma4,Sigma5,nrep){
  count <- c(0,2)
  for(i in seq(nrep)){
    n <- n1+n2+n3+n4+n5
    N <- n-5
    q <- 20
    sample1 <- mvrnorm(n=n1,mu=rep(0,5),Sigma=Sigma1,tol=1e-6)
    sample2 <- mvrnorm(n=n2,mu=rep(0,5),Sigma=Sigma2,tol=1e-6)
    sample3 <- mvrnorm(n=n3,mu=rep(0,5),Sigma=Sigma3,tol=1e-6)
    sample4 <- mvrnorm(n=n4,mu=rep(0,5),Sigma=Sigma4,tol=1e-6)
    sample5 <- mvrnorm(n=n5,mu=rep(0,5),Sigma=Sigma5,tol=1e-6)

    C <- cbind(rbind(diag(1,5),diag(1,5),diag(1,5),diag(1,5)),diag(-1,20))
    C1 <- C[,c(1,2,3,4,5)]
    C2 <- C[,c(6,7,8,9,10)]
    C3 <- C[,c(11,12,13,14,15)]
    C4 <- C[,c(16,17,18,19,20)]
    C5 <- C[,c(21,22,23,24,25)]

    muhat1 <- as.matrix(colMeans(sample1))
```

```
muhat2 <- as.matrix(colMeans(sample2))
muhat3 <- as.matrix(colMeans(sample3))
muhat4 <- as.matrix(colMeans(sample4))
muhat5 <- as.matrix(colMeans(sample5))
muhat <- rbind(muhat1,muhat2,muhat3,muhat4,muhat5)

Sigmahat1 <- cov(sample1)
Sigmahat2 <- cov(sample2)
Sigmahat3 <- cov(sample3)
Sigmahat4 <- cov(sample4)
Sigmahat5 <- cov(sample5)
zeromatrix <- matrix(0,nrow=5,ncol=5)
Sigmahat <- cbind(rbind(Sigmahat1/n1,zeromatrix,zeromatrix,
zeromatrix,zeromatrix),rbind(zeromatrix,Sigmahat2/n2,zeromatrix,
zeromatrix,zeromatrix),rbind(zeromatrix,zeromatrix,Sigmahat3/n3,
zeromatrix,zeromatrix),rbind(zeromatrix,zeromatrix,zeromatrix,
Sigmahat4/n4,zeromatrix),rbind(zeromatrix,zeromatrix,zeromatrix,
zeromatrix,Sigmahat5/n5))

T <- t(C%*%muhat)%*%solve(C%*%Sigmahat%*%t(C))%*%(C%*%muhat)

psi1hat <- N*(sum(diag(solve(C%*%Sigmahat%*%t(C))%*%C1%*%
Sigmahat1%*%t(C1)))^2/(n1^2*(n1-1))+sum(diag(solve(C%*%
Sigmahat%*%t(C))%*%C2%*%Sigmahat2%*%t(C2)))^2/(n2^2*(n2-1))+
sum(diag(solve(C%*%Sigmahat%*%t(C))%*%C3%*%Sigmahat3%*%
t(C3)))^2/(n3^2*(n3-1))+sum(diag(solve(C%*%Sigmahat%*%t(C))%*%
C4%*%Sigmahat4%*%t(C4)))^2/(n4^2*(n4-1))+sum(diag(solve(C%*%
Sigmahat%*%t(C))%*%C5%*%Sigmahat5%*%t(C5)))^2/(n5^2*(n5-1)))

J1 <- solve(C%*%Sigmahat%*%t(C))%*%C1%*%Sigmahat1%*%t(C1)
```

```
J2 <- solve(C%*%Sigmahat%*%t(C))%*%C2%*%Sigmahat2%*%t(C2)
J3 <- solve(C%*%Sigmahat%*%t(C))%*%C3%*%Sigmahat3%*%t(C3)
J4 <- solve(C%*%Sigmahat%*%t(C))%*%C4%*%Sigmahat4%*%t(C4)
J5 <- solve(C%*%Sigmahat%*%t(C))%*%C5%*%Sigmahat5%*%t(C5)
psi2hat <- N*(sum(diag(J1%*%J1))/(n1^2*(n1-1))+sum(diag(J2%*%
J2))/(n2^2*(n2-1))+sum(diag(J3%*%J3))/(n3^2*(n3-1))+sum(diag(J4%*%
J4))/(n4^2*(n4-1))+sum(diag(J5%*%J5))/(n5^2*(n5-1)))

