

# Likelihood Ratio Tests of Correlated Multivariate Samples

**Johan Lim**

Department of Statistics, Seoul National University, Seoul, 151-747, Korea

**Erning Li**

Department of Statistics, Texas A&M University, College Station, TX 77843, USA

**Shin-Jae Lee**

Department of Orthodontics, School of Dentistry and Dental Research Institute,  
Seoul National University, Seoul, 110-768, Korea

## **Abstract**

We develop methods to compare multiple multivariate normally distributed samples which may be correlated. The methods are new in the context that no assumption is made about the correlations among the samples. Three types of null hypothesis are considered: equality of mean vectors, homogeneity of covariance matrices, and equality of both mean vectors and covariance matrices. We demonstrate that the likelihood ratio test statistics have finite-sample distributions that are functions of two independent Wishart variables and dependent on the covariance matrix of the combined multiple populations. Asymptotic calculations show that the likelihood ratio test statistics converge in distribution to central Chi-squared distributions under the null hypotheses regardless how the populations are correlated. Following these theoretical findings, we propose a resampling procedure for the implementation of the likelihood ratio tests in which no restrictive assumption is imposed on the structures of the covariance matrices. The empirical size and power of the test procedure are investigated for various sample

sizes via simulations. Two examples are provided for illustration. The results show good performance of the methods in terms of test validity and power.

*Keywords:* Correlated samples; Empirical rejection probability; Equality of mean vectors; Homogeneity of covariance matrices; Multivariate analysis; Resampling.

## 1 Introduction

Multivariate data that consist of sets of measurements on a number of individuals or objects are collected in many areas of application, and comparing population (or treatment) mean vectors and/or covariance matrices is often of interest. In some cases it may not be realistic to assume that the samples are independent; one obvious example is experimental situations in which natural paired data are observed on the same set of subjects. For instance, in a dental health study on tooth size balance, measurements of tooth sizes on left side from the central incisors and those on the right side may be correlated because the observations were collected from the same group of patients. Another example is a microarray study which consists of gene expression profiles at two different stages of cancer, benign tumor and primary tumor. The two samples may be correlated due to the fact that both stages were acquired from most cancer patients in the study. Moreover, in many instances, samples are correlated due to various reasons other than an obvious “pairing” factor. For example, it is common to see within-family correlation of observations in some educational psychology studies. Problems that motivate this study of correlated samples arise in many applications such as those in the areas of medicine, psychology, environmental science, and economics.

To outline the testing problems to be investigated, we suppose samples of the same size  $n$  are drawn from  $t$  ( $> 1$ ) multivariate normal populations each with  $p$  variables. Let  $Y_{ijv}$  represent the measurement of the  $v^{\text{th}}$  variable on the  $j^{\text{th}}$  subject in sample  $i$ ,  $i = 1, \dots, t$  (samples),  $j = 1, \dots, n$  (subjects), and  $v = 1, \dots, p$  (variables). The samples are not necessarily independent from one another. For instance, part or all of the subjects can be the same across samples. For the  $i^{\text{th}}$  sample, the vectors  $\mathbf{Y}_{ij} \equiv (Y_{ij1}, \dots, Y_{ijp})^T$  for  $j = 1, \dots, n$ ,

are assumed to follow a  $p$ -variate normal distribution,  $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ , with population-specific mean vector  $\boldsymbol{\mu}_i$  and covariance matrix  $\boldsymbol{\Sigma}_i$ . We consider the following hypotheses.

(i) Testing equality of mean vectors, i.e.,

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_t \text{ versus } H_A \text{ which does not so restrict } \boldsymbol{\mu}_i\text{'s};$$

(ii) Testing equality of covariance matrices, i.e.,

$$H_0 : \boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_t \text{ versus } H_A \text{ which does not so restrict } \boldsymbol{\Sigma}_i\text{'s};$$

(iii) Testing simultaneous equality of both mean vectors and covariance matrices, i.e.,

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_t \text{ and } \boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_t \text{ versus } H_A \text{ which does not so restrict } \boldsymbol{\mu}_i\text{'s} \\ \text{and/or } \boldsymbol{\Sigma}_i\text{'s}.$$

Hypothesis tests about population means and covariance matrices for multivariate normal data have received much attention in the literature. Most of these studies assume that the samples are independent (among many others see, for example, Szatrowski, 1979; Perlman, 1980; Conover, Johnson and Johnson, 1981; Anderson, 2003, Chapters 8 and 10). Several authors have considered the problems of comparing correlated samples. Harris (1985) proposed methods for testing the equality of variances of correlated univariate normal populations, i.e., testing the null hypothesis (ii) when  $p = 1$ . Han (1968) and Choi and Wette (1972) studied special cases of the testing problem (ii) for the case  $p = 1$ , namely, that some assumptions are made on the correlation structure among the univariate normal populations.

In this article, we study the likelihood-ratio-type tests (LRTs) of aforementioned hypotheses (i)—(iii) for multivariate normal populations (i.e.,  $p \geq 1$ ). Neither independence assumption among populations nor assumption on the correlation structure among populations is imposed. Moreover, the covariance structure is not specified of any particular form for a given population. The new aspect and challenge come from the dependence among populations as well as the multivariate nature of data. We show that the finite-sample distributions of the LRT statistics are of some complicated form. Furthermore, we prove that the asymptotic distributions of the LRT statistics are central Chi-squared distributions under the null hypotheses. Besides these properties, the LRT statistics are generally not invariant

with respect to permutation transformations of the samples, and hence classical permutation technique is not applicable to the finite-sample null distributions. These considerations lead us to propose a parametric bootstrap (or resampling) procedure via substituting sample covariance matrices, which is easy to implement.

The rest of the article is organized as follows. In Section 2, we describe the LRTs for the hypotheses under consideration. There, results on the finite-sample properties are given together with results on the derivation of asymptotic distributions. The resampling procedure is provided for computing the null distributions and conducting the tests. Section 3 contains simulation studies on the performances of the proposed test procedure under a variety of null and alternative hypotheses as well as different sample sizes. In Section 4, we apply the proposed methods to a clinical study on human tooth size and a microarray data set.

