MULTIVARIATE NORMAL INFERENCE FOR HETEROGENEOUS SAMPLES AND AN APPLICATION TO META ANALYSIS

By

LI-CHI LIN

Bachelor of Science in Mathematics
Chung-Yuan Christian University
Chung-Li, Tao-Yuan
1990

Master of Science in Statistics
National Central University
Chung-Li, Tao-Yuan
1993

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2012

MULTIVARIATE NORMAL INFERENCE FOR

HETEROGENEOUS SAMPLES AND AN

APPLICATION TO META ANALYSIS

Dissertation Approved:

Dr. Ibrahim A. Ahmad

Dissertation Adviser

Dr. Carla L. Goad

Dr. Lan Zhu

Dr. Tieming Liu

Dr. Sheryl A. Tucker

Dean of the Graduate College

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

If $X_1, ..., X_n$ is a sample from a normal population, then to estimate the population mean the usual point estimator is the sample mean $\overline{X}_n$. However, if the collected data violate the assumption of "identically distributed" setting, that is, if each X_i has heterogeneous mean, estimating the "population mean" will no longer make sense, except when structuring those means. Sometimes researchers believe that their collected sample is from a single population with a common constant mean when it is not, and they want to test the "population mean" equal to a specified value μ_0 without realizing that their data has previously been polluted due to some known or unknown mechanism. Hence the chance of rejection will be affected by the degree the data are polluted. Therefore it is necessary to model the disturbance of the data caused by the external or internal mechanisms and do inference for the parameter of interest. For example, let a random sample $X_i,\ i = 1,..., n,$ be assumed independently, normally distributed with heterogeneous means $C_i\mu,\ i = 1,..., n,$ and common variance σ^2. Let $C_1,...,C_n$ be known, and assume that $X_1,..., X_n \sim indN(C_i\mu, \sigma^2)$. Although each X_i has different mean, there is still an "underlying" mean μ hidden in this model. Once μ is estimated, each mean $C_i\mu,\ i = 1,..., n,$ is obtained. Actually, this model is a linear regression model through the origin. For this univariate case, the model is very easy to estimate, while when extending it to multivariate case, the matrices $\boldsymbol{C}_i$'s become troublesome. A special case of interest for $\boldsymbol{C}_i$ is to assume it is a square matrix.

For the remainder of this chapter, a review of the literature for inferences of multivariate homogeneous mean models for single normal population is introduced in Section 1.2 as follows: Subsection 1.2.1 gives a review for inferences concerning the mean vector when the covariance matrix is unstructured. Subsections 1.2.2 to 1.2.4 are about inferences of the means assuming that the covariance matrices are patterned. Finally, Subsection 1.2.5 is about the inferences concerning both the means and covariance matrices. Section 1.3 is about the inferences for multivariate homogeneous mean model for k normal populations with $k \geq 2$. Section 1.4 gives a brief review for meta analysis. Section 1.5 formally introduces the proposed model under multivariate normal setting and gives an overall introduction for the contents of later chapters.

1.2 HOMOGENEOUS MEAN MODEL FOR SINGLE POPULATION

The p dimensional multivariate normal model has mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The basic statistical problem is to estimate the parameters with a sample of n observations $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ from the normal distribution with homogeneous mean $\boldsymbol{\mu}$ and homogeneous covariance matrix $\boldsymbol{\Sigma}$. The maximum likelihood estimator of $\boldsymbol{\mu}$ is just the sample mean and the maximum likelihood estimator of $\boldsymbol{\Sigma}$ is proportional to the matrix of sample variances and sample covariances. The sample covariance matrix is defined by

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_j - \overline{\boldsymbol{X}})', \qquad (1.1)$$

where $\overline{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i$, and $\boldsymbol{S}$ is unbiased for estimating $\boldsymbol{\Sigma}$ and follows Wishart distribution $W(\frac{1}{n-1}\boldsymbol{\Sigma}, n-1)$.

1.2.1 Inferences Concerning the Mean Vector When Covariance Matrix Is Unstructured

Tests for the mean $\boldsymbol{\mu}$ equal to a specified vector $\boldsymbol{\mu}_0$ have been discussed in many multivariate analysis textbooks (e.g. Anderson 2003, and Rencher 1998) for the cases that $\boldsymbol{\Sigma}$ is known as well as that $\boldsymbol{\Sigma}$ is unknown and unstructured. Since $\sqrt{n}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})$ is distributed according to $N(\boldsymbol{0},\boldsymbol{\Sigma})$, it follows that $n(\overline{\boldsymbol{X}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu})$ has a central chi-square distribution with p degrees of freedom for the case that $\boldsymbol{\Sigma}$ is known. For the case that $\boldsymbol{\Sigma}$ is unknown and unstructured, the likelihood of the homogeneous mean model given observations $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ is

$$\begin{aligned} L(\boldsymbol{\mu},\boldsymbol{\Sigma}\mid \boldsymbol{x}_1,\ldots,\boldsymbol{x}_n) &= \prod_{i=1}^{n}(2\pi)^{-\frac{p}{2}}\,|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\} \\ &= (2\pi)^{-\frac{np}{2}}\,|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\}, \end{aligned} \tag{1.2}$$

and the corresponding log likelihood is

$$\log L(\boldsymbol{\mu},\boldsymbol{\Sigma}\mid \boldsymbol{x}_1,\ldots,\boldsymbol{x}_n) = constant - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}),$$

where log is the logarithm taken to base *e*. Let $T^2 = n(\overline{\boldsymbol{X}}-\boldsymbol{\mu}_0)'\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu}_0)$. For the rest part of this subsection, the following theorem concerning Hotelling-T^2 distribution is stated and the likelihood ratio test for the hypothesis $H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0$ is developed and based on the T^2- statistic (Anderson 2003).

Theorem 1.1 (Anderson 2003) Let $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ be a sample from $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, and define $T^2 = n(\overline{\boldsymbol{X}}-\boldsymbol{\mu}_0)'\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu}_0)$. The distribution of $\left(\frac{T^2}{n-1}\right)\left(\frac{n-p}{p}\right)$ is noncentral F with p and $n-p$ degrees of freedom and noncentrality parameter $n(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)$. If $\boldsymbol{\mu}=\boldsymbol{\mu}_0$, then the *F*-distribution is central.

Since the T^2-statistic follows the Hotelling's T^2-distribution which is the generalized version of Student's *t* distribution, the confidence region of the mean vector can be derived on the basis of the T^2-statistic. The likelihood ratio for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is

$$\lambda = \frac{\max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} = \frac{L(\boldsymbol{\mu}_0, \hat{\boldsymbol{\Sigma}}_0)}{L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})} \tag{1.3}$$

where

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_0 &= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu}_0)(\boldsymbol{X}_j - \boldsymbol{\mu}_0)', \\ \hat{\boldsymbol{\mu}} &= \overline{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i, \text{ and} \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{X}_j - \hat{\boldsymbol{\mu}})'. \end{aligned} \tag{1.4}$$

Thus (1.3) becomes

$$\lambda = \frac{(2\pi)^{-\frac{np}{2}} |\hat{\boldsymbol{\Sigma}}_0|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu}_0)^T \hat{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}_0)\right)}{(2\pi)^{-\frac{np}{2}} |\hat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})^T \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})\right)}$$

$$= \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\left(tr\hat{\boldsymbol{\Sigma}}_0^{-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu}_0)(\boldsymbol{X}_i - \boldsymbol{\mu}_0)^T - tr\hat{\boldsymbol{\Sigma}}^{-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})^T\right)\right\}$$

$$= \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\left(tr(nI_p) - tr(nI_p)\right)\right\} = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2}(np - np)\right\} = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{\frac{n}{2}}.$$

Replacing $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_0$ using (1.4), $\lambda^{2/n}$ becomes

$$\lambda^{2/n} = \frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|} = \frac{|\sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_j - \overline{\boldsymbol{X}})'|}{|\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu}_0)(\boldsymbol{X}_j - \boldsymbol{\mu}_0)'|}.$$

Further, to derive the likelihood ratio criterion, the following Corollary is required.

Corollary 1.1 (Anderson 2003) For $\boldsymbol{C}$ nonsingular,

$$\begin{vmatrix} \boldsymbol{C} & \boldsymbol{y} \\ -\boldsymbol{y}' & 1 \end{vmatrix} = |\boldsymbol{C} + \boldsymbol{y}\boldsymbol{y}'| = \begin{vmatrix} 1 & -\boldsymbol{y}' \\ \boldsymbol{y} & \boldsymbol{C} \end{vmatrix} = |\boldsymbol{C}|(1+\boldsymbol{y}'\boldsymbol{C}^{-1}\boldsymbol{y}) .$$

Defining $\boldsymbol{A} = \sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_j - \overline{\boldsymbol{X}})'$ and using Corollary 1.1, we have

$$\lambda^{2/n} = \frac{|\boldsymbol{A}|}{|\boldsymbol{A} + n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)'|} = \frac{1}{1 + n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)'\boldsymbol{A}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)}$$
$$= \frac{1}{1 + n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)'\left[(n-1)\boldsymbol{S}\right]^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)} = \frac{1}{1 + T^2/(n-1)},$$

where T^2 is defined in Theorem 1.1. Thus the likelihood ratio test for $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ has rejection region $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_n : T^2 \geq C_0\}$ where $C_0 = \frac{(n-1)p}{n-p}F(1-\alpha, p, n-p)$ is such that $P(T^2 \geq C_0 \mid H_0) = \alpha$, the significance level of the test.

1.2.2 Inferences Concerning the Mean Vector When Covariance Matrix Has Compound Symmetry Structure

Define the p-variate mean vector $\boldsymbol{\mu} = (\mu_1, ..., \mu_p)'$. Wilks (1946) derived the exact likelihood ratio criterion for testing H_0 : equality of p entries of the mean vector $\boldsymbol{\mu}$ or $H_0 : \boldsymbol{\mu} = \mu\boldsymbol{1}_p$, where μ is an unknown real number and $\boldsymbol{1}_p$ is a $p \times 1$ vector with all entries equal to 1, when the covariance matrix $\boldsymbol{\Sigma}$ has compound symmetry structure as defined in (1.5). This could be done when the likelihood ratio criterion, which was also derived in the same paper, for testing $H_0 : \boldsymbol{\Sigma}$ has compound symmetry vs $H_a : \boldsymbol{\Sigma}$ is unstructured, does not have a significantly small value. The compound symmetry covariance matrix is of the form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \cdots & \rho\sigma^2 & \sigma^2 \end{pmatrix}, \tag{1.5}$$

where $\sigma > 0$ and $-(p-1)^{-1} < \rho < 1$ to ensure positive definiteness of the compound symmetry covariance structure of $\boldsymbol{\Sigma}$. This structure assumes that the unknown p variances are all equal through the common intra-class correlation.

Geisser (1963) derived the likelihood ratio test for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ where $\boldsymbol{\mu}_0$ is a known constant, when the underlying covariance matrix has a compound symmetry structure as shown in (1.5). In this paper, the likelihood ratio test statistic L for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ under the covariance matrix structure in (1.5) is of the form

$$L = \left(1 + \frac{1}{n-1} F_{p-1,(n-1)(p-1)}\right)^{-(p-1)} \left(1 + \frac{1}{n-1} F_{1,n-1}\right)^{-1}, \tag{1.6}$$

$$L = \left(1 + \frac{\chi^2_{p-1}}{\chi^2_{(p-1)(n-1)}}\right)^{-(p-1)} \left(1 + \frac{\chi^2_1}{\chi^2_{n-1}}\right)^{-1}, \tag{1.7}$$

or

$$L = B_1^{p-1} B_2, \tag{1.8}$$

where $F_{p-1,(n-1)(p-1)}$ and $F_{1,n-1}$ are independent F random variables with degrees of freedom indicated in subscripts and χ^2_{p-1}, χ^2_1, $\chi^2_{(p-1)(n-1)}$, and χ^2_{n-1} are independent chi-square random variables with the corresponding degrees of freedom shown in subscripts. B_1 and B_2 are independent beta variables $Beta(\frac{1}{2}(p-1)(n-1), \frac{1}{2}(p-1))$ and $Beta(\frac{1}{2}(n-1), \frac{1}{2})$, respectively, based on the following properties about beta random variables.

***Properties of beta random variables*:** (Bailey 1992) Let U and V be independent, $U \sim \chi^2(m)$, $V \sim \chi^2(n)$. Then $\frac{U}{U+V} \sim Beta(\frac{m}{2},\frac{n}{2})$.

The *r*th raw moment of L can be calculated easily and approximations to the distribution of the product has been studied by Tukey and Wilks (1946) such that finding approximate critical values for the test is feasible. The hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is rejected when L is sufficiently small.

1.2.3 Inferences Concerning the Mean Vector When Covariance Matrix Is Circulant

A *circulant matrix of order* p, or *circulant* in short, is a $p \times p$ square matrix of the form

$$\boldsymbol{A} = (a_{ij}) = \begin{pmatrix} a_0 & a_1 & \cdots & a_{p-1} \\ a_{p-1} & a_0 & \cdots & a_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \cdots & a_0 \end{pmatrix}. \qquad (1.9)$$

The elements of each row of the matrix $\boldsymbol{A}$ are identical to those of the previous row, but are moved one position to the right and wrapped around such that the last element of the previous row becomes the first element of the current row. Note that the whole circulant is evidently determined by the first row. Also we may denote the circulant $\boldsymbol{A}$ in (1.9) by

$$\boldsymbol{A} = circ(a_0, a_1, ..., a_{p-1}).$$

So $\boldsymbol{A}$ is a $p \times p$ circulant if and only if $a_{ij} = a_{(j-i)|p}$, where $(j-i)\,|\,p$ is defined as

$$(j-i)\,|\,p = \begin{cases} p+j-i & \text{when } i > j, \\ j-i & \text{when } i \le j. \end{cases}$$

For more details about circulant matrices, refer to Davis (1979) and Graybill (1983). If a positive definite covariance matrix is circulant, it must also be symmetric. Examples for circulant covariance matrices $circ(\sigma^2,\sigma^2\rho_1,...,\sigma^2\rho_{p-1})$ with $p=4$ and $p=5$ are, respectively,

$$\sigma^2\begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_1 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_1 & \rho_2 & \rho_1 & 1 \end{pmatrix}, \text{ and } \sigma^2\begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_2 & \rho_1 \\ \rho_1 & 1 & \rho_1 & \rho_2 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_1 & \rho_2 & \rho_2 & \rho_1 & 1 \end{pmatrix},$$

satisfying $\rho_j = \rho_{p-j}$ of the symmetric circulant covariance matrix $\boldsymbol{\Sigma}$ of the form

$$\boldsymbol{\Sigma} = \sigma^2\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_{p-1} & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_1 & \rho_2 & \cdots & 1 \end{pmatrix}. \tag{1.10}$$

If assuming $\rho_1 = ... = \rho_{p-1} = \rho$ in (1.10), the covariance matrix is said to be compound symmetric defined in (1.5).

Olkin and Press (1969) have found the MLEs of the mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and have derived the exact likelihood ratio criteria for testing equality of p entries of the mean vector $\boldsymbol{\mu}$ and the mean vector $\boldsymbol{\mu}$ equal to zero when the covariance matrix $\boldsymbol{\Sigma}$ has a circulant structure. Their derivations for estimation and testing started by making the transformations on $\overline{\boldsymbol{X}}$ and $\boldsymbol{S}$ such that $\boldsymbol{Y} = n^{1/2}\overline{\boldsymbol{X}}\boldsymbol{\Gamma}$, $\boldsymbol{V} = \boldsymbol{\Gamma}'\boldsymbol{S}\boldsymbol{\Gamma}$, where $\overline{\boldsymbol{X}}$ and $\boldsymbol{S}$ are sample mean and sample covariance matrix as defined in (1.1). $\boldsymbol{\Gamma}$ is orthogonal such that it transforms the circulant covariance matrix $\boldsymbol{\Sigma}$ to diagonal form. Note that $\boldsymbol{Y}$ and $\boldsymbol{V}$ are independent. They also derived the likelihood ratio tests and asymptotic approximations of the test statistics for means and covariance matrices. They simultaneously tested (i) that the mean vector $\boldsymbol{\mu}$ are zero and the covariance matrix is circulant, (ii) that the p entries of the mean vector $\boldsymbol{\mu}$ are all equal and the covariance matrix is circulant, both against general alternatives that all the entries of $\boldsymbol{\mu}$ are real numbers and the covariance matrix is positive definitive.

1.2.4 Inferences Concerning the Mean Vector When Covariance Matrix Is Block Compound Symmetry

The estimating and testing problems for block compound symmetry arising from multivariate normal distributions was first studied by Votaw (1948). He proposed twelve hypotheses and tested them using likelihood ratio method. An introduction of the six hypotheses for one sample will be mentioned in Subsection 1.2.5. The other six hypotheses for *k* samples $(k \geq 2)$ are stated in Section 1.3.

A more recent paper that estimated and tested concerning means and covariance matrices under block compound symmetry covariance structure is given by Szatrowski (1982). In his paper, two types of covariance structures – block compound symmetry of type I (BCS-I) and block compound symmetry of type II (BCS-II) were considered. The problem of testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ given that the covariance matrix has the block compound symmetry structure was also considered. In his paper, estimating and testing were based on maximum likelihood method. Null distributions of likelihood ratio statistics of the form $\lambda^{2/n} = |\hat{\boldsymbol{\Sigma}}_\Omega| / |\hat{\boldsymbol{\Sigma}}_\omega|$ were simplified for some special cases of Votaw's six hypotheses for single population, where Ω is the parameter space under the alternative hypothesis, ω is the parameter space under the null hypothesis. $\hat{\boldsymbol{\Sigma}}_\Omega$ is the MLE of covariance matrix under the alternative hypothesis and $\hat{\boldsymbol{\Sigma}}_\omega$ is the MLE of covariance matrix under the null hypothesis. Also the moments of $\lambda^{2/n}$ were obtained under the null and the approximate null distributions of $-2\log\lambda$ were found using Box's approximation (1949).

A BCS-I assumption can be illustrated by the following example. Suppose that a standard test score of college calculus is a random variable X_1 with mean μ_1. There are a set of three other alternative tests, namely X_2, X_3 ,and X_4 with means μ_2, μ_3, and μ_4, respectively. So the vector $\boldsymbol{X} = (X_1, X_2, X_3, X_4)'$ forms a 4×1 normal random vector with mean

$\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)'$. Under the block compound symmetry of type I (BCS-I) assumption, the covariance structure is of the form

$$\begin{pmatrix} A & C & C & C \\ C & B & D & D \\ C & D & B & D \\ C & D & D & B \end{pmatrix}. \qquad (1.11)$$

The hypothesis of interest is the interchangeability of variables X_2, X_3, and X_4. It is equivalent to the hypothesis that the vector $\boldsymbol{X}$ has mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_2, \mu_2)'$ and the covariance structure is of the form in (1.11). That is the random vectors $(X_1 \vdots X_2, X_3, X_4)'$, $(X_1 \vdots X_2, X_4, X_3)'$, $(X_1 \vdots X_3, X_2, X_4)'$, $(X_1 \vdots X_3, X_4, X_2)'$, $(X_1 \vdots X_4, X_2, X_3)'$, and $(X_1 \vdots X_4, X_3, X_2)'$ have the same distribution. For a more general case, consider *b* distinct standard tests and *h* sets of alternative tests, each of which measures n_i abilities. That is, $\boldsymbol{X}$ is partitioned into *b* + *h* subsets and forms a $b + \sum_{i=1}^{h} n_i = p$ -variate random vector. Under the BCS-I assumption, within each subset of variates, the means are equal, the variances are equal, and the covariances are equal and between any two distinct subsets of variates, the covariances are equal.

In regard to the BCS-II assumption, we may consider the following example. Assume that there are two types of tests of cognitive abilities. Each type of cognitive tests measures the abilities of verbal (V) and thinking (T). So the two types of test scores are assumed to be a multivariate 4×1 normal random vector $\boldsymbol{Y} = (Y_1, Y_2 \vdots Y_3, Y_4)'$ with mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)'$, where Y_1 and Y_2 are scores of verbal ability for type I and type II tests, respectively; Y_3 and Y_4 are scores of thinking ability for type I and type II tests, respectively. Under the compound symmetry of type II (CS-II) assumption, the mean of $\boldsymbol{Y}$ reduces to $\boldsymbol{\mu} = (\mu_1, \mu_1, \mu_3, \mu_3)'$, and the covariance matrix is of the form

$$\begin{pmatrix} A & C & E & F \\ C & A & F & E \\ E & F & B & D \\ F & E & D & B \end{pmatrix}. \tag{1.12}$$

The test of hypothesis of interest would be $\mu_1 = \mu_2$, $\mu_3 = \mu_4$ and that the covariance matrix has BCS-II structure shown in (1.12). Or equivalently be the test of *simultaneous* interchangeability of two types of measures for verbal and thinking abilities. For example, the distributions of $(Y_1, Y_2 \vdots Y_3, Y_4)'$ and $(Y_2, Y_1 \vdots Y_4, Y_3)'$ are the same but the distributions of $(Y_1, Y_2 \vdots Y_3, Y_4)'$ and $(Y_2, Y_1 \vdots Y_3, Y_4)'$ are not the same. These kinds of tests can also be applied to medical research especially for repeated measurements (Crowder & Hand 1990) data when comparing the effect of treatment and control groups (Morrison, 1972). For a more general case, one can consider *n* types of tests and *h* types of measures of cognitive abilities such that $\boldsymbol{Y}$ is an $n \times h$ random vector.

1.2.5 Inferences Concerning Both Means and Covariance Matrices

Wilks (1946) tested the hypothesis that a normal *p*-variate distribution has a complete symmetry covariance matrix structure as shown in (1.5) versus the hypothesis that the covariance matrix is unstructured by likelihood ratio test. In this paper, he also derived the LRT for testing $\boldsymbol{\mu} = \mu \boldsymbol{1}_p$ and $\boldsymbol{\Sigma}$ is compound symmetry simultaneously against the general alternative that all the entries of $\boldsymbol{\mu}$ are real numbers and the covariance matrix is positive definitive.

Votaw (1948) first studied the problem of estimating and testing for block compound symmetry in data arising from multivariate normal distributions. He extended Wilks' result by considering a normal *p*-variate random vector which can be partitioned in *q* mutually independent subsets of which *b* subsets contain exactly one variate each and the remaining $q - b = h$ subsets ($h \geq 1$) contain $n_1, \ldots, n_h$ variates, respectively, where $n_\alpha \geq 2$; $\alpha = 1, \ldots, h$; $b + n_1 + \ldots + n_h = p$. Let $(1^b, n_1, \ldots, n_h)$ denotes such a partition of a the *p*-variate random vector. Without loss of generality, assume $n_1 \leq \ldots \leq n_h$. A special case is that b = 0. For assumptions of block compound symmetry of type I and type II, Section 1.2.4 has given a brief introduction. In his paper, Votaw (1948) proposed 6 null hypotheses for testing the means or covariances or both based on a single sample. These hypotheses are: 1) $H_1(mvc)$, 2) $H_1(vc)$, 3) $H_1(m)$, 4) $\overline{H}_1(mvc)$, 5) $\overline{H}_1(vc)$, and 6) $\overline{H}_1(m)$. The hypotheses 1-3 are for BCS-I assumptions and the remaining three are for BCS-II assumptions. The null hypotheses 1, 2, 4, and 5 are against the alternative hypothesis that the means are real numbers and the covariance matrix is positive definite. The statements of the above six hypotheses are as follows:

- $H_1(mvc)$ is the hypothesis that within each subset of variates, the means are equal, the variances are equal, and the covariances are equal and that between any two distinct subsets of variates, the covariances are equal.
- $H_1(vc)$ is the hypothesis that within each subset of variates, the variances are equal and the covariances are equal and that between any two distinct subsets of variates, the covariances are equal.
- $H_1(m)$ is the hypothesis that within each subset of variates, the means are equal, given that the variances are equal and the covariances are equal and that between any two distinct subsets, the covariances are equal.
- $\overline{H}_1(mvc)$ is the hypothesis that within each subset of variates, the means are equal, the variances are equal, and the covariances are equal and that between any two distinct subsets of variates, the diagonal covariances are equal and the off-diagonal covariances are equal.
- $\overline{H}_1(vc)$ is the hypothesis that within each subset of variates, the variances are equal and the covariances are equal and that between any two distinct subsets of variates, the diagonal covariances are equal and the off-diagonal covariances are equal.

- $\overline{H}_1(m)$ is the hypothesis that within each subset of variates, the means are equal, given that the variances are equal and the covariances are equal and that between any two distinct subsets of variates, the diagonal covariances are equal and the off-diagonal covariances are equal.

Votaw derived the likelihood ratio for each hypothesis. In his paper, he also developed an explicit expression of the likelihood ratio criterion for each hypothesis and found its *r*th moment and approximate distribution when the corresponding hypothesis is true.

Olkin and Press (1969) have considered the problem of 1) testing the null that $\boldsymbol{\Sigma}$ has complete symmetry versus the alternative hypothesis that $\boldsymbol{\Sigma}$ is a circulant; 2) testing the null that $\boldsymbol{\Sigma} = \sigma^2 I$ versus the alternative hypothesis that $\boldsymbol{\Sigma}$ is a circulant; 3) testing the null hypothesis that $\boldsymbol{\Sigma}$ is a circulant versus the alternative hypothesis that $\boldsymbol{\Sigma}$ is positive definite.

1.3 HOMOGENEOUS MEAN MODELS FOR *k* POPULATIONS WITH $k \geq 2$

Votaw (1948) tested the following hypotheses based on *k* samples: 1*) $H_k(MVC \mid mvc)$, 2*) $H_k(VC \mid mvc)$, 3*) $H_k(M \mid mVC)$, 4*) $\overline{H}_k(MVC \mid mvc)$, 5*) $\overline{H}_k(VC \mid mvc)$, and 6*) $\overline{H}_k(M \mid mVC)$. The hypotheses 1-3 are for BCS-I assumptions and the rest three are for BCS-II assumptions. The statements of the above six hypotheses are as follows:

- $H_k(MVC \mid mvc)$ is the hypothesis that *k* normal *p*-variate distributions are the same given that they all satisfy $H_1(mvc)$ which is introduced in section 1.2.5.
- $H_k(VC \mid mvc)$ is the hypothesis that *k* normal *p*-variate distribution have the same variance-covariance matrix given that they all satisfy $H_1(mvc)$.
- $H_k(M \mid mVC)$ is the hypothesis that *k* normal *p*-variate distributions are the same given that they all satisfy $H_1(mvc)$ and that they all have the same variance-covariance matrix.

- $\overline{H}_k(MVC \mid mvc)$ is the hypothesis that *k* normal *p*-variate distributions are the same given that they all satisfy $\overline{H}_1(mvc)$ which is introduced in section 1.2.5.
- $\overline{H}_k(VC \mid mvc)$ is the hypothesis that *k* normal *p*-variate have the same variance-covariance matrix given that they all satisfy $\overline{H}_1(mvc)$.
- $\overline{H}_k(M \mid mVC)$ is the hypothesis that *k* normal *p*-variate distributions are the same given that they all satisfy $\overline{H}_1(mvc)$ and that they all have the same variance-covariance matrix.

For each of the above six hypotheses, Votaw developed the likelihood ratio test in terms of deriving the explicit expression of the likelihood ratio criteria $L = \lambda^2$, where λ is the likelihood ratio, for the hypotheses 1* – 4* and $L = \lambda^{2/N}$ for the remaining two hypotheses, where N is total number of sample sizes for all k populations. He also found the *r*th moment and approximate distribution for each test hypothesis.

Geisser (1963) compared the means of k p-variate normal populations under the assumption that the *k* normal populations have the common compound (complete) symmetry covariance structure using multivariate analysis of variance approach implemented by use of the information criterion (Chapter 9, Kullback 1959).

1.4. META ANALYSIS

Meta analysis has been widely used to synthesize results from systematic reviews of reliable research in many fields. There has been a massive growth in application of meta analysis to areas such as medical research, health care, education (Glass, 1976), criminal justice, social policy, etc. See Kulinskaya et al. (2008) and Sutton et al. (2000) for a detailed account of meta analysis. A recent development of meta analysis has been summarized by Sutton and Higgins (2008).

One uses a fixed effect model to combine treatment or parameter estimates when assuming no heterogeneity between the study results. In fact, point estimates of parameters from different studies are almost always different. If the differences of the point estimates are only simply due to sampling error, that is, the source of variation between studies is random variation, we can use a fixed effect model. Sometimes the researchers prefer to believe that the true unknown parameters from different studies vary from one study to the next, the studies represent a random sample of the parameters that could have been observed and comes from a specific distribution. Under this situation, a random effects model will be considered in the analysis.

The standard fixed effect model in meta-analysis is that if we have k independent studies, with data, each of which reports an estimate $\hat{\mu}_{(i)}$ for a common parameter μ. Each estimate $\hat{\mu}_{(i)}$ is assumed independently, normally distributed as

$$\hat{\mu}_{(i)} \sim N(\mu, \frac{\sigma_i^{\ 2}}{n_i}), \forall i = 1,...,k, \qquad (1.13)$$

where is the sample size of *i*th study and $\sigma_i^{\ 2}$ is the underlying variance parameter for ith study. Given ($\hat{\mu}_{(i)}, \sigma_i^{\ 2}, n_i$) the ML estimator for μ and its variance are, respectively,

$$\tilde{\mu} = \frac{\sum_{i=1}^{k} n_i \sigma_i^{\ -2} \hat{\mu}_{(i)}}{\sum_{i=1}^{k} n_i \sigma_i^{\ -2}}, \qquad (1.14)$$

and

$$Var(\tilde{\mu}) = \frac{1}{\sum_{i=1}^{k} n_i \sigma_i^{\ -2}}. \qquad (1.15)$$

Now consider the multivariate models with k independent samples, each of which $\boldsymbol{X}_{i1,...,}\boldsymbol{X}_{in_i}$ is from $MVN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$ population for $i = 1,...,k$. Suppose $\boldsymbol{\mu}$ is the parameter vector of interest. The ML estimator for $\boldsymbol{\mu}$ based on *i*th sample is

$$\hat{\boldsymbol{\mu}}_{(i)} = \overline{\boldsymbol{X}}_i, \forall i = 1,...,k. \qquad (1.16)$$

Here we assume $\boldsymbol{\Sigma}_i$ is known for all $i=1,...,k$. In fact, $\hat{\boldsymbol{\mu}}_{(i)}$'s are independent and

$$\hat{\boldsymbol{\mu}}_{(i)} \sim MVN(\boldsymbol{\mu}, \frac{1}{n_i}\boldsymbol{\Sigma}_i), \forall i=1,...,k. \tag{1.17}$$

Given ($\hat{\boldsymbol{\mu}}_{(i)}, \boldsymbol{\Sigma}_i^2, n_i$) for the k studies, the ML estimator for $\boldsymbol{\mu}$ and its variance-covariance matrix based on the k independent samples are respectively

$$\tilde{\boldsymbol{\mu}} = \left(\sum_{i=1}^{k}\left(n_i\boldsymbol{\Sigma}_i^{-1}\right)\right)^{-1}\sum_{i=1}^{k} n_i\boldsymbol{\Sigma}_i^{-1}\hat{\mu}_{(i)}, \tag{1.18}$$

$$Cov(\tilde{\boldsymbol{\mu}}) = \left(\sum_{i=1}^{k}\left(n_i\boldsymbol{\Sigma}_i^{-1}\right)\right)^{-1}. \tag{1.19}$$

Statistical inferences are based on the fact that

$$(\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu})'\left(\sum_{i=1}^{k}\left(n_i\boldsymbol{\Sigma}_i^{-1}\right)\right)(\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}) \sim \chi^2_{p,\alpha}. \tag{1.20}$$

Applications of the proposed heterogeneous means normal model to random and fixed effects meta analysis will be developed and presented in Chapter 4. The proposed models will be stated in the next section.

1.5 PROPOSED HETEROGENEOUS MEANS MODELS

Consider an independent sample $\boldsymbol{X}_1,...,\boldsymbol{X}_M$ such that $\boldsymbol{X}_i \sim MVN_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}_i = \boldsymbol{C}_i\boldsymbol{\mu}$ for all $i=1,...,M$, and both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. The matrices $\boldsymbol{C}_i$ are $p \times p$ for all $i=1,...,M$ and the covariance matrix $\boldsymbol{\Sigma}$ is positive definite. Some further restrictions will be considered later for $\boldsymbol{C}_i$ when necessary. The likelihood function is

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \boldsymbol{x}_1,..., \boldsymbol{x}_M) &= \prod_{i=1}^{M}(2\pi)^{-\frac{p}{2}} \mid \boldsymbol{\Sigma} \mid^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})\} \\ &= (2\pi)^{-\frac{Mp}{2}} \mid \boldsymbol{\Sigma} \mid^{-\frac{M}{2}} \exp\{-\frac{1}{2}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})\}. \end{aligned} \tag{1.21}$$

The covariance matrix $\boldsymbol{\Sigma}$ is patterned in order to make the maximum likelihood estimator (MLE) of $\boldsymbol{\mu}$ vector not involve the ML estimator of $\boldsymbol{\Sigma}$.

Based on the likelihood function for a given sample, inferences for one-sample and multisample data are presented in Chapters 2 and 3, respectively. The likelihood ratio test for one-sample case for $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is derived explicitly under some constraints on the matrices $\boldsymbol{C}_i$ and covariance matrix $\boldsymbol{\Sigma}$. Especially, $\boldsymbol{C}_i$ is assumed circulant for all *i*. $\boldsymbol{\Sigma}$ is assumed compound (complete) symmetry of the form in (1.5). The distributions of the MLEs of the intraclass correlation ρ and variance σ^2, namely $\hat{\rho}$ and $\hat{\sigma}^2$, respectively, are obtained and the behavior of $\hat{\rho}$ is investigated in terms of its mean and standard deviation by a simulation study. For the two-sample and multisample cases, the likelihood ratio test for testing $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k$ is derived exactly assuming equal compound symmetry covariance matrix for the k populations. Large sample χ^2 test is gained for each of one-sample and two-sample cases.

An application of the proposed model to meta analysis is developed in Chapter 4. In traditional meta analysis, the sample from each study is assumed independently, identically distributed, while the sample from the proposed model is not the case. In Chapter 4, applications of the proposed model to fixed and random effects models for multivariate meta analysis (Jackson et al., 2011, Nam et al., 2003) about continuous outcomes will be developed and presented. Since the outcome measures in the proposed model are non-comparative continuous, one-stage method for individual patient / participant data (IPD) random effects model is suggested by Higgins et al. (2001) to investigate the heterogeneity of the effects (parameters) among several studies.

CHAPTER II

ONE-SAMPLE INFERENCE

2.1 INTRODUCTION AND PRELIMINARY CASES

Consider an independent sample of size *M*, $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_M \sim MVN_p(\boldsymbol{\mu}_i,\boldsymbol{\Sigma})$, where $\boldsymbol{\mu}_i = \boldsymbol{C}_i\boldsymbol{\mu}$ for all $i=1,\ldots,M$, and both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. The matrices $\boldsymbol{C}_i$ are $p\times p$ for all $i=1,\ldots,M$ and the covariance matrix $\boldsymbol{\Sigma}$ is positive definite. Some further restrictions will be considered later for $\boldsymbol{C}_i$ when necessary. The likelihood function is already shown in (1.21), thus the log likelihood function is

$$\begin{aligned}\log L(\boldsymbol{\mu},\boldsymbol{\Sigma}\mid \boldsymbol{x}_1,\ldots,\boldsymbol{x}_M) &= -\frac{Mp}{2}\log(2\pi)-\frac{M}{2}\log|\boldsymbol{\Sigma}|-\frac{1}{2}\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu})\\ &= constant-\frac{M}{2}\log|\boldsymbol{\Sigma}|-\frac{1}{2}\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}).\end{aligned} \tag{2.1}$$

For simplicity, $\log L(\boldsymbol{\mu},\boldsymbol{\Sigma}\mid \boldsymbol{x}_1,\ldots,\boldsymbol{x}_M)$ will be expressed as $\log L(\boldsymbol{\mu},\boldsymbol{\Sigma}\mid \boldsymbol{x})$ from now on. Let $Q=\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu})$. Our goal is to find the MLEs for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We can start by rewriting the log likelihood function in (2.1) such that maximizing $\log L(\boldsymbol{\mu},\boldsymbol{\Sigma}\mid \boldsymbol{x})$, or equivalently minimizing Q with respect to $\boldsymbol{\mu}$, becomes easier. But Q can be expressed as

$$Q = tr\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}) = \sum_{i=1}^{M} tr\left[(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})\right]$$
$$= \sum_{i=1}^{M} tr\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\right] = tr\left[\sum_{i=1}^{M}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\right]$$
$$= tr\left[\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\right] = tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{V}],$$

where $\boldsymbol{V} = \sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T$. Define $\hat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i\right)^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i,$ then $\boldsymbol{V}$ can be expressed as

$$\boldsymbol{V} = \sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}} + \boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}} + \boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})^T$$
$$= \boldsymbol{A} + \sum_{i=1}^{M}(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})^T$$
$$+ \sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})^T + \sum_{i=1}^{M}(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T,$$

where $\boldsymbol{A} = \sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T$.

Hence we have

$$Q = tr\boldsymbol{\Sigma}^{-1}\boldsymbol{A} + tr\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{M}(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})^T$$
$$+ tr\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})^T + tr\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{M}(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T$$
$$= tr\boldsymbol{\Sigma}^{-1}\boldsymbol{A} + tr\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{M}(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})^T,$$

where the second equality is justified by

$$tr\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})^T = \sum_{i=1}^{M} tr\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T\boldsymbol{C}_i^{\,T}\right]$$
$$= (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) = 0.$$

Likewise, we have $tr\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{M}(\boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T = 0.$ Hence Q can be expressed as

$$Q = tr\boldsymbol{\Sigma}^{-1}\boldsymbol{A} + tr\left[\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{M}[\boldsymbol{C}_i(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})][\boldsymbol{C}_i(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})]^T\right]$$
$$= tr\boldsymbol{\Sigma}^{-1}\boldsymbol{A} + \sum_{i=1}^{M}[\boldsymbol{C}_i(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})]^T\boldsymbol{\Sigma}^{-1}[\boldsymbol{C}_i(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})].$$

Therefore, the log likelihood becomes

$$\log L(\boldsymbol{\mu},\boldsymbol{\Sigma}\mid\boldsymbol{x}) = constant - \frac{M}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\left\{tr\boldsymbol{\Sigma}^{-1}\boldsymbol{A} + (\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})^T\left[\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i\right](\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})\right\}, \tag{2.2}$$

where $\boldsymbol{A} = \sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T$, $\hat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i\right)^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i$. We can base on the log likelihood expressed in (2.2) to find the MLEs for $\boldsymbol{\mu}$ and/or $\boldsymbol{\Sigma}$ under some specified conditions.