#Method 1
theta1hat <- (q*psi1hat+(q-2)*psi2hat)/(q*(q+2))
theta2hat <- (psi1hat+2*psi2hat)/(q*(q+2))
vhat <- ((N-theta1hat)**2)/((N)*theta2hat-theta1hat)
Omega1hat <- (1/n1) * ((C%*%Sigmahat%*%t(C))%^%(-0.5)) %*%
(C1%*%Sigmahat1%*%t(C1)) %*% ((C%*%Sigmahat%*%t(C))%^%(-0.5))
Omega2hat <- (1/n2) * ((C%*%Sigmahat%*%t(C))%^%(-0.5)) %*%
(C2%*%Sigmahat2%*%t(C2)) %*% ((C%*%Sigmahat%*%t(C))%^%(-0.5))
Omega3hat <- (1/n3) * ((C%*%Sigmahat%*%t(C))%^%(-0.5)) %*%
(C3%*%Sigmahat3%*%t(C3)) %*% ((C%*%Sigmahat%*%t(C))%^%(-0.5))
Omega4hat <- (1/n4) * ((C%*%Sigmahat%*%t(C))%^%(-0.5)) %*%
(C4%*%Sigmahat4%*%t(C4)) %*% ((C%*%Sigmahat%*%t(C))%^%(-0.5))
Omega5hat <- (1/n5) * ((C%*%Sigmahat%*%t(C))%^%(-0.5)) %*%
(C5%*%Sigmahat5%*%t(C5)) %*% ((C%*%Sigmahat%*%t(C))%^%(-0.5))
T_F <- ((N-theta1hat)/((N)*q))*T

#Method 2
K1 <- (sum(diag(Omega1hat%*%Omega1hat))+(sum(diag(Omega1hat)))^
2)/(n1-1)
K2 <- (sum(diag(Omega2hat%*%Omega2hat))+(sum(diag(Omega2hat)))^
2)/(n2-1)
K3 <- (sum(diag(Omega3hat%*%Omega3hat))+(sum(diag(Omega3hat)))^
```

```
    2)/(n3-1)
    K4 <- (sum(diag(Omega4hat%*%Omega4hat))+(sum(diag(Omega4hat)))^
    2)/(n4-1)
    K5 <- (sum(diag(Omega5hat%*%Omega5hat))+(sum(diag(Omega5hat)))^
    2)/(n5-1)
    dhat <- (q*(q+1))/(K1+K2+K3+K4+K5)
    T_FM <- ((dhat-q+1)/(q*dhat))*T

    count[1] <- count[1]+as.integer(T_F>qf(alpha,q,vhat,lower.tail=FALSE))
    count[2] <- count[2]+as.integer(T_FM>qf(alpha,q,dhat-q+1,
    lower.tail=FALSE))
  }
  return(round(count/nrep,4))
}
```

6.4 Calculating the Empirical Sizes for the Alternative Method (Trivariate One-Way MANOVA)

```
#import necessary libraries
library(MASS)
library(rlist)

#compute empirical size for the alternative method
#(trivariate one-way MANOVA as described in Section 3.7)
multitall3custom_es <- function(alpha,p,n1,n2,n3,Sigma1,Sigma2,Sigma3,
nrep){
  count <- 0
  for(i in seq(nrep)){
    n <- n1+n2+n3
    N <- n-3
    q <- 16
    sample1 <- mvrnorm(n=n1,mu=rep(0,8),Sigma=Sigma1,tol=1e-6)
    sample2 <- mvrnorm(n=n2,mu=rep(0,8),Sigma=Sigma2,tol=1e-6)
    sample3 <- mvrnorm(n=n3,mu=rep(0,8),Sigma=Sigma3,tol=1e-6)