## 2 Likelihood Ratio Test Statistics and Properties

### 2.1 LRT Statistics

Following the notations in Section 1, the  $p$ -variate vector of observations on the  $j^{\text{th}}$  subject in the  $i^{\text{th}}$  sample is  $\mathbf{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijp})^T$ ,  $j = 1, \dots, n$ , and  $i = 1, \dots, t$ ; and the  $i^{\text{th}}$  random sample  $\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{in}$  follows the  $p$ -variate normal distribution  $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ . Based on this setting and combining the  $p$ -variate vectors of observations on the  $j^{\text{th}}$  subjects from all  $t$  samples, we let  $\mathbf{Y}_{\cdot j} = (\mathbf{Y}_{1j}^T, \dots, \mathbf{Y}_{tj}^T)^T$ ,  $j = 1, \dots, n$ , which is assumed to be independent and identically distributed from a multivariate normal distribution  $\mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Omega})$ . It is easily seen that the mean vector is a  $(tp)$ -dimensional column vector  $\boldsymbol{\eta} = (\boldsymbol{\mu}_1^T, \dots, \boldsymbol{\mu}_t^T)^T$  and the  $(tp) \times (tp)$  positive-definite covariance matrix  $\boldsymbol{\Omega}$  has diagonal  $p \times p$  matrices equal to  $\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_t$ . Notice that dependence is allowed among samples. Furthermore, no assumption is made about the correlation structure among the populations. Treating  $\mathbf{Y}_{\cdot 1}, \dots, \mathbf{Y}_{\cdot n}$  as a “new” sample, we

can represent the complete data from all  $t$  individual samples by the  $n \times (tp)$  data matrix

$$\mathbf{Y} = (\mathbf{Y}_{\cdot 1}, \dots, \mathbf{Y}_{\cdot n})^T = \begin{pmatrix} \mathbf{Y}_{11}^T & \mathbf{Y}_{21}^T & \dots & \mathbf{Y}_{t1}^T \\ \mathbf{Y}_{12}^T & \mathbf{Y}_{22}^T & \dots & \mathbf{Y}_{t2}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Y}_{1n}^T & \mathbf{Y}_{2n}^T & \dots & \mathbf{Y}_{tn}^T \end{pmatrix}.$$

The log-likelihood function is

$$\ell(\boldsymbol{\eta}, \boldsymbol{\Omega} | \mathbf{Y}) = \sum_{j=1}^n \left\{ -\frac{tp}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2} (\mathbf{Y}_{\cdot j} - \boldsymbol{\eta})^T \boldsymbol{\Omega}^{-1} (\mathbf{Y}_{\cdot j} - \boldsymbol{\eta}) \right\},$$

and the LRT statistic for testing  $H_0$  versus  $H_A$  is defined as

$$\text{LRT}_n = 2 \sup_{H_A} \ell(\boldsymbol{\eta}, \boldsymbol{\Omega} | \mathbf{Y}) - 2 \sup_{H_0} \ell(\boldsymbol{\eta}, \boldsymbol{\Omega} | \mathbf{Y}).$$

Large value of  $\text{LRT}_n$  is in favor of  $H_A$ . Some simple algebra shows that the LRT statistics for testing the hypotheses (i)—(iii) are approximately proportional to

$$\text{M}_n = n \log \left( \frac{|SS_M|}{|SS_A|} \right), \quad \text{V}_n = n \log \left( \frac{|SS_V|}{|SS_A|} \right), \quad \text{W}_n = n \log \left( \frac{|SS_W|}{|SS_A|} \right), \quad (1)$$

respectively, where  $SS_A$  is the sum of squares for the sample covariance matrix of the “new” sample  $\mathbf{Y}_{\cdot 1}, \dots, \mathbf{Y}_{\cdot n}$  when no assumption is made about  $\boldsymbol{\mu}_i$ 's or  $\boldsymbol{\Sigma}_i$ 's;  $SS_M$  is the analogous sum of squares under the assumption of equal mean vectors  $\boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_t$  (null hypothesis (i));  $SS_V$  is the analogous sum of squares under the assumption of equal covariance matrices  $\boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_t$  (null hypothesis (ii)); and  $SS_W$  is the analogous sum of squares under the assumption of both equal mean vectors and equal covariance matrices (null hypothesis (iii)). One may prefer to use sample covariance matrices instead of these sums of squares; nothing discussed here changes in any essential way and results remain the same.

First, we shall determine suitable expressions for  $SS_A$ ,  $SS_M$ ,  $SS_V$  and  $SS_W$ . When no restriction is imposed on  $\boldsymbol{\mu}_i$ 's or  $\boldsymbol{\Sigma}_i$ 's, the maximum likelihood estimator for  $E(\mathbf{Y})$  is  $n^{-1} \mathbf{J}_n \mathbf{Y}$ , where  $\mathbf{J}_n$  denotes the  $n \times n$  matrix with all entries equal to 1, and thus

$$SS_A = (\mathbf{Y} - n^{-1} \mathbf{J}_n \mathbf{Y})^T (\mathbf{Y} - n^{-1} \mathbf{J}_n \mathbf{Y}) = \mathbf{Y}^T (\mathbf{I}_n - n^{-1} \mathbf{J}_n) \mathbf{Y},$$

where  $\mathbf{I}_n$  means the  $n \times n$  identity matrix. Under the assumption  $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_t$ , one can derive that the maximum likelihood estimator for  $E(\mathbf{Y})$  is  $n^{-1}\mathbf{J}_n\mathbf{Y}\mathbf{A}$ , where  $\mathbf{A} = t^{-1}\mathbf{J}_t \otimes \mathbf{I}_p$  with  $\otimes$  as the notation for a Kronecker (or direct) product, and hence

$$\begin{aligned} SS_M &= (\mathbf{Y} - n^{-1}\mathbf{J}_n\mathbf{Y}\mathbf{A})^\top (\mathbf{Y} - n^{-1}\mathbf{J}_n\mathbf{Y}\mathbf{A}) \\ &= \mathbf{Y}^\top (\mathbf{I}_n - n^{-1}\mathbf{J}_n) \mathbf{Y} + n^{-1}(\mathbf{I}_{tp} - \mathbf{A})\mathbf{Y}^\top \mathbf{J}_n \mathbf{Y}(\mathbf{I}_{tp} - \mathbf{A}). \end{aligned}$$