2.1.1 Inference for $\boldsymbol{\mu}$ When $\boldsymbol{\Sigma}$ Is Known

From the log likelihood derived in (2.2), we can see that the third term of the right-hand side is the only one involving $\boldsymbol{\mu}$. If $\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i$ is a positive definite matrix, the minimum of Q, w.r.t. $\boldsymbol{\mu}$, occurs at

$$\hat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i\right)^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i, \tag{2.3}$$

which is the MLE of $\boldsymbol{\mu}$, a linear combination of $\boldsymbol{X}_i$'s. Note that $\hat{\boldsymbol{\mu}}$ is normally distributed with mean

$$E(\hat{\boldsymbol{\mu}}) = E\left\{\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i\right)^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i\right\}$$
$$= \left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i\right)^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i\boldsymbol{\mu} = \boldsymbol{\mu},$$

and the covariance matrix $Cov(\hat{\boldsymbol{\mu}})$ obtained in the following way. Since $\hat{\boldsymbol{\mu}}$ satisfies the identity

$\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i \hat{\boldsymbol{\mu}} = \sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_i$, taking covariance on both sides yields

$$
\begin{aligned}
& Cov\left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i \hat{\boldsymbol{\mu}}\right) = Cov\left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_i\right) \\
\Rightarrow & \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i\right) Cov(\hat{\boldsymbol{\mu}}) \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i\right)^T = \sum_{i=1}^{M} (\boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1}) Cov(\boldsymbol{X}_i) (\boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1})^T \\
& = \sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \, \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i = \sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i \quad (\boldsymbol{\Sigma} \text{ is positive definite}) \\
\Rightarrow & \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i\right) Cov(\hat{\boldsymbol{\mu}}) \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i\right) = \sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i \\
\Rightarrow & Cov(\hat{\boldsymbol{\mu}}) = \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i\right)^{-1}.
\end{aligned}
$$

Hence $\hat{\boldsymbol{\mu}}$ is normally distributed as

$$
\hat{\boldsymbol{\mu}} \sim MVN_p\left(\boldsymbol{\mu}, \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i\right)^{-1}\right), \tag{2.4}
$$

which leads to the result

$$
(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i\right) (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \sim \chi_p^2 .
$$

Therefore, for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ we reject H_0 if

$$
(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^T \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_i\right) (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0) > \chi_{p,\alpha}^2 .
$$

2.1.2 Inference for $\boldsymbol{\mu}$ When $\boldsymbol{\Sigma}$ Is Unknown without Pattern

When $\boldsymbol{\Sigma}$ is unknown we have the MLE of $\boldsymbol{\mu}$ which has the same form as that in (2.3) with $\boldsymbol{\Sigma}$ replaced by $\hat{\boldsymbol{\Sigma}}$, the MLE of $\boldsymbol{\Sigma}$. Hence the MLE of $\boldsymbol{\mu}$ is

$$\hat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{C}_i \right)^{-1} \sum_{i=1}^{M} \boldsymbol{C}_i^T \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{X}_i, \tag{2.5}$$

where $\hat{\boldsymbol{\Sigma}}$ is the MLE of $\boldsymbol{\Sigma}$. Therefore based on a result of Anderson (2003, Lemma 3.2.2, p. 69) in connection with (2.2), we have the MLE of $\boldsymbol{\Sigma}$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{M} \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})(\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})^T. \tag{2.6}$$

We should note that the expression of $\hat{\boldsymbol{\mu}}$ in (2.5) involves $\hat{\boldsymbol{\Sigma}}$. Recall that in the iid case, if $\boldsymbol{\mu}_i = \boldsymbol{\mu}$ for all *i*, the MLE of $\boldsymbol{\mu}$ does not involve $\boldsymbol{\Sigma}$ at all. In general, there are no explicit solutions for $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ and the equations in (2.5) and (2.6) need to be solved iteratively for $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$. Thus the approximation $(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{C}_i \right) (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \xrightarrow{D} \chi_p^2$ (Crowder and Hand, 1990) is still attainable such that testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ asymptotically can be done. Nevertheless, to remove $\hat{\boldsymbol{\Sigma}}$ in (2.5) such that the MLEs $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ can be gained explicitly, we should consider a patterned covariance matrix $\boldsymbol{\Sigma}$ with details about inference for $\boldsymbol{\mu}$ covered in Section 2.2. Before doing so, let us consider another structure of $\boldsymbol{\Sigma}$ in the next subsection.

2.1.3 Inference for $\boldsymbol{\mu}$ When $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{V}$, σ^2 Unknown, $\boldsymbol{V}$ Known

Recall that $\boldsymbol{X}_i \sim MVN_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$. In this subsection, we consider the case that $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{V}$, where $\sigma^2 > 0$ is an unknown constant and $\boldsymbol{V}$ is a known positive definite matrix. So $\boldsymbol{\mu}$ and σ^2 are the only unknown parameters. Therefore, the maximum likelihood estimator of $\boldsymbol{\mu}$ is

$$\hat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{V}^{-1} \boldsymbol{C}_i \right)^{-1} \sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{V}^{-1} \boldsymbol{X}_i \, .$$

To find the MLE of σ^2, let us consider the log likelihood function first. Define $\theta = \sigma^2$, the log likelihood function is

$$\ln L(\boldsymbol{\mu},\theta \mid \boldsymbol{x}) = constant - \frac{Mp}{2}\ln\theta - \frac{1}{2\theta}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{V}^{-1}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}),$$

which yields

$$\frac{\partial \ln L}{\partial\theta} = \frac{Mp}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{V}^{-1}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}).$$

Setting the above equation zero and solving for θ, the MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{Mp}\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}).$$

Since $\hat{\boldsymbol{\mu}}$ is a linear combination of $\boldsymbol{X}_i$'s, the distribution of $\hat{\boldsymbol{\mu}}$ can be found as

$$\hat{\boldsymbol{\mu}} \sim MVN_p(\boldsymbol{\mu}, \sigma^2\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right)^{-1}).$$

Next, the distribution of $Mp\hat{\sigma}^2/\sigma^2$ can be shown to follow χ^2 distribution with $p(M-1)$ degrees of freedom. We may also show that $\hat{\boldsymbol{\mu}}$ and $\hat{\sigma}^2$ are independent. To proceed, partition the quantity $\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})$. That is,

$$\begin{aligned}
&\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}) \\
&\quad = \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}} + \boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}} + \boldsymbol{C}_i\hat{\boldsymbol{\mu}} - \boldsymbol{C}_i\boldsymbol{\mu}) \\
&\quad = \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) + 2\sum_{i=1}^{M}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T\boldsymbol{C}_i^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) \\
&\qquad + (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T\left[\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right](\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \\
&\quad = \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) + (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T\left[\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right](\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})
\end{aligned}$$

We need to note that $\sum_{i=1}^{M}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T\boldsymbol{C}_i^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})$ equals 0 due to the fact that

$\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})=0$. Therefore we have $\sigma^{-2}\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu})$ equal to

$$\sigma^{-2}\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})+\sigma^{-2}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})^T\left[\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right](\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}).$$

We can show that both terms of the above quantity are independent by showing that each pair of $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}$ for all $i=1,\ldots,M$ are independent. Since both $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}$ are normally distributed, we can show that they are statistically independent by just showing that their covariance matrix is zero. That is,

$$\begin{aligned}Cov(\hat{\boldsymbol{\mu}},\boldsymbol{X}_k-\boldsymbol{C}_k\hat{\boldsymbol{\mu}})&=Cov(\hat{\boldsymbol{\mu}},\boldsymbol{X}_k)-Cov(\hat{\boldsymbol{\mu}},\boldsymbol{C}_k\hat{\boldsymbol{\mu}})\\&=Cov\left(\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right)^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{X}_i,\boldsymbol{X}_k\right)-Cov(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\mu}})\boldsymbol{C}_k^T\\&=\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right)^{-1}\boldsymbol{C}_k^T\boldsymbol{V}^{-1}Cov(\boldsymbol{X}_k,\boldsymbol{X}_k)-\sigma^2\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right)^{-1}\boldsymbol{C}_k^T\\&=\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right)^{-1}\boldsymbol{C}_k^T\boldsymbol{V}^{-1}(\sigma^2\boldsymbol{V})-\sigma^2\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right)^{-1}\boldsymbol{C}_k^T\\&=\sigma^2\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right)^{-1}\boldsymbol{C}_k^T-\sigma^2\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right)^{-1}\boldsymbol{C}_k^T=0.\end{aligned}$$

This implies that $\sigma^{-2}\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})$ and $\sigma^{-2}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})^T\left[\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right](\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})$ are statistically independent. In addition, using the result of sum of two independent chi-square random variables (Bain & Engelhardt 1992, page 284), we have

$$\sigma^{-2}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})^T\left[\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{V}^{-1}\boldsymbol{C}_i\right](\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})\sim\chi_p^2$$

implying that

$$Mp\hat{\sigma}^2/\sigma^2=\sigma^{-2}\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{V}^{-1}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})\sim\chi_{(M-1)p}^2.$$

Therefore, under H_0: $\boldsymbol{\mu}=\boldsymbol{\mu}_0$, we have

$$\frac{(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)^T\left[\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{V}^{-1}\boldsymbol{C}_i\right](\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)\Big/p}{M\hat{\sigma}^2\big/(M-1)} \sim F_{p,(M-1)p},$$

which can be used for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$.

2.2 MAINSTREAM: INFERENCE FOR $\boldsymbol{\mu}$ WHEN $\boldsymbol{\Sigma}$ HAS COMPOUND SYMMETRY STRUCTURE AND $\boldsymbol{C}_i$ ARE CIRCULANT

2.2.1 Maximum Likelihood Estimators

There are three conditions considered before deriving the MLEs for the unknown parameters. The theories developed later for section 2.2 are based on these three assumptions stated below.

Condition (1). If $\boldsymbol{C}_i^{\,T}\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i^{\,T}$ for all *i*, the MLE for $\boldsymbol{\mu}$ in (2.3) reduces to

$$\hat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i\right)^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{X}_i. \qquad (2.7)$$

Condition (2). To guarantee $\boldsymbol{C}_i^{\,T}\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i^{\,T}$ in Condition (1), we assume that $\boldsymbol{C}_i$ is a circulant matrix for every *i* and $\boldsymbol{\Sigma}$ has a compound symmetry structure. The following theorem will be applied to this condition.

Theorem 2.0: (Schott (1997): Theorem 7.58, page 303) Suppose that *A* and *B* are $m\times m$ circulant matrices. Then their product commutes; That is, $AB = BA$.

Let $\boldsymbol{\Sigma}$ have the structure

$$\boldsymbol{\Sigma} = \sigma^2[(1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{J}_p], \qquad (2.8)$$

by rewriting the covariance matrix defined in (1.5), where $\rho > -(p-1)^{-1}$ to ensure positive definiteness of $\boldsymbol{\Sigma}$. Note that the eigenvalues for $\boldsymbol{\Sigma}$ in (2.8) are $\sigma^2[1+(p-1)\rho]$ with multiplicity 1 and $\sigma^2(1-\rho)$ with multiplicity $p-1$. Thus $\boldsymbol{\Sigma}$ is a symmetric circulant matrix and we say $\boldsymbol{\Sigma}$ has compound symmetry which has been introduced in Subsection 1.2.2. For each $i=1,\ldots,M$, if $\boldsymbol{C}_i$ is also a circulant matrix, then we have $\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i^T$ which results in the reduced form of $\hat{\boldsymbol{\mu}}$ shown in (2.7). Working on the log likelihood function in (2.1) with $\boldsymbol{\Sigma}$ of the form in (2.8), we may get the MLEs for ρ and σ^2. To find the MLEs for ρ and σ^2, first note that the determinant and inverse of $\boldsymbol{\Sigma}$ are, respectively,

$$|\boldsymbol{\Sigma}| = (\sigma^2)^p(1-\rho)^{p-1}[1+(p-1)\rho], \text{ and } \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2(1-\rho)}[\boldsymbol{I}_p - \frac{\rho}{1+(p-1)\rho}\boldsymbol{J}_p] \ .$$

(cf. Graybill, 1983, Theorem 8.34, page 190.)

Let $\theta = \sigma^2$, the log likelihood function in (2.1) becomes

$$\begin{aligned}\log L(\boldsymbol{\mu},\theta,\rho \mid \boldsymbol{x}) = constant - \frac{M}{2}\{p\log\theta + (p-1)\log(1-\rho) + \log[1+(p-1)\rho]\} \\ -\frac{1}{2}\frac{1}{\theta(1-\rho)}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}) \\ +\frac{1}{2}\frac{1}{\theta(1-\rho)}\frac{\rho}{1+(p-1)\rho}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{J}_p(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}).\end{aligned} \tag{2.9}$$

Let $B1 = \sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})$, and $B2 = \sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})$, where $\boldsymbol{J}_p$ is a $p \times p$ square matrix with all elements equal to 1. To find the maximum likelihood estimators for $\theta\,(=\sigma^2)$ and ρ, we take the first partial derivative of the log likelihood function in (2.9) with respect to θ and ρ separately. So we have

$$\frac{\partial \log L(\boldsymbol{\mu},\theta,\rho \mid \boldsymbol{x})}{\partial\theta} = -\frac{M}{2}\frac{p}{\theta} + \frac{1}{2\theta^2(1-\rho)}B1 - \frac{1}{2\theta^2(1-\rho)}\frac{\rho}{1+(p-1)\rho}B2$$

and

$$\frac{\partial \log L(\boldsymbol{\mu},\theta,\rho\,|\,\boldsymbol{x})}{\partial \rho} = -\frac{M}{2}\left\{\frac{-(p-1)}{1-\rho}+\frac{p-1}{1+(p-1)\rho}\right\}-\frac{1}{2}\frac{1}{\theta(1-\rho)^2}B1$$
$$-\frac{(1-\rho)[1+(p-1)\rho]-\{(1-\rho)(p-1)-[1+(p-1)\rho]\}\rho}{2\theta(1-\rho)^2[1+(p-1)\rho]^2}B2$$
$$= -\frac{M}{2}\left\{\frac{-(p-1)}{1-\rho}+\frac{p-1}{1+(p-1)\rho}\right\}-\frac{1}{2}\frac{1}{\theta(1-\rho)^2}B1$$
$$+\frac{1+(p-1)\rho^2}{2\theta(1-\rho)^2[1+(p-1)\rho]^2}B2.$$

Setting $\frac{\partial \log L(\hat{\boldsymbol{\mu}},\theta,\rho\,|\,\boldsymbol{x})}{\partial \theta}=0$ and $\frac{\partial \log L(\hat{\boldsymbol{\mu}},\theta,\rho\,|\,\boldsymbol{x})}{\partial \rho}=0$ and solving for θ and ρ, we have

$$\hat{\theta}=\frac{1}{Mp(1-\hat{\rho})}\left\{B1-\frac{\hat{\rho}}{1+(p-1)\hat{\rho}}B2\right\} \tag{2.10}$$

and

$$-M\left\{\frac{-(p-1)}{1-\hat{\rho}}+\frac{p-1}{1+(p-1)\hat{\rho}}\right\}-\frac{1}{\hat{\theta}(1-\hat{\rho})^2}\left\{B1-\frac{1+(p-1)\hat{\rho}^2}{[1+(p-1)\hat{\rho}]^2}B2\right\}=0. \tag{2.11}$$

Note that $\hat{\theta}$ in (2.10) can also be expressed as

$$\hat{\theta}=\frac{1}{Mp(1-\hat{\rho})}\left\{\frac{[1+(p-1)\hat{\rho}]B1-\hat{\rho}B2}{1+(p-1)\hat{\rho}}\right\}.$$

Inserting $\hat{\theta}$ in (2.10) into (2.11) and solving for $\hat{\rho}$ yields

$$\frac{Mp(p-1)\hat{\rho}}{(1-\hat{\rho})[1+(p-1)\hat{\rho}]}-\frac{1}{\hat{\theta}(1-\hat{\rho})^2}\left\{\frac{[1+(p-1)\hat{\rho}]^2B1-[1+(p-1)\hat{\rho}^2]B2}{[1+(p-1)\hat{\rho}]^2}\right\}=0$$
$$\Rightarrow\ Mp(p-1)\hat{\rho}-\frac{1}{\hat{\theta}(1-\hat{\rho})}\left\{\frac{[1+(p-1)\hat{\rho}]^2B1-[1+(p-1)\hat{\rho}^2]B2}{1+(p-1)\hat{\rho}}\right\}=0$$
$$\Rightarrow\ (p-1)\hat{\rho}\{[1+(p-1)\hat{\rho}]B1-\hat{\rho}B2\}-[1+(p-1)\hat{\rho}]^2B1+[1+(p-1)\hat{\rho}^2]B2=0$$
$$\Rightarrow\ [1+(p-1)\hat{\rho}][(p-1)\hat{\rho}-1-(p-1)\hat{\rho}]B1+[-(p-1)\hat{\rho}^2+1+(p-1)\hat{\rho}^2]B2=0$$
$$\Rightarrow\ -[1+(p-1)\hat{\rho}]B1+B2=0,$$

which implies

$$\hat{\rho} = \frac{1}{p-1}\left(\frac{B2}{B1} - 1\right). \tag{2.12}$$

Substituting $\hat{\rho}$ in (2.12) into (2.10) , the MLE for σ^2 is

$$\hat{\sigma}^2 = \hat{\theta} = \frac{1}{Mp(1-\hat{\rho})}\left\{B1 - \frac{\hat{\rho}}{1+(p-1)\hat{\rho}}B2\right\} = \frac{1}{Mp(1-\hat{\rho})}\left\{B1 - \frac{\hat{\rho}}{1+(B2/B1-1)}B2\right\}$$
$$= \frac{1}{Mp(1-\hat{\rho})}\{B1 - \hat{\rho}B1\} = \frac{1}{Mp}B1.$$

Hence we arrive at the following lemma.

***Lemma 2.1*:** Let $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_M \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}_i = \boldsymbol{C}_i\boldsymbol{\mu}$ for all $i = 1,\ldots,M$, $\boldsymbol{C}_i$ is circulant and $\boldsymbol{\Sigma} = \sigma^2[(1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{J}_p]$ defined in (2.8) such that $\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i^T$. Then the MLEs for $\boldsymbol{\mu}$, σ^2, and ρ are, respectively,

$$\hat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{X}_i, \ \hat{\sigma}^2 = \frac{1}{Mp}B1, \text{ and } \ \hat{\rho} = \frac{1}{p\text{-}1}\left(\frac{B2}{B1} - 1\right),$$

where $B1 = \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})$, and $B2 = \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})$.

2.2.2 Hypothesis Testing for $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ Using LR Test

In this subsection, the likelihood ratio test will be derived for $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$. Restrictions $\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i^T$ for all $i = 1,\ldots,M$ are still valid here and we also assume that $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{R}$, where $\boldsymbol{R} = (1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{J}_p$, and both σ^2 and ρ are unknown. The following theorem states the likelihood ratio test for $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ under the above assumptions.

***Theorem 2.1*:** Let $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_M \sim N_p(\boldsymbol{\mu}_i,\boldsymbol{\Sigma})$, where $\boldsymbol{\mu}_i = \boldsymbol{C}_i\boldsymbol{\mu}$ for all $i=1,\ldots,M$, $\boldsymbol{C}_i$ is circulant and $\boldsymbol{\Sigma} = \sigma^2[(1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{J}_p]$ defined in (2.8). The likelihood ratio test for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is to reject H_0 if $W < C_\alpha$, where C_α is such that $P(W < C_\alpha \mid H_0) = \alpha$, and W is defined as:

$$W = \left[\frac{p\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) - \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})}{p\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0) - \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)}\right]^{p-1} \cdot \frac{\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})}{\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)},$$

where $\hat{\boldsymbol{\mu}}$ is defined in (2.7).

Proof:

The likelihood ratio λ for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is

$$\lambda = \frac{\max_{\boldsymbol{\theta}\in\Omega_0} L(\boldsymbol{\mu},\boldsymbol{\Sigma})}{\max_{\boldsymbol{\theta}\in\Omega} L(\boldsymbol{\mu},\boldsymbol{\Sigma})},$$

which can be further derived as follows:

$$\lambda = \frac{\max_{\boldsymbol{\theta}\in\Omega_0} L(\boldsymbol{\mu},\boldsymbol{\Sigma})}{\max_{\boldsymbol{\theta}\in\Omega} L(\boldsymbol{\mu},\boldsymbol{\Sigma})} = \frac{(2\pi)^{-\frac{Mp}{2}}|\hat{\boldsymbol{\Sigma}}_0|^{-\frac{M}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T\hat{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)\right)}{(2\pi)^{-\frac{Mp}{2}}|\hat{\boldsymbol{\Sigma}}|^{-\frac{M}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})\right)}$$

$$= \frac{|\hat{\boldsymbol{\Sigma}}_0|^{-\frac{M}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T\hat{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)\right)}{|\hat{\boldsymbol{\Sigma}}|^{-\frac{M}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})\right)}$$

$$= \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{\frac{M}{2}}\exp\left\{-\frac{1}{2}\left(tr\hat{\boldsymbol{\Sigma}}_0^{-1}\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T - tr\hat{\boldsymbol{\Sigma}}^{-1}\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\right)\right\},$$

where $\Omega = \{(\boldsymbol{\mu},\boldsymbol{\Sigma}) \mid \boldsymbol{\Sigma} \text{ is pd}\}$, $\Omega_0 = \{(\boldsymbol{\mu},\boldsymbol{\Sigma}) \mid \boldsymbol{\mu} = \boldsymbol{\mu}_0, \boldsymbol{\Sigma} \text{ is pd}\}$. Showing

$$tr\hat{\boldsymbol{\Sigma}}_0^{-1}\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T = Mp$$

and

$$tr\hat{\boldsymbol{\Sigma}}^{-1}\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T = Mp$$

from Appendix A.1 we obtain

$$\lambda=\frac{\left|\hat{\boldsymbol{\Sigma}}_0\right|^{-\frac{M}{2}}}{\left|\hat{\boldsymbol{\Sigma}}\right|^{-\frac{M}{2}}}\exp-\frac{1}{2}\left(Mp-Mp\right)=\frac{\left|\hat{\boldsymbol{\Sigma}}_0\right|^{-\frac{M}{2}}}{\left|\hat{\boldsymbol{\Sigma}}\right|^{-\frac{M}{2}}}=\left(\frac{(\hat{\sigma}^2)^p(1-\hat{\rho})^{p-1}[1+(p-1)\hat{\rho}]}{(\hat{\sigma}_0{}^2)^p(1-\hat{\rho}_0)^{p-1}[1+(p-1)\hat{\rho}_0]}\right)^{\frac{M}{2}}, \qquad (2.13)$$

where

$$\hat{\sigma}_0{}^2=\frac{1}{Mp}\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0),$$

$$\hat{\rho}_0=\frac{1}{p-1}\left[\frac{\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)}{\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)}-1\right],$$

$$\hat{\sigma}^2=\frac{1}{Mp}\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}),$$

and

$$\hat{\rho}=\frac{1}{p-1}\left[\frac{\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})}{\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})}-1\right].$$

Using the above expressions for $\hat{\sigma}_0{}^2$, $\hat{\rho}_0$, $\hat{\sigma}^2$, and $\hat{\rho}$ in (2.13), we gain the likelihood ratio test as stated in this theorem (detail shown in Appendix A.2). The proof is complete.

Although the likelihood ratio has been derived in Theorem 2.1, the null distribution of W in Theorem 2.1 is still not derived yet. Define

$$B1_0 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}_0)^T (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}_0) \text{ and } B2_0 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}_0)^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}_0).$$

$B1$ and $B2$ have been defined in Lemma 2.1. Hence W can be expressed as

$$W = \frac{[pB1 - B2]^{p-1} B2}{[pB1_0 - B2_0]^{p-1} B2_0} = \frac{\left[B1 - \frac{1}{p} B2\right]^{p-1} B2}{\left[B1_0 - \frac{1}{p} B2_0\right]^{p-1} B2_0}.$$

Under the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, the exact, asymptotic, or approximate distributions of W is of our great interest. To find the exact null distribution of W, the following propositions are needed.

<u>Proposition 2.1</u>: Under $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, $B1_0 - \frac{1}{p} B2_0$ is distributed as a chi-square random variable with $M(p-1)$ degrees of freedom times a constant $\sigma^2(1-\rho)$; that is,

$B1_0 - \frac{1}{p} B2_0 \stackrel{d}{=} \sigma^2 (1-\rho) \chi^2_{M(p-1)}$. In addition, $\frac{1}{M}\left(B1_0 - \frac{1}{p} B2_0\right)$ is strongly convergent to a constant $(p-1)\sigma^2(1-\rho)$, that is,

$$\frac{1}{M}\left(B1_0 - \frac{1}{p} B2_0\right) \xrightarrow{wp1} (p-1)\sigma^2(1-\rho).$$

Proof:

Under $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, the distribution of $\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}_0$ is $N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, where the covariance matrix $\boldsymbol{\Sigma} = \sigma^2 [(1-\rho)\boldsymbol{I}_p + \rho \boldsymbol{J}_p]$. It follows from Box (1954) that the quantities

$$Q_i = (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}_0)^T (\boldsymbol{I}_p - \frac{1}{p} \boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}_0), \; i = 1, \ldots, M$$

are independently, identically distributed like a $\sum_{j=1}^{p}\lambda_j\chi_1^2$ random variable, where λ_j 's are the latent roots of

$$\begin{aligned}\boldsymbol{P}_1 &= \boldsymbol{\Sigma}(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p) = \sigma^2\left[(1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{J}_p\right](\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p) \\ &= \sigma^2\left[(1-\rho)\boldsymbol{I}_p - \frac{(1-\rho)}{p}\boldsymbol{J}_p + \rho\boldsymbol{J}_p - \rho\boldsymbol{J}_p\right] \\ &= \sigma^2(1-\rho)\left(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p\right),\end{aligned} \tag{2.14}$$

and χ_1^2 's are independent chi-square random variables with 1 degree of freedom. Hence

$$B1_0 - \frac{1}{p}B2_0 = \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0) = \sum_{i=1}^{M}Q_i$$

is distributed as sum of M independent $\sum_{j=1}^{p}\lambda_j\chi_1^2$ random variables.

Because $\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p$ is symmetric idempotent, the latent roots of $\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p$ are 0's or 1's. In fact, the latent roots of $\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p$ are 1 with multiplicity $p-1$ and 0 with multiplicity 1. Therefore the latent roots of $\boldsymbol{P}_1$ are $\sigma^2(1-\rho)$ with multiplicity $p-1$ and 0 with multiplicity 1. So we have

$$Q_i \overset{iid}{\sim} \sigma^2(1-\rho)\chi_{p-1}^2 \text{ distribution} \tag{2.15}$$

for all $i = 1,\ldots,M$, implying that

$$B1_0 - \frac{1}{p}B2_0 \overset{d}{=} \sigma^2(1-\rho)\sum_{i=1}^{M}\chi_{p-1}^2 \overset{d}{=} \sigma^2(1-\rho)\chi_{M(p-1)}^2.$$

Moreover, based on SLLN in connection with (2.15), we have

$$\frac{1}{M}\left(B1_0 - \frac{1}{p}B2_0\right) = \frac{1}{M}\left[\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)\right]$$
$$\xrightarrow{wp1} E\left[\sigma^2(1-\rho)\chi^2(p-1)\right] = (p-1)\sigma^2(1-\rho).$$

The proof is complete.

<u>Proposition 2.2</u>: Under $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, $B2_0$ is distributed as a chi-square random variable with M degrees of freedom times a constant $p\sigma^2\left[1+(p-1)\rho\right]$, that is, $B2_0 \overset{d}{=} p\sigma^2\left[1+(p-1)\rho\right]\chi_M^2$. In addition, $\frac{1}{M}B2_0$ is strongly convergent to a constant $p\sigma^2\left[1+(p-1)\rho\right]$, that is,

$$\frac{1}{M}B2_0 \xrightarrow{wp1} p\sigma^2\left[1+(p-1)\rho\right].$$

Proof:

Recall that $B2_0 = \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T \boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)$, where $(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T \boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)$'s are iid random variables. First we have

$$(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T \boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0) \overset{d}{=} \sum_{j=1}^{p}\lambda_j\chi_1^2, \qquad (2.16)$$

where λ_j 's are latent roots of

$$\boldsymbol{P}_2 = \boldsymbol{\Sigma}\boldsymbol{J}_p = \sigma^2\left[(1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{J}_p\right]\boldsymbol{J}_p = \sigma^2\left[1+(p-1)\rho\right]\boldsymbol{J}_p.$$

Note that the latent roots of $\boldsymbol{J}_p$ is p with multiplicity 1 and 0 with multiplicity $p-1$. So λ_j 's are $p\sigma^2\left[1+(p-1)\rho\right]$ with multiplicity 1 and 0 with multiplicity $p-1$. Hence (2.16) becomes

$$(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T \boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0) \overset{d}{=} p\sigma^2\left[1+(p-1)\rho\right]\chi_1^2,$$

implying that

$$B2_0 \overset{d}{=} p\sigma^2\left[1+(p-1)\rho\right]\chi_M^2, \qquad (2.17)$$

and

$$\frac{1}{M}B2_0 = \frac{1}{M}\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T \boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)$$
$$\xrightarrow{wp1} E\left[p\sigma^2\left(1+(p-1)\rho\right)\chi^2(1)\right] = p\sigma^2\left[1+(p-1)\rho\right].$$

The proof is complete.

<u>Proposition 2.3</u>: $B1 - \frac{1}{p}B2$ is distributed as a chi-square random variable with $(M-1)(p-1)$ degrees of freedom times a constant $\sigma^2(1-\rho)$, that is,

$$B1 - \frac{1}{p}B2 \overset{d}{=} \sigma^2(1-\rho)\chi^2_{(M-1)(p-1)}\,.$$

Proof:

Assume that $E(\hat{\boldsymbol{\mu}}) = \boldsymbol{\mu}$. So $\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})$ can be written as

$$\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})$$
$$= B1 - \frac{1}{p}B2 + (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})$$
$$+ 2\sum_{i=1}^{M}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T\boldsymbol{C}_i^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}).$$

Because both $\boldsymbol{C}_i$ and $\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p$ are circulant matrices, $\boldsymbol{C}_i$ and $\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p$ are commutable such that $\boldsymbol{C}_i^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p) = (\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)\boldsymbol{C}_i^T$. In connection with the fact that $\sum_{i=1}^{M}\boldsymbol{C}_i^T(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) = 0$, we have $\sum_{i=1}^{M}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T\boldsymbol{C}_i^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) = 0$ implying that

$$B1-\frac{1}{p}B2=\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu})$$
$$-(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}). \tag{2.18}$$

From Proposition 2.1 we have

$$\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu})\overset{d}{=}\sigma^2(1-\rho)\chi^2_{M(p-1)}.$$

Also we need to know the distribution of the second term of the last expression in (2.18). Because we have that

$$\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}\sim MVN(\boldsymbol{0},\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)^{-1}\boldsymbol{\Sigma}),$$

the quadratic form

$$(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})$$

is distributed like the quantity $\sum_{j=1}^{p}\lambda_j\chi^2_1$, where λ_j's are the latent roots of the matrix

$$\boldsymbol{P}_3=\left[\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)^{-1}\boldsymbol{\Sigma}\right]\cdot\left[\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)\right].$$

Note that $\boldsymbol{\Sigma}$ and $\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i$ commute so $\boldsymbol{P}_3=\boldsymbol{\Sigma}\cdot(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)=\boldsymbol{P}_1$, where $\boldsymbol{P}_1$ is defined in (2.14). Hence we have

$$(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})\overset{d}{=}\sigma^2(1-\rho)\chi^2_{p-1}.$$

Note that $B1-\frac{1}{p}B2$ and $(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)$ are independent chi-square random variables since $B1-\frac{1}{p}B2$ is a function of $\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}$'s , also $\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\mu}}$ are independent due to the fact that $Cov(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}) = 0$. In addition, using the result of sum of two independent chi-square random variables, we have

$$B1 - \frac{1}{p} B2 \overset{d}{=} \sigma^2 (1-\rho)\chi^2_{(M-1)(p-1)}.$$

The proof is complete.

<u>Proposition 2.4</u>: $B2$ is distributed as a chi-square random variable with $M-1$ degrees of freedom times a constant $p\sigma^2[1+(p-1)\rho]$; that is,

$$B2 \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2_{M-1}.$$

Proof:

Assume that $E(\hat{\boldsymbol{\mu}}) = \boldsymbol{\mu}$. Using the fact that $\sum_{i=1}^{M} \boldsymbol{C}_i^T (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) = 0$, we have the expression for $B2$ that

$$\begin{aligned} B2 &= \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) \\ &= \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}) - (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \right) \boldsymbol{J}_p (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}). \end{aligned}$$

The second term $(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \right) \boldsymbol{J}_p (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})$ of the last expression above is distributed as the quantity $\sum_{j=1}^{p} \lambda_j \chi^2_1$ where λ_j's are the latent roots of the matrix

$$\boldsymbol{P}_4 = \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \right)^{-1} \boldsymbol{\Sigma} \cdot \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \right) \boldsymbol{J}_p = \boldsymbol{\Sigma} \boldsymbol{J}_p = \boldsymbol{P}_2,$$

where $\boldsymbol{P}_2$ is previously defined in the proof of Proposition 2.2. Hence

$$(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \right) \boldsymbol{J}_p (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2_1.$$

Since $B2$ and $(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i\right)\boldsymbol{J}_p(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})$ are independent and we have from Proposition 2.2 that

$$\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu})\overset{d}{=}p\sigma^2[1+(p-1)\rho]\chi_M^2,$$

$B2$ is distributed as the quantity

$$p\sigma^2[1+(p-1)\rho]\chi_{M-1}^2.$$

The proof of Proposition 2.4 is complete.

The following proposition can be used to show independence of $B1-\frac{1}{p}B2$ and $B2$ required when finding the exact null distribution of the likelihood ratio test statistic W stated in Theorem 2.1.

***Proposition 2.5*:** Let $\boldsymbol{Y}_i=\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\mu}}=\left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i\right)^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{X}_i$, $\boldsymbol{A}=\boldsymbol{I}-\frac{1}{p}\boldsymbol{J}_p$, $\boldsymbol{B}=\frac{1}{p}\boldsymbol{J}_p$, $S_{\boldsymbol{A}}=\sum_{i=1}^{M}\boldsymbol{Y}_i^{\,T}\boldsymbol{A}\boldsymbol{Y}_i$ and $S_{\boldsymbol{B}}=\sum_{i=1}^{M}\boldsymbol{Y}_i^{\,T}\boldsymbol{B}\boldsymbol{Y}_i$. Then $\boldsymbol{Y}_i^{\,T}A\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j^{\,T}\boldsymbol{B}\boldsymbol{Y}_j$ are independent for all *i* and *j* and hence $S_{\boldsymbol{A}}$ and $S_{\boldsymbol{B}}$ are independent.

Remark 2.1: Since $\boldsymbol{Y}_i$ is a linear combination of $\boldsymbol{X}=vec(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_M)$ it can be expressed as $\boldsymbol{Y}_i=\boldsymbol{M}_i\boldsymbol{X}$, where $\boldsymbol{M}_i$ is a $p\times Mp$ matrix with the structure

$$\boldsymbol{M}_i=\begin{pmatrix}-\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_1^{\,T} & \cdots & -\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_{i-1}^{T} & \boldsymbol{I}_p-\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_i^{\,T} & -\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_{i+1}^{T} & \cdots & -\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_M^{\,T}\end{pmatrix},$$

where $\boldsymbol{Q}=\left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i\right)$.

Rewrite both $S_{\boldsymbol{A}}$ and $S_{\boldsymbol{B}}$ we have

$$S_{\boldsymbol{A}} = \sum_{i=1}^{M} (\boldsymbol{M}_i \boldsymbol{X})^T \boldsymbol{A} (\boldsymbol{M}_i \boldsymbol{X}) \quad \text{and} \quad S_{\boldsymbol{B}} = \sum_{i=1}^{M} (\boldsymbol{M}_i \boldsymbol{X})^T \boldsymbol{B} (\boldsymbol{M}_i \boldsymbol{X}),$$

$\boldsymbol{X}$ is distributed as a multivariate normal $MVN_{Mp}(vec(\boldsymbol{C}_1\boldsymbol{\mu},\ldots,\boldsymbol{C}_M\boldsymbol{\mu}), I_M \otimes \boldsymbol{\Sigma})$. And we have

$$S_{\boldsymbol{A}} = \boldsymbol{X}^T \left(\sum_{i=1}^{M} \boldsymbol{M}_i^T \boldsymbol{A} \boldsymbol{M}_i \right) \boldsymbol{X} \quad \text{and} \quad S_{\boldsymbol{B}} = \boldsymbol{X}^T \left(\sum_{i=1}^{M} \boldsymbol{M}_i^T \boldsymbol{B} \boldsymbol{M}_i \right) \boldsymbol{X} .$$

To show that $S_{\boldsymbol{A}}$ and $S_{\boldsymbol{B}}$ are independent, it suffices to show that

$$\left(\sum_{i=1}^{M} \boldsymbol{M}_i^T \boldsymbol{A} \boldsymbol{M}_i \right) (I_M \otimes \boldsymbol{\Sigma}) \left(\sum_{i=1}^{M} \boldsymbol{M}_i^T \boldsymbol{B} \boldsymbol{M}_i \right) = \boldsymbol{0} , \tag{2.19}$$

where $\boldsymbol{\Sigma}$ has compound symmetry with the structure $\boldsymbol{\Sigma} = \sigma^2[(1-\rho)\boldsymbol{I}_p + \rho \boldsymbol{J}_p]$ and is circulant. The calculation of the matrix $\left(\sum_{i=1}^{M} \boldsymbol{M}_i^T \boldsymbol{A} \boldsymbol{M}_i \right) \boldsymbol{\Sigma} \left(\sum_{i=1}^{M} \boldsymbol{M}_i^T \boldsymbol{B} \boldsymbol{M}_i \right)$ is complicated so another way to prove Proposition 2.5 is to show first that $\boldsymbol{Y}_i^T A \boldsymbol{Y}_i$ and $\boldsymbol{Y}_j^T \boldsymbol{B} \boldsymbol{Y}_j$ are independent for all *i* and *j*.