    C <- cbind(rbind(diag(8),diag(8)),-1*diag(16))
    C1 <- C[,c(1,2,3,4,5,6,7,8)]
    C2 <- C[,c(9,10,11,12,13,14,15,16)]
    C3 <- C[,c(17,18,19,20,21,22,23,24)]

    muhat1 <- as.matrix(colMeans(sample1))
    muhat2 <- as.matrix(colMeans(sample2))
    muhat3 <- as.matrix(colMeans(sample3))
    muhat <- rbind(muhat1,muhat2,muhat3)
```

```
    Sigmahat1 <- cov(sample1)
    Sigmahat2 <- cov(sample2)
    Sigmahat3 <- cov(sample3)
    zeromatrix <- matrix(0,nrow=8,ncol=8)
    Sigmahat <- cbind(rbind(Sigmahat1/n1,zeromatrix,zeromatrix),
    rbind(zeromatrix,Sigmahat2/n2,zeromatrix),rbind(zeromatrix,
    zeromatrix,Sigmahat3/n3))

    D <- cbind(rbind(diag(rep(1/n1,8)),zeromatrix,zeromatrix),
    rbind(zeromatrix,diag(rep(1/n2,8)),zeromatrix),rbind(zeromatrix,zeromat
    T <- t(C%*%muhat)%*%solve(C%*%D%*%t(C))%*%(C%*%muhat)

    X <- solve(C%*%D%*%t(C))%*%C%*%Sigmahat%*%t(C)
    beta <- sum(diag(X%*%X))/sum(diag(X))
    d <- (sum(diag(X)))^2/sum(diag(X%*%X))
    count <- count+as.integer(T/beta>qchisq(alpha,d,lower.tail=FALSE))
  }
  return(round(count/nrep,4))
}
```

References

[1] Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A* 160:268-282.

[2] Fang, K. T., Kotz, S., Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. London: Chapman & Hall/CRC.

[3] Fujikoshi, Y. (2000). Transformations with improved chi-squared approximations. *Journal of Multivariate Analysis* 72:249-263.

[4] Gupta, A. K., Nagar, D. K. (1999). *Matrix Variate Distributions.* Florida: Chapman & Hall/CRC.

[5] Hanson, T. (2014). *STAT 730: Multivariate Analysis, chapter 3 notes* [LaTeX slides]. Retrieved from `http://people.stat.sc.edu/hansont/stat730/chapter3.pdf`.

[6] James, G. S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika* 41:19-43.

[7] Johansen, S. (1980). The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika* 67:85-92.

[8] Jung, S. (2013). *STAT 2221: Advanced Applied Multivariate Analysis, lecture 2 notes* [LaTeX document]. Retrieved from `https://www.stat.pitt.edu/sungkyu/course/2221Fall13/lec2.pdf`.

[9] Krishnamoorthy, K., Lu, F. (2010). A parametric bootstrap solution to the MANOVA under heteroscedasticity. *Journal of Statistical Computation and Simulation* 80:873-887.

[10] Krishnamoorthy, K., Yu, J. (2004). Modified Nel and van der Merwe test for the multivariate Behrens-Fisher problem. *Statistics & Probability Letters* 66:161-169.

[11] Nel, D. G., van der Merwe, C. A. (1986). A solution to the multivariate Behrens-Fisher problem. *Communications in Statistics - Theory and Methods* 15:3719-3735.

[12] Nel, D. G., van der Merwe, C. A., Moser, B. K. (1990). The exact distributions of the univariate and multivariate Behrens-Fisher statistics with a comparison of several solutions in the univariate case. *Communications in Statistics - Theory and Methods* 19:279-298.

[13] Rencher, A. C., Schaalje, G. B. (2008). *Linear Models in Statistics*. New Jersey: Wiley.

[14] Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika* 29:350-362.

[15] Yanagihara, H., Yuan, K-H. (2005). Three approximate solutions to the multivariate Behrens-Fisher problem. *Communications in Statistics - Simulation and Computation* 34:975-988.

[16] Yao, Y. (1965). An approximate degrees of freedom solution to the multivariate Behrens-Fisher problem. *Biometrika* 52:139-147.

[17] Zhang, J-T. (2012). An approximate Hotelling T^2-test for heteroscedastic one-way MANOVA. *Open Journal of Statistics* 2:1-11.