Introduce the operator  $\mathbf{h}(\cdot)$  such that for any  $(tp) \times (tp)$  matrix  $SS$ ,  $\mathbf{h}(SS)$  has  $t$  diagonal  $p \times p$  matrices that are identical and equal to the arithmetic average of the  $t$  diagonal  $p \times p$  matrices of  $SS$ , i.e., equal to  $t^{-1} \sum_{k=1}^t SS^{[k,k]}$  where  $SS^{[k,k]}$  denotes the  $k^{\text{th}}$   $p \times p$  diagonal matrix of  $SS$ ; and the rest components in  $\mathbf{h}(SS)$  are the same as those in  $SS$ . Under the assumption  $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_t$ , we let

$$SS_V = \mathbf{h}(SS_A).$$

Similarly, when assuming  $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_t$  and  $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_t$ , we express

$$SS_W = \mathbf{h}(SS_M).$$

One remark is that in (1), each of  $V_n$  and  $W_n$  has an additional random term that is trace  $\{SS_A \mathbf{h}(SS_A)^{-1}\} - tp$  and trace  $\{SS_M \mathbf{h}(SS_M)^{-1}\} - tp$ , respectively. We ignore these terms in the test statistics because they are negligible in comparison to the term remained and they converge in probability to zero under the null hypotheses. Another remark is that a sum of squares is not necessarily positive-definite and when this happens, the procedure given by Bock and Petersen (1975) is used to construct an estimated covariance matrix that is at least positive-semidefinite. Also, notice that under the null hypotheses, the probability that  $SS_A$  or  $SS_M$  is singular decreases to 0 as  $n \rightarrow \infty$ .

## 2.2 Properties of LRT Statistics

We now investigate the finite-sample and asymptotic properties of the LRT statistics by demonstrating two theorems. Note that  $SS_A$ ,  $SS_M$ ,  $SS_V$ , and  $SS_W$  are all invariant with

respect to  $\boldsymbol{\eta}$  under the equal mean vectors assumption  $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_t$ . Consequently, without loss of generality, we assume  $\boldsymbol{\eta} = \mathbf{0}$  in the remainder of this section.

The following theorem characterizes the finite-sample distributions of the LRT statistics  $M_n$ ,  $V_n$  and  $W_n$ . The theorem involves Wishart distributions. A positive definite  $q \times q$  symmetric matrix of random variables, denoted by  $\mathbf{X}$  is said to have the Wishart distribution with parameters  $\boldsymbol{\Psi}$ ,  $n$ , and  $q$ , if the probability density function of  $\mathbf{X}$  is

$$f(\mathbf{x}) = \frac{|\mathbf{x}|^{(n-q-1)/2} \exp\{-\text{tr}(\mathbf{x}\boldsymbol{\Psi}^{-1})/2\}}{2^{nq/2} \pi^{q(q-1)/4} |\mathbf{x}|^{n/2} \prod_{k=1}^q \Gamma\{(n+1-k)/2\}}, \quad n \geq q,$$

where  $\boldsymbol{\Psi}$  is a fixed positive definite matrix of size  $q \times q$ , and  $\Gamma(\cdot)$  stands for the gamma function. In short, we write  $\mathbf{X} \sim \text{Wishart}(\boldsymbol{\Psi}, n, q)$ .

**Theorem 1.** *The LRT statistic  $M_n$  for testing the equality of mean vectors in (i) is distributed as*

$$n (\log |\mathbf{C}_n + \mathbf{D}| - \log |\mathbf{C}_n|),$$

where  $\mathbf{C}_n$  and  $\mathbf{D}$  are independently distributed as  $\text{Wishart}(\mathbf{I}_{tp}, n-1, tp)$  and  $\text{Wishart}\{\boldsymbol{\Omega}^{-1/2}(\mathbf{I}_{tp} - \mathbf{A})^T \boldsymbol{\Omega}(\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega}^{-1/2}, 1, tp\}$ , respectively. The LRT statistic  $V_n$  for testing the equality of covariance matrices in (ii) is distributed as

$$n \left\{ \log |\boldsymbol{\Omega}^{-1/2} \mathbf{h}(\boldsymbol{\Omega}^{1/2} \mathbf{C}_n \boldsymbol{\Omega}^{1/2}) \boldsymbol{\Omega}^{-1/2}| - \log |\mathbf{C}_n| \right\}.$$

The LRT statistic  $W_n$  for testing the simultaneous equality of mean vectors and equality of covariance matrices in (iii) is distributed as

$$n \left[ \log |\boldsymbol{\Omega}^{-1/2} \{\mathbf{h}(\boldsymbol{\Omega}^{1/2} \mathbf{C}_n \boldsymbol{\Omega}^{1/2}) + \mathbf{h}(\boldsymbol{\Omega}^{1/2} \mathbf{D} \boldsymbol{\Omega}^{1/2})\} \boldsymbol{\Omega}^{-1/2}| - \log |\mathbf{C}_n| \right],$$

where  $\mathbf{h}(\cdot)$  is the operator defined in Section 2.1.  $\mathbf{h}(\boldsymbol{\Omega}^{1/2} \mathbf{C}_n \boldsymbol{\Omega}^{1/2})$  and  $\mathbf{h}(\boldsymbol{\Omega}^{1/2} \mathbf{D} \boldsymbol{\Omega}^{1/2})$  are independent from each other followed by the fact that  $\mathbf{C}_n$  and  $\mathbf{D}$  are independent.