Proof of Proposition 2.5:

Because both $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric and idempotent, we may rewrite $\boldsymbol{Y}_i^T \boldsymbol{A} \boldsymbol{Y}_i$ and $\boldsymbol{Y}_j^T \boldsymbol{B} \boldsymbol{Y}_j$ respectively by

$$\boldsymbol{Y}_i^T \boldsymbol{A} \boldsymbol{Y}_i = \boldsymbol{Y}_i^T \boldsymbol{A} \boldsymbol{A} \boldsymbol{Y} = (\boldsymbol{A} \boldsymbol{Y}_i)^T \boldsymbol{A} \boldsymbol{Y}_i ,$$

and

$$\boldsymbol{Y}_i^T \boldsymbol{B} \boldsymbol{Y}_i = \boldsymbol{Y}_i^T \boldsymbol{B} \boldsymbol{B} \boldsymbol{Y} = (\boldsymbol{B} \boldsymbol{Y}_i)^T \boldsymbol{B} \boldsymbol{Y}_i .$$

Note that $\boldsymbol{Y}_i^T \boldsymbol{A} \boldsymbol{Y}_i$ and $\boldsymbol{Y}_j^T \boldsymbol{B} \boldsymbol{Y}_j$ are squared lengths of $\boldsymbol{A} \boldsymbol{Y}_i$ and $\boldsymbol{B} \boldsymbol{Y}_j$, respectively. So we only have to show that $\boldsymbol{A} \boldsymbol{Y}_i$ and $\boldsymbol{B} \boldsymbol{Y}_j$ are independent.

Consider the distribution of the random vector

$$\begin{pmatrix} \boldsymbol{AY}_i \\ \boldsymbol{BY}_j \end{pmatrix} = \begin{pmatrix} \boldsymbol{AM}_i\boldsymbol{X} \\ \boldsymbol{BM}_j\boldsymbol{X} \end{pmatrix} = \begin{pmatrix} \boldsymbol{AM}_i \\ \boldsymbol{BM}_j \end{pmatrix}\boldsymbol{X} ,$$

where $\boldsymbol{M}_i$ is defined in Remark 2.1. $\begin{pmatrix} \boldsymbol{AY}_i \\ \boldsymbol{BY}_j \end{pmatrix}$ is a linear combination of $\boldsymbol{X}$ which is normal so

$\begin{pmatrix} \boldsymbol{AY}_i \\ \boldsymbol{BY}_j \end{pmatrix}$ is normal. Thus showing $Cov(\boldsymbol{AY}_i, \boldsymbol{BY}_j) = \boldsymbol{0}$ implies that $\boldsymbol{AY}_i$ and $\boldsymbol{BY}_j$ are independent normal random vectors then the proof is done. Since $\boldsymbol{B}$ is symmetric, we have

$$Cov(\boldsymbol{AY}_i, \boldsymbol{BY}_j) = \boldsymbol{A}Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j)\boldsymbol{B}^T = \boldsymbol{A}Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j)\boldsymbol{B} .$$

Thus it suffices to show that $Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$ is a circulant matrix so $\boldsymbol{A}$, $\boldsymbol{B}$, and $Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$ commute implying that $\boldsymbol{A}Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j)\boldsymbol{B} = Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j)\boldsymbol{AB} = \boldsymbol{0}$, using the fact that $\boldsymbol{AB} = \boldsymbol{0}$.

To show that $Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$ is a circulant matrix, we may use a direct proof. We have

$$\begin{aligned}
Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j) &= Cov(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}, \boldsymbol{X}_j - \boldsymbol{C}_j\hat{\boldsymbol{\mu}}) \\
&= Cov(\boldsymbol{X}_i, \boldsymbol{X}_j) - Cov(\boldsymbol{X}_i, \boldsymbol{C}_j\hat{\boldsymbol{\mu}}) - Cov(\boldsymbol{C}_i\hat{\boldsymbol{\mu}}, \boldsymbol{X}_j) + Cov(\boldsymbol{C}_i\hat{\boldsymbol{\mu}}, \boldsymbol{C}_j\hat{\boldsymbol{\mu}}) \\
&= Cov(\boldsymbol{X}_i, \boldsymbol{X}_j) - Cov(\boldsymbol{X}_i, \hat{\boldsymbol{\mu}})\boldsymbol{C}_j^T - \boldsymbol{C}_i Cov(\hat{\boldsymbol{\mu}}, \boldsymbol{X}_j) + \boldsymbol{C}_i Cov(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}})\boldsymbol{C}_j^T \\
&= Cov(\boldsymbol{X}_i, \boldsymbol{X}_j) - Cov(\boldsymbol{X}_i, \boldsymbol{Q}^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{X}_i)\boldsymbol{C}_j^T - \boldsymbol{C}_i Cov(\boldsymbol{Q}^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{X}_i, \boldsymbol{X}_j) + \boldsymbol{C}_i Var(\hat{\boldsymbol{\mu}})\boldsymbol{C}_j^T \\
&= Cov(\boldsymbol{X}_i, \boldsymbol{X}_j) - Var(\boldsymbol{X}_i)(\boldsymbol{Q}^{-1}\boldsymbol{C}_i^{\,T})^T\boldsymbol{C}_j^T - \boldsymbol{C}_i(\boldsymbol{Q}^{-1}\boldsymbol{C}_j^{\,T})Var(\boldsymbol{X}_j) + \boldsymbol{C}_i Var(\hat{\boldsymbol{\mu}})\boldsymbol{C}_j^T \\
&= Cov(\boldsymbol{X}_i, \boldsymbol{X}_j) - Var(\boldsymbol{X}_i)\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j^T - \boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j^{\,T}Var(\boldsymbol{X}_j) + \boldsymbol{C}_i Var(\hat{\boldsymbol{\mu}})\boldsymbol{C}_j^T ,
\end{aligned}$$

where $\boldsymbol{Q} = \sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i$. Note that $Cov(\boldsymbol{X}_i, \boldsymbol{X}_j) = \boldsymbol{\Sigma}$ if $i = j$ and $\boldsymbol{0}$ otherwise. Also from Section 2.3.1 we have $Var(\hat{\boldsymbol{\mu}}) = \boldsymbol{Q}^{-1}\boldsymbol{\Sigma}$ and the fact that $\boldsymbol{\Sigma}$, $\boldsymbol{C}_i$, and $\boldsymbol{Q}$ are circulant matrices implying that their inverse and transpose are also circulant so the commutability holds. Hence $Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$ becomes

$$Cov(\boldsymbol{Y}_i,\boldsymbol{Y}_j)=\begin{cases}\boldsymbol{\Sigma}-\boldsymbol{\Sigma C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_i^T-\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_i^{\,T}\boldsymbol{\Sigma}+\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{\Sigma C}_i^{\,T}, \text{if } i=j\\ \boldsymbol{0}-\boldsymbol{\Sigma C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j^T-\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j^{\,T}\boldsymbol{\Sigma}+\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{\Sigma C}_j^{\,T}, \text{if } i\neq j\end{cases}$$

$$=\begin{cases}(I-\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_i^T)\boldsymbol{\Sigma}, \text{if } i=j\\ -(\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j^{\,T})\boldsymbol{\Sigma}, \text{if } i\neq j.\end{cases}$$

Therefore $Cov(\boldsymbol{Y}_i,\boldsymbol{Y}_j)$ is circulant. The proof of Proposition 2.5 is complete.

Now, it is time to state and prove the following main results using Propositions 2.1 – 2.5.

***Theorem 2.2*:** The likelihood ratio test for testing $H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0$ in Theorem 2.1 is to reject H_0 if $W<C_\alpha$, where C_α is such that $P(W<C_\alpha\mid H_0)=\alpha$, and W is expressed as

$$W=\frac{(pB1-B2)^{p-1}B2}{(pB1_0-B2_0)^{p-1}B2_0}=\frac{\left(B1-\frac{1}{p}B2\right)^{p-1}B2}{\left(B1_0-\frac{1}{p}B2_0\right)^{p-1}B2_0}=\frac{B^{p-1}D}{A^{p-1}C},$$

where

$$B1=\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}) \text{ and } B2=\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}),$$

$$B1_0=\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0) \text{ and } B2_0=\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0),$$

$$A=B1_0-\frac{1}{p}B2_0,\ B=B1-\frac{1}{p}B2,\ C=B2_0,\ D=B2.$$

Furthermore, under $H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0$, W is distributed as the random variable

$$\frac{1}{\left(1+\frac{1}{M-1}F^{*}\right)^{p-1}\left(1+\frac{1}{M-1}F^{**}\right)},$$

where F^* and F^{**} are independent and distributed as $F_{p-1,(M-1)(p-1)}$, and $F_{1,M-1}$ random variables, respectively.

Proof:

Recall from the proofs of propositions 2.1 - 2.4 that

$$A = B + R, \tag{2.20}$$

where

$$R = (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^T \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \right) (\boldsymbol{I}_p - \frac{1}{p} \boldsymbol{J}_p)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0) .$$

Also

$$C = D + S, \tag{2.21}$$

where

$$S = (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^T \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \right) \boldsymbol{J}_p (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0) .$$

If we can show that *B*, *R*, *D*, and *S* are mutually independent, combined with the following facts (7), (8), (9), and (10), then the proof is done. Note that Facts (1) - (6) for showing pairwise independence among *B*, *R*, *D*, and *S* are sufficient for showing mutual independence among them. The facts needed to prove this theorem are shown below:

(1) *B* and *R* are independent,

(2) *D* and *S* are independent,

(3) *B* and *D* are independent (Proposition 2.5),

(4) *B* and *S* are independent,

(5) *R* and *D* are independent,

(6) *R* and *S* are independent,

(7) $B = B1 - \frac{1}{p}B2 = \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) \overset{d}{=} \sigma^2(1-\rho)\chi^2((M-1)(p-1))$,

(8) $R = (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0) \overset{d}{=} \sigma^2(1-\rho)\chi^2(p-1)$,

(9) $D = B2 = \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2(M-1)$, and

(10) $S = (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)\boldsymbol{J}_p(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0) \sim p\sigma^2[1+(p-1)\rho]\chi^2(1)$.

First, Facts (1), (2), (4), and (5) hold due to the facts that $\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\mu}}$ are independent for each *i*. Fact (6) is true because $(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)\boldsymbol{J}_p = \boldsymbol{0}$. Fact (3) is the result of Proposition 2.5. Facts (7) and (9) are direct results of Propositions 2.3 and 2.4, respectively. Facts (8) and (10) are shown in the proofs of Proposition 2.3 and 2.4, respectively. Hence the result that R, S, B, and D are independent in connection with the expressing of W

$$W = \frac{B^{p-1}D}{(B+R)^{p-1}(D+S)} = \frac{1}{\left(1+\frac{R}{B}\right)^{p-1}\left(1+\frac{S}{D}\right)}$$

fulfills the proof of Theorem 2.2.

2.2.3 Properties and Useful Results about ML Estimators

In addition to the likelihood ratio test for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, the null distribution of the statistic

$$(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)'[\widehat{Var}(\hat{\boldsymbol{\mu}})]^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0) \tag{2.22}$$

also draws our attention. The exact null distribution of (2.22) is not easy to obtain while we may at least find its asymptotic distribution. First note that

$$Var(\hat{\boldsymbol{\mu}}) = (\sum_{i=1}^{M} \boldsymbol{C}_i{}'\boldsymbol{C}_i)^{-1}\hat{\boldsymbol{\Sigma}} ,$$

and

$$\hat{\boldsymbol{\Sigma}}^{-1} = \frac{1}{\hat{\sigma}^2(1-\hat{\rho})}\left[\boldsymbol{I}_p - \frac{\hat{\rho}}{1+(p-1)\hat{\rho}}\boldsymbol{J}_p\right].$$

The quadratic form (2.22) can be phrased as:

$$\frac{(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)^T\left[\sum_{i=1}^{M}\boldsymbol{C}_i{}^T\hat{\boldsymbol{R}}^{-1}\boldsymbol{C}_i\right](\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)\Big/p}{M\hat{\sigma}^2\big/(M-1)}, \tag{2.23}$$

where

$$\hat{\boldsymbol{R}}^{-1} = \frac{1}{(1-\hat{\rho})}\left[\boldsymbol{I}_p - \frac{\hat{\rho}}{1+(p-1)\hat{\rho}}\boldsymbol{J}_p\right].$$

The following propositions are helpful for developing an approximate distribution of the statistic in (2.22) under the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$. Details of the derivation of the approximate null distribution of (2.22) will be shown in Subsection 2.2.4. Before deriving the approximate null distribution of the statistic in (2.22), let us first look at the following proposition about the MLE of σ^2.

<u>Proposition 2.6</u>**:** Let $B1 = \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})$. Then $\hat{\sigma}^2 = \frac{1}{Mp}B_1$ is the MLE of σ^2, and $\hat{\hat{\sigma}}^2 = \frac{M}{M-1}\hat{\sigma}^2$ is an unbiased estimator for σ^2. In addition the following results hold.

a) $E(\hat{\sigma}^2) = E\frac{1}{Mp}B_1 = (1-\frac{1}{M})\sigma^2$, and $E(\hat{\hat{\sigma}}^2) = E\frac{1}{(M-1)p}B_1 = \sigma^2$.

b) $V(\hat{\sigma}^2) = \frac{2(M-1)}{M^2 p}[1+(p-1)\rho^2]\sigma^4 = O(\frac{1}{M})$, and

$$V(\hat{\hat{\sigma}}^2) = \frac{2}{(M-1)p}[1+(p-1)\rho^2]\sigma^4 = O(\frac{1}{M}).$$

c) Both $\hat{\sigma}^2$ and $\hat{\hat{\sigma}}^2$ are consistent estimators of σ^2; that is, $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ and $\hat{\hat{\sigma}}^2 \xrightarrow{p} \sigma^2$.

Proof:

Recall that the MLE of $\boldsymbol{\mu}$ is

$$\hat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \right)^{-1} \sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{X}_i .$$

To find $E(B1)$, $E(\hat{\sigma}^2)$, $E(B1^2)$, and $V(\hat{\sigma}^2)$, recall that $B1$ can be expressed as

$$B1 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}) - (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'(\sum_{i=1}^{M} \boldsymbol{C}_i'\boldsymbol{C}_i)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}),$$

and the following result (cf. Christenson (2002), Theorem 1.3.2) is needed.

$$\text{If } E(\boldsymbol{Y}) = \boldsymbol{\mu} \text{ and } Cov(\boldsymbol{Y}) = \boldsymbol{V} \text{ then } E(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}) = tr(\boldsymbol{A}\boldsymbol{V}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}.$$

So we have

$$\begin{aligned} E(B1) &= \sum_{i=1}^{M} E(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}) - E(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'(\sum_{i=1}^{M} \boldsymbol{C}_i'\boldsymbol{C}_i)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \\ &= \sum_{i=1}^{M} [tr(\boldsymbol{\Sigma}) + 0] - tr\left[(\sum_{i=1}^{M} \boldsymbol{C}_i'\boldsymbol{C}_i)(\sum_{i=1}^{M} \boldsymbol{C}_i'\boldsymbol{C}_i)^{-1}\boldsymbol{\Sigma} + 0 \right] \\ &= Mtr(\boldsymbol{\Sigma}) - tr(\boldsymbol{\Sigma}) = (M-1)p\sigma^2, \end{aligned}$$

which implies that $E(\hat{\sigma}^2) = E\frac{1}{Mp}B1 = (1 - \frac{1}{M})\sigma^2$. Next,

$$\begin{aligned} E(B1^2) &= E\left[\sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}) \right]^2 \\ &\quad - 2E\left[\sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}) \right]\left[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'(\sum_{i=1}^{M} \boldsymbol{C}_i'\boldsymbol{C}_i)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \right] \\ &\quad + E\left[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'(\sum_{i=1}^{M} \boldsymbol{C}_i'\boldsymbol{C}_i)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \right]^2 = A - 2B + C, \end{aligned}$$

where A, B, and C are, respectively, given by

$$
\begin{aligned}
A &= E\left[\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})\right]^2 \\
&= \sum_{i=1}^{M} E\left[(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})\right]^2 \quad \text{(Neudecker \& Magnus (1979), Theorem 4.2)} \\
&\quad + 2\sum_{i<j} E(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}) \cdot E(\boldsymbol{X}_j - \boldsymbol{C}_j\boldsymbol{\mu})^T(\boldsymbol{X}_j - \boldsymbol{C}_j\boldsymbol{\mu}) \\
&= M\left[(tr\boldsymbol{\Sigma})^2 + 2tr(\boldsymbol{\Sigma}^2)\right] + (M^2 - M)(tr\boldsymbol{\Sigma})^2 = M^2(tr\boldsymbol{\Sigma})^2 + 2Mtr(\boldsymbol{\Sigma}^2),
\end{aligned}
$$

$$
B = E\left[\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})\right]\left[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'(\sum_{i=1}^{M}\boldsymbol{C}_i'\boldsymbol{C}_i)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\right],
$$

and

$$
\begin{aligned}
C &= E\left[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'(\sum_{i=1}^{M}\boldsymbol{C}_i'\boldsymbol{C}_i)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\right]^2 \\
&= (tr\boldsymbol{\Sigma})^2 + 2tr(\boldsymbol{\Sigma}^2).
\end{aligned}
$$

Let us attend to the representation of B. Define $\boldsymbol{Q} = \sum_{i=1}^{M}\boldsymbol{C}_i'\boldsymbol{C}_i$, we have $\hat{\boldsymbol{\mu}} = \boldsymbol{Q}^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{X}_i$. Thus the quadratic form $(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})$ can be rewritten as:

$$
\begin{aligned}
(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) &= (\boldsymbol{Q}^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{X}_i - \boldsymbol{Q}^{-1}\boldsymbol{Q}\boldsymbol{\mu})^T\boldsymbol{Q}(\boldsymbol{Q}^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{X}_i - \boldsymbol{Q}^{-1}\boldsymbol{Q}\boldsymbol{\mu}) \\
&= (\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{X}_i - \boldsymbol{Q}\boldsymbol{\mu})^T\boldsymbol{Q}^{-1}(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{X}_i - \boldsymbol{Q}\boldsymbol{\mu}) \quad (\boldsymbol{Q}\text{ is symmetric, circulant}) \\
&= (\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{X}_i - \sum_{i=1}^{M}\boldsymbol{C}_i^T C_i\boldsymbol{\mu})^T\boldsymbol{Q}^{-1}(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{X}_i - \sum_{i=1}^{M}\boldsymbol{C}_i^T C_i\boldsymbol{\mu}) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{M}(\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j^T(\boldsymbol{X}_j - C_j\boldsymbol{\mu}).
\end{aligned}
$$

So B can be expressed as

$$B = E\left[\sum_{k=1}^{M}(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})\right]\cdot\left[\sum_{i=1}^{M}\sum_{j=1}^{M}(\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j{}^T(\boldsymbol{X}_j - C_j\boldsymbol{\mu})\right]$$
$$= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}E\left[(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})\right]\cdot\left[(\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j{}^T(\boldsymbol{X}_j - C_j\boldsymbol{\mu})\right]. \quad (2.24)$$

Consider the term in (2.24):

$$E\left[(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})\right]\cdot\left[(\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j{}^T(\boldsymbol{X}_j - C_j\boldsymbol{\mu})\right]. \quad (2.25)$$

To calculate (2.25), the results of Magnus (1979) can be applied to the following two cases.

Case 1: $i = j$,

- $i = j = k$

$$E\left[(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})\right]\cdot\left[(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})^T\boldsymbol{C}_k\boldsymbol{Q}^{-1}\boldsymbol{C}_k{}^T(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})\right]$$
$$= tr\boldsymbol{\Sigma}\cdot tr\left((\boldsymbol{C}_k\boldsymbol{Q}^{-1}\boldsymbol{C}_k{}^T)\boldsymbol{\Sigma}\right) + 2\cdot tr\left(\boldsymbol{\Sigma}(\boldsymbol{C}_k\boldsymbol{Q}^{-1}\boldsymbol{C}_k{}^T)\boldsymbol{\Sigma}\right) \quad (2.26)$$

- For *i*, *j* such that $i = j \neq k$

$$E\left[(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})\right]\cdot\left[(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_i{}^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})\right]$$
$$= E\left[(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})\right]\cdot E\left[(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_i{}^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})\right] \quad (2.27)$$
$$= tr\boldsymbol{\Sigma}\cdot tr\left((\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_i{}^T)\boldsymbol{\Sigma}\right)$$

Case 2: $i \neq j$,

In this case, only one of *i* and *j* equal to *k*, or neither of them equal to *k*. For these two scenarios, (2.25) is equal to zero. That is,

$$E\left[(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - \boldsymbol{C}_k\boldsymbol{\mu})\right]\cdot\left[(\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j{}^T(\boldsymbol{X}_j - C_j\boldsymbol{\mu})\right] = 0\,. \quad (2.28)$$

Thus (2.24) becomes

$$B = \left[(tr\boldsymbol{\Sigma})^2 + 2\cdot tr(\boldsymbol{\Sigma}^2)\right] + \left[(M-1)\cdot(tr\boldsymbol{\Sigma})^2\right] = M\left(tr\boldsymbol{\Sigma}\right)^2 + 2\cdot tr(\boldsymbol{\Sigma}^2) \quad (2.29)$$

So we have

$$
\begin{aligned}
E(B1^2) &= A-2B+C \\
&= \left[M^2(tr\boldsymbol{\Sigma})^2 + 2M\cdot tr(\boldsymbol{\Sigma}^2)\right] - 2\left[M(tr\boldsymbol{\Sigma})^2 + 2\cdot tr(\boldsymbol{\Sigma}^2)\right] + \left[(tr\boldsymbol{\Sigma})^2 + 2tr(\boldsymbol{\Sigma}^2)\right] \\
&= (M-1)^2(tr\boldsymbol{\Sigma})^2 + 2(M-1)tr(\boldsymbol{\Sigma}^2) \\
&= (M-1)^2[p\sigma^2]^2 + 2(M-1)p[1+(p-1)\rho^2]\sigma^4,
\end{aligned}
$$

and

$$
\begin{aligned}
V(B1) &= E(B1^2) - E^2(B1) \\
&= \left[(M-1)^2(tr\boldsymbol{\Sigma})^2 + 2(M-1)tr(\boldsymbol{\Sigma}^2)\right] - (M-1)^2(tr\boldsymbol{\Sigma})^2 \\
&= 2(M-1)tr(\boldsymbol{\Sigma}^2) \\
&= 2(M-1)p[1+(p-1)\rho^2]\sigma^4.
\end{aligned}
$$

Hence we have

$$
E[(\hat{\sigma}^2)^2] = E\left(\frac{B1^2}{M^2p^2}\right) = (1-\frac{1}{M})^2\sigma^4 + 2\frac{M-1}{M^2p}[1+(p-1)\rho^2]\sigma^4,
$$

yielding

$$
\begin{aligned}
V(\hat{\sigma}^2) &= E[(\hat{\sigma}^2)^2] - \left[E[\hat{\sigma}^2]\right]^2 \\
&= (1-\frac{1}{M})^2\sigma^4 + 2\frac{M-1}{M^2p}[1+(p-1)\rho^2]\sigma^4 - (1-\frac{1}{M})^2\sigma^4 \\
&= \frac{2(M-1)}{M^2p}[1+(p-1)\rho^2]\sigma^4 = O(\frac{1}{M}).
\end{aligned}
$$

Therefore we have that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$. The proof is complete.

Remark 2.2: Proposition 2.6 (a) and (b) can be shown more effortlessly by using the results about the distribution of $B1$ which will be stated in Theorem 2.3 later in this subsection. Theorem 2.3 (a) states that $B1$ is distributed as the quantity $\sigma^2(1-\rho)\chi^2_{(M-1)(p-1)} + \sigma^2[1+(p-1)\rho]\chi^2_{(M-1)}$, where $\chi^2_{(M-1)(p-1)}$ and $\chi^2_{(M-1)}$ are independent chi-squared random variables with $(M-1)(p-1)$ and $(M-1)$ degrees of freedom, respectively. Hence the results

$$
\begin{aligned}
E(B_1) &= E[\sigma^2(1-\rho)\chi^2_{(M-1)(p-1)} + \sigma^2[1+(p-1)\rho]\chi^2_{(M-1)}] \\
&= (M-1)(p-1)(1-\rho)\sigma^2 + (M-1)[1+(p-1)\rho]\sigma^2 \\
&= (M-1)p\sigma^2,
\end{aligned}
$$

and

$$
\begin{aligned}
Var(B_1) &= Var[\sigma^2(1-\rho)\chi^2_{(M-1)(p-1)} + \sigma^2[1+(p-1)\rho]\chi^2_{(M-1)}] \\
&= 2(M-1)(p-1)(1-\rho)^2\sigma^4 + 2(M-1)[1+(p-1)\rho]^2\sigma^4 \\
&= 2(M-1)\sigma^4\{(p-1)(1-\rho)^2 + [1+(p-1)\rho]^2\} \\
&= 2(M-1)p[1+(p-1)\rho^2]\sigma^4
\end{aligned}
$$

obtained from Theorem 2.3 (a) are exactly the results of Proposition 2.6 (a) and (b), respectively.

The following proposition is helpful to prove Theorem 2.3 (a).

***Proposition 2.7*:**

(a) If $A_i \overset{iid}{\sim} \sum_{j=1}^{p} \lambda_j X_{ij}$, $i = 1,\dots,M$, where X_{ij} are independent χ^2 random variables with 1 degree of freedom. Then $\sum_{i=1}^{M} A_i \overset{d}{=} \sum_{j=1}^{p} \lambda_j \chi^2_M$.

(b) If $A \overset{d}{=} \sigma^2(1-\rho)\chi^2_{M(p-1)} + \sigma^2[1+(p-1)\rho]\chi^2_M$, where $\chi^2_{M(p-1)}$ and χ^2_M are independent, $C \overset{d}{=} \sigma^2(1-\rho)\chi^2_{p-1} + \sigma^2[1+(p-1)\rho]\chi^2_1$, where χ^2_{p-1} and χ^2_1 are independent, $A = B + C$, where B and C are independent, then B is distributed as the quantity

$$
\sigma^2(1-\rho)\chi^2_{(M-1)(p-1)} + \sigma^2[1+(p-1)\rho]\chi^2_{M-1} .
$$

Proof:

(a) Let Y_j be independent chi-squared random variables with *M* degrees of freedom. The moment generating function of $\sum_{i=1}^{M} A_i$ is

$$M_{\sum_{i=1}^{M} A_i}(t) = \prod_{i=1}^{M} M_{A_i}(t) = \prod_{i=1}^{M} E(e^{\sum_{j=1}^{p} \lambda_j X_{ij}}) = \left[E(e^{\sum_{j=1}^{p} t\lambda_j X_{ij}}) \right]^M$$
$$= \left[\prod_{j=1}^{p} M_{X_{ij}}(t\lambda_j) \right]^M = \left[\prod_{j=1}^{p} (1-2t\lambda_j)^{-1/2} \right]^M$$
$$= \prod_{j=1}^{p} \left[(1-2t\lambda_j)^{-M/2} \right] = \prod_{j=1}^{p} M_{Y_j}(t\lambda_j) = \prod_{j=1}^{p} E(e^{t\lambda_j Y_j}) = E(e^{\sum_{j=1}^{p} t\lambda_j Y_j}) = M_{\sum_{j=1}^{p} \lambda_j Y_j}(t),$$

which is the moment generating function of the random variable $\sum_{j=1}^{p} \lambda_j Y_j$.

(b) Since B and C are independent, we have the moment generating of A which can be expressed as $M_A(t) = M_{B+C}(t) = M_B(t) M_C(t)$. The mgf of A is

$$M_A(t) = M_{\sigma^2(1-\rho)\chi^2_{M(p-1)} + \sigma^2[1+(p-1)\rho]\chi^2_M}(t) = M_{\sigma^2(1-\rho)\chi^2_{M(p-1)}}(t) M_{\sigma^2[1+(p-1)\rho]\chi^2_M}(t)$$
$$= M_{\chi^2_{M(p-1)}}(t\sigma^2(1-\rho)) \cdot M_{\chi^2_M}(t\sigma^2[1+(p-1)\rho])$$
$$= \left(1-2t\sigma^2(1-\rho)\right)^{-M(p-1)/2} \cdot \left(1-2t\sigma^2[1+(p-1)\rho]\right)^{-M/2}.$$

The mgf of C is

$$M_C(t) = M_{\sigma^2(1-\rho)\chi^2_{p-1} + \sigma^2[1+(p-1)\rho]\chi^2_1}(t) = M_{\sigma^2(1-\rho)\chi^2_{p-1}}(t) M_{\sigma^2[1+(p-1)\rho]\chi^2_1}(t)$$
$$= M_{\chi^2_{p-1}}(t\sigma^2(1-\rho)) \cdot M_{\chi^2_1}(t\sigma^2[1+(p-1)\rho])$$
$$= \left(1-2t\sigma^2(1-\rho)\right)^{-(p-1)/2} \cdot \left(1-2t\sigma^2[1+(p-1)\rho]\right)^{-1/2}.$$

Thus the mgf of B is

$$M_B(t) = \frac{M_A(t)}{M_C(t)} = \frac{\left(1-2t\sigma^2(1-\rho)\right)^{-M(p-1)/2} \cdot \left(1-2t\sigma^2[1+(p-1)\rho]\right)^{-M/2}}{\left(1-2t\sigma^2(1-\rho)\right)^{-(p-1)/2} \cdot \left(1-2t\sigma^2[1+(p-1)\rho]\right)^{-1/2}}$$
$$= \left(1-2t\sigma^2(1-\rho)\right)^{-(M-1)(p-1)/2} \cdot \left(1-2t\sigma^2[1+(p-1)\rho]\right)^{-(M-1)/2}$$

which is the mgf of $\sigma^2(1-\rho)\chi^2_{(M-1)(p-1)} + \sigma^2[1+(p-1)\rho]\chi^2_{M-1}$ random variable, where $\chi^2_{(M-1)(p-1)}$ and χ^2_{M-1} are independent chi-squared random variables with $(M-1)(p-1)$ and $M-1$ degrees of freedom, respectively. The proof is complete.

Proposition 2.7 will be used to prove the following theorem.

<u>Theorem 2.3</u>:

(a) $B1 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})^T (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})$ is distributed as the quantity

$$\sigma^2 (1-\rho) \chi^2_{(M-1)(p-1)} + \sigma^2 [1 + (p-1)\rho] \chi^2_{M-1},$$

where $\chi^2_{(M-1)(p-1)}$ and χ^2_{M-1} are independent chi-squared random variables with $(M-1)(p-1)$ and $(M-1)$ degrees of freedom, respectively.

(b) $B1$ has an approximate $\sigma^2 g \chi^2_{(M-1)h}$ distribution, where

(1) $g = \dfrac{(p-1)(1-\rho)^2 + [1+(p-1)\rho]^2}{p}$, and

(2) $h = \dfrac{p^2}{(p-1)(1-\rho)^2 + [1+(p-1)\rho]^2}$.

Proof:

(a) Recall that $B1$ can be expressed as

$$B1 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu})^T (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}) - (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' (\sum_{i=1}^{M} \boldsymbol{C}_i ' \boldsymbol{C}_i)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}). \tag{2.30}$$

The first term of the last expression in (2.30) has the same distribution as that of sum of M independent random variables $\sum_{j=1}^{p} \lambda_j \chi^2_1$, where χ^2_1 are independent chi-square random variables with 1 degree of freedom. λ_j 's are eigenvalues of $\boldsymbol{\Sigma} = \sigma^2 [(1-\rho) I_p + \rho J_p]$. The eigenvalues of $\boldsymbol{\Sigma}$ are $\sigma^2 (1-\rho)$ with multiplicity $p-1$ and $\sigma^2 [1+(p-1)\rho]$ with multiplicity 1. Thus $\sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu})^T (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu})$ is distributed as M independent random variables each of which is $\sigma^2 (1-\rho) \chi^2_{p-1} + \sigma^2 [1+(p-1)\rho] \chi^2_1$; that is,

$$\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu})\overset{d}{=}\sigma^2(1-\rho)\chi^2_{M(p-1)}+\sigma^2[1+(p-1)\rho]\chi^2_M .$$

Similarly for the second term of the last expression in (2.30), $(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})'(\sum_{i=1}^{M}\boldsymbol{C}_i'\boldsymbol{C}_i)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})$ is distributed like the quantity $\sigma^2(1-\rho)\chi^2_{p-1}+\sigma^2[1+(p-1)\rho]\chi^2_1$ because $\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}$ is distributed as $N(\boldsymbol{0},(\sum_{i=1}^{M}\boldsymbol{C}_i'\boldsymbol{C}_i)^{-1}\boldsymbol{\Sigma})$, implying that $(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})'(\sum_{i=1}^{M}\boldsymbol{C}_i'\boldsymbol{C}_i)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})$ is distributed as the quantity $\sum_{j=1}^{p}\lambda_j\chi^2_1$,where λ_j 's are eigenvalues of $(\sum_{i=1}^{M}\boldsymbol{C}_i'\boldsymbol{C}_i)(\sum_{i=1}^{M}\boldsymbol{C}_i'\boldsymbol{C}_i)^{-1}\boldsymbol{\Sigma}=\boldsymbol{\Sigma}$. Since $B1$ and $(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})'(\sum_{i=1}^{M}\boldsymbol{C}_i'\boldsymbol{C}_i)(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})$ are independent, based on Proposition 2.7 $B1$ is distributed as the quantity

$$\sigma^2(1-\rho)\chi^2_{(M-1)(p-1)}+\sigma^2[1+(p-1)\rho]\chi^2_{M-1},$$

where $\chi^2_{(M-1)(p-1)}$ and χ^2_{M-1} are independent chi-squared random variables with $(M-1)(p-1)$ and $(M-1)$ degrees of freedom, respectively. The proof of part (a) is complete.

(b) Box (1954, Theorem 3.1) showed that $g=\sum_{j=1}^{p}\lambda_j{}^2/\sum_{i=1}^{p}\lambda_i$ and $h=\left(\sum_{j=1}^{p}\lambda_j\right)^2/\sum_{i=1}^{p}\lambda_i{}^2$ are chosen so that the distribution of $\sum_{j=1}^{p}\lambda_j\chi^2_1$ has the same first two moments as $\sigma^2 g\chi^2_h$. Since $B1$ is distributed like sum of $M-1$ independent random variables $\sum_{j=1}^{p}\lambda_j\chi^2_1$, $B1$ has an approximate $\sigma^2 g\chi^2_{(M-1)h}$ distribution. The proof of Theorem 2.3 (b) is complete.

Corollary 2.1: The test statistic for testing $H_0:\sigma^2=\sigma_0{}^2$ is

$$Mp\hat{\sigma}^2 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}).$$

Under H_0, $\frac{Mp\hat{\sigma}^2}{\sigma_0^2} \dot{\sim} \hat{g}\chi^2_{(M-1)\hat{h}}$, where $\hat{g} = \frac{(p-1)(1-\hat{\rho})^2 + [1+(p-1)\hat{\rho}]^2}{p}$, and

$$\hat{h} = \frac{p^2}{(p-1)(1-\hat{\rho})^2 + [1+(p-1)\hat{\rho}]^2}.$$

Proof:

It follows directly from Theorem 2.3.

The maximum likelihood estimator of ρ is biased while its approximate mean is ρ and the approximate variance can also be obtained. Some results about the maximum likelihood estimator of ρ are shown in the next proposition.

<u>Proposition 2.8</u>**:** The MLE of ρ, namely $\hat{\rho} = \frac{1}{p-1}(\frac{B2}{B1} - 1)$, where

$B1 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})$ and $B2 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})$, has the following properties:

a) $Corr(B1, B2) = \frac{1+(p-1)\rho}{\sqrt{p[1+(p-1)\rho^2]}}$,

b) $E(\hat{\rho}) = E\left[\frac{1}{p-1}(\frac{B2}{B1} - 1)\right] \approx \rho$, $V(\hat{\rho}) \approx \frac{2}{M-1} \cdot \frac{[1+(p-1)\rho]^2 (1-\rho)^2}{p(p-1)}$, and

c) $\hat{\rho} \to \rho$ in probability.