*Proof.* See the Appendix. □

It is straightforward to see from Theorem 1 that the only case the LRT statistics are invariant with respect to  $\boldsymbol{\Omega}$  is when  $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_t \equiv \boldsymbol{\Sigma}$  and  $\boldsymbol{\Omega} = \mathbf{I}_t \otimes \boldsymbol{\Sigma}$ ; that is, when the  $t$

populations are uncorrelated and share homogeneous covariance matrix. The next theorem studies the asymptotic null distributions of the LRT statistics.

**Theorem 2.** *Let  $\mathbf{Z}$  be a  $(tp)$ -variate standard normal random variable.*

(a) *Under the hypothesis (i) of equal mean vectors, as  $n \rightarrow \infty$ ,  $\mathbf{M}_n$  converges in distribution to*

$$\mathbf{M}_\infty = \mathbf{Z}\boldsymbol{\Omega}^{1/2}(\mathbf{I}_{tp} - \mathbf{A})\boldsymbol{\Omega}^{-1}(\mathbf{I}_{tp} - \mathbf{A})\boldsymbol{\Omega}^{1/2}\mathbf{Z}^T, \quad (2)$$

*which has the Chi-squared distribution with  $(t-1)p$  degrees of freedom (d.f.).*

(b) *Under the hypothesis (ii) of homogenous covariance matrices, the asymptotic distribution of  $\mathbf{V}_n$  as  $n \rightarrow \infty$  is the Chi-squared distribution with  $(t-1)p(p+1)/2$  d.f.*

(c) *Under the hypothesis (iii) of both equal mean vectors and equal covariance matrices, as  $n \rightarrow \infty$ ,  $\mathbf{W}_n$  converges in distribution to the Chi-squared distribution with  $(t-1)p(p+3)/2$  d.f.*

*Proof.* See the Appendix. □

Theorems 1 points out that the finite-sample distributions of the LRT statistics depend on the true covariance matrix  $\boldsymbol{\Omega}$  unless  $\boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_t \equiv \boldsymbol{\Sigma}$  and  $\boldsymbol{\Omega} = \mathbf{I}_t \otimes \boldsymbol{\Sigma}$ . However, Theorem 2 indicates that in asymptotic, the test statistics have the Chi-squared distributions, in spite of the correlations among the populations.

## 2.3 Test Procedure

In addition to the theoretical findings in Section 2.2, the LRT statistics are not invariant with respect to permutation transformations of the samples that leave the null hypothesis invariant. Consequently a permutation method (e.g., Box and Anderson, 1955) is not applicable for null distribution computation. The characteristic of the LRT statistics in Theorem 1 motivates us to apply a bootstrap method to estimate the finite-sample null distributions without making any assumptions about the correlations among samples. In particular, to implement the LRT for testing the null hypothesis (i), the following procedure is proposed.



- (1) Use the observed samples to formulate the  $n \times (tp)$  data matrix  $\mathbf{Y}$  as described in Section 2.1. Compute the sums of squares  $SS_A$  and  $SS_M$  and then calculate the value of the LRT statistic  $M_n$ , labeled by  $M_n^{\text{obs}}$  to indicate that this is the observed test statistic.
- (2) Simulate  $K$  random samples of size  $n$  from the  $(tp)$ -variate normal distribution  $\mathcal{N}(\mathbf{0}, SS_M)$ . For each simulated random sample, compute the LRT statistic  $M_n$ , labeled by  $M_n^{(k)}$  for the  $k^{\text{th}}$  sample,  $k = 1, \dots, K$ .
- (3) Count the number of  $M_n^{(k)}$ 's that are greater than and equal to  $M_n^{\text{obs}}$ . The  $p$ -value of the test is then estimated by this number divided by the total number of simulated random samples, i.e.,  $\sum_{k=1}^K I(M_n^{(k)} \geq M_n^{\text{obs}})/K$  where  $I(\cdot)$  stands for the indicator function.

Note that the empirical distribution of  $M_n^{(1)}, \dots, M_n^{(K)}$  is the bootstrap estimate of the null distribution of the LRT statistic  $M_n$ . Because  $H_0$  is rejected for large  $M_n$ , the bootstrap level  $\alpha$  critical value is the  $(1 - \alpha)^{\text{th}}$  percentile of the  $M_n^{(k)}$  empirical distribution and the bootstrap  $p$ -value is the proportion of the  $M_n^{(k)}$  that are at least as large as  $M_n^{\text{obs}}$ , the value of  $M_n$  based on the original data.

The algorithm for implementation of the LRT for testing the null hypothesis (ii) is virtually the same as the above steps except that  $SS_V$  is in place of  $SS_M$  and  $V_n$  is in place of  $M_n$ . Similarly, the implementation of the LRT for testing the null hypothesis (iii) is the same except that  $SS_W$  and  $W_n$  are used accordingly. The larger the number of simulated random samples  $K$  is, the better approximation for the null distribution. In practice, we find  $K \geq 5000$  is reliable and adequate.

The proposed resampling procedure is basically a parametric bootstrap. An alternative would be the use of a nonparametric bootstrap to benefit from its robustness against distributional assumptions. That is, instead of resampling from a multivariate normal distribution in step (2), one forms random bootstrap samples through drawing random subjects from the centered original data set (the data are centered at the null hypothesis). However, there are two main issues that make such alternative less favorable. First, although one can easily cen-

ter the original data at the average of the sample mean vectors for hypothesis (i), it is fairly difficult to center the original data at the null hypotheses (ii) and (iii) without affecting the inter-population correlations. Secondly, a simulation we report in Section 3.3 exposes that the nonparametric bootstrap has weaker power for testing hypothesis (i) than the proposed parametric bootstrap, especially when sample sizes are not large.

### 3 Simulation

To evaluate the performances of  $\mathbf{M}_n$ ,  $\mathbf{V}_n$  and  $\mathbf{W}_n$  and the proposed finite-sample implementation procedure, we carried out Monte Carlo studies on the size (i.e., type I error probability) and power (i.e., one minus type II error probability) of the tests. In addition, we also provide a simulation to compare the proposed parametric bootstrap procedure with its nonparametric bootstrap counterpart.

In particular, we considered  $t = 2$  populations and let  $\mathbf{Y}_{1j} = (Y_{1j1}, \dots, Y_{1jp})^T$ ,  $j = 1, \dots, n$ , be the sample of size  $n$  from  $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$  distribution, and  $\mathbf{Y}_{2j} = (Y_{2j1}, \dots, Y_{2jp})^T$ ,  $j = 1, \dots, n$ , be the other sample of size  $n$  that comes from  $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$  distribution. The samples were generated according to  $\mathbf{Y}_{\cdot j} = (\mathbf{Y}_{1j}^T, \mathbf{Y}_{2j}^T)^T \sim \mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Omega})$ ,  $j = 1, \dots, n$ , where  $\boldsymbol{\eta} = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T)^T$  with  $\boldsymbol{\mu}_1 = \mu_1 \mathbf{1}_p$  and  $\boldsymbol{\mu}_2 = \mu_2 \mathbf{1}_p$  for  $\mathbf{1}_p$  denoting the  $p \times 1$  column vector of 1's; and  $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{pmatrix}$  with  $\boldsymbol{\Sigma}_1 = \sigma_1^2 \{(1 - \rho_1)\mathbf{I}_p + \rho_1 \mathbf{J}_p\}$ ,  $\boldsymbol{\Sigma}_2 = \sigma_2^2 \{(1 - \rho_2)\mathbf{I}_p + \rho_2 \mathbf{J}_p\}$ ,  $\boldsymbol{\Sigma}_{12} = \rho_{12} \mathbf{J}_p$  and  $\boldsymbol{\Sigma}_{21} = \rho_{21} \mathbf{J}_p$ . That is,  $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  if  $\mu_1 = \mu_2$ . Both  $\boldsymbol{\Sigma}_1$  and  $\boldsymbol{\Sigma}_2$  have compound symmetry structures, and their within-population correlation is controlled by  $\rho_1$  and  $\rho_2$ , respectively. Also, notice that  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$  if  $\sigma_1^2 = \sigma_2^2$  and  $\rho_1 = \rho_2$ . The dependency between the two samples is governed by the parameters  $\rho_{12}$  and  $\rho_{21}$ ; the samples are uncorrelated if  $\rho_{12} = \rho_{21} = 0$ .

#### 3.1 Size of the Tests

Simulated samples were generated under the null hypotheses in order to evaluate the size of the LRTs implemented using the proposed test procedure. Meanwhile, we compare their

type I error probabilities with those of the asymptotic Chi-squared tests. The simulation presented in Table 1 was set up according to  $p = 5$ ,  $\mu_1 = \mu_2 = 0$ ,  $\sigma_1^2 = \sigma_2^2 = 1$ , and  $\rho_1 = \rho_2 = 0.5$ . That is the null conditions  $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  and  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$  are simultaneously satisfied for all three tests. The performances of the LRTs were investigated for different sample sizes. We let the two populations to be independent by setting  $\rho_{12} = \rho_{21} = 0$  or correlated by setting  $\rho_{12} = \rho_{21} = 0.3$ . 500 Monte Carlo data sets were simulated for each setting with a choice of sample size  $n$  ( $= 25, 50, 100$ , or  $500$ ) and a choice of between-population correlation  $\rho_{12} = \rho_{21}$  ( $= 0$  or  $0.3$ ). For each data set, the LRT statistics  $M_n$ ,  $V_n$ ,  $W_n$ , and their estimated  $p$ -values were computed using the procedure described in Section 2.3. The asymptotic Chi-squared test for mean equivalence is  $\chi_5^2$ , the asymptotic Chi-squared test for covariance equivalence is  $\chi_{15}^2$ , and the asymptotic Chi-squared test for both mean equivalence and covariance equivalence is  $\chi_{20}^2$ . Whether a null hypothesis was rejected or not at nominal significance level  $\alpha = 0.05$  for a data set was recorded by checking if a  $p$ -value is less than  $0.05$ . The proportion of rejections from 500 replications was calculated for each test. Note that when the power is near  $0.05$ , the estimated power (rejection rate) from 500 Monte Carlo replications has an approximate standard deviation of  $0.0098$ .

From the results in Table 1, we can see that regardless of the sample size and the between-population correlation, the empirical type I error probabilities of  $M_n$ ,  $V_n$ , and  $W_n$  implemented using the proposed test procedure are close to or attain the nominal  $\alpha = 0.05$  level. However, when the sample size  $n$  is small, all Chi-squared tests are fairly liberal with empirical type I error probabilities well above  $0.05$ , no matter the populations are independent or correlated. The proposed test procedure and the Chi-squared tests provide almost identical type I error probabilities when  $n$  is large enough.

We further investigated how the test  $M_n$  performs when its null hypothesis of equal mean vectors is met for the two correlated populations by setting  $\mu_1 = \mu_2 = 0$  and  $\rho_{12} = \rho_{21} = 0.3$ , but their covariance matrices differ due to the change of  $\sigma_2^2$  ( $= 1, 1.5$ , or  $2$ ) and the change of  $\rho_2$  ( $= 0.5$  or  $0$ ). The data generating scheme was the same as the preceding simulation study, and results are displayed in Table 2. In all cases,  $M_n$  implemented using the proposed

procedure demonstrates good performance through achieving approximately valid nominal  $\alpha = 0.05$  level and is not influenced by the heterogeneity of the covariance matrices, while the Chi-squared test shows degradation of performance which worsens when the within-population correlations are different ( $\rho_1 \neq \rho_2$ ) and remains quite liberal even for large sample sizes (e.g.,  $n = 500$ ), although as expected the situation becomes better when  $n$  increases. We obtained similar results for  $V_n$  and the corresponding Chi-squared test when the null hypothesis of equal covariance matrices is met for the two populations but their mean vectors differ. Moreover, similar behaviors of these tests were observed when the two populations are uncorrelated ( $\rho_{12} = \rho_{21} = 0$ ). Overall, the proposed test procedure offers assurance of credible estimation of type I error probability. The asymptotic Chi-squared tests are valid only if  $n$  is very large. These observations are consistent with our theory.