Proof:

Recall that

$$B1 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})^T (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})$$
$$= \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu})^T (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}) - (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \boldsymbol{Q} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}), \tag{2.31}$$

$$B2 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})$$
$$= \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu})^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}) - (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \boldsymbol{Q} \boldsymbol{J}_p (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}), \tag{2.32}$$

and

$$(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \boldsymbol{Q} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) = \sum_{i=1}^{M} \sum_{j=1}^{M} (\boldsymbol{X}_i - C_i \boldsymbol{\mu})^T \boldsymbol{C}_i \boldsymbol{Q}^{-1} \boldsymbol{C}_j^{\,T} (\boldsymbol{X}_j - C_j \boldsymbol{\mu}), \tag{2.33}$$

where $\boldsymbol{Q} = \sum_{i=1}^{M} \boldsymbol{C}_i^{\,T} \boldsymbol{C}_i$. Similarly,

$$(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \boldsymbol{Q} \boldsymbol{J}_p (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) = \sum_{i=1}^{M} \sum_{j=1}^{M} (\boldsymbol{X}_i - C_i \boldsymbol{\mu})^T \boldsymbol{C}_i \boldsymbol{Q}^{-1} \boldsymbol{C}_j^{\,T} \boldsymbol{J}_p (\boldsymbol{X}_j - C_j \boldsymbol{\mu}). \tag{2.34}$$

Note that $\boldsymbol{C}_i$, $\boldsymbol{Q}^{-1}$, $\boldsymbol{C}_j^{\,T}$, and $\boldsymbol{J}_p$ commute with each other since all of them are circulant matrices. The commutation property will be used when necessary in calculations. So $Cov(B1, B2)$ can be expressed as

$$Cov(B1, B2) = Cov\left(\sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu})^T (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}), \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu})^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}) \right)$$
$$- Cov\left(\sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu})^T (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}), (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \boldsymbol{Q} \boldsymbol{J}_p (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \right)$$
$$- Cov\left((\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \boldsymbol{Q} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}), \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu})^T J_p (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}) \right)$$
$$+ Cov\left((\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \boldsymbol{Q} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}), (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \boldsymbol{Q} \boldsymbol{J}_p (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \right)$$
$$= D - E - F + G,$$

The derivations of *D*, *E*, *F*, and *G* are shown below.

$$D = Cov\left(\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}), \sum_{j=1}^{M}(\boldsymbol{X}_j - \boldsymbol{C}_j\boldsymbol{\mu})^T\boldsymbol{J}_p(\boldsymbol{X}_j - \boldsymbol{C}_j\boldsymbol{\mu})\right)$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M}Cov\left((\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}), (\boldsymbol{X}_j - \boldsymbol{C}_j\boldsymbol{\mu})^T\boldsymbol{J}_p(\boldsymbol{X}_j - \boldsymbol{C}_j\boldsymbol{\mu})\right)$$

$$= \sum_{i=1}^{M}Cov\left((\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}), (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})\right) \quad \text{if } i \neq j, \text{covariance is zero.}$$

$$= \sum_{i=1}^{M}\left\{\begin{aligned}&E\left[(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})\cdot(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})\right]\\ &-E\left[(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})\right]\cdot E\left[(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T\boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})\right]\end{aligned}\right\}$$

$$= \sum_{i=1}^{M}\left\{tr(\boldsymbol{\Sigma})tr(\boldsymbol{J}_p\boldsymbol{\Sigma}) + 2tr(\boldsymbol{\Sigma}\cdot\boldsymbol{J}_p\boldsymbol{\Sigma}) - tr(\boldsymbol{\Sigma})tr(\boldsymbol{J}_p\boldsymbol{\Sigma})\right\} = 2M\cdot tr(\boldsymbol{J}_p\boldsymbol{\Sigma}^2) = 2Mp[1+(p-1)\rho]^2\sigma^4,$$

$$G = Cov\left((\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}), (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}\boldsymbol{J}_p(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\right)$$

$$= E\left[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\cdot(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}\boldsymbol{J}_p(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\right]$$

$$- E\left[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\right]E\left[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}\boldsymbol{J}_p(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\right]$$

$$= \left[tr(\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{\Sigma})tr(\boldsymbol{Q}\boldsymbol{J}_p\boldsymbol{Q}^{-1}\boldsymbol{\Sigma}) + 2tr(\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{\Sigma}\boldsymbol{Q}\boldsymbol{J}_p\boldsymbol{Q}^{-1}\boldsymbol{\Sigma})\right] - tr(\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{\Sigma})tr(\boldsymbol{Q}\boldsymbol{J}_p\boldsymbol{Q}^{-1}\boldsymbol{\Sigma})$$

$$= 2tr(\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{\Sigma}\boldsymbol{Q}\boldsymbol{J}_p\boldsymbol{Q}^{-1}\boldsymbol{\Sigma}) = 2tr(\boldsymbol{\Sigma}\boldsymbol{J}_p\boldsymbol{\Sigma}) = 2tr(\boldsymbol{J}_p\boldsymbol{\Sigma}^2),$$

$$E = Cov\left(\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}), (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}\boldsymbol{J}_p(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\right)$$

$$= Cov\left(\sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}), \sum_{i=1}^{M}\sum_{j=1}^{M}(\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j^{\;T}\boldsymbol{J}_p(\boldsymbol{X}_j - C_j\boldsymbol{\mu})\right)$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}Cov\left((\boldsymbol{X}_k - C_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - C_k\boldsymbol{\mu}), (\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j^{\;T}\boldsymbol{J}_p(\boldsymbol{X}_j - C_j\boldsymbol{\mu})\right)$$

$$= \sum_{\substack{i,j,k=1\\ i=j=k}}^{M}Cov\left((\boldsymbol{X}_k - C_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - C_k\boldsymbol{\mu}), (\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j^{\;T}\boldsymbol{J}_p(\boldsymbol{X}_j - C_j\boldsymbol{\mu})\right)$$

$$+ \sum_{\substack{i,j,k=1\\ i=j\neq k}}^{M}Cov\left((\boldsymbol{X}_k - C_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - C_k\boldsymbol{\mu}), (\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T\boldsymbol{C}_i\boldsymbol{Q}^{-1}\boldsymbol{C}_j^{\;T}\boldsymbol{J}_p(\boldsymbol{X}_j - C_j\boldsymbol{\mu})\right)$$

if $i \neq j$, covariance is zero

$$
\begin{aligned}
&= \sum_{i=1}^{M} Cov\Big((\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - C_i\boldsymbol{\mu}), (\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T \boldsymbol{C}_i^T \boldsymbol{C}_i \boldsymbol{Q}^{-1}\boldsymbol{J}_p(\boldsymbol{X}_i - C_i\boldsymbol{\mu})\Big) \\
&\quad + \sum_{\substack{i,k=1 \\ i\neq k}}^{M} Cov\Big((\boldsymbol{X}_k - C_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - C_k\boldsymbol{\mu}), (\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T \boldsymbol{C}_i^T \boldsymbol{C}_i \boldsymbol{Q}^{-1}\boldsymbol{J}_p(\boldsymbol{X}_i - C_i\boldsymbol{\mu})\Big) \\
&= \sum_{i=1}^{M} \left\{ \begin{aligned} &E\Big[(\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - C_i\boldsymbol{\mu}) \cdot (\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T \boldsymbol{C}_i^T \boldsymbol{C}_i \boldsymbol{Q}^{-1}\boldsymbol{J}_p(\boldsymbol{X}_i - C_i\boldsymbol{\mu})\Big] \\ &- E\Big[(\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T(\boldsymbol{X}_i - C_i\boldsymbol{\mu})\Big] \cdot E\Big[(\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T \boldsymbol{C}_i^T \boldsymbol{C}_i \boldsymbol{Q}^{-1}\boldsymbol{J}_p(\boldsymbol{X}_i - C_i\boldsymbol{\mu})\Big] \end{aligned} \right\} \\
&\quad + \sum_{\substack{i,k=1 \\ i\neq k}}^{M} \left\{ \begin{aligned} &E\Big[(\boldsymbol{X}_k - C_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - C_k\boldsymbol{\mu}) \cdot (\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T \boldsymbol{C}_i^T \boldsymbol{C}_i \boldsymbol{Q}^{-1}\boldsymbol{J}_p(\boldsymbol{X}_i - C_i\boldsymbol{\mu})\Big] \\ &- E\Big[(\boldsymbol{X}_k - C_k\boldsymbol{\mu})^T(\boldsymbol{X}_k - C_k\boldsymbol{\mu})\Big] \cdot E\Big[(\boldsymbol{X}_i - C_i\boldsymbol{\mu})^T \boldsymbol{C}_i^T \boldsymbol{C}_i \boldsymbol{Q}^{-1}\boldsymbol{J}_p(\boldsymbol{X}_i - C_i\boldsymbol{\mu})\Big] \end{aligned} \right\} \\
&= \sum_{i=1}^{M} \Big\{\Big[tr(\boldsymbol{\Sigma})tr(\boldsymbol{C}_i^T \boldsymbol{C}_i \boldsymbol{Q}^{-1}\boldsymbol{J}_p\boldsymbol{\Sigma}) + 2tr(\boldsymbol{\Sigma} \cdot \boldsymbol{C}_i^T \boldsymbol{C}_i \boldsymbol{Q}^{-1}\boldsymbol{J}_p\boldsymbol{\Sigma})\Big] - \Big[tr(\boldsymbol{\Sigma})tr(\boldsymbol{C}_i^T \boldsymbol{C}_i \boldsymbol{Q}^{-1}\boldsymbol{J}_p\boldsymbol{\Sigma})\Big]\Big\} + 0 \\
&= 2tr(\boldsymbol{\Sigma} \cdot \sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \boldsymbol{Q}^{-1}\boldsymbol{J}_p\boldsymbol{\Sigma}) = 2tr(\boldsymbol{\Sigma} \cdot \boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{J}_p\boldsymbol{\Sigma}) = 2tr(\boldsymbol{J}_p\boldsymbol{\Sigma}^2).
\end{aligned}
$$

Similarly,

$$
F = Cov\left((\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}), \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T J_p (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}) \right) = 2tr(\boldsymbol{J}_p\boldsymbol{\Sigma}^2).
$$

So D, E, F, and G are, respectively,

$$
D = 2M \cdot tr(\boldsymbol{J}_p\boldsymbol{\Sigma}^2) = 2Mp[1 + (p-1)\rho]^2\sigma^4, \text{ and}
$$

$$
E = F = G = 2 \cdot tr(\boldsymbol{J}_p\boldsymbol{\Sigma}^2) = 2p[1 + (p-1)\rho]^2\sigma^4.
$$

Therefore, $Cov(B1, B2)$ and $Corr(B1, B2)$ are respectively

$$
\begin{aligned}
&Cov(B1, B2) = D - E - F + G \\
&= 2(M-1)tr(\boldsymbol{J}_p\boldsymbol{\Sigma}^2) = 2(M-1)p[1 + (p-1)\rho]^2\sigma^4,
\end{aligned} \tag{2.35}
$$

and

$$
\begin{aligned}
Corr(B1, B2) &= \frac{Cov(B1, B2)}{\sqrt{Var(B1)}\sqrt{Var(B2)}} \\
&= \frac{2(M-1)p[1 + (p-1)\rho]^2\sigma^4}{\sqrt{2(M-1)p[1 + (p-1)\rho^2]\sigma^4}\sqrt{2(M-1) \cdot p^2[1 + (p-1)\rho]^2\sigma^4}} \\
&= \frac{1 + (p-1)\rho}{\sqrt{p[1 + (p-1)\rho^2]}}.
\end{aligned}
$$

Finally, we may compute the approximate mean and variance of B_2 / B_1 using the first-order Taylor's series in two variables $f(x, y) = x / y$, $y \neq 0$. Hence we have

$$E(\frac{B2}{B1}) \approx \frac{E(B2)}{E(B1)} = \frac{(M-1)tr(\boldsymbol{J}_p \boldsymbol{\Sigma})}{(M-1)tr(\boldsymbol{\Sigma})} = \frac{p[1+(p-1)\rho]\sigma^2}{p\sigma^2} = 1+(p-1)\rho, \qquad (2.36)$$

and

$$\begin{aligned} V(\frac{B2}{B1}) &\approx \frac{E^2(B2)}{E^2(B1)} \left\{ \frac{V(B2)}{E^2(B2)} + \frac{V(B1)}{E^2(B1)} - 2\frac{Cov(B2,B1)}{E(B2)E(B1)} \right\} \\ &= [1+(p-1)\rho]^2 \left\{ \frac{2}{M-1} + \frac{2[1+(p-1)\rho^2]}{(M-1)p} - \frac{4[1+(p-1)\rho]}{(M-1)p} \right\} \\ &= \frac{2[1+(p-1)\rho]^2}{(M-1)p} \cdot \left\{ p + [1+(p-1)\rho^2] - 2[1+(p-1)\rho] \right\} \\ &= \frac{2}{M-1} \cdot \frac{[1+(p-1)\rho]^2 (p-1)(1-\rho)^2}{p}, \end{aligned} \qquad (2.37)$$

implying that $\hat{\rho} \to \rho$ in probability. The proof is complete.

The following theorem states the exact distribution of the MLE of ρ .

<u>Theorem 2.4</u>**:** The MLE of ρ , say $\hat{\rho} = \frac{1}{p-1}(\frac{B2}{B1} - 1)$ with $B1 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})^T (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})$ and $B2 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}})$ is distributed as the quantity

$$\frac{1}{p-1} \left[\frac{p}{\frac{(1-\rho)(p-1)}{1+(p-1)\rho} F_{(M-1)(p-1), M-1} + 1} - 1 \right].$$

***Remark 2.3*:** $\hat{\rho}$ is between $-(p-1)^{-1}$ and 1 since the ratio $\frac{B2}{B1}$ is between 0 and *p*. To show this, first we have that $\frac{B2}{B1} > 0$ implying $\rho > -(p-1)^{-1}$ since $B1 > 0$ and $B2 > 0$ for nonzero vectors $\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}$. Secondly, consider the identity $x_i'x_i = x_i'(p^{-1}\boldsymbol{J}_p)x_i + x_i'(\boldsymbol{I}_p - p^{-1}\boldsymbol{J}_p)x_i$. Since all the three quantities $x_i'x_i$, $x_i'(p^{-1}\boldsymbol{J}_p)x_i$, and $x_i'(\boldsymbol{I}_p - p^{-1}\boldsymbol{J}_p)x_i$ are positive for nonzero vectors x_i, the inequality $\sum_{i=1}^{M} x_i'x_i > \sum_{i=1}^{M} x_i'(p^{-1}\boldsymbol{J}_p)x_i$ holds and it implies that $\sum_{i=1}^{M} x_i'\boldsymbol{J}_p x_i / \sum_{i=1}^{M} x_i'x_i < p$. Hence $\rho < 1$.

Proof of Theorem 2.4:

Recall from (2.31) and (2.32) that

$$B1 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}) - (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \boldsymbol{Q}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}),$$

and

$$B2 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu})^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}) - (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \boldsymbol{Q}\boldsymbol{J}_p (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}),$$

where $\boldsymbol{Q} = \sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i$. And we have $B1 = [B1 - (1/p)B2] + (1/p)B2$. Since from Proposition 2.5 $B1 - (1/p)B2$ and $(1/p)B2$ are independent, and from Propositions 2.3 and 2.4 $B1 - (1/p)B2$ and $(1/p)B2$ are distributed as $\sigma^2(1-\rho)\chi^2((M-1)(p-1))$ and $\sigma^2[1+(p-1)\rho]\chi^2(M-1)$ random variables, respectively, we have

$$\frac{B2}{B1} = \frac{B2}{[B1 - (1/p)B2] + (1/p)B2} = \frac{p}{\frac{B1 - (1/p)B2}{(1/p)B2} + 1},$$

which is distributed as $p\left[\frac{1-\rho}{1+(p-1)\rho}(p-1)F_{(M-1)(p-1), M-1} + 1\right]^{-1}$ random variable where

$F_{(M-1)(p-1),M-1}$ is F random variable with $(M-1)(p-1)$ and $M-1$ as the numerator and denominator degrees of freedom, respectively. Thus $\hat{\rho}$ is distributed as the quantity

$$\frac{1}{p-1}\left[\frac{p}{\frac{(1-\rho)(p-1)}{1+(p-1)\rho}F_{(M-1)(p-1),M-1}+1}-1\right].$$

The proof is complete.

For the rest of this subsection, a simulation study is performed to investigate the behavior of $\hat{\rho}$ based on the distribution of $\hat{\rho}$ obtained from Theorem 2.4. Figure 1 and Figure 2 show the expectation and the standard deviation of $\hat{\rho}$, the MLE of ρ, for each value $\rho \in (-(p-1)^{-1},1)$ via a simulation study with various combinations of dimensions $p = 2, 3, 4, 5, 6, 7$ and sample sizes $M = 2, 3, 5, 10, 20, 50, 100$. Note that the starting points of ρ on the x-axis are different for various p values since the restriction on ρ is $\rho > -(p-1)^{-1}$ due to the requirement of a positive definite compound symmetry covariance matrix structure. Summarizing the information provided from Figure 1 and Figure 2 we have the following results:

About the expectation of $\hat{\rho}$:

(1) When $\rho = 0$, the MLE $\hat{\rho}$ is unbiased. This can also be verified by looking at the pdf of $\hat{\rho}$ stated in Theorem 2.4 for the special case that $\rho = 0$. With $\rho = 0$, $\hat{\rho}$ is distributed like the random variable $\frac{1}{p-1}\left[p \cdot Beta(\alpha,\beta)-1\right]$, where $\alpha = (M-1)/2$, $\beta = (M-1)(p-1)/2$, and $Beta(\alpha,\beta)$ is the beta random variable . Therefore $\hat{\rho}$ is unbiased since

$$E(\hat{\rho}) = \frac{1}{p-1}\left[p \cdot EBeta(\alpha,\beta) - 1\right] = \frac{1}{p-1}\left[p \cdot \frac{\alpha}{\alpha+\beta} - 1\right] = \frac{1}{p-1}\left[p \cdot p^{-1} - 1\right] = 0 .$$

(2) When ρ is close to one of the end points $-(p-1)^{-1}$ and 1, $\hat{\rho}$ tends to be unbiased. Otherwise, when $\rho < 0$, $\hat{\rho}$ overestimates ρ; when $\rho > 0$, $\hat{\rho}$ underestimates ρ.

(3) When the sample size M increases, $\hat{\rho}$ becomes more accurate. Actually from the results of Proposition 2.8, $\hat{\rho}$ converges in probability to ρ.

About the standard deviation of $\hat{\rho}$:

(1) When $p = 2$, the function of the standard deviation of $\hat{\rho}$ is like an upside-down bathtub when M is small. When the same size increases, the bathtub shape become flatter.

(2) When $p > 2$, the bathtub shape is not symmetric and shrinks to the right.

(3) Basically, with fixed p and ρ, the standard deviation decreases when the sample size increases.

Figures 3 and 4 illustrate the simulated probability density functions for the MLE of ρ for the cases $p = 2$ and $p = 3$, respectively. Various sample sizes 2, 5, 20, and 40 are considered for each figure. Summarizing the information provided from these two figures we have the following results about the probability density function of $\hat{\rho}$:

(1) With fixed p, when sample size is very small ($M = 2$), the probability density function is bimodal. Otherwise it is unimodal.

(2) With fixed p, when sample size becomes larger, the pdf of $\hat{\rho}$ becomes more concentrated and symmetric.

(3) With fixed sample size, when ρ is less than 0, the pdf is skewed to the right; otherwise it is skewed to the left.

(4) With fixed sample size, when ρ is more extreme, the pdf of $\hat{\rho}$ is steeper.

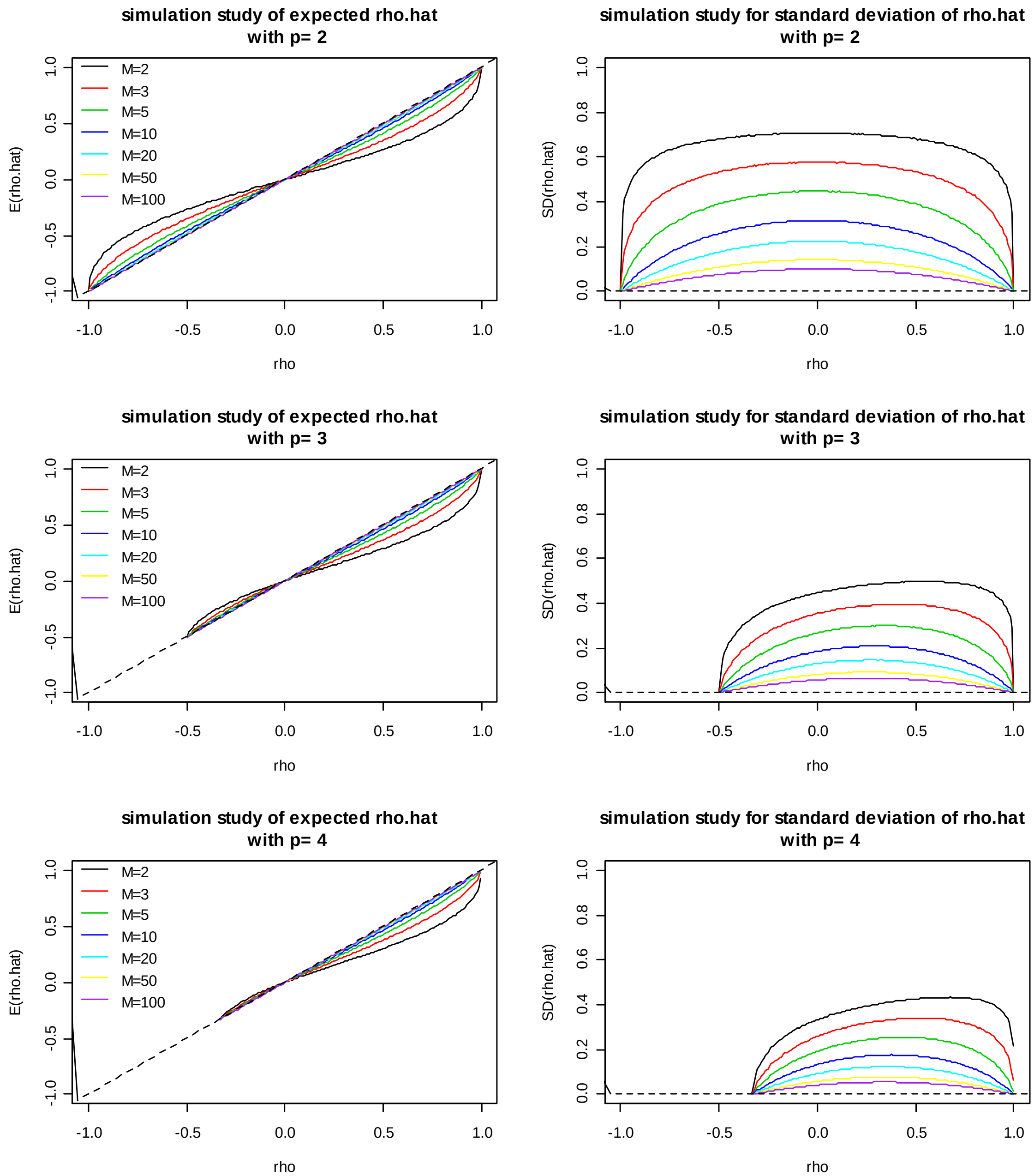

Figure 1

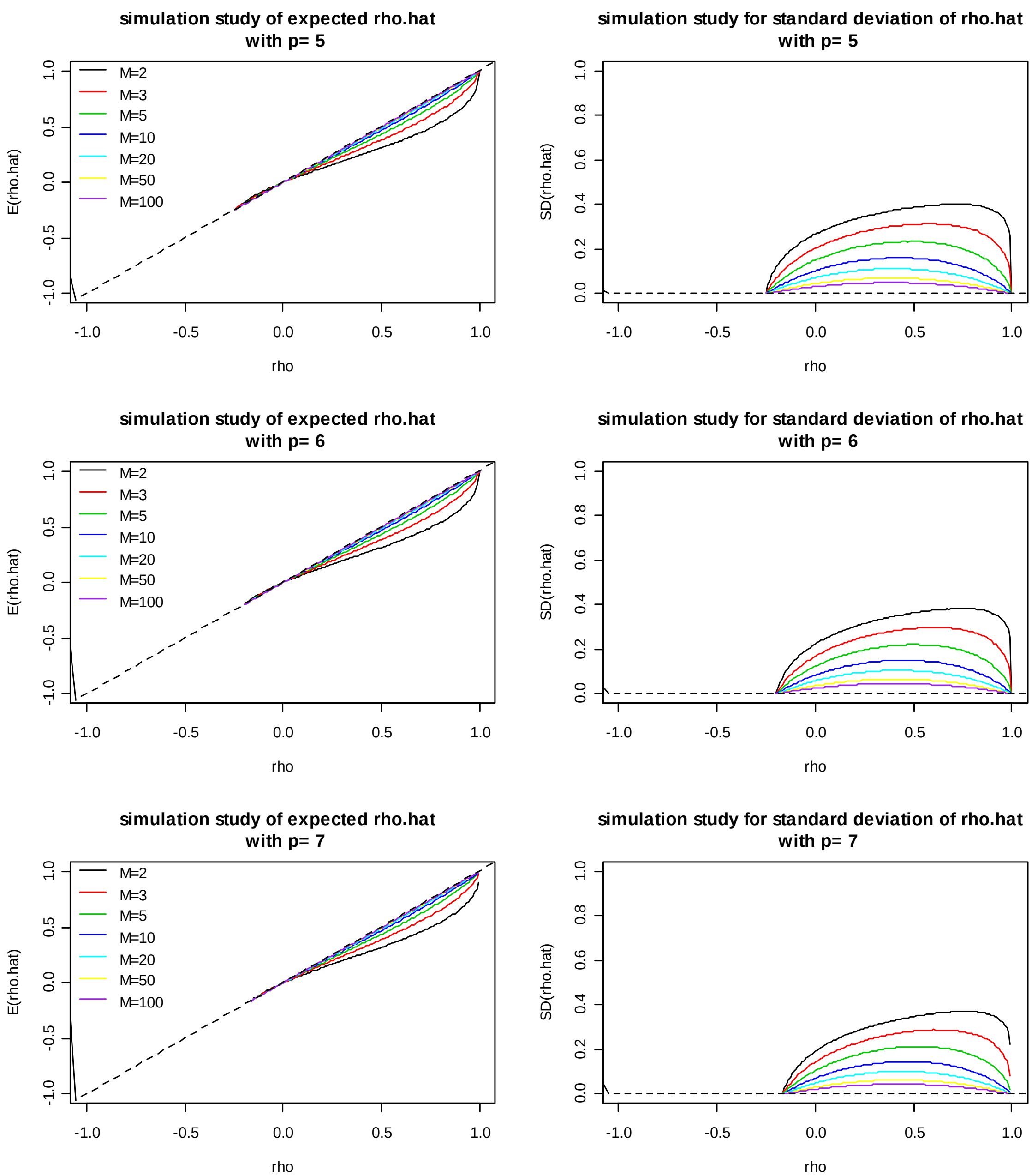
simulation study of expected rho.hat
with p= 5
simulation study for standard deviation of rho.hat
with p= 5
simulation study of expected rho.hat
with p= 6
simulation study for standard deviation of rho.hat
with p= 6
simulation study of expected rho.hat
with p= 7
simulation study for standard deviation of rho.hat
with p= 7
E(rho.hat)
SD(rho.hat)
rho
M=2
M=3
M=5
M=10
M=20
M=50
M=100
-1.0
-0.5
0.0
0.5
1.0
0.2
0.4
0.6
0.8

Figure 2

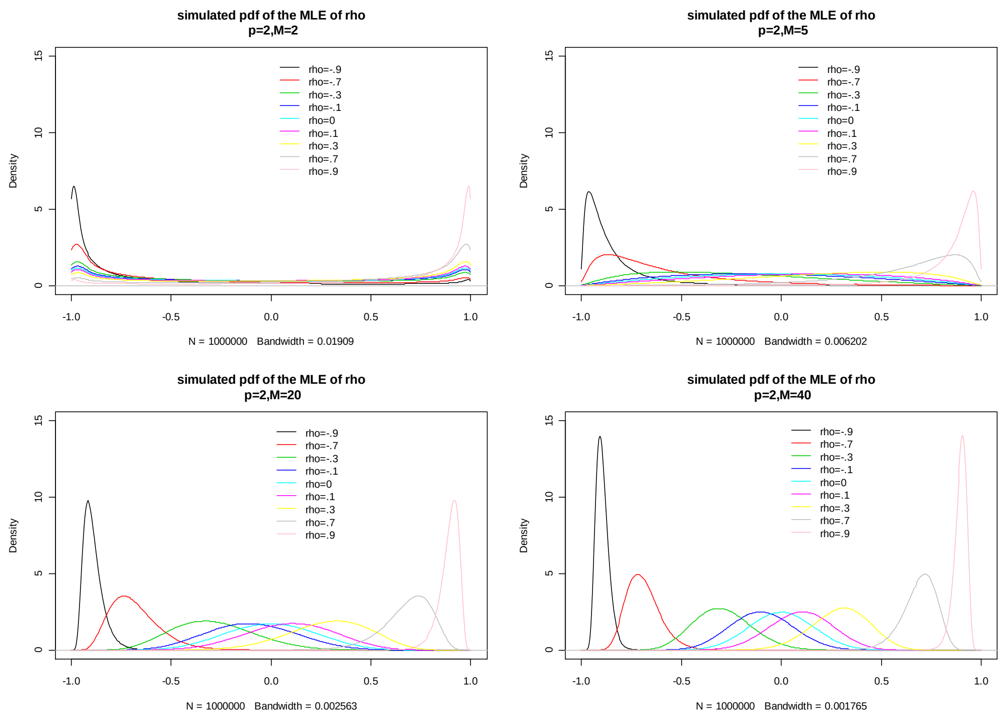

Figure 3

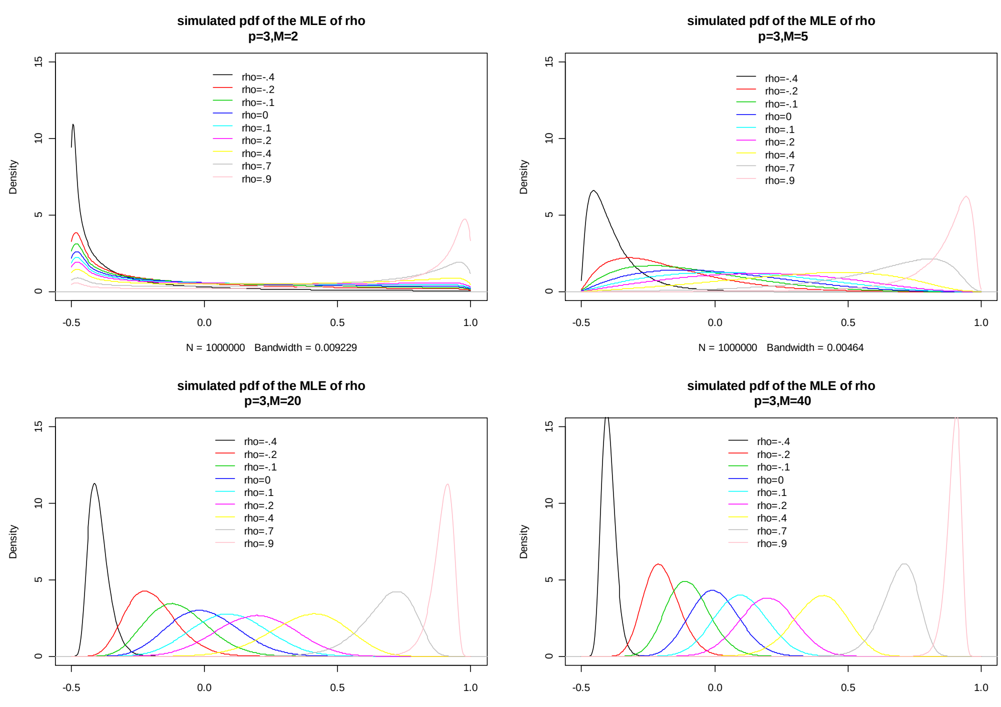

Figure 4

2.2.4 Hypothesis Testing for $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ Using Approximate χ^2 Test

Using the results from Subsection 2.2.3 that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ and $\hat{\rho} \xrightarrow{p} \rho$, we arrive at the following approximation theorem which can be used to test the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$.

***Theorem 2.5*:** $(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'[\hat{Var}(\hat{\boldsymbol{\mu}})]^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{d} \chi_p^2$.

Proof:

Recall that

$$[\hat{Var}(\hat{\boldsymbol{\mu}})]^{-1} = \hat{\boldsymbol{\Sigma}}^{-1}(\sum_{i=1}^{M} \boldsymbol{C}_i{}'\boldsymbol{C}_i)^{-1}, \ \hat{\boldsymbol{\Sigma}}^{-1} = \frac{1}{\hat{\sigma}^2(1-\hat{\rho})}\left(\boldsymbol{I}_p - \frac{\hat{\rho}}{1+(p-1)\hat{\rho}}\boldsymbol{J}_p\right).$$

Also we have the expression

$$\begin{aligned}
&(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'[\hat{Var}(\hat{\boldsymbol{\mu}})]^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \\
&= (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \frac{1}{\hat{\sigma}^2(1-\hat{\rho})}\left(\boldsymbol{I}_p - \frac{\hat{\rho}}{1+(p-1)\hat{\rho}}\boldsymbol{J}_p\right)\left(\sum_{i=1}^{M} \boldsymbol{C}_i{}'\boldsymbol{C}_i\right)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \\
&= \frac{(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \frac{1}{\sigma^2(1-\rho)}\left(\boldsymbol{I}_p - \frac{\rho}{1+(p-1)\rho}\boldsymbol{J}_p\right)\left(\sum_{i=1}^{M} \boldsymbol{C}_i{}'\boldsymbol{C}_i\right)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})}{\frac{\hat{\sigma}^2(1-\hat{\rho})}{\sigma^2(1-\rho)}} + \\
&\quad \frac{(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \frac{1}{\sigma^2(1-\rho)}\left(\frac{\hat{\rho}}{1+(p-1)\hat{\rho}} - \frac{\rho}{1+(p-1)\rho}\right)\boldsymbol{J}_p\left(\sum_{i=1}^{M} \boldsymbol{C}_i{}'\boldsymbol{C}_i\right)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})}{\frac{\hat{\sigma}^2(1-\hat{\rho})}{\sigma^2(1-\rho)}}.
\end{aligned}$$

Since $\frac{\hat{\sigma}^2(1-\hat{\rho})}{\sigma^2(1-\rho)} \xrightarrow{p} 1$ and $\frac{\hat{\rho}}{1+(p-1)\hat{\rho}} - \frac{\rho}{1+(p-1)\rho} \xrightarrow{p} 0$, we have by Slutsky's theorem that

$$(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'[\hat{Var}(\hat{\boldsymbol{\mu}})]^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{d} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\sum_{i=1}^{M} \boldsymbol{C}_i{}'\boldsymbol{C}_i)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}),$$

which follows a χ_p^2 distribution. The proof is complete.

2.3 SIMULATION STUDY FOR MISUSE OF HOMOGENEOUS MEAN MODELS

In this section, power under $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ based on two test procedures, each of which corresponds to the same hypothesis but different model setting, will be compared for the purpose of showing that the usual test procedure for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is not appropriate when our data are polluted by some reasons but ignored by researchers.

In each simulation, a sample of independent bivariate normal data $X_1, ..., X_m$, $m = 100$ is generated from $MVN_2(\boldsymbol{C}_i\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$, where $\boldsymbol{C}_i = \boldsymbol{I}_2 + \boldsymbol{C}_{i0}$, where

$$\boldsymbol{C}_{i0} = \begin{pmatrix} a_{i0} & b_{i0} \\ b_{i0} & a_{i0} \end{pmatrix}.$$

Note that $\boldsymbol{C}_{i0}$ is (symmetric) circulant, and thus so is $\boldsymbol{C}_i$. Two likelihood ratio tests are denoted by LRT$_{\boldsymbol{C\mu}}$ and LRT$_{\boldsymbol{\mu}}$ which are stated below respectively:

- LRT$_{\boldsymbol{C\mu}}$: LRT for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ for homogeneous mean model $\boldsymbol{X}_i \sim N_2(\boldsymbol{C}_i\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and
- LRT$_{\boldsymbol{\mu}}$: LRT for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ for heterogeneous means model $\boldsymbol{X}_i \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown but $\boldsymbol{\Sigma}$ has compound symmetry structure. Recall from Theorem 2.2 that the test statistics for LRT$_{\boldsymbol{C\mu}}$ is

$$\text{LRT}_{\boldsymbol{C\mu}} \text{ statistic} = \frac{[pB1 - B2]^{p-1} B2}{[pB1_0 - B2_0]^{p-1} B2_0},$$

where

$$B1 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}),\ B2 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}),$$

$$B1_0 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0),\ B2_0 = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0),$$

and

$$\hat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{M} \boldsymbol{C}_i^{\,T} \boldsymbol{C}_i\right)^{-1} \sum_{i=1}^{M} \boldsymbol{C}_i^{\,T} \boldsymbol{X}_i \ .$$

When $\boldsymbol{C}_{i0} = \boldsymbol{0}$ for all *i*, the two test statistics are the same. Under $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, both of the test statistics are distributed as the random variable stated in Theorem 2.2. We reject the null hypothesis when the test statistics are sufficiently small.

The simulation study is described as follows.

Data: Data are generated from $N_2(\boldsymbol{C}_i \boldsymbol{\mu}_0, \boldsymbol{\Sigma})$, where $\boldsymbol{C}_i = \boldsymbol{I}_2 + \boldsymbol{C}_{i0}$, $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}$, and $\boldsymbol{C}_{i0}$ are shown in the first four columns in Table 1.

Hypotheses: Both tests correspond to the hypothesis of interest $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$.

Tests and critical value: Two likelihood ratio tests are performed based on the generated data. The critical value for the two tests is the same since the null distribution of both tests are the same. As we can see in Theorem 2.2, the null distribution of the test statistic of $\text{LRT}_{\boldsymbol{C\mu}}$ does not depend on the matrices $\boldsymbol{C}_i$.

Number of simulations: The number of LRT values needed to compute the empirical alpha of the test $\text{LRT}_{\boldsymbol{C\mu}}$ or the rejection probability of the test $\text{LRT}_{\boldsymbol{\mu}}$ is 10000.

Interpretation of the simulation study: Column 4 of Table 1 shows the diagonal elements a_{i0} of the matrices $\boldsymbol{C}_i$. For instance, a_{i0} = - .99(.02) means that the first value of a_{i0} is $a_{10} = -.99$, then increases by 0.02 for each one unit increase of *i*. As denoted in column 5 from Table 1, the value (probability) in each cell is the empirical α for the test $\text{LRT}_{\boldsymbol{C\mu}}$ given the generated data from the heterogeneous means models. All the values in column 5 are close to 0.05, the significant level specified for the test and is as expected. On the other hand, since the data are polluted, adopting the test $\text{LRT}_{\boldsymbol{\mu}}$ does not make sense and is not appropriate. If we still consider that the generated data are from the homogeneous mean model $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the rejection probability for each scenario is shown in column 6 of Table 1. As we can see, the values of this column vary from one scenario to another. Some achieve the probability of 1 and some is less than 0.05. Generally, when the pollution of the data becomes more severe, that is when matrices $\boldsymbol{C}_{i0}$ is far away from zero matrix with a faster rate, the rejection probability is larger. Under the scenario $\sum \boldsymbol{C}_{i0} = \boldsymbol{0}$, all the three rejection probabilities are less than 0.05 and one of them is even 0. Lastly, the two rejection probabilities of column 6 are 1 even when data suffer only slight contamination (a_{i0} =.001(.001) and $b_{i0} = -a_{i0}$ for both of the two cases about $\boldsymbol{\Sigma}$).