### 3.2 Power of the Tests

We carried out simulations to detect a variety of alternative hypotheses using the proposed test procedure: difference between mean vectors only, or between covariance matrices only, or both. In the simulation results reported in Table 3, the parameter configurations are  $p = 5$ ,  $\mu_1 = 0$ ,  $\sigma_1^2 = 1$ ,  $\rho_{12} = \rho_{21} = 0.3$ ; that is, the two normal populations are correlated. We vary the values of  $\mu_2$ ,  $\sigma_2^2$  to create various null and alternative situations and sample size  $n$  is changed as well. We also let the within-population correlations to be the same ( $\rho_1 = \rho_2 = 0.5$ ) or different ( $\rho_1 = 0.5$ ,  $\rho_2 = 0$ ). For each setting with a choice of  $\mu_2$  ( $= 0, 0.5$ , or  $1$ ), a choice of  $\sigma_2^2$  ( $= 1, 1.5$ , or  $2$ ), a choice of  $\rho_2$  ( $= 0.5$ , or  $0$ ), and a choice of  $n$  ( $= 25, 50, 100$ , or  $500$ ), 500 Monte Carlo data sets were simulated using the same scheme as the preceding simulations.

Table 3 displays the empirical rejection probabilities (the number of rejections divided by the number of Monte Carlo data sets). Results show that  $M_n$ ,  $V_n$  and  $W_n$  are powerful for detecting mean vector inequality, covariance matrix inequality, and simultaneous mean vector and covariance matrix inequalities, respectively. As expected, the larger difference between population mean vectors, the more powerful  $M_n$  becomes; the greater difference

between population covariance matrices, the more powerful  $V_n$  is; and the more difference between population mean vectors and/or between variance matrices, the more power  $W_n$  improves. This indicates the power functions for these LRTs are monotone since their power increases as the distance between the alternative hypothesis and the null hypothesis increases. The power of any of the three LRTs increases with sample size  $n$ ; and they are substantially powerful when  $n$  is large, say  $n \geq 100$ . Moreover, from the lower part of Table 3,  $V_n$  and  $W_n$  are powerful for detecting the difference between within-population correlations ( $\rho_1$  and  $\rho_2$ ) regardless whether the variance components ( $\sigma_1^2$  and  $\sigma_2^2$ ) differ or not.

When the two populations have the same covariance matrix (i.e., when  $\sigma_1^2 = \sigma_2^2$  and  $\rho_1 = \rho_2$ ),  $M_n$  is slightly more likely to detect mean vector difference than  $W_n$ . However, when the population variances are different,  $W_n$  exhibits better power than  $M_n$ . Likewise, when there is no difference between the population mean vectors (i.e., when  $\mu_1 = \mu_2$ ),  $V_n$  is slightly more powerful than  $W_n$  to detect the difference between population covariance matrices; while  $W_n$  is more powerful than  $V_n$  when the difference between population mean vectors is present. Moreover, when the two population variances differ more sizeably, the power of  $M_n$  for detecting mean equivalence declines mildly, but this is the price one has to pay for  $M_n$  if the two populations are heterogenous. This loss of power is less pronounced and  $M_n$  achieves good power similar to  $W_n$  when the presence of unequal population mean vectors is large enough, e.g., when  $\mu_2 - \mu_1 = 1$ .

The empirical power of the asymptotic Chi-squared tests are quite liberal with larger rejection rates (not reported here), similar to their behaviors seen in Section 3.1. Again, the simulation results coincide with our theoretical findings.

### 3.3 Comparison to Nonparametric Bootstrap

As discussed in Section 2.3, a nonparametric bootstrap approach can be used to implement  $M_n$ , in place of the proposed parametric bootstrap procedure. To compare the performances of these two procedures, we conducted a simulation using the same set up as in Section 3.2 and Table 4 gives the results. When sample sizes are small, the proposed parametric bootstrap

procedure provides a more powerful test for detecting mean vector inequalities. The two procedures exhibit similar performances when sample sizes are sufficiently large. Therefore, we recommend the parametric bootstrap procedure for the hypothesis tests under consideration.

## 4 Applications

To illustrate the proposed methods in practice, we applied them to two data examples.

### 4.1 Tooth Size Study

A numerical example is taken from a dental health study conducted during 1999—2002 in Seoul, Korea, which was part of a standard occlusion study that has been undergoing since 1997 (Kim *et al.*, 2005; Lee *et al.*, 2007). Human adults have 28 permanent teeth (central incisors to second molars), or up to 32 including the third molars (or wisdom teeth). The permanent tooth sizes (mesiodistal diameters of teeth) of patients were measured using digital vernier calipers. The data set contains mandibular (lower jaw) teeth size measurements obtained from 115 young female adults who had natural normal occlusion and whose age range was between 17 and 24 years old with an average age of 20 years. Tooth size imbalances can complicate harmonious intercuspation and often justify extraction treatment modalities. Assessing whether tooth size profile is identical for the left and right sides around central incisors would be helpful for understanding the biological phenomena of teeth. It can provide important insight into normative data of human tooth size, diagnostic criteria for malocclusion, and dental treatment planning such as orthodontic correction and rapid palatal expansion in these women (Uysal *et al.*, 2005). To address this issue, we apply the proposed tests to the data set.

Let  $\mathbf{Y}_{Lj} = (Y_{Lj1}, \dots, Y_{Lj7})^T$  and  $\mathbf{Y}_{Rj} = (Y_{Rj1}, \dots, Y_{Rj7})^T$  denote the tooth size measurements on the left and those on the right from central incisors to second molars in the mandible of the  $j^{\text{th}}$  woman, respectively,  $j = 1, \dots, 115$ . It has been previously reported that the tooth sizes have a multivariate normal distribution (Wang *et al.*, 2006) and normality

checking performed on the data agreed with this observation, thus we assume

$$(\mathbf{Y}_{Lj}^T, \mathbf{Y}_{Rj}^T)^T \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_L \\ \boldsymbol{\mu}_R \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_L & \boldsymbol{\Sigma}_{LR} \\ \boldsymbol{\Sigma}_{RL} & \boldsymbol{\Sigma}_R \end{pmatrix}\right).$$

Because  $\mathbf{Y}_{Lj}$  and  $\mathbf{Y}_{Rj}$  consisted of observations from the same woman, it is very likely that they are correlated, i.e.,  $\boldsymbol{\Sigma}_{LR}$  and  $\boldsymbol{\Sigma}_{RL}$  have non-zero components. The two samples that consist measurements on the left side and those on the right side, respectively, can be regarded as coming from two correlated 7-variate normal distributions. To test the sameness of tooth size profiles on the two sides, we applied the proposed methods to analyze the data. Data analysis indicates  $M_n = 11.8571$  with an estimated  $p$ -value = 0.2305, which suggests  $\boldsymbol{\mu}_L = \boldsymbol{\mu}_R$ ;  $V_n = 73.3759$  with an estimated  $p$ -value = 0.0000, which indicates  $\boldsymbol{\Sigma}_L \neq \boldsymbol{\Sigma}_R$ ; and  $W_n = 84.7504$  with estimated  $p$ -value = 0.0000, complementing the result of  $V_n$ .

We also conducted a simulation study based on the design of tooth size data set for further checking the validity of the analysis results. 500 independent Monte Carlo random samples were simulated; each data set  $(\mathbf{Y}_{Lj}^T, \mathbf{Y}_{Rj}^T)^T$ ,  $j = 1, \dots, 115$ , were generated from 14-variate normal distribution with mean and covariance matrix set equal to the overall sample mean and sample covariance matrix of the tooth size data set, respectively. The three proposed LRT statistics and their estimated  $p$ -values were computed. Results show that, at nominal significance level 0.05, the empirical rejection probability of  $M_n$  for detecting mean vector inequality is 0.02, the empirical power of  $V_n$  for detecting difference between covariance matrices is 0.56, and the empirical power of  $W_n$  for detecting inequality of mean vectors and/or covariance matrices is 0.54. These simulation results are consistent with the data analysis results; both data analysis and simulation suggest equal mean vectors but unequal covariance matrices. Therefore, we conclude that the mean tooth size profiles on the two sides are the same, but the dispersion of tooth size differs between the left and right sides in the mandible of these women.

The asymptotic Chi-squared tests suggest similar conclusions, but with more significant  $p$ -values 0.1054, 0.0000, and 0.0000 for testing the hypotheses (i)—(iii), respectively. This is not surprising because  $n = 115$  is relatively large, although the results obtained using the

proposed methods are more reliable.

## 4.2 Microarray Study

Our second example contains microarray data from a study of head and neck squamous cell carcinoma (HNSCC) (Kuriakose et al., 2004; Irizarry et al., 2003), which is composed of gene expression profiles at two stages of the cancer, benign tumor and primary tumor. For each stage, the sample consists of 16 measurements of a gene set containing four co-expressed genes (named 32242\_at, 38625\_g\_at, 36334\_at, and AFFX-HSAC07/X00351\_M\_at); and 10 patients were involved in both samples (stages). These four genes are known to scientists that they are differentially co-expressed. We now further investigate whether the differential co-expressions of the gene set are due to mean profile difference, or heterogeneous covariance matrices, or both.

Let  $Y_{Nj} = (Y_{Nj1}, \dots, Y_{Nj4})^T, j = 1, \dots, 16$ , denote the gene expression profiles of the four genes of normal tissue from the  $j^{\text{th}}$  patient in the benign tumor stage, and  $Y_{Tj} = (Y_{Tj1}, \dots, Y_{Tj4})^T, j = 1, \dots, 16$ , be those of the tumor tissue from the  $j^{\text{th}}$  patient in the primary tumor stage. Since the measurements are continuous, we use the common assumptions  $Y_{Nj} \sim \mathcal{N}(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N)$  and  $Y_{Tj} \sim \mathcal{N}(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T)$ . Taking into account that the two samples are only partially correlated, we apply the LRTs using a slightly modified test procedure. That is, we modify only the calculation of  $SS_A$  and  $SS_M$  in the procedure. In particular, the  $i^{\text{th}}$  diagonal  $p \times p$  (here,  $p = 4$ ) matrix of  $SS_A$  is the sum of squares obtained using all the observations in the  $i^{\text{th}}$  sample. The  $(i, i')^{\text{th}}$  off-diagonal  $p \times p$  matrix of  $SS_A$  is the off-diagonal  $p \times p$  matrix of the sum of squares obtained using the observations that are correlated in both  $i^{\text{th}}$  and  $i'^{\text{th}}$  samples.  $SS_M$  is computed in a similar way but after all the samples are centered at the average of the individual sample mean vectors.

The analysis results show  $M_n = 44.8895$  with an estimated  $p$ -value = 0.0004, which suggests  $\boldsymbol{\mu}_N \neq \boldsymbol{\mu}_T$ ;  $V_n = 23.5296$  with an estimated  $p$ -value = 0.08532, which conveys  $\boldsymbol{\Sigma}_N \approx \boldsymbol{\Sigma}_T$ ; and  $W_n = 62.6172$  with estimated  $p$ -value = 0.0033, in agreement with the result of  $M_n$ . Therefore, we may conclude that the differential co-expressions of the four genes are



mainly contributed by the mean vector difference between the two stages of the cancer.

In contrast, the asymptotic Chi-squared tests have  $p$ -values that are all highly significant: 0.0000, 0.0090, and 0.0000 for testing the hypotheses (i)—(iii), respectively. Such significant results are likely to be caused by the very liberal behaviors of the Chi-squared tests for finite samples, thus their credibility is questionable for these small samples.