TABLE 1: Result of simulation study for misuse of homogeneous mean model

(1)	(2)	(3)	(4)	(5)	(6)
$\boldsymbol{\mu}_0$	$\boldsymbol{\Sigma}$	$\boldsymbol{C}_{i0} = \begin{pmatrix} a_{i0} & b_{i0} \\ b_{i0} & a_{i0} \end{pmatrix}$ $\boldsymbol{C}_i = \boldsymbol{I}_2 + \boldsymbol{C}_{i0}$ i=1,…, m	Values of a_{i0}	Testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ $\text{LRT}_{\boldsymbol{C\mu}}$ (Empirical α)	$\text{LRT}_{\boldsymbol{\mu}}$ (Rejection Probability)
$\begin{pmatrix} 10 \\ 30 \end{pmatrix}$	$\begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$	$b_{i0} = a_{i0}$	a_{i0} =.02(.02)	.055	1
		$b_{i0} = 0$		.045	1
		$b_{i0} = -a_{i0}$		.051	1
		$b_{i0} = a_{i0}$	$\Sigma \boldsymbol{C}_{i0} = \boldsymbol{0}$ a_{i0} = - .99(.02)	.049	.015
		$b_{i0} = 0$		.043	0
		$b_{i0} = -a_{i0}$		.047	.014
		$b_{i0} = -a_{i0}$	a_{i0} = .00001(.00001)	.050	.0528
			a_{i0} =.0001(.0001)	.06	.407
			a_{i0} =.001(.001)	.047	1
	$\begin{pmatrix} 1 & .2 \\ .2 & 1 \end{pmatrix}$	$b_{i0} = -a_{i0}$	a_{i0} =.00001(.00001)	.056	.057
			a_{i0} =.0001(.0001)	.053	.18
			a_{i0} =.001(.001)	.046	1

CHAPTER III

MULTISAMPLE INFERENCE

3.1 INTRODUCTION

In this chapter, we move on to the inference for multisample case when the heterogeneous means models are adopted. Two-sample inference will be the starting point. Consider two independent samples $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_M \sim MVN_p(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_x),\ \boldsymbol{\mu}_i = \boldsymbol{C}_i\boldsymbol{\mu}_x$, for all $i = 1,\ldots,M$, and

$\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_N \sim MVN_p(\boldsymbol{v}_j,\boldsymbol{\Sigma}_y),\ \boldsymbol{v}_j = \boldsymbol{D}_j\boldsymbol{\mu}_y$, for all $j = 1,\ldots,N$. Both $\boldsymbol{C}_i$ and $\boldsymbol{D}_j$ are known $p\times p$ matrices. The hypotheses of interest are $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$ versus $H_a : \boldsymbol{\mu}_x \neq \boldsymbol{\mu}_y$. The likelihood function is

$$L(\boldsymbol{\mu}_x,\boldsymbol{\mu}_y,\boldsymbol{\Sigma}_x,\boldsymbol{\Sigma}_y) = constant\cdot|\boldsymbol{\Sigma}_x|^{-\frac{M}{2}}\cdot|\boldsymbol{\Sigma}_y|^{-\frac{N}{2}} \cdot\exp\left\{-\frac{1}{2}\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_x)-\frac{1}{2}\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_y)^T\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_y)\right\}.$$

The corresponding log likelihood function is

$$\begin{aligned}\log L(\boldsymbol{\mu}_x,\boldsymbol{\mu}_y,\boldsymbol{\Sigma}_x,\boldsymbol{\Sigma}_y) &= constant-\frac{M}{2}\log|\boldsymbol{\Sigma}_x|-\frac{N}{2}\log|\boldsymbol{\Sigma}_y|-\frac{1}{2}\left\{\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_x)\right. \\ &\quad\left.+\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_y)^T\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_y)\right\}\end{aligned} \tag{3.1}$$

First consider the simple case where both $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$ are known. The MLEs for $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$

are, respectively,

$$\hat{\boldsymbol{\mu}}_x = \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{C}_i \right)^{-1} \sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{X}_i ,$$

$$\hat{\boldsymbol{\mu}}_y = \left(\sum_{j=1}^{N} \boldsymbol{D}_i^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{D}_j \right)^{-1} \sum_{j=1}^{N} \boldsymbol{D}_j^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{Y}_j .$$

$\hat{\boldsymbol{\mu}}_x$ and $\hat{\boldsymbol{\mu}}_y$ are independent and

$$\hat{\boldsymbol{\mu}}_x \sim MVN_p\left(\boldsymbol{\mu}_x, \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}_x^{-1} \boldsymbol{C}_i \right)^{-1} \right) ,$$

$$\hat{\boldsymbol{\mu}}_y \sim MVN_p\left(\boldsymbol{\mu}_y, \left(\sum_{j=1}^{N} \boldsymbol{D}_i^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{D}_j \right)^{-1} \right),$$

$$\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y \sim MVN_p\left(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y, \left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}_x^{-1} \boldsymbol{C}_i \right)^{-1} + \left(\sum_{j=1}^{N} \boldsymbol{D}_i^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{D}_j \right)^{-1} \right).$$

Define the statistic

$$T_0 = (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \left(\left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}_x^{-1} \boldsymbol{C}_i \right)^{-1} + \left(\sum_{j=1}^{N} \boldsymbol{D}_i^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{D}_j \right)^{-1} \right)^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y).$$

Under the null hypothesis $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$, $T_0 \sim \chi_p^2$. Thus we reject H_0 if $T_0 > \chi_{p,\alpha}^2$.

For the case that both $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$ are unknown but equal, likelihood approach is used to test $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$ in Section 3.2. In Section 3.3, the asymptotic χ^2 test for testing $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$ is derived. Finally in Section 3.4 the LR test for two-sample case is extended to *k*-sample case and the exact distribution of the LRT statistic for $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k$ is derived.

3.2 LIKELIHOOD RATIO TEST FOR TWO-SAMPLE CASE

In this section, the case that $\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}$ unknown is considered. We also assume that $\boldsymbol{\Sigma}$ has compound symmetry with the form in (2.8), $\boldsymbol{C}_i$, $\boldsymbol{D}_j$, and $\boldsymbol{\Sigma}$ commute with each other; that is, $\boldsymbol{C}_i^T\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_i^T$, $\boldsymbol{D}_j^T\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{D}_j^T$ for all i and j. Before deriving the likelihood ratio test for $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$, it is necessary to find the MLEs of the parameters under the null and alternative hypotheses separately.

3.2.1 Estimation Under $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$

Assume that $\boldsymbol{\mu}_x = \boldsymbol{\mu}_y = \boldsymbol{\mu}_0$ under H_0. Using the same technique as shown in one-sample case, the MLE of $\boldsymbol{\mu}_0$, say $\hat{\boldsymbol{\mu}}_0$, can be derived as

$$\hat{\boldsymbol{\mu}}_0 = \left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\hat{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{C}_i + \sum_{j=1}^{N}\boldsymbol{D}_j^T\hat{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{D}_j\right)^{-1}\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\hat{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{X}_i + \sum_{j=1}^{N}\boldsymbol{D}_j^T\hat{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{Y}_j\right),$$

where $\hat{\boldsymbol{\Sigma}}_0$ is the MLE for $\boldsymbol{\Sigma}$ under H_0. Since $\boldsymbol{C}_i^T\hat{\boldsymbol{\Sigma}}_0^{-1} = \hat{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{C}_i^T$ and $\boldsymbol{D}_j^T\hat{\boldsymbol{\Sigma}}_0^{-1} = \hat{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{D}_j^T$ for all i and j, $\hat{\boldsymbol{\mu}}_0$ reduces to

$$\hat{\boldsymbol{\mu}}_0 = \left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i + \sum_{j=1}^{N}\boldsymbol{D}_j^T\boldsymbol{D}_j\right)^{-1}\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{X}_i + \sum_{j=1}^{N}\boldsymbol{D}_j^T\boldsymbol{Y}_j\right). \tag{3.2}$$

Therefore $\hat{\boldsymbol{\Sigma}}_0$ can be obtained using the reduced log likelihood function

$$\begin{aligned}\log L(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}) = constant - \frac{M+N}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\Big\{&\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0) \\ &+ \sum_{j=1}^{N}(\boldsymbol{y}_j - \boldsymbol{D}_j\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_j - \boldsymbol{D}_j\boldsymbol{\mu}_0)\Big\},\end{aligned} \tag{3.3}$$

The MLE for $\boldsymbol{\Sigma}$ under H_0 is thus

$$\hat{\boldsymbol{\Sigma}}_0 = \hat{\sigma}_0^{\,2}[(1-\hat{\rho}_0)\boldsymbol{I}_p + \hat{\rho}_0\boldsymbol{J}_p],$$

where

$$\hat{\sigma}_0^{\ 2} = \frac{1}{(M+N)p}\left(B1^{(0)} + E1^{(0)}\right), \ \ \hat{\rho}_0 = \frac{1}{p-1}\left(\frac{B2^{(0)} + E2^{(0)}}{B1^{(0)} + E1^{(0)}} - 1\right), \tag{3.4}$$

where

$$\begin{aligned} B1^{(0)} &= \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_0)^T (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_0), B2^{(0)} = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_0)^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_0), \\ E1^{(0)} &= \sum_{j=1}^{N} (\boldsymbol{Y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_0)^T (\boldsymbol{Y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_0), \ E2^{(0)} = \sum_{j=1}^{N} (\boldsymbol{Y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_0)^T \boldsymbol{J}_p (\boldsymbol{Y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_0). \end{aligned} \tag{3.5}$$

3.2.2 Estimation Under $H_a : \boldsymbol{\mu}_x \neq \boldsymbol{\mu}_y$

Under $H_a : \boldsymbol{\mu}_x \neq \boldsymbol{\mu}_y$, the log likelihood function is

$$\begin{aligned} \log L(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}) &= constant - \frac{M+N}{2} \log|\boldsymbol{\Sigma}| \\ &- \frac{1}{2}\left\{ \sum_{i=1}^{M} (\boldsymbol{x}_i - \boldsymbol{C}_i \boldsymbol{\mu}_x)^T \boldsymbol{\Sigma}_x^{\ -1} (\boldsymbol{x}_i - \boldsymbol{C}_i \boldsymbol{\mu}_x) + \sum_{j=1}^{N} (\boldsymbol{y}_j - \boldsymbol{D}_j \boldsymbol{\mu}_y)^T \boldsymbol{\Sigma}_y^{\ -1} (\boldsymbol{y}_j - \boldsymbol{D}_j \boldsymbol{\mu}_y) \right\}. \end{aligned}$$

Using a similar approach as shown in Section 2.2.3, the MLEs for $\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$, and $\boldsymbol{\Sigma}$ are, respectively,

$$\hat{\boldsymbol{\mu}}_x = \left(\sum_{i=1}^{M} \boldsymbol{C}_i^{\ T} \boldsymbol{C}_i \right)^{-1} \sum_{i=1}^{M} \boldsymbol{C}_i^{\ T} \boldsymbol{X}_i \ , \ \hat{\boldsymbol{\mu}}_y = \left(\sum_{j=1}^{N} \boldsymbol{D}_j^{\ T} \boldsymbol{D}_j \right)^{-1} \sum_{j=1}^{N} \boldsymbol{D}_j^{\ T} \boldsymbol{Y}_j \ ,$$

and

$$\hat{\boldsymbol{\Sigma}} = \hat{\sigma}^2 \boldsymbol{I}_p + \hat{\rho}\hat{\sigma}^2 (\boldsymbol{J}_p - \boldsymbol{I}_p) = \hat{\sigma}^2 [(1-\hat{\rho}) \boldsymbol{I}_p + \hat{\rho} \boldsymbol{J}_p],$$

where

$$\hat{\sigma}^2 = \frac{1}{(M+N)p}\left(B1^{(a)} + E1^{(a)}\right), \ \ \hat{\rho} = \frac{1}{p-1}\left(\frac{B2^{(a)} + E2^{(a)}}{B1^{(a)} + E1^{(a)}} - 1\right), \tag{3.6}$$

where

$$B1^{(a)} = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_x)^T (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_x),\ B2^{(a)} = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_x)^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_x),$$

$$E1^{(a)} = \sum_{j=1}^{N} (\boldsymbol{Y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_y)^T (\boldsymbol{Y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_y),\ E2^{(a)} = \sum_{j=1}^{N} (\boldsymbol{Y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_y)^T \boldsymbol{J}_p (\boldsymbol{Y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_y). \tag{3.7}$$

3.2.3 Likelihood Ratio Test for Testing $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$

Subsections 3.2.1 and 3.2.2 derived the MLE's for parameters under both null and alternative hypotheses. The likelihood ratio test can now be developed. The likelihood ratio is

$$\lambda = \frac{\max_{\boldsymbol{\theta}\in\Omega_0} L(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\theta}\in\Omega} L(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\Sigma})} = \frac{(2\pi)^{-\frac{(M+N)p}{2}} |\hat{\boldsymbol{\Sigma}}_0|^{-\frac{M+N}{2}}}{(2\pi)^{-\frac{(M+N)p}{2}} |\hat{\boldsymbol{\Sigma}}|^{-\frac{M+N}{2}}} \times$$

$$\frac{\exp -\frac{1}{2}\left(\sum_{i=1}^{M} (\boldsymbol{x}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_0)^T \hat{\boldsymbol{\Sigma}}_0^{-1} (\boldsymbol{x}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_0) + \sum_{j=1}^{N} (\boldsymbol{y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_0)^T \hat{\boldsymbol{\Sigma}}_0^{-1} (\boldsymbol{y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_0) \right)}{\exp -\frac{1}{2}\left(\sum_{i=1}^{M} (\boldsymbol{x}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_x)^T \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{x}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_x) + \sum_{j=1}^{N} (\boldsymbol{y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_y)^T \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_y) \right)},$$

where $\boldsymbol{\theta} = (\boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\Sigma})$,

$$\Omega = \{(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}) \mid \boldsymbol{\Sigma} = \sigma^2[(1-\rho)\boldsymbol{I}_p + \rho \boldsymbol{J}_p]\}, \text{ and}$$

$$\Omega_0 = \{(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}) \mid \boldsymbol{\mu}_x = \boldsymbol{\mu}_y, \boldsymbol{\Sigma} = \sigma^2[(1-\rho)\boldsymbol{I}_p + \rho \boldsymbol{J}_p]\}.$$

Hence the results (from Appendix A.3)

$$\sum_{i=1}^{M} (\boldsymbol{x}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_0)^T \hat{\boldsymbol{\Sigma}}_0^{-1} (\boldsymbol{x}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_0) + \sum_{j=1}^{N} (\boldsymbol{y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_0)^T \hat{\boldsymbol{\Sigma}}_0^{-1} (\boldsymbol{y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_0) = (M+N)p$$

and

$$\sum_{i=1}^{M} (\boldsymbol{x}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_x)^T \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{x}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_x) + \sum_{j=1}^{N} (\boldsymbol{y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_y)^T \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_y) = (M+N)p$$

imply that the likelihood ratio is

$$\lambda = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|} \right)^{\frac{M+N}{2}} = \left(\frac{(\hat{\sigma}^2)^p (1-\hat{\rho})^{p-1}[1+(p-1)\hat{\rho}]}{(\hat{\sigma}_0{}^2)^p (1-\hat{\rho}_0)^{p-1}[1+(p-1)\hat{\rho}_0]} \right)^{\frac{M+N}{2}}$$

$$= \left(\frac{\left[(B1^{(a)} + E1^{(a)}) - \frac{1}{p}(B2^{(a)} + E2^{(a)}) \right]^{p-1} (B2^{(a)} + E2^{(a)})}{\left[(B1^{(0)} + E1^{(0)}) - \frac{1}{p}(B2^{(0)} + E2^{(0)}) \right]^{p-1} (B2^{(0)} + E2^{(0)})} \right)^{\frac{M+N}{2}} .$$

Thus we arrive at the following theorem.

Theorem 3.1: The likelihood ratio test for testing $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$ is to reject H_0 if $L < C_\alpha$, where C_α is such that $P(L < C_\alpha \mid H_0) = \alpha$, and L is defined as:

$$L = \lambda^{2/(M+N)} = \frac{\left[(B1^{(a)} + E1^{(a)}) - \frac{1}{p}(B2^{(a)} + E2^{(a)}) \right]^{p-1} (B2^{(a)} + E2^{(a)})}{\left[(B1^{(0)} + E1^{(0)}) - \frac{1}{p}(B2^{(0)} + E2^{(0)}) \right]^{p-1} (B2^{(0)} + E2^{(0)})},$$

where λ is the likelihood ratio and $B1^{(a)}$, $B2^{(a)}$, $B1^{(0)}$, and $B2^{(0)}$ are defined in (3.5) and (3.7).

To show the null distribution of L, the following propositions are needed.

Proposition 3.1: Under $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y (= \boldsymbol{\mu}_0)$, $(B1^{(0)} + E1^{(0)}) - \frac{1}{p}(B2^{(0)} + E2^{(0)})$ is distributed as the quantity $\sigma^2(1-\rho)\chi^2_{(M+N-1)(p-1)}$.

Proof:

First rewrite $(B1^{(0)} + E1^{(0)}) - \frac{1}{p}(B2^{(0)} + E2^{(0)})$ as

$$(B1^{(0)} + E1^{(0)}) - \frac{1}{p}(B2^{(0)} + E2^{(0)}) = (B1^{(0)} - \frac{1}{p}B2^{(0)}) + (E1^{(0)} - \frac{1}{p}E2^{(0)})$$

$$= \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T (\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0) + \sum_{j=1}^{N} (\boldsymbol{Y}_j - \boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T (\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j - \boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0),$$

where $\hat{\boldsymbol{\mu}}_0 = \left(\sum_{i=1}^{M} \boldsymbol{C}_i^{\ T} \boldsymbol{C}_i + \sum_{j=1}^{N} \boldsymbol{D}_j^{\ T} \boldsymbol{D}_j \right)^{-1} \left(\sum_{i=1}^{M} \boldsymbol{C}_i^{\ T} \boldsymbol{X}_i + \sum_{j=1}^{N} \boldsymbol{D}_j^{\ T} \boldsymbol{Y}_j \right)$.

Appendix A.4 shows that

$$\begin{aligned}
&(B1^{(0)} + E1^{(0)}) - \frac{1}{p}(B2^{(0)} + E2^{(0)}) \\
&\quad = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}_0)^T (\boldsymbol{I}_p - \frac{1}{p} \boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}_0) \\
&\qquad + \sum_{j=1}^{N} (\boldsymbol{Y}_j - \boldsymbol{D}_j \boldsymbol{\mu}_0)^T (\boldsymbol{I}_p - \frac{1}{p} \boldsymbol{J}_p)(\boldsymbol{Y}_j - \boldsymbol{D}_j \boldsymbol{\mu}_0) \\
&\qquad - (\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0)^T \left[\left(\sum_{i=1}^{M} \boldsymbol{C}_i^{\ T} \boldsymbol{C}_i + \sum_{j=1}^{N} \boldsymbol{D}_j^{\ T} \boldsymbol{D}_j \right) (\boldsymbol{I}_p - \frac{1}{p} \boldsymbol{J}_p) \right] (\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0).
\end{aligned} \tag{3.8}$$

Under $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y (= \boldsymbol{\mu}_0)$ we have

$$\hat{\boldsymbol{\mu}}_0 \sim N(\boldsymbol{\mu}_0, \left(\sum_{i=1}^{M} \boldsymbol{C}_i^{\ T} \boldsymbol{C}_i + \sum_{j=1}^{N} \boldsymbol{D}_j^{\ T} \boldsymbol{D}_j \right)^{-1} \boldsymbol{\Sigma}) .$$

Thus the quadratic form

$$(\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0)^T \left[\left(\sum_{i=1}^{M} \boldsymbol{C}_i^{\ T} \boldsymbol{C}_i + \sum_{j=1}^{N} \boldsymbol{D}_j^{\ T} \boldsymbol{D}_j \right) (\boldsymbol{I}_p - \frac{1}{p} \boldsymbol{J}_p) \right] (\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0) \tag{3.9}$$

is distributed as the quantity $\sum_{j=1}^{p} \lambda_j \chi_1^2$, where λ_j 's are the latent roots of $\boldsymbol{P}_1$ defined in (2.14).

Using the results in the proof of Proposition 2.3, expression in (3.9) is distributed as a $\sigma^2(1-\rho)\chi_{p-1}^2$ random variable, and the random variable

$$\sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}_0)^T (\boldsymbol{I}_p - \frac{1}{p} \boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i \boldsymbol{\mu}_0) + \sum_{j=1}^{N} (\boldsymbol{Y}_j - \boldsymbol{D}_j \boldsymbol{\mu}_0)^T (\boldsymbol{I}_p - \frac{1}{p} \boldsymbol{J}_p)(\boldsymbol{Y}_j - \boldsymbol{D}_j \boldsymbol{\mu}_0)$$

are distributed as chi-square random variables with $(M+N)(p-1)$ degrees of freedom times a constant $\sigma^2(1-\rho)$. Hence, using the result of sum of independent chi-square random variables, it follows that $(B_1^{(0)}+E_1^{(0)})-\frac{1}{p}(B_2^{(0)}+E_2^{(0)})$ is distributed as the random variable

$\sigma^2(1-\rho)\chi^2_{(M+N-1)(p-1)}$. The proof is complete.

Proposition 3.2: Under $H_0:\boldsymbol{\mu}_x=\boldsymbol{\mu}_y(=\boldsymbol{\mu}_0)$, $B2^{(0)}+E2^{(0)}$ is distributed as the quantity

$p\sigma^2[1+(p-1)\rho]\chi^2_{M+N-1}$.

Proof:

$$\begin{aligned} B2^{(0)}+E2^{(0)} &= \sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)+\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{J}_p(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0) \\ &= \sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)+\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_0) \\ &\quad -(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T\left[\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i+\sum_{j=1}^{N}\boldsymbol{D}_j^T\boldsymbol{D}_j\right)\boldsymbol{J}_p\right](\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0) \qquad \text{Appendix A.4} \end{aligned}$$

Under $H_0:\boldsymbol{\mu}_x=\boldsymbol{\mu}_y(=\boldsymbol{\mu}_0)$, referring to the proof of Proposition 2.4,

$$\begin{aligned} &\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)+\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_0) \\ &\qquad \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2_{M+N} \end{aligned} \tag{3.10}$$

and

$$(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T\left[\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i+\sum_{j=1}^{N}\boldsymbol{D}_j^T\boldsymbol{D}_j\right)\boldsymbol{J}_p\right](\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)\overset{d}{=}p\sigma^2[1+(p-1)\rho]\chi^2_1, \tag{3.11}$$

implies that

$$B2^{(0)}+E2^{(0)}\overset{d}{=}p\sigma^2[1+(p-1)\rho]\chi^2_{M+N-1}$$

by using the result of sum of two independent chi-square random variables. The proof is complete.

<u>Proposition 3.3</u>: $(B1^{(a)}+E1^{(a)})-\frac{1}{p}(B2^{(a)}+E2^{(a)})$ is distributed as the quantity

$\sigma^2(1-\rho)\chi^2_{(M+N-2)(p-1)}$.

Proof:

Assume that $E(\hat{\boldsymbol{\mu}}_x)=\boldsymbol{\mu}_x$ and $E(\hat{\boldsymbol{\mu}}_y)=\boldsymbol{\mu}_y$. So we have

$$
\begin{aligned}
(B1^{(a)}+E1^{(a)})-\frac{1}{p}(B2^{(a)}+E2^{(a)}) &= (B1^{(a)}-\frac{1}{p}B2^{(a)})+(E1^{(a)}-\frac{1}{p}E2^{(a)}) \\
&= \sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x)+\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y) \\
&= \sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_x)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_x)-(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_x)^T\left[\left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\;T}\boldsymbol{C}_i\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)\right](\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_x) \\
&\quad+\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_y)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_y)-(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_y)^T\left[\left(\sum_{j=1}^{N}\boldsymbol{D}_j^{\;T}\boldsymbol{D}_j\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)\right](\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_y).
\end{aligned}
$$

Applying Proposition 2.3 we have that

$$B1^{(a)}-\frac{1}{p}B2^{(a)}\overset{d}{=}\sigma^2(1-\rho)\chi^2_{(M-1)(p-1)}$$

and

$$E1^{(a)}-\frac{1}{p}E2^{(a)}\overset{d}{=}\sigma^2(1-\rho)\chi^2_{(N-1)(p-1)}.$$

Since $B1^{(a)}-\frac{1}{p}B2^{(a)}$ and $E1^{(a)}-\frac{1}{p}E2^{(a)}$ are independent, we have

$$(B1^{(a)}+E1^{(a)})-\frac{1}{p}(B2^{(a)}+E2^{(a)})\overset{d}{=}\sigma^2(1-\rho)\chi^2_{(M+N-1)(p-1)}.$$

The proof is complete.

<u>Proposition 3.4</u>: $B2^{(a)}+E2^{(a)}$ is distributed as the random variable

$$p\sigma^2[1+(p-1)\rho]\chi^2_{M+N-2}.$$

Proof:

$$B2^{(a)} + E2^{(a)} = \sum_{i=1}^{M} (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_x)^T \boldsymbol{J}_p (\boldsymbol{X}_i - \boldsymbol{C}_i \hat{\boldsymbol{\mu}}_x) + \sum_{j=1}^{N} (\boldsymbol{Y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_y)^T \boldsymbol{J}_p (\boldsymbol{Y}_j - \boldsymbol{D}_j \hat{\boldsymbol{\mu}}_y)$$

Applying Proposition 2.4, $B2^{(a)} + E2^{(a)}$ is distributed as the sum of two independent random variables $p\sigma^2[1+(p-1)\rho]\chi^2_{M-1}$ and $p\sigma^2[1+(p-1)\rho]\chi^2_{N-1}$. Therefore

$$B2^{(a)} + E2^{(a)} \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2_{M+N-2}.$$

The proof is complete.

Now we arrive at the following theorem.

<u>Theorem 3.2</u>: The likelihood ratio test statistic in Theorem 3.1 for testing $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$ is L defined as:

$$L = \frac{\left[(B1^{(a)} + E1^{(a)}) - \frac{1}{p}(B2^{(a)} + E2^{(a)})\right]^{p-1} (B2^{(a)} + E2^{(a)})}{\left[(B1^{(0)} + E1^{(0)}) - \frac{1}{p}(B2^{(0)} + E2^{(0)})\right]^{p-1} (B2^{(0)} + E2^{(0)})} = \frac{B^{p-1}D}{A^{p-1}C},$$

where $A = (B1^{(0)} + E1^{(0)}) - \frac{1}{p}(B2^{(0)} + E2^{(0)})$, $B = (B1^{(a)} + E1^{(a)}) - \frac{1}{p}(B2^{(a)} + E2^{(a)})$, $C = B2^{(0)} + E2^{(0)}$, and $D = B2^{(a)} + E2^{(a)}$.

(a) B and D are distributed respectively as the following:

$$B \overset{d}{=} \sigma^2(1-\rho)\chi^2_{(M+N-2)(p-1)} \text{ and } D \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2_{M+N-2}.$$

Under $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$, A and C are distributed respectively as the following:

$$A \overset{d}{=} \sigma^2(1-\rho)\chi^2_{(M+N-1)(p-1)} \text{ and } C \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2_{M+N-1}.$$

(b) A-B, B, C-D, and D are mutually independent weighted chi-square random variables.

(c) Furthermore, under $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$, L is distributed as the random variable

$$\frac{1}{\left(1+\frac{1}{M+N-2}F^*\right)^{p-1}\left(1+\frac{1}{M+N-2}F^{**}\right)},$$

where F^* and F^{**} are independent and distributed like $F_{p-1,(M+N-2)(p-1)}$, and $F_{1,M+N-2}$, respectively.

Proof of (a): Results are obtained directly from Propositions 3.1 to 3.4.

Proof of (b) and (c):

First rewrite *A* and *C* as follows. *A* can be expressed as

$$\begin{aligned} A &= \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0) + \sum_{j=1}^{N}(\boldsymbol{Y}_j - \boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j - \boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0) \\ &= B + R, \end{aligned}$$

where B and R are, respectively,

$$B = \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x)^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x) + \sum_{j=1}^{N}(\boldsymbol{Y}_j - \boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y)^T(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j - \boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y)$$

and

$$\begin{aligned} R = {} & (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_0)^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^T\boldsymbol{C}_i\right)(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_0) \\ & + (\hat{\boldsymbol{\mu}}_y - \hat{\boldsymbol{\mu}}_0)^T\left(\sum_{j=1}^{N}\boldsymbol{D}_j^T\boldsymbol{D}_j\right)(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_y - \hat{\boldsymbol{\mu}}_0). \end{aligned} \tag{3.12}$$

Similarly, *C* can be expressed as

$$\begin{aligned} C &= \sum_{i=1}^{M}(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{J}_p(\boldsymbol{X}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0) + \sum_{j=1}^{N}(\boldsymbol{Y}_j - \boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{J}_p(\boldsymbol{Y}_j - \boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0) \\ &= D + S, \end{aligned}$$

where

$$D=\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x)^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x)+\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y)^T\boldsymbol{J}_p(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y)$$

and

$$\begin{aligned} S &= (\hat{\boldsymbol{\mu}}_x-\hat{\boldsymbol{\mu}}_0)^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i\right)\boldsymbol{J}_p(\hat{\boldsymbol{\mu}}_x-\hat{\boldsymbol{\mu}}_0) \\ &\quad+(\hat{\boldsymbol{\mu}}_y-\hat{\boldsymbol{\mu}}_0)^T\left(\sum_{j=1}^{N}\boldsymbol{D}_j^{\,T}\boldsymbol{D}_j\right)\boldsymbol{J}_p(\hat{\boldsymbol{\mu}}_y-\hat{\boldsymbol{\mu}}_0). \end{aligned} \tag{3.13}$$

Some other facts necessary to prove (b) are stated below.

(1) B and R are independent

(2) D and S are independent

(3) $R\overset{d}{=}\sigma^2(1-\rho)\chi^2_{p-1}$ and $S\overset{d}{=}p\sigma^2[1+(p-1)\rho]\chi^2_1$.

(4) B and D are independent

(5) B and S are independent

(6) R and D are independent

(7) R and S are independent

Facts (1) and (2) are true because both B and D are functions of $\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x$ and $\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y$ for all $i=1,\dots,M$, $j=1,\dots,N$, also R and S are functions of $\hat{\boldsymbol{\mu}}_x$ and $\hat{\boldsymbol{\mu}}_y$ since $\hat{\boldsymbol{\mu}}_0$ in (3.2) can be expressed as a linear combination of $\hat{\boldsymbol{\mu}}_x$ and $\hat{\boldsymbol{\mu}}_y$ as follows:

$$\hat{\boldsymbol{\mu}}_0=(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}[\boldsymbol{C}^*\hat{\boldsymbol{\mu}}_x+\boldsymbol{D}^*\hat{\boldsymbol{\mu}}_y], \tag{3.14}$$

where $\boldsymbol{C}^*=\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i$, $\boldsymbol{D}^*=\sum_{j=1}^{N}\boldsymbol{D}_j^{\,T}\boldsymbol{D}_j$, $\hat{\boldsymbol{\mu}}_x=(\boldsymbol{C}^*)^{-1}\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{X}_i$ and $\hat{\boldsymbol{\mu}}_y=(\boldsymbol{D}^*)^{-1}\sum_{j=1}^{N}\boldsymbol{D}_j^{\,T}\boldsymbol{Y}_j$.

Combining the facts that $\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x$ and $\hat{\boldsymbol{\mu}}_x$ are independent as well as $\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y$ and $\hat{\boldsymbol{\mu}}_y$ are independent, Facts (1) and (2) are shown.

Fact (3) can be shown using the results in part (a) in conjunction with Facts (1) and (2), and the result about sum of independent chi-square random variables. More clearly, the results

$$A \stackrel{d}{=} \sigma^2(1-\rho)\chi^2_{(M+N-1)(p-1)} \text{ and } B \stackrel{d}{=} \sigma^2(1-\rho)\chi^2_{(M+N-2)(p-1)}$$

combined with Fact (1) imply $R \stackrel{d}{=} \sigma^2(1-\rho)\chi^2_{p-1}$. In addition, the results

$$C \stackrel{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2_{M+N-1} \text{ and } D \stackrel{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2_{M+N-2}$$

in connection with fact (2) implies $S \stackrel{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2_1$.

Fact (4) can be shown by applying Proposition 2.5. $(B1^{(a)} - \frac{1}{p}B2^{(a)})$ and $B2^{(a)}$ are independent, $(E1^{(a)} - \frac{1}{p}E2^{(a)})$ and $E2^{(a)}$ are independent as well. As a matter of fact, $(B1^{(a)} - \frac{1}{p}B2^{(a)})$, $B2^{(a)}$, $(E1^{(a)} - \frac{1}{p}E2^{(a)})$ and $E2^{(a)}$ are mutually independent so fact (4) is shown.

Facts (5) and (6) are true using the same argument when Facts (1) and (2) were shown.

To show Fact (7), it is necessary to rewrite *R* and *S* in (3.12) and (3.13), respectively. In (3.12) the two terms on the right-hand side can be expressed respectively as

$$\begin{aligned}
&(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_0)^T \boldsymbol{C}^*(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_0) \\
&= (\hat{\boldsymbol{\mu}}_x - \boldsymbol{\mu}_0)^T \boldsymbol{C}^*(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{\mu}_x - \hat{\boldsymbol{\mu}}_0) - (\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0)^T \boldsymbol{C}^*(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0) \\
&- 2(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_0)^T \boldsymbol{C}^*(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0),
\end{aligned}$$

and

$$(\hat{\boldsymbol{\mu}}_y-\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{D}^*(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_y-\hat{\boldsymbol{\mu}}_0)=(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0)^T\boldsymbol{D}^*(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{\mu}_y-\hat{\boldsymbol{\mu}}_0)$$
$$-(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T\boldsymbol{D}^*(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)-2(\hat{\boldsymbol{\mu}}_y-\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{D}^*(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0).$$

We should note that

$$(\hat{\boldsymbol{\mu}}_x-\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{C}^*(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)+(\hat{\boldsymbol{\mu}}_y-\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{D}^*(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)=0$$

by substituting (3.14) into the left-hand side of the above equation. Therefore, R becomes

$$\begin{aligned}R&=(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)^T[\boldsymbol{C}^*(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)](\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)\\&+(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0)^T[\boldsymbol{D}^*(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)](\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0)\\&-(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T[(\boldsymbol{C}^*+\boldsymbol{D}^*)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)](\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0).\end{aligned}\tag{3.15}$$

Likewise, S can be written as

$$\begin{aligned}S&=(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)^T[\boldsymbol{C}^*\boldsymbol{J}_p](\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)+(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0)^T[\boldsymbol{D}^*\boldsymbol{J}_p](\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0)\\&-(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T[(\boldsymbol{C}^*+\boldsymbol{D}^*)\boldsymbol{J}_p](\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0).\end{aligned}\tag{3.16}$$

Since $\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0$ can be written as

$$\begin{aligned}\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0&=(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}(\boldsymbol{C}^*\hat{\boldsymbol{\mu}}_x+\boldsymbol{D}^*\hat{\boldsymbol{\mu}}_y)-(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}(\boldsymbol{C}^*+\boldsymbol{D}^*)\boldsymbol{\mu}_0\\&=(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}[\boldsymbol{C}^*(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)+\boldsymbol{D}^*(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0)],\end{aligned}$$

the last term of (3.15) becomes

$$(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)^T\boldsymbol{C}^*(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)\boldsymbol{C}^*(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)+$$
$$(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0)^T\boldsymbol{D}^*(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)\boldsymbol{D}^*(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0)-$$
$$2(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)^T\boldsymbol{C}^*(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)\boldsymbol{D}^*(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0).$$

Hence R can be expressed as

$$R=(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)^T[\boldsymbol{C}^*-\boldsymbol{C}^{*2}(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}](\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)$$
$$+(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0)^T[\boldsymbol{D}^*-\boldsymbol{D}^{*2}(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}](\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0) \quad (3.17)$$
$$-2(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)^T[\boldsymbol{C}^*\boldsymbol{D}^*(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}](\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0),$$

Likewise, S can be expressed as

$$S=(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)^T[\boldsymbol{C}^*-\boldsymbol{C}^{*2}(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}]\boldsymbol{J}_p(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)$$
$$+(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0)^T[\boldsymbol{D}^*-\boldsymbol{D}^{*2}(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}]\boldsymbol{J}_p(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0) \quad (3.18)$$
$$-2(\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0)^T[\boldsymbol{C}^*\boldsymbol{D}^*(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}]\boldsymbol{J}_p(\hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0).$$

Now rewrite R and S in (3.17) and (3.18) respectively in matrix forms as

$$R=\begin{pmatrix}\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0\\ \hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0\end{pmatrix}^T\begin{pmatrix}\boldsymbol{\Psi}_1 & -\boldsymbol{\Psi}_3\\ -\boldsymbol{\Psi}_3 & \boldsymbol{\Psi}_2\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0\\ \hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0\end{pmatrix}$$

and

$$S=\begin{pmatrix}\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0\\ \hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0\end{pmatrix}^T\begin{pmatrix}\boldsymbol{\Psi}_4 & -\boldsymbol{\Psi}_6\\ -\boldsymbol{\Psi}_6 & \boldsymbol{\Psi}_5\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0\\ \hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0\end{pmatrix},$$

where

$$\boldsymbol{\Psi}_1=[\boldsymbol{C}^*-\boldsymbol{C}^{*2}(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}](\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p),\boldsymbol{\Psi}_2=[\boldsymbol{D}^*-\boldsymbol{D}^{*2}(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}](\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p),$$

$$\boldsymbol{\Psi}_3=[\boldsymbol{C}^*\boldsymbol{D}^*(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}](\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p),\ \boldsymbol{\Psi}_4=[\boldsymbol{C}^*-\boldsymbol{C}^{*2}(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}]\boldsymbol{J}_p,$$

$$\boldsymbol{\Psi}_5=[\boldsymbol{D}^*-\boldsymbol{D}^{*2}(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}]\boldsymbol{J}_p,\ \text{and}\ \boldsymbol{\Psi}_6=[\boldsymbol{C}^*\boldsymbol{D}^*(\boldsymbol{C}^*+\boldsymbol{D}^*)^{-1}]\boldsymbol{J}_p.$$

Since $\begin{pmatrix}\hat{\boldsymbol{\mu}}_x-\boldsymbol{\mu}_0\\ \hat{\boldsymbol{\mu}}_y-\boldsymbol{\mu}_0\end{pmatrix}$ are distributed as $N_{2p}(\boldsymbol{0},\begin{pmatrix}\boldsymbol{C}^*\boldsymbol{\Sigma} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{D}^*\boldsymbol{\Sigma}\end{pmatrix})$, if we can show

$$\begin{pmatrix}\boldsymbol{\Psi}_1 & -\boldsymbol{\Psi}_3\\ -\boldsymbol{\Psi}_3 & \boldsymbol{\Psi}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{C}^*\boldsymbol{\Sigma} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{D}^*\boldsymbol{\Sigma}\end{pmatrix}\begin{pmatrix}\boldsymbol{\Psi}_4 & -\boldsymbol{\Psi}_6\\ -\boldsymbol{\Psi}_6 & \boldsymbol{\Psi}_5\end{pmatrix}=0, \quad (3.19)$$

then the proof of fact (7) is done. Expression (3.19) is true because of the facts

$(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)\boldsymbol{J}_p = \boldsymbol{0}$ and commutability of circulant matrices. Thus *R* and *S* are independent.

Therefore,

$$L = \frac{B^{p-1}D}{A^{p-1}C} = \frac{B^{p-1}D}{(B+R)^{p-1}(D+S)} = \frac{1}{\left(1+\frac{R}{B}\right)^{p-1}\left(1+\frac{S}{D}\right)}$$

$$\stackrel{d}{=} \frac{1}{\left(1+\frac{1}{M+N-2}F^{*}\right)^{p-1}\left(1+\frac{1}{M+N-2}F^{**}\right)},$$

where F^{*} and F^{**} are independent and distributed like $F_{p-1,(M+N-2)(p-1)}$, and $F_{1,M+N-2}$, respectively. The proof of Theorem 3.2 is complete.

3.3 APPROXIMATE χ^2 TEST FOR $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$

Referring to the beginning of Section 3.2, we have the quadratic form

$$T_0 = (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \left(\left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}_x^{-1} \boldsymbol{C}_i \right)^{-1} + \left(\sum_{j=1}^{N} \boldsymbol{D}_j^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{D}_j \right)^{-1} \right)^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y).$$

Under the hypothesis $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$,

$$T_0 = (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \left(\left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{\Sigma}_x^{-1} \boldsymbol{C}_i \right)^{-1} + \left(\sum_{j=1}^{N} \boldsymbol{D}_j^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{D}_j \right)^{-1} \right)^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y) \sim \chi_p^2,$$

where $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$ are known. What if both $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$ are unknown? In this section, the assumption $\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}$ is assumed, where $\boldsymbol{\Sigma}$ is unknown and with compound symmetry structure for testing $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$ using approximate χ^2 test.

First note that

$$\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y \sim N_p(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y, \left[\left(\sum_{i=1}^{M} \boldsymbol{C}_i^{\ T} \boldsymbol{C}_i\right)^{-1} + \left(\sum_{j=1}^{N} \boldsymbol{D}_j^{\ T} \boldsymbol{D}_j\right)^{-1}\right] \boldsymbol{\Sigma}).$$

The test statistic for testing $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$ is defined as

$$\begin{aligned} T_1 &= (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \left[\hat{Var}(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)\right]^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y) \\ &= (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \hat{\boldsymbol{\Sigma}}^{-1} \left[\left(\sum_{i=1}^{M} \boldsymbol{C}_i^{\ T} \boldsymbol{C}_i\right)^{-1} + \left(\sum_{j=1}^{N} \boldsymbol{D}_j^{\ T} \boldsymbol{D}_j\right)^{-1}\right]^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y), \end{aligned} \tag{3.20}$$

where $\hat{\boldsymbol{\Sigma}}^{-1}$ is

$$\hat{\boldsymbol{\Sigma}}^{-1} = \frac{1}{\hat{\sigma}^2(1-\hat{\rho})}[\boldsymbol{I}_p - \frac{\hat{\rho}}{1+(p-1)\hat{\rho}} \boldsymbol{J}_p],$$

where $\hat{\sigma}^2$ and $\hat{\rho}$ are defined in (3.6).

***Theorem 3.3*:** Under $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$,

$$T_1 = (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \left[\hat{Var}(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)\right]^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y) \xrightarrow{d} \chi_p^2 .$$

Proof:

Recall that

$$\left[\hat{Var}(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)\right]^{-1} = \hat{\boldsymbol{\Sigma}}^{-1} \left[\left(\sum_{i=1}^{M} \boldsymbol{C}_i^{\ T} \boldsymbol{C}_i\right)^{-1} + \left(\sum_{j=1}^{N} \boldsymbol{D}_j^{\ T} \boldsymbol{D}_j\right)^{-1}\right]^{-1}.$$

Extending the result from (2.35) we have

$$\begin{aligned} &Cov(B1^{(a)} + E1^{(a)}, B2^{(a)} + E2^{(a)}) = Cov(B1^{(a)}, B2^{(a)}) + Cov(E1^{(a)}, E2^{(a)}) \\ &= 2(M+N-2) tr(\boldsymbol{J}_p \boldsymbol{\Sigma}^2) = 2(M+N-2) p[1+(p-1)\rho]^2 \sigma^4. \end{aligned} \tag{3.21}$$

From the result of Theorem 2.3 and the fact that $B1^{(a)}$ and $E1^{(a)}$ defined in (3.7) are independent, we have

$$B1^{(a)} + E1^{(a)} \overset{d}{=} \sigma^2(1-\rho)\chi^2_{(M+N-2)(p-1)} + \sigma^2[1+(p-1)\rho]\chi^2_{M+N-2} \ , \tag{3.22}$$

which, after doing some calculation, implies that

$$E(B1^{(a)} + E1^{(a)}) = (M+N-2)p\sigma^2,$$
$$V(B1^{(a)} + E1^{(a)}) = 2(M+N-2)p\left[1+(p-1)\rho^2\right]\sigma^4 . \tag{3.23}$$

Similarly, from the result of Proposition 2.4 and the fact that $B2^{(a)}$ and $E2^{(a)}$ defined in (3.7) are independent, we have

$$B2^{(a)} + E2^{(a)} \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2_{M+N-2} \ , \tag{3.24}$$

which implies that

$$E(B2^{(a)} + E2^{(a)}) = (M+N-2)p[1+(p-1)\rho]\sigma^2,$$
$$V(B2^{(a)} + E2^{(a)}) = 2(M+N-2)p^2\left[1+(p-1)\rho\right]^2\sigma^4 . \tag{3.25}$$

Using the results from (2.36), (2.37), (3.22), and (3.24) we have

$$E\left(\frac{B2^{(a)} + E2^{(a)}}{B1^{(a)} + E1^{(a)}}\right) \approx \frac{E(B2^{(a)} + E2^{(a)})}{E(B1^{(a)} + E1^{(a)})} = 1+(p-1)\rho, \tag{3.26}$$

and

$$V\left(\frac{B2^{(a)} + E2^{(a)}}{B1^{(a)} + E1^{(a)}}\right) \approx \frac{2}{M+N-2} \cdot \frac{[1+(p-1)\rho]^2(p-1)(1-\rho)^2}{p} \ . \tag{3.27}$$

Therefore, $\hat{\rho} \rightarrow \rho$ in probability. In addition, $\hat{\sigma}^2 = \frac{1}{(M+N)p}\left(B1^{(a)} + E1^{(a)}\right)$ converges in probability to σ^2 due to the facts from (3.23). Hence using the decompositions

$$\boldsymbol{I}_p - \frac{\hat{\rho}}{1+(p-1)\hat{\rho}}\boldsymbol{J}_p = (\boldsymbol{I}_p - \frac{\rho}{1+(p-1)\rho}\boldsymbol{J}_p) + (\frac{\rho}{1+(p-1)\rho} - \frac{\hat{\rho}}{1+(p-1)\hat{\rho}})\boldsymbol{J}_p$$

and

$$\frac{1}{\hat{\sigma}^2(1-\hat{\rho})} = \frac{\sigma^2(1-\rho)}{\hat{\sigma}^2(1-\hat{\rho})}\frac{1}{\sigma^2(1-\rho)} \ ,$$

and mimicking the proof in Theorem 2.5, we have under $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$ that

$$(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \left[\hat{Var}(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y) \right]^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)$$

$$= (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \frac{1}{\hat{\sigma}^2(1-\hat{\rho})} \left(\boldsymbol{I}_p - \frac{\hat{\rho}}{1+(p-1)\hat{\rho}} \boldsymbol{J}_p \right) (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)$$

$$= \frac{(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \frac{1}{\sigma^2(1-\rho)} \left(\boldsymbol{I}_p - \frac{\rho}{1+(p-1)\rho} \boldsymbol{J}_p \right) \left[\left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \right)^{-1} + \left(\sum_{j=1}^{N} \boldsymbol{D}_j^T \boldsymbol{D}_j \right)^{-1} \right]^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)}{\frac{\hat{\sigma}^2(1-\hat{\rho})}{\sigma^2(1-\rho)}} +$$

$$(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \frac{1}{\hat{\sigma}^2(1-\hat{\rho})} \left(\frac{\hat{\rho}}{1+(p-1)\hat{\rho}} - \frac{\rho}{1+(p-1)\rho} \right) \boldsymbol{J}_p \left[\left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \right)^{-1} + \left(\sum_{j=1}^{N} \boldsymbol{D}_j^T \boldsymbol{D}_j \right)^{-1} \right]^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y).$$

Since $\frac{\hat{\sigma}^2(1-\hat{\rho})}{\sigma^2(1-\rho)} \xrightarrow{p} 1$ and $\frac{\hat{\rho}}{1+(p-1)\hat{\rho}} - \frac{\rho}{1+(p-1)\rho} \xrightarrow{p} 0$, we have by Slutsky's

Theorem and the fact $(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)^T \left[Var(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y) \right]^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y) \overset{d}{=} \chi_p^2$ that

$$(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)' [\hat{Var}(\hat{\boldsymbol{\mu}})]^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y) \xrightarrow{d}$$

$$(\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y)' \boldsymbol{\Sigma}^{-1} \left[\left(\sum_{i=1}^{M} \boldsymbol{C}_i^T \boldsymbol{C}_i \right)^{-1} + \left(\sum_{j=1}^{N} \boldsymbol{D}_j^T \boldsymbol{D}_j \right)^{-1} \right]^{-1} (\hat{\boldsymbol{\mu}}_x - \hat{\boldsymbol{\mu}}_y),$$

which follows a χ_p^2 distribution under $H_0 : \boldsymbol{\mu}_x = \boldsymbol{\mu}_y$. The proof is complete.

3.4 LRT FOR *k*-SAMPLE CASE

Consider *k* independent samples each with sample size N_i from *p*-variate multivariate normal distributions with heterogeneous mean vectors $\boldsymbol{C}_{ij}\boldsymbol{\mu}_i$, where $j = 1,..., N_i$, $i = 1,..., k$. The homoscedastic case is considered in this section such that all covariance matrices of the *k* populations are the same. For the *i*th sample, we have $\boldsymbol{X}_{i1},..., \boldsymbol{X}_{iN_i} \sim MVN_p(\boldsymbol{C}_{ij}\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$. All $\boldsymbol{C}_{ij}$ are known $p \times p$ circulant matrices, $\boldsymbol{\Sigma}$ has compound symmetry structure defined in (2.8) such that $\boldsymbol{C}_{ij}$ and $\boldsymbol{\Sigma}$ commute for all *i* and *j*, and $\sum_{i=1}^{k} N_i = N$. The hypotheses of our interest are $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k$ versus $H_a : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$ for some $i \neq j$. The likelihood function is

$$L(\boldsymbol{\mu}_1,...,\boldsymbol{\mu}_k,\boldsymbol{\Sigma}) = constant \cdot |\boldsymbol{\Sigma}|^{-\frac{N}{2}} \cdot \exp\left\{-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_i)\right\}.$$

The corresponding log likelihood function is

$$\log L(\boldsymbol{\mu}_1,...,\boldsymbol{\mu}_k,\boldsymbol{\Sigma}) = constant - \frac{N}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_i). \quad (3.28)$$

We skip the trivial case that $\boldsymbol{\Sigma}$ is known. Before deriving the likelihood ratio test for $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k (= \boldsymbol{\mu}_0)$, it is necessary to find the MLEs of the parameters $\boldsymbol{\mu}_1,...,\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}$ under the null and alternative hypotheses separately.

3.4.1 Estimation Under $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k$

Assume that $\boldsymbol{\mu}_i = \boldsymbol{\mu}_0$ for all $i = 1,...,k$ under H_0, the MLE of $\boldsymbol{\mu}_0$, say $\hat{\boldsymbol{\mu}}_0$, can be derived as

$$\hat{\boldsymbol{\mu}}_0 = \left(\sum_{i=1}^{k}\sum_{j=1}^{N_i}\boldsymbol{C}_{ij}^{\ T}\boldsymbol{C}_{ij}\right)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{N_i}\boldsymbol{C}_{ij}^{\ T}\boldsymbol{X}_{ij} . \quad (3.29)$$

Hence, the MLE for $\boldsymbol{\Sigma}$ under H_0, namely $\hat{\boldsymbol{\Sigma}}_0$, can be obtained using the reduced log likelihood function

$$L(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}) = constant - \frac{N}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_0), \quad (3.30)$$

which yields

$$\hat{\boldsymbol{\Sigma}}_0 = \hat{\sigma}_0^{\ 2}[(1-\hat{\rho}_0)\boldsymbol{I}_p + \hat{\rho}_0\boldsymbol{J}_p],$$

where

$$\hat{\sigma}_0^{\ 2} = \frac{1}{Np}\sum_{i=1}^{k} B_{i\boldsymbol{I}}^{(0)}\ ,\ \ \hat{\rho}_0 = \frac{1}{p-1}\left(\frac{\sum_{i=1}^{k} B_{i\boldsymbol{J}}^{(0)}}{\sum_{i=1}^{k} B_{i\boldsymbol{I}}^{(0)}} - 1\right), \tag{3.31}$$

where

$$B_{i\boldsymbol{I}}^{(0)} = \sum_{j=1}^{N_i} (\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0)^T (\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0)\,,$$

$$B_{i\boldsymbol{J}}^{(0)} = \sum_{j=1}^{N_i} (\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0)^T \boldsymbol{J} (\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0)\,, \tag{3.32}$$

by extending the results of two-sample case in (3.4) and (3.5).

3.4.2 Estimation Under $H_a : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$ for some $i \neq j$

Consider the case that the $\boldsymbol{\mu}_i$'s are distinct, the log likelihood function is

$$\begin{aligned}\log L(\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma}) &= constant - \frac{N}{2}\log|\boldsymbol{\Sigma}| \\ &\quad - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{N_i} (\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_i).\end{aligned} \tag{3.33}$$

Hence the MLEs for $\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}$ are, respectively,

$$\hat{\boldsymbol{\mu}}_i = (\sum_{j=1}^{N_i} \boldsymbol{C}_{ij}{}^T \boldsymbol{C}_{ij})^{-1} \sum_{j=1}^{N_i} \boldsymbol{C}_{ij}{}^T \boldsymbol{X}_{ij}\ , \text{ for } i = 1, ..., k\,, \tag{3.34}$$

and

$$\hat{\boldsymbol{\Sigma}} = \hat{\sigma}^2[(1-\hat{\rho})\boldsymbol{I}_p + \hat{\rho}\boldsymbol{J}_p]\,,$$

where

$$\hat{\sigma}^2 = \frac{1}{Np}\sum_{i=1}^{k} B_{i\boldsymbol{I}}^{(a)}\ ,\ \ \hat{\rho} = \frac{1}{p-1}\left(\frac{\sum_{i=1}^{k} B_{i\boldsymbol{J}}^{(a)}}{\sum_{i=1}^{k} B_{i\boldsymbol{I}}^{(a)}} - 1\right), \tag{3.35}$$

where

$$B_{i\boldsymbol{I}}^{(a)} = \sum_{j=1}^{N_i} (\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_i)^T (\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_i) \ ,$$

$$B_{i\boldsymbol{J}}^{(a)} = \sum_{j=1}^{N_i} (\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_i)^T \boldsymbol{J} (\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_i) \,. \tag{3.36}$$

3.4.3 Likelihood Ratio Test for Testing $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k$

Sections 3.4.1 and 3.4.2 derived the MLE's for parameters under both hypotheses. The likelihood ratio test now can be developed. The likelihood ratio is

$$\lambda = \frac{\max_{\boldsymbol{\theta} \in \Omega_0} L(\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\theta} \in \Omega} L(\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma})} = \frac{(2\pi)^{-\frac{Np}{2}} |\hat{\boldsymbol{\Sigma}}_0|^{-\frac{N}{2}}}{(2\pi)^{-\frac{Np}{2}} |\hat{\boldsymbol{\Sigma}}|^{-\frac{N}{2}}} \times \frac{\exp -\frac{1}{2}\left(\sum_{i=1}^{k} \sum_{j=1}^{N_i} (\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0)^T \hat{\boldsymbol{\Sigma}}_0^{-1} (\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0)\right)}{\exp -\frac{1}{2}\left(\sum_{i=1}^{k} \sum_{j=1}^{N_i} (\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_i)^T \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_i)\right)},$$

where $\boldsymbol{\theta} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma})$,

$$\Omega = \{(\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma}) \mid \boldsymbol{\Sigma} = \sigma^2[(1-\rho)\boldsymbol{I}_p + \rho \boldsymbol{J}_p]\}, \text{ and}$$

$$\Omega_0 = \{(\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma}) \mid \boldsymbol{\mu}_i = \boldsymbol{\mu}_j \ , \forall i \neq j; \boldsymbol{\Sigma} = \sigma^2[(1-\rho)\boldsymbol{I}_p + \rho \boldsymbol{J}_p]\}.$$

Hence the results (by mimicking Appendix A.3)

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0)^T \hat{\boldsymbol{\Sigma}}_0^{-1} (\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0) = Np$$

and

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_i)^T \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{x}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_i) = Np$$

imply that

$$\lambda = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{\frac{N}{2}} = \left(\frac{(\hat{\sigma}^2)^p (1-\hat{\rho})^{p-1}[1+(p-1)\hat{\rho}]}{(\hat{\sigma}_0{}^2)^p (1-\hat{\rho}_0)^{p-1}[1+(p-1)\hat{\rho}_0]}\right)^{N/2}$$
$$= \left(\frac{\{\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(a)} - (1/p)B_{i\boldsymbol{J}}^{(a)}]\}^{p-1}(\sum_{i=1}^{k} B_{i\boldsymbol{J}}^{(a)})}{\{\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(0)} - (1/p)B_{i\boldsymbol{J}}^{(0)}]\}^{p-1}(\sum_{i=1}^{k} B_{i\boldsymbol{J}}^{(0)})}\right)^{N/2} .$$

Thus we arrive at the following theorem.

Theorem 3.4: The likelihood ratio test for testing $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k$ is to reject H_0 if $L \le C_\alpha$, where C_α is such that $P(L \le C_\alpha) = \alpha$, and L is defined as:

$$L = \lambda^{2/N} = \frac{\{\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(a)} - (1/p)B_{i\boldsymbol{J}}^{(a)}]\}^{p-1}(\sum_{i=1}^{k} B_{i\boldsymbol{J}}^{(a)})}{\{\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(0)} - (1/p)B_{i\boldsymbol{J}}^{(0)}]\}^{p-1}(\sum_{i=1}^{k} B_{i\boldsymbol{J}}^{(0)})},$$

where λ is the likelihood ratio and $B_{i\boldsymbol{I}}^{(a)}$, $B_{i\boldsymbol{J}}^{(a)}$, $B_{i\boldsymbol{I}}^{(0)}$, and $B_{i\boldsymbol{J}}^{(0)}$ are defined in (3.32) and (3.36).

Before deriving the exact null distribution of the LRT statistic L, we need the following propositions.

Proposition 3.5: Under $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k$, $\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(0)} - (1/p)B_{i\boldsymbol{J}}^{(0)}]$ is distributed as the random variable $\sigma^2(1-\rho)\chi^2_{(N-1)(p-1)}$.

Proof:

By extending the result in (3.8) for two-sample case we obtain

$$\begin{aligned}\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(0)} - (1/p)B_{i\boldsymbol{J}}^{(0)}] &= \sum_{i=1}^{k}\sum_{j=1}^{N_i}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{I} - \frac{1}{p}\boldsymbol{J})(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{N_i}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_0)^T(\boldsymbol{I} - \frac{1}{p}\boldsymbol{J})(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_0) \\ &\quad - (\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0)^T(\sum_{i=1}^{k}\boldsymbol{C}_i^*)(\boldsymbol{I} - \frac{1}{p}\boldsymbol{J})(\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0),\end{aligned} \tag{3.37}$$

where $\boldsymbol{C}_i^* = \sum_{j=1}^{N_i}\boldsymbol{C}_{ij}{}^T\boldsymbol{C}_{ij}$. Under $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k (= \boldsymbol{\mu}_0)$ we have

$$\hat{\boldsymbol{\mu}}_0 \sim N(\boldsymbol{\mu}_0, (\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}\boldsymbol{\Sigma}).$$

Hence the quadratic form

$$(\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0)^T\left[(\sum_{i=1}^{k}\boldsymbol{C}_i^*)(\boldsymbol{I} - \frac{1}{p}\boldsymbol{J})\right](\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0) \tag{3.38}$$

is distributed as sum of weighted chi-square random variable $\sum_{j=1}^{p}\lambda_j\chi_1^2$, where λ_j's are 0 with multiplicity 1 and $\sigma^2(1-\rho)$ with multiplicity $p-1$, the latent roots of the matrix $\boldsymbol{P}_1$ as defined in (2.14), using the results in the proof of Proposition 2.1. Hence (3.38) is distributed as a χ_{p-1}^2 random variable times a constant $\sigma^2(1-\rho)$.

Since each pair of $\boldsymbol{X}_{ij}-\boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\mu}}_0$ are independent for all *i* and *j*, the quantities $\sum_{j=1}^{N_i}(\boldsymbol{X}_{ij}-\boldsymbol{C}_{ij}\boldsymbol{\mu}_0)^T(\boldsymbol{I}-\frac{1}{p}\boldsymbol{J})(\boldsymbol{X}_{ij}-\boldsymbol{C}_{ij}\boldsymbol{\mu}_0)$, $i=1,...,k$, are independently distributed as chi-square random variables with $N_i(p-1)$ degrees of freedom times a constant $\sigma^2(1-\rho)$ respectively, so the sum of them are distributed as $\sigma^2(1-\rho)\chi^2_{(\sum_{i=1}^k N_i)(p-1)}$ random variable. It follows that $\sum_{i=1}^k[B_{i\boldsymbol{I}}^{(0)}-(1/p)B_{i\boldsymbol{J}}^{(0)}]$ is distributed as a $\sigma^2(1-\rho)\chi^2_{(N-1)(p-1)}$ random variable, where $\sum_{i=1}^k N_i=N$, by the result of the sum of two independent chi-square random variables. The proof is complete.

<u>Proposition 3.6</u>: Under $H_0:\boldsymbol{\mu}_1=...=\boldsymbol{\mu}_k$, $\sum_{i=1}^k B_{i\boldsymbol{J}}^{(0)}$ is distributed as the quantity $p\sigma^2[1+(p-1)\rho]\chi^2_{N-1}$.

Proof:

Rewrite $\sum_{i=1}^k B_{i\boldsymbol{J}}^{(0)}$ as

$$\begin{aligned}\sum_{i=1}^k B_{i\boldsymbol{J}}^{(0)} &= \sum_{i=1}^k\sum_{j=1}^{N_i}(\boldsymbol{X}_{ij}-\boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{J}(\boldsymbol{X}_{ij}-\boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0)\\ &= \sum_{i=1}^k\sum_{j=1}^{N_i}(\boldsymbol{X}_{ij}-\boldsymbol{C}_{ij}\boldsymbol{\mu}_0)^T\boldsymbol{J}(\boldsymbol{X}_{ij}-\boldsymbol{C}_{ij}\boldsymbol{\mu}_0)\\ &\quad -(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T(\sum_{i=1}^k\boldsymbol{C}_i^*)\boldsymbol{J}(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0).\end{aligned}$$

Under $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k (= \boldsymbol{\mu}_0)$, extending the results of two-sample case in (3.10) and (3.11) to *k*-sample case we have

$$\sum_{i=1}^{k}\sum_{j=1}^{N_i}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_0)^T \boldsymbol{J}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_0) \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi_N^2$$

and

$$(\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0)^T[(\sum_{i=1}^{k}\boldsymbol{C}_i^*)\boldsymbol{J}_p](\hat{\boldsymbol{\mu}}_0 - \boldsymbol{\mu}_0) \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi_1^2 .$$

Connecting the facts that each pair of $\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\mu}}_0$ are independent for all *i* and *j*, and the result of sum of two independent chi-square random variables, we have

$$\sum_{i=1}^{k} B_{i\boldsymbol{J}}^{(0)} \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi_{N-1}^2 .$$

The proof is complete.

Proposition 3.7: $\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(a)} - (1/p)B_{i\boldsymbol{J}}^{(a)}]$ is distributed as the quantity $\sigma^2(1-\rho)\chi_{(N-k)(p-1)}^2$.

Proof:

Assume that $E(\hat{\boldsymbol{\mu}}_i) = \boldsymbol{\mu}_i$ for all $i = 1,...,k$. So we have

$$\begin{aligned}\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(a)} - (1/p)B_{i\boldsymbol{J}}^{(a)}] &= \sum_{i=1}^{k}\sum_{j=1}^{N_i}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_i)^T(\boldsymbol{I} - \frac{1}{p}\boldsymbol{J})(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_i)\\ &= \sum_{i=1}^{k}\sum_{j=1}^{N_i}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_i)^T(\boldsymbol{I} - \frac{1}{p}\boldsymbol{J})(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}_i)\\ &\quad - \sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i)^T(\boldsymbol{C}_i^*)(\boldsymbol{I} - \frac{1}{p}\boldsymbol{J})(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i),\end{aligned}$$

Hence, under $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k (= \boldsymbol{\mu}_0)$, we have that $B_{i\boldsymbol{I}}^{(a)} - (1/p)B_{i\boldsymbol{J}}^{(a)}$ are independently distributed and

$$B_{i\boldsymbol{I}}^{(a)} - (1/p)B_{i\boldsymbol{J}}^{(a)} \overset{d}{=} \sigma^2(1-\rho)\chi_{(N_i-1)(p-1)}^2 \text{ for all } i = 1,...,k ,$$

by applying Proposition 2.3. Therefore,

$$\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(a)}-(1/p)B_{i\boldsymbol{J}}^{(a)}]\overset{d}{=}\sigma^2(1-\rho)\chi^2_{(N-k)(p-1)}\text{, where } N=\sum_{i=1}^{k}N_i\,.$$

The proof is complete.

Proposition 3.8: $\sum_{i=1}^{k}B_{i\boldsymbol{J}}^{(a)}$ is distributed as the quantity $p\sigma^2[1+(p-1)\rho]\chi^2_{N-k}$.

Proof:

Applying Proposition 2.4, $B_{i\boldsymbol{J}}^{(a)}$ are independently distributed and each of which is distributed as $p\sigma^2[1+(p-1)\rho]\chi^2_{N_i-1}$ random variable, so $\sum_{i=1}^{k}B_{i\boldsymbol{J}}^{(a)}$ is distributed as the sum of *k* independent chi-square random variables $\chi^2_{N_i-1}$ times a constant $p\sigma^2[1+(p-1)\rho]$. Therefore we have $\sum_{i=1}^{k}B_{i\boldsymbol{J}}^{(a)}\overset{d}{=}p\sigma^2[1+(p-1)\rho]\chi^2_{N-k}$, where $N=\sum_{i=1}^{k}N_i$. The proof is complete.

Theorem 3.5: The likelihood ratio test statistic in Theorem 3.4 for testing $H_0:\boldsymbol{\mu}_1=\ldots=\boldsymbol{\mu}_k$ is L which is defined as:

$$L=\frac{\{\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(a)}-(1/p)B_{i\boldsymbol{J}}^{(a)}]\}^{p-1}(\sum_{i=1}^{k}B_{i\boldsymbol{J}}^{(a)})}{\{\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(0)}-(1/p)B_{i\boldsymbol{J}}^{(0)}]\}^{p-1}(\sum_{i=1}^{k}B_{i\boldsymbol{J}}^{(0)})}=\frac{B^{p-1}D}{A^{p-1}C},$$

where $A=\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(0)}-(1/p)B_{i\boldsymbol{J}}^{(0)}]$, $B=\sum_{i=1}^{k}[B_{i\boldsymbol{I}}^{(a)}-(1/p)B_{i\boldsymbol{J}}^{(a)}]$, $C=\sum_{i=1}^{k}B_{i\boldsymbol{J}}^{(0)}$, and $D=\sum_{i=1}^{k}B_{i\boldsymbol{J}}^{(a)}$.

(a) *B* and *D* are distributed, respectively, as the following:

$$B\overset{d}{=}\sigma^2(1-\rho)\chi^2_{(N-k)(p-1)}\text{ and } D\overset{d}{=}p\sigma^2[1+(p-1)\rho]\chi^2_{N-k}.$$

Under $H_0:\boldsymbol{\mu}_1=\ldots=\boldsymbol{\mu}_k$, *A* and *C* are distributed respectively as the following:

$$A\overset{d}{=}\sigma^2(1-\rho)\chi^2_{(N-1)(p-1)}\text{ and } C\overset{d}{=}p\sigma^2[1+(p-1)\rho]\chi^2_{N-1},$$

(b) *A*-*B*, *B*, *C*-*D*, and *D* are mutually independent weighted chi-square random variables.

(c) Furthermore, under $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k$, L is distributed as the random variable

$$\frac{1}{\left(1+\frac{k-1}{N-k}F^*\right)^{p-1}\left(1+\frac{k-1}{N-k}F^{**}\right)},$$

where F^* and F^{**} are independent and distributed like $F_{(k-1)(p-1),\,(N-k)(p-1)}$, and $F_{k-1,N-k}$, respectively.

Proof of (a): Results can be obtained by directly applying Propositions 3.5 to 3.8.

Proof of (b) and (c):

First rewrite *A* and *C* as follows. *A* can be expressed as $A = B + R,$ where

$$R = A - B = \sum_{i=1}^{k} (\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_0)^T \boldsymbol{C}_i^* (\boldsymbol{I} - \frac{1}{p}\boldsymbol{J})(\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_0). \tag{3.39}$$

C can be expressed as $C = D + S,$ where

$$S = C - D = \sum_{i=1}^{k} (\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_0)^T \boldsymbol{C}_i^* \boldsymbol{J} (\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_0). \tag{3.40}$$

Some other facts necessary to prove part (b) are stated below.

(1) *B* and *R* are independent

(2) *D* and *S* are independent

(3) $R \overset{d}{=} \sigma^2(1-\rho)\chi^2_{(k-1)(p-1)}$ and $S \overset{d}{=} p\sigma^2[1+(p-1)\rho]\chi^2_{k-1}$.

(4) *B* and *D* are independent

(5) *B* and *S* are independent

(6) *R* and *D* are independent

(7) *R* and *S* are independent

Facts (1), (2), (5), and (6) are true because both B and D are functions of $\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_i$, and R and S are functions of $\hat{\boldsymbol{\mu}}_i$, due to the fact that $\hat{\boldsymbol{\mu}}_0$ in (3.39) can be expressed as a linear combination of $\hat{\boldsymbol{\mu}}_i$ which is

$$\hat{\boldsymbol{\mu}}_0 = (\sum_{i=1}^{k} \boldsymbol{C}_i^*)^{-1} \sum_{i=1}^{k} \boldsymbol{C}_i^* \hat{\boldsymbol{\mu}}_i \, , \tag{3.41}$$

by using the relation $\hat{\boldsymbol{\mu}}_i = (\boldsymbol{C}_i^*)^{-1} \sum_{j=1}^{N_i} \boldsymbol{C}_{ij}^T \boldsymbol{X}_{ij}$.

Combining the facts that each pair of $\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\mu}}_0$ are independent for all *i* and *j*, Facts (1), (2), (5), and (6) are shown.

Fact (3) can be shown using the results in part (a) in conjunction with facts (1) and (2) and the result about sum of independent chi-square random variables. More clearly, the results

$$A \stackrel{d}{=} \sigma^2 (1-\rho) \chi^2_{(N-1)(p-1)} \text{ and } B \stackrel{d}{=} \sigma^2 (1-\rho) \chi^2_{(N-k)(p-1)}$$

combined with fact (1) imply $R \stackrel{d}{=} \sigma^2 (1-\rho) \chi^2_{(k-1)(p-1)}$. In addition, the results

$$C \stackrel{d}{=} p\sigma^2 [1+(p-1)\rho] \chi^2_{N-1} \text{ and } D \stackrel{d}{=} p\sigma^2 [1+(p-1)\rho] \chi^2_{N-k}$$

in connection with fact (2) implies $S \stackrel{d}{=} p\sigma^2 [1+(p-1)\rho] \chi^2 (k-1)$.

Fact (4) can be shown using the result in Proposition 2.5 that $B_{i\boldsymbol{I}}^{(a)} - (1/p) B_{i\boldsymbol{J}}^{(a)}$ and $B_{i\boldsymbol{J}}^{(a)}$ for all $i = 1, \ldots, k$ are independent.

To show Fact (7), it is necessary to rewrite *R* and *S* in (3.39) and (3.40), respectively. *R* in (3.39) can be expressed as

$$\sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)^T\boldsymbol{C}_i^*(\boldsymbol{I}-\frac{1}{p}\boldsymbol{J})(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)$$
$$-\sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T\boldsymbol{C}_i^*(\boldsymbol{I}-\frac{1}{p}\boldsymbol{J})(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)$$
$$-2\sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_i-\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{C}_i^*(\boldsymbol{I}-\frac{1}{p}\boldsymbol{J})(\hat{\boldsymbol{\mu}}_0-\hat{\boldsymbol{\mu}}_0).$$

Note that the identity $\sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_i-\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{C}_i^*=0$ is true based on the substitution of the expression of $\hat{\boldsymbol{\mu}}_0$ in (3.41). Hence we have

$$R=\sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)^T\boldsymbol{C}_i^*(\boldsymbol{I}-\frac{1}{p}\boldsymbol{J})(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)$$
$$-(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T(\sum_{i=1}^{k}\boldsymbol{C}_i^*)(\boldsymbol{I}-\frac{1}{p}\boldsymbol{J})(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0). \tag{3.42}$$

Similarly,

$$S=\sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)^T\boldsymbol{C}_i^*\boldsymbol{J}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)-\sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T\boldsymbol{C}_i^*\boldsymbol{J}(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0). \tag{3.43}$$

Since $\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0$ can be written as

$$\begin{aligned}\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0 &=(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}(\sum_{i=1}^{k}\boldsymbol{C}_i^*\hat{\boldsymbol{\mu}}_i)-(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}(\sum_{i=1}^{k}\boldsymbol{C}_i^*)\boldsymbol{\mu}_0\\ &=(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}[\sum_{i=1}^{k}\boldsymbol{C}_i^*(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)],\end{aligned}$$

the second term of (3.42) of the right-hand side becomes

$$\sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)^T\boldsymbol{C}_i^*(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}(\sum_{i=1}^{k}\boldsymbol{C}_i^*)(\boldsymbol{I}-\frac{1}{p}\boldsymbol{J})(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}\sum_{j=1}^{k}\boldsymbol{C}_j^*(\hat{\boldsymbol{\mu}}_j-\boldsymbol{\mu}_0)$$
$$=\sum_{i=1}^{k}\sum_{j=1}^{k}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)^T\boldsymbol{C}_i^*\boldsymbol{C}_j^*(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}(\boldsymbol{I}-\frac{1}{p}\boldsymbol{J})(\hat{\boldsymbol{\mu}}_j-\boldsymbol{\mu}_0).$$

Hence R in (3.42) can be expressed as

$$R=\sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)^T[\boldsymbol{C}_i^*-\boldsymbol{C}_i^{*2}(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}](\boldsymbol{I}-\frac{1}{p}\boldsymbol{J})(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)$$
$$-\sum\sum_{i\neq j}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)^T\boldsymbol{C}_i^*\boldsymbol{C}_j^*(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}(\boldsymbol{I}-\frac{1}{p}\boldsymbol{J})(\hat{\boldsymbol{\mu}}_j-\boldsymbol{\mu}_0). \tag{3.44}$$

Likewise, S can be expressed as

$$S=\sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)^T[\boldsymbol{C}_i^*-\boldsymbol{C}_i^{*2}(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}]\boldsymbol{J}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)$$
$$-\sum\sum_{i\neq j}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_0)^T\boldsymbol{C}_i^*\boldsymbol{C}_j^*(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}\boldsymbol{J}(\hat{\boldsymbol{\mu}}_j-\boldsymbol{\mu}_0). \tag{3.45}$$

Now rewrite R and S in (3.44) and (3.45) respectively in matrix forms. R can be written as

$$R=\begin{pmatrix}\hat{\boldsymbol{\mu}}_1-\boldsymbol{\mu}_0\\ \vdots\\ \hat{\boldsymbol{\mu}}_k-\boldsymbol{\mu}_0\end{pmatrix}^T\left(\boldsymbol{\Psi}_{ij}\right)\begin{pmatrix}\hat{\boldsymbol{\mu}}_1-\boldsymbol{\mu}_0\\ \vdots\\ \hat{\boldsymbol{\mu}}_k-\boldsymbol{\mu}_0\end{pmatrix},$$

where

$$\left(\boldsymbol{\Psi}_{ij}\right)=\begin{pmatrix}\boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} & \cdots & \boldsymbol{\Psi}_{1k}\\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} & \cdots & \boldsymbol{\Psi}_{2k}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{\Psi}_{k1} & \boldsymbol{\Psi}_{k2} & \cdots & \boldsymbol{\Psi}_{kk}\end{pmatrix},$$

where

$$\boldsymbol{\Psi}_{ij}=\begin{cases}[\boldsymbol{C}_i^*-\boldsymbol{C}_i^{*2}(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}](\boldsymbol{I}-\dfrac{1}{p}\boldsymbol{J}), & i=j,\\ -\boldsymbol{C}_i^*\boldsymbol{C}_j^*(\sum_{i=1}^{k}\boldsymbol{C}_i^*)^{-1}(\boldsymbol{I}-\dfrac{1}{p}\boldsymbol{J}), & i\neq j.\end{cases}$$

Likewise, S can be written as

$$S=\begin{pmatrix}\hat{\boldsymbol{\mu}}_1-\boldsymbol{\mu}_0\\ \vdots\\ \hat{\boldsymbol{\mu}}_k-\boldsymbol{\mu}_0\end{pmatrix}^T\left(\boldsymbol{\Phi}_{ij}\right)\begin{pmatrix}\hat{\boldsymbol{\mu}}_1-\boldsymbol{\mu}_0\\ \vdots\\ \hat{\boldsymbol{\mu}}_k-\boldsymbol{\mu}_0\end{pmatrix},$$

where

$$\left(\boldsymbol{\Phi}_{ij}\right)=\begin{pmatrix}\boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} & \cdots & \boldsymbol{\Phi}_{1k}\\ \boldsymbol{\Phi}_{21} & \boldsymbol{\Phi}_{22} & \cdots & \boldsymbol{\Phi}_{2k}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{\Phi}_{k1} & \boldsymbol{\Phi}_{k2} & \cdots & \boldsymbol{\Phi}_{kk}\end{pmatrix},$$

where

$$\boldsymbol{\Phi}_{ij} = \begin{cases} [\boldsymbol{C}_i^* - \boldsymbol{C}_i^{*2}(\sum_{i=1}^k \boldsymbol{C}_i^*)^{-1}]\boldsymbol{J}, & i = j, \\ -\boldsymbol{C}_i^*\boldsymbol{C}_j^*(\sum_{i=1}^k \boldsymbol{C}_i^*)^{-1}\boldsymbol{J}, & i \neq j. \end{cases}$$

Note that the vector $(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_0, ..., \hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_0)'$ is distributed as $N_{kp}(\boldsymbol{0}, \text{diag}(\boldsymbol{C}_i^*\boldsymbol{\Sigma}))$, where

$$\text{diag}(\boldsymbol{C}_i^*\boldsymbol{\Sigma}) = \begin{pmatrix} \boldsymbol{C}_1^*\boldsymbol{\Sigma} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{C}_k^*\boldsymbol{\Sigma} \end{pmatrix}_{kp\times kp}.$$

If we can show

$$\left(\boldsymbol{\Psi}_{ij}\right)\left(\text{diag}(\boldsymbol{C}_i^*\boldsymbol{\Sigma})\right)\left(\boldsymbol{\Phi}_{ij}\right) = 0, \tag{3.46}$$

then the proof of Fact (8) is done. The expression (3.46) is true because we have

$$\begin{aligned} &\left(\boldsymbol{\Psi}_{ij}\right)\left(\text{diag}(\boldsymbol{C}_i^*\boldsymbol{\Sigma})\right)\left(\boldsymbol{\Phi}_{ij}\right) \\ &= \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} & \cdots & \boldsymbol{\Psi}_{1k} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} & \cdots & \boldsymbol{\Psi}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{k1} & \boldsymbol{\Psi}_{k2} & \cdots & \boldsymbol{\Psi}_{kk} \end{pmatrix} \begin{pmatrix} \boldsymbol{C}_1^*\boldsymbol{\Sigma} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{C}_k^*\boldsymbol{\Sigma} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} & \cdots & \boldsymbol{\Phi}_{1k} \\ \boldsymbol{\Phi}_{21} & \boldsymbol{\Phi}_{22} & \cdots & \boldsymbol{\Phi}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}_{k1} & \boldsymbol{\Phi}_{k2} & \cdots & \boldsymbol{\Phi}_{kk} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Psi}_{11}\boldsymbol{C}_1^*\boldsymbol{\Sigma} & \boldsymbol{\Psi}_{12}\boldsymbol{C}_2^*\boldsymbol{\Sigma} & \cdots & \boldsymbol{\Psi}_{1k}\boldsymbol{C}_k^*\boldsymbol{\Sigma} \\ \boldsymbol{\Psi}_{21}\boldsymbol{C}_1^*\boldsymbol{\Sigma} & \boldsymbol{\Psi}_{22}\boldsymbol{C}_2^*\boldsymbol{\Sigma} & \cdots & \boldsymbol{\Psi}_{2k}\boldsymbol{C}_k^*\boldsymbol{\Sigma} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{k1}\boldsymbol{C}_1^*\boldsymbol{\Sigma} & \boldsymbol{\Psi}_{k2}\boldsymbol{C}_2^*\boldsymbol{\Sigma} & \cdots & \boldsymbol{\Psi}_{kk}\boldsymbol{C}_k^*\boldsymbol{\Sigma} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} & \cdots & \boldsymbol{\Phi}_{1k} \\ \boldsymbol{\Phi}_{21} & \boldsymbol{\Phi}_{22} & \cdots & \boldsymbol{\Phi}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}_{k1} & \boldsymbol{\Phi}_{k2} & \cdots & \boldsymbol{\Phi}_{kk} \end{pmatrix} \\ &= \left(\sum_{m=1}^k \boldsymbol{\Psi}_{im}\boldsymbol{C}_m^*\boldsymbol{\Sigma}\boldsymbol{\Phi}_{mj}\right), \end{aligned}$$

which is zero matrix because of validity of the identities

$$\sum_{m=1}^k \boldsymbol{\Psi}_{im}\boldsymbol{C}_m^*\boldsymbol{\Sigma}\boldsymbol{\Phi}_{mj} = \boldsymbol{0} \text{ for all } i \text{ and } j,$$

due to the fact $(\boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p)\boldsymbol{J}_p = \boldsymbol{0}$ and commutability of circulant matrices. Thus *R* and *S* are independent. Therefore,

$$L = \frac{B^{p-1}D}{A^{p-1}C} = \frac{B^{p-1}D}{(B+R)^{p-1}(D+S)} = \frac{1}{\left(1+\frac{R}{B}\right)^{p-1}\left(1+\frac{S}{D}\right)}$$

$$\overset{d}{=} \frac{1}{\left(1+\frac{k-1}{N-k}F^{*}\right)^{p-1}\left(1+\frac{k-1}{N-k}F^{**}\right)},$$

where F^{*} and F^{**} are independent and distributed like $F_{(k-1)(p-1),\,(N-k)(p-1)}$ and $F_{k-1,N-k}$, respectively. The proof of Theorem 3.5 is complete.

CHAPTER IV

APPLICATION TO META ANALYSIS

4.1 INTRODUCTION AND PRELIMINARY UNIVARIATE CASE

In this chapter, an application of the proposed model to fixed and random effects multivariate meta analysis (Jackson et al. 2011, Nam et al. 2003) will be introduced and developed. Individual patient / participant data (IPD) are assumed available in the whole chapter. Since the outcome measures under the assumption of the proposed heterogeneous means model are continuous, one-stage method for IPD random effects model is suggested by Higgins et al. (2001) to investigate the parameter of interest for each study. In the analysis, both fixed and random effects models are focused.

First consider the fixed effect model of univariate case. Let X_{ij} denote the outcome measure of subject j in study i generated from k independent studies, where $i = 1,...,k$, $j = 1,...,n_i$, and $X_{ij} \sim N_p(C_{ij}\mu_{(i)}, \sigma_i^2)$, where $\mu_{(i)}$ is an unknown constant, C_{ij} and σ_i^2 are known. The ML estimator for $\mu_{(i)}$ based on the *i*th study is

$$\hat{\mu}_{(i)} = \left(\sum_{j=1}^{n_i} C_{ij}^2 \right)^{-1} \sum_{j=1}^{n_i} C_{ij} X_{ij}, i = 1,...,k.$$

Since the estimates are derived from different individual participant data sets, $\hat{\mu}_{(i)}$ are conditionally independent given $\mu_{(i)}$. Hence $\hat{\mu}_{(i)}$ given $\mu_{(i)}$ are independent and exact normal with

$$\hat{\mu}_{(i)} \sim N(\mu_{(i)}, \sigma_i^{*2}), i = 1,...,k, \tag{4.1}$$

where $\sigma_i^{*2} = (\sum_{j=1}^{n_i} C_{ij}^{\ 2})^{-1} \sigma_i^{\ 2}$.

In traditional meta analysis, results from several studies are combined. For a fixed effect model, if we tacitly assume that the true value $\mu_{(i)} = \mu$ is the same for the k studies, given (4.1) the ML estimator for μ and its variance based on the k independent samples are, respectively,

$$\tilde{\mu} = (\sum_{i=1}^{k} \sigma_i^{*-2})^{-1} \sum_{i=1}^{k} \sigma_i^{*-2} \hat{\mu}_{(i)} = \left(\sum_{i=1}^{k} \frac{\sum_{j=1}^{n_i} C_{ij}^{\ 2}}{\sigma_i^{\ 2}} \right)^{-1} \sum_{i=1}^{k} \frac{\sum_{j=1}^{n_i} C_{ij}^{\ 2} \hat{\mu}_{(i)}}{\sigma_i^{\ 2}},$$

and

$$Var(\tilde{\mu}) = (\sum_{i=1}^{k} \sigma_i^{*-2})^{-1} = \left(\sum_{i=1}^{k} \frac{\sum_{j=1}^{n_i} C_{ij}^{\ 2}}{\sigma_i^{\ 2}} \right)^{-1}.$$

Hence, statistical inference for μ is based on the fact that

$$\frac{\tilde{\mu} - \mu}{\sqrt{Var(\tilde{\mu})}} \sim N(0,1).$$

Some other fixed effect models of multivariate case will be considered in Section 4.2.

4.2 FIXED EFFECT MODEL

Let $\boldsymbol{X}_{ij}$ denote the outcome measure of subject j in study i generated from k independent studies, where $i = 1,...,k$, $j = 1,...,n_i$, and $\boldsymbol{X}_{ij} \sim N_p(\boldsymbol{C}_{ij} \boldsymbol{\mu}_{(i)}, \boldsymbol{\Sigma}_i)$. The ML estimator for $\boldsymbol{\mu}_{(i)}$ based on ith sample is

$$\hat{\boldsymbol{\mu}}_{(i)} = \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{C}_{ij} \right)^{-1} \sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{X}_{ij}, i = 1, \ldots, k. \quad (4.2)$$

Here we assume that $\boldsymbol{\Sigma}_i$ is known for all $i = 1, \ldots, k$. In fact, $\hat{\boldsymbol{\mu}}_{(i)}$'s are independent and

$$\hat{\boldsymbol{\mu}}_{(i)} \sim MVN(\boldsymbol{\mu}_{(i)}, \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{C}_{ij} \right)^{-1}), i = 1, \ldots, k. \quad (4.3)$$

For traditional meta analysis, the true core mean vector $\boldsymbol{\mu}_{(i)} = \boldsymbol{\mu}$ is the same for the k studies, given (4.3) for the k studies, the ML estimator for $\boldsymbol{\mu}$ and its variance-covariance matrix based on the k independent samples are, respectively,

$$\tilde{\boldsymbol{\mu}} = \left(\sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{C}_{ij} \right) \right)^{-1} \sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{C}_{ij} \right) \hat{\boldsymbol{\mu}}_{(i)},$$

and

$$Cov(\tilde{\boldsymbol{\mu}}) = \left(\sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{C}_{ij} \right) \right)^{-1}.$$

Statistical inference for $\boldsymbol{\mu}$ is based on the fact

$$\tilde{\boldsymbol{\mu}} \sim N_p(\boldsymbol{\mu}, \left(\sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{C}_{ij} \right) \right)^{-1}),$$

which results in

$$(\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu})' \left(\sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{C}_{ij} \right) \right) (\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}) \sim \chi_p^2.$$

A great difficulty for multivariate meta analysis in practice is that the within-study covariance matrix $\left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{C}_{ij} \right)^{-1}$ is usually unknown since $\boldsymbol{\Sigma}_i$ is unknown. If IPD are available, we may use an appropriate estimate $\boldsymbol{S}_i$ from IPD data for the *i*th study to replace $\boldsymbol{\Sigma}_i$. Hence the estimator of $\boldsymbol{\mu}_{(i)}$ becomes

$$\hat{\boldsymbol{\mu}}_{(i)} = \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{S}_i^{-1} \boldsymbol{C}_{ij} \right)^{-1} \sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{S}_i^{-1} \boldsymbol{X}_{ij}, i = 1,...,k, \tag{4.4}$$

where $\boldsymbol{S}_i$ is a consistent estimator of $\boldsymbol{\Sigma}_i$ it follows that $\hat{\boldsymbol{\mu}}_{(i)}$ is approximate normal with covariance matrix estimated by

$$\left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{S}_i^{-1} \boldsymbol{C}_{ij} \right)^{-1}, i = 1,...,k. \tag{4.5}$$

There must be some impact on the analysis using estimate $\boldsymbol{S}_i$ instead of true $\boldsymbol{\Sigma}_i$. Basically, when the *i*th sample size is large, $\boldsymbol{S}_i$ is a good substitute for $\boldsymbol{\Sigma}_i$, $i = 1,...,k$. If $\boldsymbol{\mu}_{(i)} = \boldsymbol{\mu}$ is the same for the k studies, the estimator of $\boldsymbol{\mu}$ can be obtained by

$$\tilde{\boldsymbol{\mu}} = \left(\sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{S}_i^{-1} \boldsymbol{C}_{ij} \right) \right)^{-1} \sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{S}_i^{-1} \boldsymbol{C}_{ij} \right) \hat{\boldsymbol{\mu}}_{(i)}, \tag{4.6}$$

with the following approximation:

$$(\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu})' \left(\sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}' \boldsymbol{S}_i^{-1} \boldsymbol{C}_{ij} \right) \right) (\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}) \overset{\bullet}{\sim} \chi_p^2. \tag{4.7}$$

Inference about $\boldsymbol{\mu}$ can be based on the above approximation by means of meta-analysis.

There are two conditions about the structure of $\boldsymbol{\Sigma}_i$ for the *i*th study considered here.

Condition 1: If $\boldsymbol{\Sigma}_i$ is unknown and unstructured, $\boldsymbol{S}_i$ is obtained from IPD data assuming that $\boldsymbol{\Sigma}_i$ is positive definite. As mentioned in Chapter 2, $\boldsymbol{S}_i$ needs to be assessed iteratively based on the *i*th study.

Condition 2: $\boldsymbol{\Sigma}_i$ is has compound symmetric structure, that is $\boldsymbol{\Sigma}_i = \sigma_i^2[(1-\rho_i)\boldsymbol{I}_p + \rho_i \boldsymbol{J}_p]$.

Assume that $\boldsymbol{C}_{ij}$ and $\boldsymbol{\Sigma}_i$ commute for all i and j. Hence $\boldsymbol{S}_i$ is obtained based on maximum likelihood method in Chapter 2 from *i*th study with estimators

$$\hat{\sigma}_i^{\,2} = \frac{1}{n_i p}\sum_{j=1}^{n_i}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_{(i)})^T(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_{(i)}),$$

and

$$\hat{\rho}_i = \frac{1}{p-1}\left[\frac{\sum_{j=1}^{n_i}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_{(i)})^T\boldsymbol{J}_p(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_{(i)})}{\sum_{j=1}^{n_i}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_{(i)})^T(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\hat{\boldsymbol{\mu}}_{(i)})} - 1\right], \text{ for all } i = 1,\ldots,k,$$

where $\hat{\boldsymbol{\mu}}_{(i)}$ in (4.4) reduces to

$$\hat{\boldsymbol{\mu}}_{(i)} = \left(\sum_{j=1}^{n_i}\boldsymbol{C}_{ij}'\boldsymbol{C}_{ij}\right)^{-1}\sum_{j=1}^{n_i}\boldsymbol{C}_{ij}'\boldsymbol{X}_{ij}, \tag{4.8}$$

for all $i = 1,\ldots,k$.

4.3 RANDOM EFFECTS MODEL

4.3.1 Two-Stage Method

The two-stage meta analysis still can be used when IPD are available. We denote the estimated core mean vector for the *i*th study as $\hat{\boldsymbol{\mu}}_{(i)}$ and it is assumed that

$$\hat{\boldsymbol{\mu}}_{(i)} \mid \boldsymbol{\mu}_{(i)} \overset{ind}{\sim} MVN(\boldsymbol{\mu}_{(i)}, \boldsymbol{\Psi}_i), i = 1,\ldots,k,$$

which is referred as the within-study model. The entries of the matrices $\boldsymbol{\Psi}_i$ for each study are estimated from IPD data and are usually assumed known and fixed. If the covariance matrix $\boldsymbol{\Sigma}_i$

of $\boldsymbol{X}_{ij}$ is known for each study, for example, $\boldsymbol{\Psi}_i = \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}{}' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{C}_{ij} \right)^{-1}$ is known and there is no need to estimate it.

The multivariate random effect model allows $\boldsymbol{\mu}_{(i)}$ to vary from one study to the next. So we can further assume that the between-study normal assumption is

$$\boldsymbol{\mu}_{(i)} \overset{iid}{\sim} MVN(\boldsymbol{\mu}, \boldsymbol{\Gamma}), i = 1, \ldots, k.$$

The resulting two-stage marginal model is obtained by

$$\hat{\boldsymbol{\mu}}_{(i)} \overset{ind}{\sim} MVN(\boldsymbol{\mu}, \boldsymbol{\Psi}_i + \boldsymbol{\Gamma}), i = 1, \ldots, k \,,$$

with corresponding log likelihood

$$\log L(\boldsymbol{\mu}, \boldsymbol{\Gamma}) = -\frac{kp}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{k}\log|(\boldsymbol{\Psi}_i + \boldsymbol{\Gamma})| - \frac{1}{2}\sum_{i=1}^{k}(\hat{\boldsymbol{\mu}}_{(i)} - \boldsymbol{\mu})^T(\boldsymbol{\Psi}_i + \boldsymbol{\Gamma})^{-1}(\hat{\boldsymbol{\mu}}_{(i)} - \boldsymbol{\mu}).$$

The parameters of interest which need to be estimated would be $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$. The MLE of $\boldsymbol{\mu}$ is

$$\tilde{\boldsymbol{\mu}} = \left[\sum_{i=1}^{k} (\boldsymbol{\Psi}_i + \hat{\boldsymbol{\Gamma}})^{-1} \right]^{-1} \sum_{i=1}^{k} (\boldsymbol{\Psi}_i + \hat{\boldsymbol{\Gamma}})^{-1} \hat{\boldsymbol{\mu}}_{(i)} \,,$$

where $\tilde{\boldsymbol{\mu}}$ is approximate normal with the variance

$$Var(\tilde{\boldsymbol{\mu}}) = \left[\sum_{i=1}^{k} (\boldsymbol{\Psi}_i + \hat{\boldsymbol{\Gamma}})^{-1} \right]^{-1}.$$

The main statistical difficulty is to estimate the between-study covariance matrix $\boldsymbol{\Gamma}$. A few methods can be used to obtain the estimated $\hat{\boldsymbol{\Gamma}}$. They are maximum likelihood (ML) estimation, restricted maximum likelihood (REML) estimation, method of moment (MM), and some alternative procedures such as profile likelihood and Bayesian analyses which have been reviewed by Jackson et al. (2011).

4.3.2 One-Stage Method

If individual participant data are available, we have the assumption

$$\boldsymbol{X}_{ij} \mid \boldsymbol{\mu}_{(i)} \overset{ind}{\sim} MVN(\boldsymbol{C}_{ij}\boldsymbol{\mu}_{(i)}, \boldsymbol{\Sigma}_i),\ i=1,\dots,k,\ j=1,\dots,n_i,$$

where $\boldsymbol{\Sigma}_i$ is unknown, and $\boldsymbol{\mu}_{(i)} \overset{iid}{\sim} MVN(\boldsymbol{\mu}, \boldsymbol{\Gamma}), i=1,\dots,k$, which implies that

$$\boldsymbol{C}_{ij}\boldsymbol{\mu}_{(i)} \overset{ind}{\sim} MVN(\boldsymbol{C}_{ij}\boldsymbol{\mu}, \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T),\ i=1,\dots,k,\ j=1,\dots,n_i.$$

The unknown parameters are $\boldsymbol{\mu}, \boldsymbol{\Sigma}_1,\dots,\boldsymbol{\Sigma}_k$, and $\boldsymbol{\Gamma}$. With the above assumptions,

$$\boldsymbol{X}_{ij} \overset{ind}{\sim} MVN(\boldsymbol{C}_{ij}\boldsymbol{\mu}, \boldsymbol{\Sigma}_i + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T), i=1,\dots,k, j=1,\dots,n_i.$$

We may rewrite the model in matrix notation as:

$$\boldsymbol{X}_{ij} = \boldsymbol{C}_{ij}(\boldsymbol{\mu} + \boldsymbol{F}_{ij}) + \boldsymbol{E}_i,$$

where $\boldsymbol{E}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$ and $\boldsymbol{F}_{ij} \sim N(\boldsymbol{0}, \boldsymbol{\Gamma})$. The corresponding likelihood is

$$\begin{aligned}\log L(&\boldsymbol{\mu}, \boldsymbol{\Sigma}_1,\dots,\boldsymbol{\Sigma}_k, \boldsymbol{\Gamma}) \\ &= -\frac{np}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log\left|\boldsymbol{\Sigma}_i + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T\right| \\ &\quad -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu})^T(\boldsymbol{\Sigma}_i + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T)^{-1}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}),\end{aligned} \tag{4.9}$$

where $n = \sum_{i=1}^{k} n_i$ is the total number of observations.

The MLEs of $\boldsymbol{\mu}, \boldsymbol{\Sigma}_1,\dots,\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Gamma}$, say $\tilde{\tilde{\boldsymbol{\mu}}}, \tilde{\tilde{\boldsymbol{\Sigma}}}_1,\dots,\tilde{\tilde{\boldsymbol{\Sigma}}}_k$ and $\tilde{\tilde{\boldsymbol{\Gamma}}}$, can be found iteratively. The MLE of $\boldsymbol{\mu}$ is

$$\tilde{\tilde{\boldsymbol{\mu}}} = \left[\sum_{i=1}^{k}\left(\sum_{j=1}^{n_i}\boldsymbol{C}_{ij}'(\tilde{\tilde{\boldsymbol{\Sigma}}}_i + \boldsymbol{C}_{ij}\tilde{\tilde{\boldsymbol{\Gamma}}}\boldsymbol{C}_{ij}^T)^{-1}\boldsymbol{C}_{ij}\right)\right]^{-1}\sum_{i=1}^{k}\left(\sum_{j=1}^{n_i}\boldsymbol{C}_{ij}'(\tilde{\tilde{\boldsymbol{\Sigma}}}_i + \boldsymbol{C}_{ij}\tilde{\tilde{\boldsymbol{\Gamma}}}\boldsymbol{C}_{ij}^T)^{-1}X_{ij}\right), \tag{4.10}$$

where $\tilde{\tilde{\boldsymbol{\mu}}}$ is approximate normal with the variance estimated by:

$$\hat{Var}(\tilde{\tilde{\boldsymbol{\mu}}}) = \left[\sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij}'(\tilde{\tilde{\boldsymbol{\Sigma}}}_i + \boldsymbol{C}_{ij}\tilde{\tilde{\boldsymbol{\Gamma}}}\boldsymbol{C}_{ij}^T)^{-1}\boldsymbol{C}_{ij} \right) \right]^{-1}.$$

Finding the MLEs for $\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Gamma}$ is not easy work, a large amount of computations will be needed to assess the result. If we make some assumptions on the forms of the matrices $\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Gamma}$, we may get some explicit results for the MLEs. Consider the following assumptions:

Assumption 1. $\boldsymbol{C}_{ij}$, $\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Gamma}$ are symmetric regular circulant matrices providing that they commute with each other. Note that $\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Gamma}$ are also positive definite.

Assumption 2. $\boldsymbol{\Sigma}_i = \sigma_i^{\,2}\left[(1-\rho_i)\boldsymbol{I} + \rho_i\boldsymbol{J}\right]$, $\boldsymbol{\Gamma} = \omega^2\left[(1-\tau)\boldsymbol{I} + \tau\boldsymbol{J}\right]$, where $1+(p-1)\rho_i > 0$ and $1+(p-1)\tau > 0$.

Assumption 3. $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ for all $i = 1,...,k$.

Assumption 4. $\boldsymbol{C}_{ij}$ is a $p \times p$ circular matrix for all i and j with $p \geq 2$ and $\boldsymbol{C}_{ij} = (s_{ij} - t_{ij})\boldsymbol{I} + t_{ij}\boldsymbol{J}$.

Assumption 5. $p = 2$.

The following proposition can be used to simplify the log likelihood function in (4.9) under Assumption 4 stated above.

<u>Proposition 4.1</u>: Suppose $\boldsymbol{C}_{ij}$ is a $p \times p$ circular matrix for all i and j with $p \geq 2$ and $\boldsymbol{C}_{ij} = (s_{ij} - t_{ij})\boldsymbol{I} + t_{ij}\boldsymbol{J}$. Then $\boldsymbol{C}_{ij}^2 = (a_{ij} - b_{ij})\boldsymbol{I} + b_{ij}\boldsymbol{J}$, where $a_{ij} = s_{ij}^2 + (p-1)t_{ij}^2$, $b_{ij} = 2s_{ij}t_{ij} + (p-2)t_{ij}^2$, and $a_{ij} \geq b_{ij}$.

***Proof*:**

Since $\boldsymbol{C}_{ij} = (s_{ij} - t_{ij})\boldsymbol{I} + t_{ij}\boldsymbol{J}$, $\boldsymbol{C}_{ij}^2$ becomes

$$\begin{aligned}\boldsymbol{C}_{ij}^2 &= \left[(s_{ij} - t_{ij})\boldsymbol{I} + t_{ij}\boldsymbol{J}\right]\cdot\left[(s_{ij} - t_{ij})\boldsymbol{I} + t_{ij}\boldsymbol{J}\right] = (s_{ij} - t_{ij})^2\boldsymbol{I} + 2(s_{ij} - t_{ij})t_{ij}\boldsymbol{J} + t_{ij}^2\boldsymbol{J}^2 \\ &= (s_{ij} - t_{ij})^2\boldsymbol{I} + \left[2s_{ij}t_{ij} + (p-2)t_{ij}^2\right]\boldsymbol{J} = (a_{ij} - b_{ij})\boldsymbol{I} + b_{ij}\boldsymbol{J},\end{aligned}$$

where $a_{ij} = s_{ij}^2 + (p-1)t_{ij}^2 > 0$, $b_{ij} = 2s_{ij}t_{ij} + (p-2)t_{ij}^2$, and $a_{ij} \geq b_{ij}$. The proof is complete.

We should first expand the matrix $\boldsymbol{\Sigma}_i + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T$ under the above Assumptions 1 - 3.

$$\boldsymbol{\Sigma}_i + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T = \boldsymbol{\Sigma} + \boldsymbol{C}_{ij}\boldsymbol{C}_{ij}^T\boldsymbol{\Gamma} = \boldsymbol{\Sigma} + \boldsymbol{C}_{ij}^2\boldsymbol{\Gamma} \tag{4.11}$$

Furthermore, with Assumption 4, (4.11) becomes

$$\begin{aligned}&\sigma^2[(1-\rho)\boldsymbol{I} + \rho\boldsymbol{J}] + \boldsymbol{C}_{ij}^2\cdot\omega^2[(1-\tau)\boldsymbol{I} + \tau\boldsymbol{J}] \\ &= \sigma^2[(1-\rho)\boldsymbol{I} + \rho\boldsymbol{J}] + [(a_{ij} - b_{ij})\boldsymbol{I} + b_{ij}\boldsymbol{J}]\cdot\omega^2[(1-\tau)\boldsymbol{I} + \tau\boldsymbol{J}] \\ &= [\sigma^2(1-\rho) + \omega^2(1-\tau)(a_{ij} - b_{ij})]\boldsymbol{I} + \left\{\sigma^2\rho + \omega^2[\tau(a_{ij} - b_{ij}) + (1-\tau)b_{ij} + p\tau b_{ij}]\right\}\boldsymbol{J} \\ &= [\sigma^2(1-\rho) + \omega^2(1-\tau)(a_{ij} - b_{ij})]\boldsymbol{I} + \left\{\sigma^2\rho + \omega^2[\tau a_{ij} + (1+(p-2)\tau)b_{ij}]\right\}\boldsymbol{J} \\ &= (A_{ij} - B_{ij})\boldsymbol{I} + B_{ij}\boldsymbol{J},\end{aligned} \tag{4.12}$$

where $a_{ij} = s_{ij}^2 + (p-1)t_{ij}^2 \geq 0$, $b_{ij} = 2s_{ij}t_{ij} + (p-2)t_{ij}^2$, $a_{ij} \geq b_{ij}$, and

$$A_{ij} = \sigma^2 + \omega^2[a_{ij} + (p-1)\tau b_{ij}] \text{ , } B_{ij} = \sigma^2\rho + \omega^2[\tau a_{ij} + (1+(p-2)\tau)b_{ij}]. \tag{4.13}$$

Also note that $A_{ij} > B_{ij}$ since $a_{ij} \geq b_{ij}$. Hence (4.11) can be expressed as a symmetric p by p matrix with diagonals equal to A_{ij} and off-diagonals equal to B_{ij} as shown in (4.12). Therefore, under Assumptions 1-4 we have (Graybill (1983), Theorem 8.3.4)

$$\left|\boldsymbol{\Sigma}_i + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T\right| = \left|(A_{ij} - B_{ij})\boldsymbol{I} + B_{ij}\boldsymbol{J}\right| = (A_{ij} - B_{ij})^{p-1}\left[A_{ij} + (p-1)B_{ij}\right], \tag{4.14}$$

and

$$(\boldsymbol{\Sigma}_i + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T)^{-1} = \frac{1}{A_{ij} - B_{ij}}\left[\boldsymbol{I} - \frac{B_{ij}}{A_{ij} + (p-1)B_{ij}}\boldsymbol{J}\right]. \tag{4.15}$$

Consider a special case that $p = 2$ (Assumption 5), (4.12) reduces to

$$\begin{pmatrix} \sigma^2 + \omega^2(a_{ij} + \tau b_{ij}) & \sigma^2\rho + \omega^2(\tau a_{ij} + b_{ij}) \\ \sigma^2\rho + \omega^2(\tau a_{ij} + b_{ij}) & \sigma^2 + \omega^2(a_{ij} + \tau b_{ij}) \end{pmatrix} = \begin{pmatrix} A_{ij(2)} & B_{ij(2)} \\ B_{ij(2)} & A_{ij(2)} \end{pmatrix}$$

where $A_{ij(2)} = \sigma^2 + \omega^2(a_{ij} + \tau b_{ij})$ and $B_{ij(2)} = \sigma^2\rho + \omega^2(\tau a_{ij} + b_{ij})$.

Thus, under Assumptions 1-5, we have from (4.14) and (4.15) that

$$\begin{aligned} \left|\boldsymbol{\Sigma}_i + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T\right| &= A_{ij(2)}^2 - B_{ij(2)}^2 \\ &= \sigma^4(1-\rho^2) + 2\sigma^2\omega^2\left[(1-\rho\tau)a_{ij} + (\tau-\rho)b_{ij}\right] + \omega^4(1-\tau^2)(a_{ij}^2 - b_{ij}^2) \end{aligned} \tag{4.16}$$

and

$$\begin{aligned} (\boldsymbol{\Sigma}_i + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T)^{-1} &= \frac{1}{A_{ij(2)} - B_{ij(2)}}\left[\boldsymbol{I}_2 - \frac{B_{ij(2)}}{A_{ij(2)} + (p-1)B_{ij(2)}}\boldsymbol{J}_2\right] \\ &\overset{p=2}{=} \frac{1}{\sigma^2(1-\rho) + \omega^2(1-\tau)(a_{ij} - b_{ij})}\left[\boldsymbol{I}_2 - \frac{\sigma^2\rho + \omega^2(\tau a_{ij} + b_{ij})}{\sigma^2(1+\rho) + \omega^2(1+\tau)(a_{ij} + b_{ij})}\boldsymbol{J}_2\right]. \end{aligned} \tag{4.17}$$

Note that $\left|\boldsymbol{\Sigma}_i + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T\right| > 0$ since $A_{ij} > B_{ij}$.

Now we arrive at the following theorem. This theorem states the marginal log likelihood of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\Gamma}$ based on one-stage meta analysis of the random effect multivariate heterogeneous means model. Inference about the overall core mean vector $\boldsymbol{\mu}$ for the k studies can be obtained using the log likelihood stated in the following theorem.

Theorem 4.1: Under Assumptions 1-4, and $a_{ij} = s_{ij}^2 + (p-1)t_{ij}^2 > 0$, $b_{ij} = 2s_{ij}t_{ij} + (p-2)t_{ij}^2$, the corresponding likelihood of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\Gamma}$ is

$$\begin{aligned} &\log L(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}) \\ &\quad = -\frac{np}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log\left|\boldsymbol{\Sigma} + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T\right| \\ &\quad\quad - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu})^T(\boldsymbol{\Sigma} + \boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^T)^{-1}(\boldsymbol{X}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{\mu}), \end{aligned} \tag{4.18}$$

where $\left|\boldsymbol{\Sigma}+\boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^{T}\right|$ and $\left(\boldsymbol{\Sigma}+\boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^{T}\right)^{-1}$ are stated in (4.14) and (4.15), respectively, and $n=\sum_{i=1}^{k} n_i$ is the total number of observations. For the special case where p = 2 (Assumption 5), the determinant and the inverse of the matrix $\boldsymbol{\Sigma}+\boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^{T}$,$\left|\boldsymbol{\Sigma}+\boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^{T}\right|$ and $\left(\boldsymbol{\Sigma}+\boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^{T}\right)^{-1}$, are stated in (4.16) and (4.17) , respectively.

In the next subsection, a simulation study for finding the estimates of the unknown parameters based on the marginal log likelihood function stated in Theorem 4.1 is investigated.

4.3.3 One-Stage Method – Simulation Study

The main purpose of this simulation study is to maximize the log likelihood function in (4.18) with respect to the unknown parameters such that the inference for the overall core mean $\boldsymbol{\mu}$ can be obtained using Quasi-Newton optimization method. In this simulation study, bivariate data based on the marginal model

$$\boldsymbol{X}_{ij} \overset{ind}{\sim} MVN(\boldsymbol{C}_{ij}\boldsymbol{\mu}, \boldsymbol{\Sigma}_i+\boldsymbol{C}_{ij}\boldsymbol{\Gamma}\boldsymbol{C}_{ij}^{T}), i=1,...,k,\ j=1,...,n_i$$

are generated. $\boldsymbol{C}_{ij}$ is circulant with the form $\boldsymbol{C}_{ij}=I+\begin{pmatrix} a_{ij} & b_{ij} \\ b_{ij} & a_{ij} \end{pmatrix}$, where a_{ij} and b_{ij} are independently generated from Uniform (-1, 1) distribution for each simulation study. We assume $\boldsymbol{\Sigma}_i=\boldsymbol{\Sigma}$ for all i and $\boldsymbol{\Sigma}=\sigma^2\left[(1-\rho)\boldsymbol{I}+\rho\boldsymbol{J}\right]$, $\boldsymbol{\Gamma}=\omega^2\left[(1-\tau)\boldsymbol{I}+\tau\boldsymbol{J}\right]$. We also assume equal sample size $N_i=N$ for all the k studies and consider four cases $N=5, 10, 20, 40$ for each of the k studies.

Define the unknown parameter vector $\vartheta=(\sigma^2,\rho,\omega^2,\tau,\mu_1,\mu_2)'$, where $\boldsymbol{\mu}=(\mu_1,\mu_2)'$. Let the data of the k studies be generated based on the given true vector $\vartheta_{True}=(2,.1,3,.5,10,2)'$.

The initial guesses of the vector of parameter estimators which maximize the log likelihood function in (4.18) are:

$\vartheta_{Guess1} = \vartheta_{True}$ for the first simulation study, and

$\vartheta_{Guess2} = (20, 0, 6, 0, 1, 20)'$ for the second simulation study.

We shall see the impact of the two sets of initial guesses on the estimates of the parameters. It can be seen from the left-hand column of Figures 5-8 about the behaviors of the parameter estimates based on 100 generated data sets. Each graph of the right-hand column illustrates the boxplots corresponding to the 100 sets of parameter estimates on its left. From the left-hand column of Figures 5 and 6, we can see that the use of the true parameter vector ϑ_{True} as the initial guess produces more stable parameter estimates, while the use of second initial guess ϑ_{Guess2} produces some estimates of parameter vector falling outside the main trail of most of the estimated parameter vectors as seen from the left-hand column of Figures 7 and 8. Therefore our conclusion about parameter estimates will be based on the results of the initial guess ϑ_{True}.

For each case of equal sample size for the k = 4 studies, 100 data sets are generated and each of which is based on the same sets of $\boldsymbol{C}_{ij}$ matrices generated from Uniform (-1, 1) distribution stated previously. In the left-hand column of Figures 5 and 6, each line denotes the estimates for the six parameters expressed as the vector form ϑ. When the equal sample size is 5, the estimates look unstable and vary more dramatically than the estimates with larger equal sample size. This can be seen from the boxplots of the estimates on the right columns. The estimates seem accurate especially for the two elements of the overall core mean vector μ_1 and μ_2.

This simulation study is an introduction to application to meta analysis when we have heterogeneous data and want to do inference for the overall core mean $\boldsymbol{\mu}$ of the k studies. We could use the same computational technique to deal with more general cases when necessary without being subject to specific conditions.

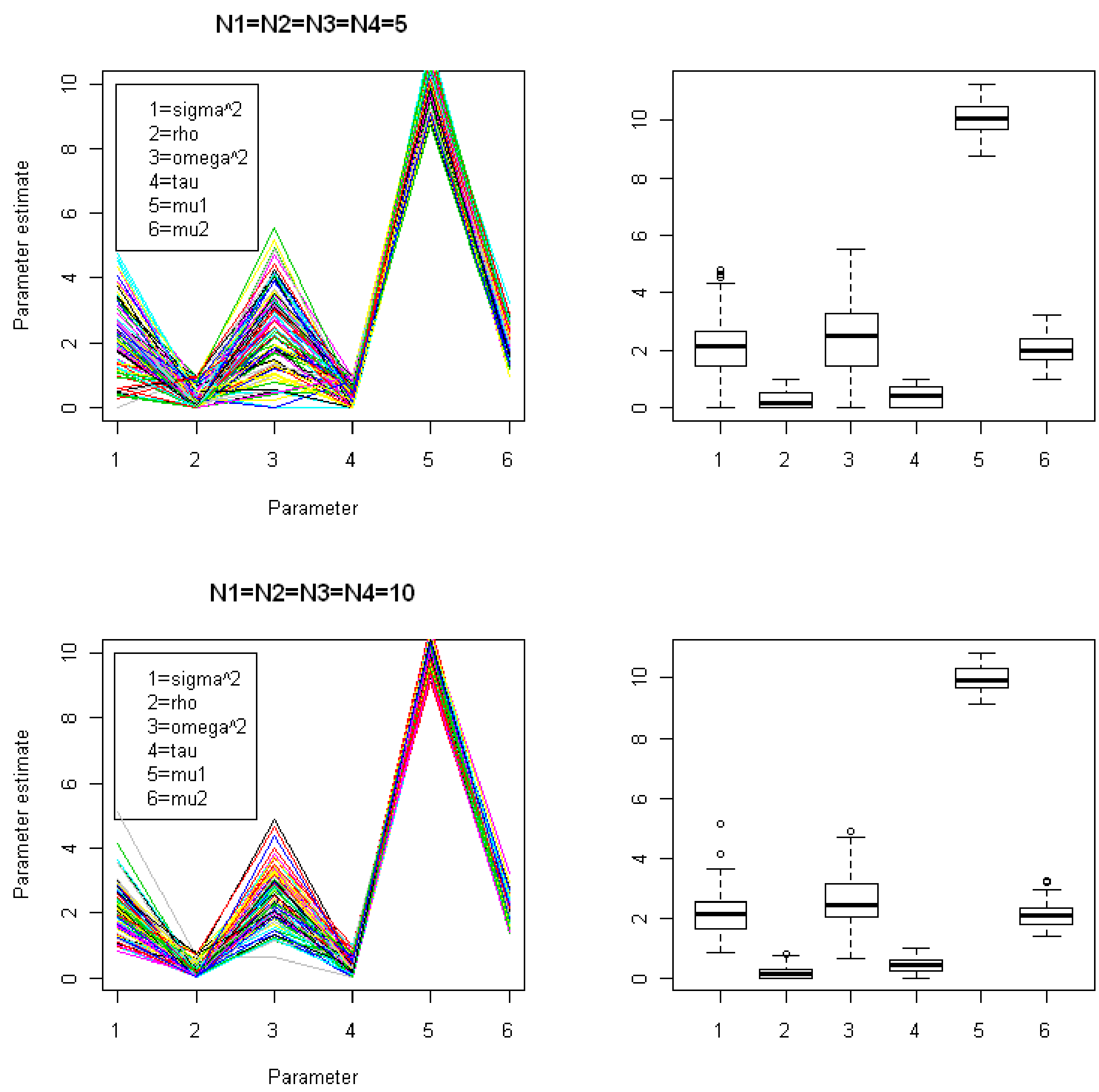

Figure 5: First simulation study using initial guess

$\vartheta_{Guess1} = \vartheta_{True} = (2, .1, 3, .5, 10, 2)'$ for $N = 5$ and 10

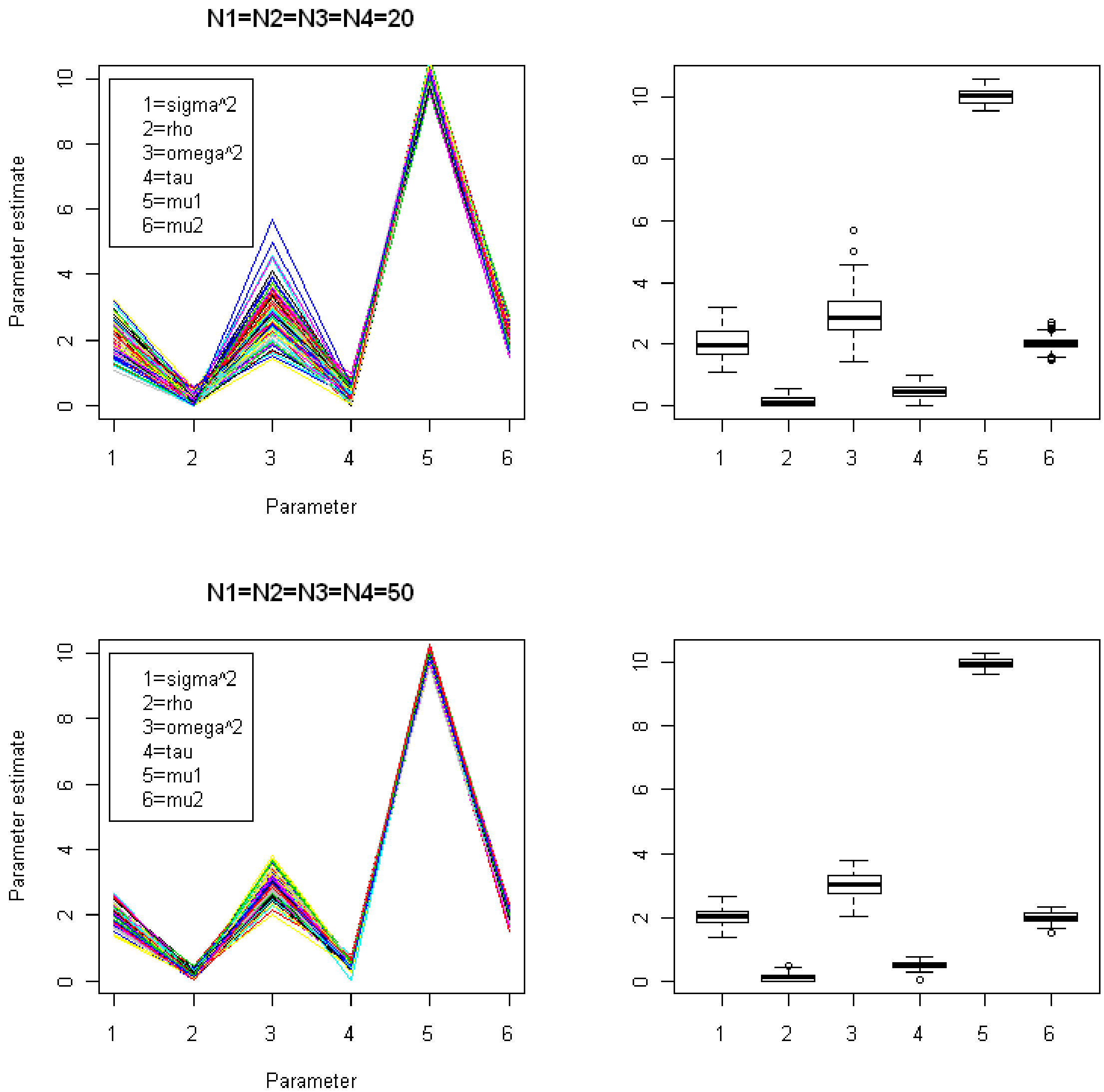

Figure 6: First simulation study using initial guess

$\vartheta_{Guess1} = \vartheta_{True} = (2, .1, 3, .5, 10, 2)'$ for $N = 20$ and 50

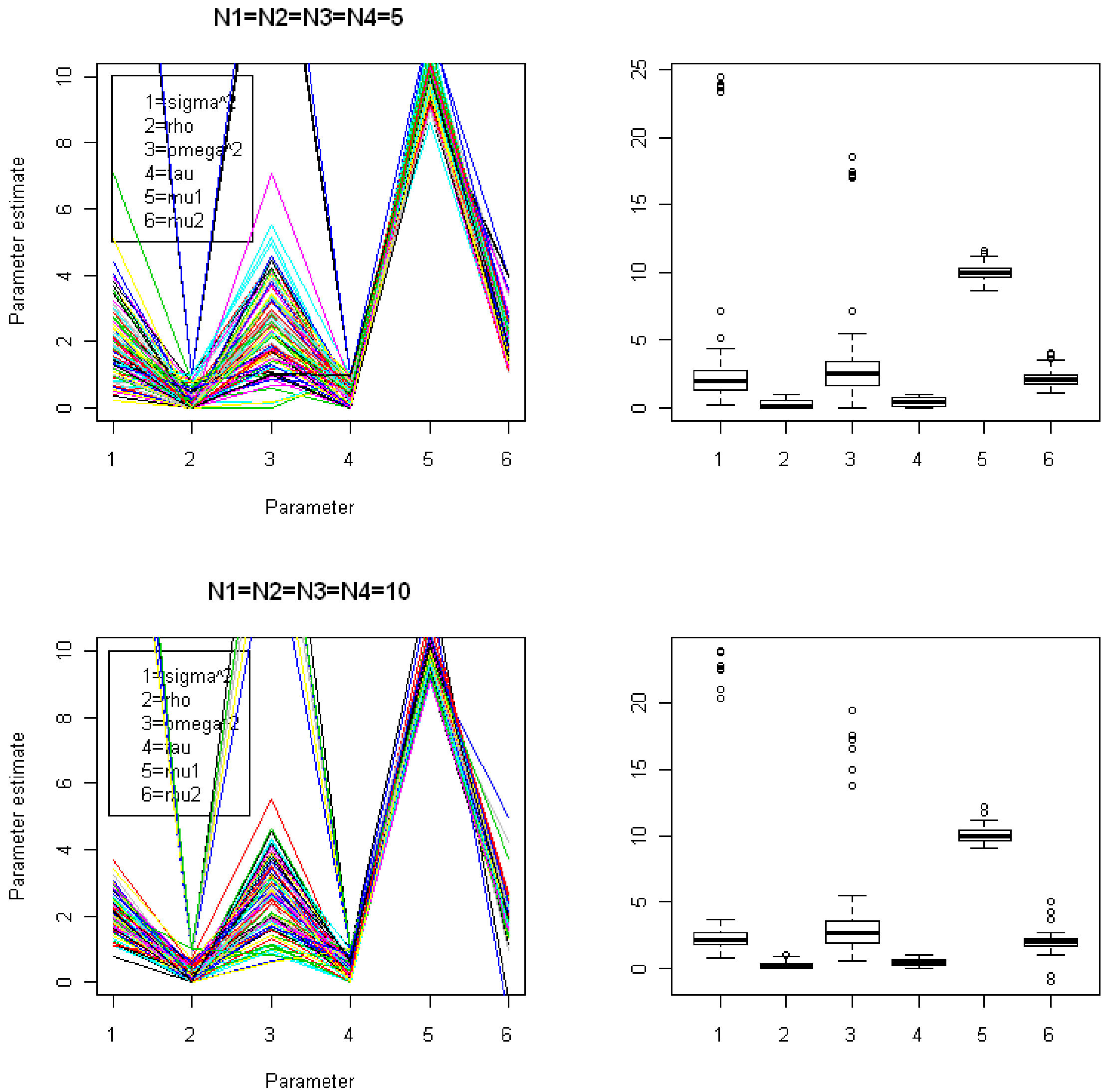

Figure 7: Second simulation study using initial guess

$\vartheta_{Guess2} = (20, 0, 6, 0, 1, 20)'$ for $N = 5$ and 10

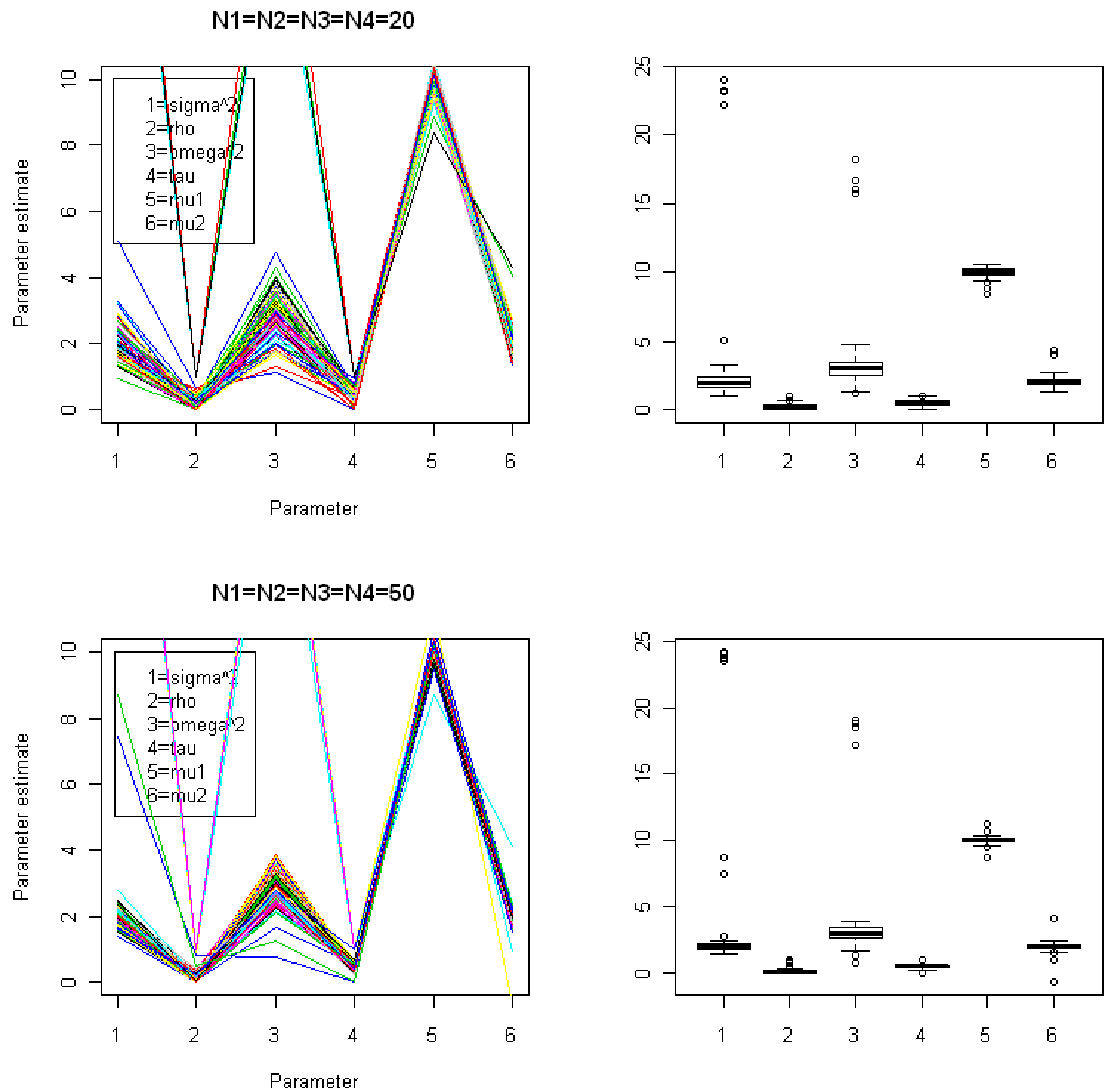

Figure 8: Second simulation study using initial guess

$\vartheta_{Guess2} = (20, 0, 6, 0, 1, 20)'$ for N = 20 and 50

CHAPTER V

CONCLUSIONS AND FUTURE WORK

5.1 CONCLUSIONS

For multivariate estimating and testing procedures, one uses the sample mean of the data $\boldsymbol{X}_1,...,\boldsymbol{X}_n$ to estimate the population mean assuming, for example, that the given sample is from a p variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; that is to assume that the sample is a set of independently, identically distributed $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ random vectors. When testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, one uses Hotelling's T^2 statistic without doubts. These are standard procedures for estimating and testing for the mean, while most of the time we obtain data that violate the "identically distributed" assumption. Some known or unknown disturbances may exist in the data which may be caused by specific mechanisms that are often neglected by analyzers. When this situation occurs, a heterogeneous means model should be employed. If the standard Hotelling's T^2 procedure is adopted in this case to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, one might expect that the rejection probability would be quite high even if the data just suffer slightly disturbed noise.

To remedy the violation of assumptions of the data stated above, assume that each $\boldsymbol{X}_i$ follows $N_p(\boldsymbol{C}_i\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution for $i = 1,...,n$. The disturbance of the data can be structured by the matrix $\boldsymbol{C}_i$ which is a $p \times p$ square matrix such that the vector $\boldsymbol{\mu}$ is meaningful. We name the vector $\boldsymbol{\mu}$ the core mean. To further look at the model, we can assume that the square matrix $\boldsymbol{C}_i$ is known or unknown, fixed or random.

The current research deals with the case that $\boldsymbol{C}_i$ is a known square matrix. When $\boldsymbol{C}_i$ is circulant for $i = 1,...,n$ and $\boldsymbol{\Sigma}$ has a compound symmetry structure, we can still do inference by generalizing the Hotelling's T^2 statistic.

The exact distribution of the ML estimator $\hat{\rho}$ of the intra-correlation is derived and it is distributed as a function of *F* random variable. In addition, it is not unbiased in general, yet it is unbiased when ρ is 0 and tends to be unbiased when ρ is close to $-(p-1)^{-1}$ or 1. So there is a need to do bias correction on $\hat{\rho}$ when $\rho \neq 0$. As for the inference about the MLE of σ^2, $\hat{\sigma}^2$ is exactly distributed like sum of two weighted chi-square random variables. An approximate χ^2 test is also derived for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$.

When extending the inference about the core mean for one sample case to comparing two core means for the two-sample case, i.e. to testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, the test statistic has been proved to be distributed as the random variable similar to the one in one sample case when testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$. An approximate χ^2 test is also derived for testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. When comparing k core means $\boldsymbol{\mu}_i$, $i = 1,...,k$ for k independent studies, the likelihood ratio test statistic for testing $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k$ has a null distribution which is analogous to the ones for one-sample and two-samples cases.

It is believed that the effect of commutability of the covariance matrix $\boldsymbol{\Sigma}$ and the $\boldsymbol{C}_i$ play a crucial role on the absence of the $\boldsymbol{C}_i$ in the null distribution of the LRT statistic for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ for one-sample data and analogous effect for $H_0 : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_k$ multi-sample data.

An application to meta analysis of the data from heterogeneous means models for *k* independent studies is addressed. Some preliminary results are presented as well. Results are synthesized from *k* studies, each study reports an estimate for the parameters of interest – the core mean. The fixed effect model assumes that the true core mean vector is the same for the *k* studies, so the ML estimator for the common core mean is obtained when the covariance matrix for each within-study model is known. If the covariance matrix for each study is unknown, a consistent estimator of the corresponding covariance matrix could be used to estimate the true unknown one, so the inference for the common core mean can be done, for both general case and the case that the covariance matrix for each study is compound symmetric and $\boldsymbol{C}_i$ is circulant for all *i*.

Two methods for random effects model are considered in the current research. The two-stage meta analysis considers both the within- and between- study models. The within-study model requires an estimate for the covariance matrix of the estimator of the core mean for each study, which is assumed fixed and known in the resulting two-stage marginal model involving the unknown common core mean and the unknown between-study covariance matrix. Unlike the two-stage meta analysis model, the one-stage meta analysis using the individual participant data (original data) to simultaneously do inference on the estimates for the common core mean and both within and between studies covariance matrices. A simulation study for a special case of the one-stage meta analysis is performed for finding the estimates for all the unknown parameters based on the derived marginal log likelihood function.

5.2 FUTURE WORK

For the future research, the case that the matrix $\boldsymbol{C}_i$ is random will be the starting point because it is more apt for real data. Consider the univariate case that $X_i \mid C_i \sim N(C_i\mu, \sigma^2)$, where μ, σ^2 are unknown parameters, C_i is random and $C_i \sim N(D_i, \nu^2)$, where D_i and ν^2 are known. Then the marginal pdf of X_i is

$$f(x_i) = \frac{1}{\sqrt{2\pi}\sqrt{\nu^2\mu^2+\sigma^2}} \exp\left\{-\frac{(x_i - D_i\mu)^2}{2(\nu^2\mu^2+\sigma^2)}\right\};$$

that is, $X_i \sim N(D_i\mu, \mu^2\nu^2+\sigma^2)$. Hence the marginal likelihood function is

$$L(\mu,\sigma^2) = \prod_{i=1}^{M} f(x_i)$$

and the corresponding log likelihood function is

$$\log L(\mu,\sigma^2) = -\frac{M}{2}\ln\left[2\pi(\nu^2\mu^2+\sigma^2)\right] - \frac{\sum_{i=1}^{M}(x_i - D_i\mu)^2}{2(\nu^2\mu^2+\sigma^2)},$$

on which the inference about μ could be based.

When extending it to multivariate case, one would expect that a great amount of computational calculation should be done when moving on to the general setting. The starting point for this would be the case that $\boldsymbol{C}_i$ are diagonal. Suppose that a sample $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_M$ is from $N_p(\boldsymbol{C}_i\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution. Define $\boldsymbol{X}_i = (X_{i1},\ldots,X_{ip})'$, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_p)'$, $\boldsymbol{C}_i = diag(c_{i1},\ldots,c_{ip})$, where $\boldsymbol{\Sigma}$ is unknown , c_{ij} are independently distributed like a normal random variable, for example, $N(d_{ij}, \nu^2)$, where ν^2 is specified. Note that $\boldsymbol{C}_i\boldsymbol{\mu}$ can be re-parameterized by $\Theta\boldsymbol{c}_i^*$, where $\Theta = diag(\mu_1,\ldots,\mu_p)$, $\boldsymbol{c}_i^* = (c_{i1},\ldots,c_{ip})'$, and $\boldsymbol{c}_i^* \sim N(\boldsymbol{d}_i, \nu^2\boldsymbol{I})$ implying that $\Theta\boldsymbol{c}_i^* \sim N(\Theta\boldsymbol{d}_i, \nu^2\Theta^2)$, where $\boldsymbol{d}_i = (d_{i1},\ldots,d_{ip})'$. Assume that the conditional pdf of $\boldsymbol{X}_i \mid \Theta\boldsymbol{c}_i^*$ is

$N(\Theta \boldsymbol{c}_i^*, \boldsymbol{\Sigma})$, then the marginal distribution of $\boldsymbol{X}_i$ is $N(\Theta \boldsymbol{d}_i, \nu^2\Theta^2 + \boldsymbol{\Sigma})$ which can be used to do inference for $\boldsymbol{\mu}$. The general case when $\boldsymbol{c}_i^*$ is multivariate normal is working in progress.

For a special bivariate case, let $\boldsymbol{C}_i = \begin{pmatrix} a_i & b_i \\ b_i & a_i \end{pmatrix}$ and assume that both a_i and b_i have a prior distribution, for instance, a Uniform (*a*, *b*) distribution. In addition to the normal case, some other continuous models like finite mixture normal models also will be considered in my future research. Testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{G\mu}_2$ for two heterogeneous normal samples will also be my interest.

REFERENCES

1. Anderson, T. W. (2003), "An Introduction to Multivariate Statistical Analysis", Third edition, *Wiley*.
2. Bailey, Ralph W. (1992), "Distribution Identities of Beta and Chi-Squared Variates: A Geometrical Interpretation", *The American Statistician*, Vol. 46, 2, 117-120.
3. Bain, Lee J., Engelhardt, Max (1992), "Introduction to Probability and Mathematical Statistics", 2nd edition, Duxbury.
4. Box, G. E. P. (1949), "A General Distribution Theory for a Class of Likelihood Criteria", *Biometrika*, 36, 317-346.
5. Box, G. E. P. (1954), "Some theorems on quadratic forms applied in the study if analysis of variance problems, I. Effect of Inequality of Variance In the One-Way Classification" *The Annals of Mathematical Statistics*, Vol. 25, 2, pp290-302.
6. Christensen, Ronald (2003), "Plane Answers to Complex Questions: The Theory of Linear Models", 3rd edition, Springer.
7. Crowder, M. J., Hand, D. J. (1990) "Analysis of Repeated Measures", *Chapman & Hall*.
8. Davis, Philip J. (1979), "Circulant Matrices", *New York: Wiley*.
9. Geisser, Seymour (1963), "Multivariate Analysis of Variance for A Special Covariance Case", *Journal of the American Statistical Association*, Vol. 58, 303, 660-669
10. Glass, G. V. (1976), "Primary, Secondary, and Meta-Analysis of Research", *Educational Researcher*, 5, 3-8.
11. Graybill, Franklin A. (1983), "Matrices with Applications in Statistics", 2nd edition, *Duxbury*

12. Higgins, Julian P. T.; Whitehead, Anne; Turner, Rebecca M.; Omar, Rumana Z.; Thompson, Simon G. (2001), "Meta-Analysis of Continuous Outcome Data From Individual Patients", *Statistics in Medicine*, 20, 2219-2241.

13. Jackson, Dan; Riley, Richard; White, Ian R. (2011), "Multivariate Meta-Analysis: Potential and Promise", *Statistics in Medicine, Willy Online Library*.

14. Kulinskaya, Elena; Morgenthaler, Stephan; Staudte, Robert G. (2008), "Meta analysis: A Guide to Calibrating and Combining Statistical Evidence". *Chichester: Wiley*.

15. Kullback, Solomon (1959), "Information Theory and Statistics", *New York: John Wiley & Sons, Inc*.

16. Magnus, Jan R.; Neudecker, H. (1979), "The Commutation Matrix: Some Properties and Applications", *The Annals of Statistics*, Vol. 7, 2, 381-394.

17. Morrison, Donald F. (1972), "The Analysis of A Single Sample of Repeated Measurements", *Biometrika*, 28, 55-71.

18. Nam, In-Sun; Mengersen, Kerri; Garthwaite, Paul (2003), "Multivariate Meta-Analysis", *Statistics in Medicine*, 22, 2309-2333.

19. Olkin, Ingram; Press, S. James (1969), "Testing and Estimation for a Circular Stationary Model", *The Annals o f Mathematical Statistics*, Vol. 40, 4, 1358-1373.

20. Rencher, Alvin C. (1998), "Multivariate Statistical Inference and Applications", *Wiley*.

21. Schott, James R. (1997), "Matrix Analysis for Statistics", *New York: John Wiley & Sons, Inc*.

22. Sutton, A. J.; Abrams, K. R.; Jones, D. R.; Sheldon, T. A.; Song, F. (2000), "Methods for Meta-analysis in Medical Research". *Chichester: Wiley*.

23. Sutton, Alexander J.; Higgins, Julian P. T. (2008), "Recent Developments In Meta-analysis", *Statistics in Medicine*, 27, 625-650.

24. Szatrowski, Ted H. (1982), "Testing and Estimation in the Block Compound Symmetry Problem", *Journal of Educational Statistics*, Vol. 7, 1, 3-18.

25. Tukey, J. W., Wilks, S. S. (1946), "Approximation of the Distribution of the Product of Beta Variables by A Single Beta Variable", *Annals of Mathematical Statistics*, Vol. 17, 318-324.

26. Votaw, David F., JR. (1948), "Testing Compound Symmetry in a Normal Multivariate Distribution", *Annals o f Mathematical Statistics*, Vol. 19, 447-473.

27. Wilks, Samuel. S. (1946), "Sample Criteria for Testing Equality of Means, Equality of Variances, and Equality of Covariances In a Normal Multivariate Distribution", *Annals of Mathematical Statistics*, Vol. 17, 3, 257-281.

APPPENDICES

A.1

We need to show that

(i) $tr\hat{\boldsymbol{\Sigma}}_0^{-1}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T = Mp$, and

(ii) $tr\hat{\boldsymbol{\Sigma}}^{-1}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T = Mp$

First show (i). Based on theories of trace of a matrix, we have

$$tr\hat{\boldsymbol{\Sigma}}_0^{-1}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T = tr\frac{1}{\hat{\sigma}_0^{\ 2}(1-\hat{\rho}_0)}[\boldsymbol{I}_p - \frac{\hat{\rho}_0}{1+(p-1)\hat{\rho}_0}\boldsymbol{J}_p]\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T$$

$$= \frac{1}{\hat{\sigma}_0^{\ 2}(1-\hat{\rho}_0)}tr\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T - \frac{1}{\hat{\sigma}_0^{\ 2}(1-\hat{\rho}_0)}\frac{\hat{\rho}_0}{1+(p-1)\hat{\rho}_0}tr\boldsymbol{J}_p\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T$$

$$= \frac{1}{\hat{\sigma}_0^{\ 2}(1-\hat{\rho}_0)}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0) - \frac{1}{\hat{\sigma}_0^{\ 2}(1-\hat{\rho}_0)}\frac{\hat{\rho}_0}{1+(p-1)\hat{\rho}_0}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0).$$

Since from Section 2.2.2 we have

$$\hat{\sigma}_0^{\ 2} = \frac{1}{Mp}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0),$$

and

$$\hat{\rho}_0 = \frac{1}{p-1}\left[\frac{\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)}{\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)} - 1\right],$$

we have

$$\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)=Mp\hat{\sigma}_0{}^2,$$

and

$$\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)=\left[1+(p-1)\hat{\rho}_0\right]Mp\hat{\sigma}_0{}^2.$$

Therefore, $tr\hat{\boldsymbol{\Sigma}}_0{}^{-1}\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T$ becomes

$$\begin{aligned}
&tr\hat{\boldsymbol{\Sigma}}_0{}^{-1}\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T\\
&=\frac{1}{\hat{\sigma}_0{}^2(1-\hat{\rho}_0)}\cdot Mp\hat{\sigma}_0{}^2-\frac{1}{\hat{\sigma}_0{}^2(1-\hat{\rho}_0)}\frac{\hat{\rho}_0}{1+(p-1)\hat{\rho}_0}\left[1+(p-1)\hat{\rho}_0\right]Mp\hat{\sigma}_0{}^2\\
&=\frac{Mp}{(1-\hat{\rho}_0)}-\frac{Mp\hat{\rho}_0}{(1-\hat{\rho}_0)}=Mp.
\end{aligned}$$

Similarly, the identity $tr\hat{\boldsymbol{\Sigma}}^{-1}\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T=Mp$ can be obtained in the same way.

A.2

The likelihood ratio for testing $H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0$ in (2.13) can also be expressed as

$$\lambda=\left(\frac{[\hat{\sigma}^2(1-\hat{\rho})]^{p-1}\left\{\hat{\sigma}^2[1+(p-1)\hat{\rho}]\right\}}{[\hat{\sigma}_0{}^2(1-\hat{\rho}_0)]^{p-1}\left\{\hat{\sigma}_0{}^2[1+(p-1)\hat{\rho}_0]\right\}}\right)^{\frac{M}{2}}.$$

Since we have from Section 2.2.2 that

$$\hat{\sigma}_0{}^2=\frac{1}{Mp}\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0),$$

$$\hat{\rho}_0=\frac{1}{p-1}\left[\frac{\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)}{\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{x}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)}-1\right],$$

$$\hat{\sigma}^2 = \frac{1}{Mp}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}),$$

and

$$\hat{\rho} = \frac{1}{p-1}\left[\frac{\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})}{\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})} - 1\right],$$

Substituting the above expressions for $\hat{\sigma}_0{}^2$, $\hat{\rho}_0$, $\hat{\sigma}^2$, and $\hat{\rho}$ into (2.13) we have

$$\begin{aligned}\hat{\sigma}^2(1-\hat{\rho}) = \hat{\sigma}^2 - \hat{\sigma}^2\hat{\rho} &= \frac{1}{Mp}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) \\ &\quad - \frac{1}{Mp(p-1)}\left[\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) - \sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})\right] \\ &= \frac{1}{Mp}\left[p\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) - \sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})\right]\end{aligned}$$

and

$$\begin{aligned}\hat{\sigma}^2[1+(p-1)\hat{\rho}] &= \hat{\sigma}^2 + (p-1)\hat{\sigma}^2\hat{\rho} \\ &= \frac{1}{Mp}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) \\ &\quad + \frac{1}{Mp}\left[\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}) - \sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})\right] \\ &= \frac{1}{Mp}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{x}_i - \boldsymbol{C}_i\hat{\boldsymbol{\mu}}).\end{aligned}$$

Likewise, we have

$$\hat{\sigma}_0{}^2(1-\hat{\rho}_0) = \frac{1}{Mp}\left[p\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0) - \sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)\right]$$

and

$$\hat{\sigma}_0{}^2[1+(p-1)\hat{\rho}_0] = \frac{1}{Mp}\sum_{i=1}^{M}(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{x}_i - \boldsymbol{C}_i\boldsymbol{\mu}_0).$$

Therefore the likelihood ratio test is to reject H_0 if

$$\frac{\left[p\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})-\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})\right]^{p-1}\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}})}{\left[p\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)-\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)\right]^{p-1}\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T\boldsymbol{J}_p(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)}<C_\alpha.$$

A.3

We need to show that

$$\exp\left(-\frac{1}{2}\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T\hat{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)-\frac{1}{2}\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T\hat{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)\right)$$
$$=\exp\left\{-\frac{1}{2}(M+N)p\right\},$$

and

$$\exp\left(-\frac{1}{2}\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x)^T\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x)-\frac{1}{2}\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y)^T\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y)\right)$$
$$=\exp\left\{-\frac{1}{2}(M+N)p\right\}.$$

Proof:

$$\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T\hat{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)+\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T\hat{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)$$
$$=tr\hat{\boldsymbol{\Sigma}}_0^{-1}\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T+tr\hat{\Sigma}_0^{-1}\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T$$
$$=tr\hat{\boldsymbol{\Sigma}}_0^{-1}\left\{\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T+\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T\right\}$$
$$=tr\,\frac{1}{\hat{\sigma}_0{}^2(1-\hat{\rho}_0)}[\boldsymbol{I}_p-\frac{\hat{\rho}_0}{1+(p-1)\hat{\rho}_0}\boldsymbol{J}_p]\cdot$$
$$\left\{\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T+\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T\right\}$$
$$=\frac{1}{\hat{\sigma}_0{}^2(1-\hat{\rho}_0)}\cdot tr\left\{\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T+\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T\right\}$$
$$-\frac{1}{\hat{\sigma}_0{}^2(1-\hat{\rho}_0)}\frac{\hat{\rho}_0}{1+(p-1)\hat{\rho}_0}\cdot tr\boldsymbol{J}_p\left\{\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T+\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T\right\}$$

$$=\frac{1}{\hat{\sigma}_0{}^2(1-\hat{\rho}_0)}\left\{\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)+\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)\right\}$$

$$-\frac{1}{\hat{\sigma}_0{}^2(1-\hat{\rho}_0)}\frac{\hat{\rho}_0}{1+(p-1)\hat{\rho}_0}\left\{\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{J}_p(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)+\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T\boldsymbol{J}_p(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)\right\}$$

$$=\frac{1}{\hat{\sigma}_0{}^2(1-\hat{\rho}_0)}\cdot(M+N)p\hat{\sigma}_0{}^2-\frac{1}{\hat{\sigma}_0{}^2(1-\hat{\rho}_0)}\frac{\hat{\rho}_0}{1+(p-1)\hat{\rho}_0}\left[1+(p-1)\hat{\rho}_0\right](M+N)p\hat{\sigma}_0{}^2$$

$$=\frac{(M+N)p}{(1-\hat{\rho}_0)}-\frac{(M+N)p\hat{\rho}_0}{(1-\hat{\rho}_0)}=(M+N)p.$$

Similar argument can be used when showing

$$\exp\left(-\frac{1}{2}\sum_{i=1}^{M}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x)^T\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_x)-\frac{1}{2}\sum_{j=1}^{N}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y)^T\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_y)\right)$$
$$=\exp\left\{-\frac{1}{2}(M+N)p\right\}.$$

A.4

We need to show that

$$\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)+\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_0)$$
$$=\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)$$
$$+\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)$$
$$+(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T\left[\left(\sum_{i=1}^{M}\boldsymbol{C}_i{}^T\boldsymbol{C}_i+\sum_{j=1}^{N}\boldsymbol{D}_j{}^T\boldsymbol{D}_j\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)\right](\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0).$$

where $\hat{\boldsymbol{\mu}}_0=\left(\sum_{i=1}^{M}\boldsymbol{C}_i{}^T\boldsymbol{C}_i+\sum_{j=1}^{N}\boldsymbol{D}_j{}^T\boldsymbol{D}_j\right)^{-1}\left(\sum_{i=1}^{M}\boldsymbol{C}_i{}^T\boldsymbol{X}_i+\sum_{j=1}^{N}\boldsymbol{D}_j{}^T\boldsymbol{Y}_j\right).$

We first rewrite $\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)$ as

$$\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)$$
$$+2(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)$$
$$+(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0).$$

Note that the last quantity is expressed so because $\boldsymbol{C}_i$ and $(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)$ commute due to the fact that both of them are circulant matrices. Similarly, we have

$$\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_0)=\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)$$
$$+2(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)\sum_{j=1}^{N}\boldsymbol{D}_j^{\,T}(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)$$
$$+(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T\left(\sum_{j=1}^{N}\boldsymbol{D}_j^{\,T}\boldsymbol{D}_j\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0).$$

Therefore, we arrive at

$$\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i-\boldsymbol{C}_i\boldsymbol{\mu}_0)+\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j-\boldsymbol{D}_j\boldsymbol{\mu}_0)$$
$$=\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)+\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)$$
$$+2(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)\left[\left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{X}_i+\sum_{j=1}^{N}\boldsymbol{D}_j^{\,T}\boldsymbol{Y}_j\right)-\left(\sum_{jj=1}^{N}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i+\boldsymbol{D}_j\boldsymbol{D}_j^{\,T}\right)\hat{\boldsymbol{\mu}}_0\right]$$
$$+(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i+\sum_{j=1}^{N}\boldsymbol{D}_j^{\,T}\boldsymbol{D}_j\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)$$
$$=\sum_{i=1}^{M}(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{X}_i-\boldsymbol{C}_i\hat{\boldsymbol{\mu}}_0)+\sum_{j=1}^{N}(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)^T(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\boldsymbol{Y}_j-\boldsymbol{D}_j\hat{\boldsymbol{\mu}}_0)$$
$$+(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)^T\left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i+\sum_{j=1}^{N}\boldsymbol{D}_j^{\,T}\boldsymbol{D}_j\right)(\boldsymbol{I}_p-\frac{1}{p}\boldsymbol{J}_p)(\hat{\boldsymbol{\mu}}_0-\boldsymbol{\mu}_0)$$

due to the fact that $\left(\sum_{i=1}^{M}\boldsymbol{C}_i^{\,T}\boldsymbol{X}_i+\sum_{j=1}^{N}\boldsymbol{D}_j^{\,T}\boldsymbol{Y}_j\right)-\left(\sum_{jj=1}^{N}\boldsymbol{C}_i^{\,T}\boldsymbol{C}_i+\boldsymbol{D}_j\boldsymbol{D}_j^{\,T}\right)\hat{\boldsymbol{\mu}}_0=0$.

VITA

LI-CHI LIN

Candidate for the Degree of

Doctor of Philosophy

Thesis: MULTIVARIATE NORMAL INFERENCE FOR HETEROGENEOUS SAMPLES AND AN APPLICATION TO META ANALYSIS

Major Field: Statistics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Statistics at Oklahoma State University, Stillwater, Oklahoma, July 2012.

Completed the requirements for the Master of Science in Statistics at National Central University, Chungli, Taoyuan, Taiwan, June 1993.

Completed the requirements for the Bachelor of Science/Arts in Mathematics at Chungyuan Christian University, Chungli, Taoyuan, Taiwan, June 1990.

Experience:

- Instructor at Providence University, Taiwan, Jan. 2005 – Dec. 2006
  Teaching courses: Biostatistics

- President of Statistics Club, Oklahoma State University, 2008 – 2009

- Teaching Assistant: Aug. 2007 – May 2010
  SAS programming and Linear Models I and II

- Instructor of Record at Oklahoma State University, June 2010 – July 2012
  Teaching courses:
    - Elementary Statistics for Business and Economics
    - Elementary Statistics
    - Applied Regression Analysis

- Invited speaker at AMS meeting, Lawrence, Kansas, March 2012

Professional Memberships:

American Statistical Association, 2010-present

Name: Li-Chi Lin Date of Degree: July, 2012

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: MULTIVARIATE NORMAL INFERENCE FOR HETEROGENEOUS SAMPLES AND AN APPLICATION TO META ANALYSIS

Pages in Study: 128 Candidate for the Degree of Doctor of Philosophy

Major Field: Statistics

Scope and Method of Study:

When extending likelihood inference in the case of the normal distribution to heterogeneous samples, one discovers that this is easily done in the univariate case but is prohibitive in the multivariate cases.

In the current work, the exact maximum likelihood estimates for the core mean and the covariance matrix are obtained for samples of different means but with core parameter vector and an unknown covariance matrix but a structured one. Then the celebrated Hotelling's T-square statistic is generalized to this case where the exact null distribution is derived. The approximate Chi-Square test is then obtained as well.

Next, we derive analogous results in the *k*-sample situation. The generalized Hotelling T-square statistic developed allows us to proceed to testing hypotheses in the one-way multivariate ANOVA when samples are heterogeneous. A cutting edge application of this work is its introduction to multivariate meta analysis approaches for multivariate heterogeneous data for the first time.

Findings and Conclusions:

The one-sample and multi-sample inferences for testing the core parameter vector(s) using likelihood ratio test approaches when the covariance matrix has compound symmetry are obtained. The null distribution of the LRT statistic for each case is derived exactly. Both of them follow a distribution of a powered function of two independent beta random variables but different forms. The exact distribution of the ML estimator of the intra-correlation is derived for one-sample case and it is distributed as a function of *F* random variable. As for the inference about the MLE of the common variance, it is exactly distributed as sum of two weighted chi-square random variables. An approximate Chi-square test is also derived for testing the core parameter vector(s) for each of one-sample and two-sample cases.

An application to multivariate meta analysis of the heterogeneous data for *k* independent studies is addressed. Both fixed and random effects models are investigated. A simulation study for inference of the overall core parameter vector is performed when data are heterogeneous for one-stage random effects model.

ADVISER'S APPROVAL: Dr. Ibrahim A. Ahmad