## 5 Discussion

We focused on correlated multiple multivariate samples and required no assumption on how the samples are correlated. We presented likelihood ratio tests for equality of mean vectors, equality of covariance matrices, and simultaneous equality of mean vectors and covariance matrices of the populations. Derivation of the finite-sample distributions of the LRT statistics shows that these distributions depend essentially on the true overall covariance matrix  $\mathbf{\Omega}$ , whose diagonal square matrices are the population-specific covariance matrices and rest elements represent the covariances among populations. Another finding is that the asymptotic null distributions are Chi-squared distributions regardless how the populations are correlated. A resampling procedure is proposed to numerically estimate the null distributions for finite samples. The test statistics are easy to calculate and the implementation of the test procedure is straightforward. Simulation studies demonstrate that, under the null hypotheses, the proposed tests approximate valid nominal level, unlike the asymptotic Chi-squared tests which may provide misleading type I error probabilities; meanwhile for a variety of alternative hypotheses, the tests demonstrate desired performance attaining appreciable power. The application to the dental data and the microarray data further illustrates the usefulness of adapting all three LRTs for data analysis so that comprehensive and credible conclusions can be drawn. These features provide a strong case for the use of the proposed methods.

In theory the proposed methods can be applied to any number of multiple samples that are correlated and multivariate normally distributed. However, when the sample sizes and multivariate dimension are significantly large, computation will be a nontrivial issue and new methods may be needed; this is included in our future work. The proposed approach

assumes multivariate normality, which is reasonable for many data in practice, perhaps on a transformed scale, but one should take care to check this assumption. In practice, multivariate samples may be correlated due to various factors. We feel that the proposed methods can be recommended for data analysis when the normality assumption looks realistic and correlations among samples is evident or suspicious.

It is well known that most likelihood ratio tests based on the limit Chi-squared approximation have large rejection probabilities. A Bartlett correction (Bartlett, 1937) is a modification applied to likelihood ratio statistics that may improve the Chi-squared null distributions to order  $O(n^{-1})$  from the original order  $O(1)$ ; refer to Carbari-Neto and Cordeiro (1996) and references therein for more details. The extension of the Bartlett correction to the correlated multivariate samples we consider in this article would be involving, and moreover, the concern for the small-sample performances of the modified Chi-squared tests may still remain. For example, Table 2 shows that the Chi-squared test for mean vector equality easily rejects the null hypothesis even for fairly large samples, especially when the samples have heterogeneous covariance matrices and are correlated. Thus for such kinds of situations, we conjecture that there would be an improvement over the usual Chi-squared tests with a Bartlett correction, but the proposed methods always provide reliable results no matter the samples are large or small.

## Acknowledgments

We thank two referees and the Editor for their helpful and insightful comments and suggestions.

## References

- [1] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd edition. New York: Wiley.

- [2] Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London A* **160**, 268–282.
- [3] Bock, R.D. and Petersen, A.C. (1975). A multivariate correction for attenuation. *Biometrika* **62**, 673–678.
- [4] Box, G.E.P. and Anderson, S.L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption. *Journal of the Royal Statistical Society, Series B* **17**, 1–26.
- [5] Cribari-Neto, F. and Cordeiro, G.M. (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews* **15**, 339–367.
- [6] Choi, S.C. and Wette, R. (1972). A test for homogeneity of variances among correlated variables. *Biometrics* **28**, 589–592.
- [7] Conover, W.J., Johnson, M.E. and Johnson, M.M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* **23**, 351–361.
- [8] Eaton, M.R. (1983). *Multivariate Statistics: A Vector Space Approach*. New York: Wiley.
- [9] Han, C.P. (1968). Testing the homogeneity of a test of correlated variances. *Biometrika* **55**, 317–326.
- [10] Harris, P. (1985). Testing for variance homogeneity of correlated variables. *Biometrika* **72**, 103–107.
- [11] Irizarry, R., Hobbs, B., Collin, F., Beazer-Barclay, Y., Antonellis, K., Scherf, U., and Speed, T. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264.
- [12] Kim, J.Y., Lee, S.J., Kim, T.W., Nahm, D.S., and Chang Y.I. (2005). Classification of the skeletal variation in normal occlusion. *Angle Orthodontics* **75**, 303–311.

- [13] Kuriakose<sup>1</sup>, M., Chen, W., He<sup>1</sup>, Z., Sikora<sup>1</sup>, A., Zhang, P., Zhang, Z., Qiu, W., Hsu, D., McMunn-Coffran, C., Brown, S., Elango, E., Delacure<sup>1</sup>, M., and Chen, F. (2004). Selection and validation of differentially expressed genes in head and neck cancer. *Cellular and Molecular Life Sciences*, **61**, 1372–1383.
- [14] Lee, S.J., Lee, S., Lim, J., Ahn, S.J., and Kim, T.W. (2007). Cluster analysis of human tooth size in subjects with normal occlusion. *American Journal of Orthodontics & Dentofacial Orthopedics*, **132**, 796–800.
- [15] Martin, R.J. (1993). Approximations to the determinant term in Gaussian maximum likelihood estimation of some spatial models. *Comm. Statist. Theory Methods* **22**, 189–205.
- [16] Perlman, M.D. (1980). Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. *The Annals of Statistics* **8**, 247–263.
- [17] Szatrowski, T.H. (1979). Asymptotic nonnull distributions for likelihood ratio statistics in the multivariate normal patterned mean and covariance matrix testing problem. *The Annals of Statistics* **7**, 823–837.
- [18] Uysal, T., Sari, Z., Basciftci, F.A., and Memili, B. (2005). Intermaxillary tooth size discrepancy and malocclusion: is there a relation? *Angle Orthodontics* **75**, 204–209.
- [19] Wang, X., Stokes, L., Lim, J., and Chen, M. (2006). Concomitant of multivariate order statistics with application to judgment post-stratification. *Journal of the American Statistical Association*, **101**, 1693–1704.

## Appendix

### Proof of Theorem 1

We start with the distributions of sums of squares matrixes.

First, we show  $\boldsymbol{\Omega}^{-1/2}\mathbf{Y}^T\{\mathbf{I}_n-(1/n)\mathbf{J}_n\}\mathbf{Y}\boldsymbol{\Omega}^{-1/2}$  follows Wishart( $\mathbf{I}_{tp}, n-1, tp$ ) distribution. The matrix  $\{\mathbf{I}_n-(1/n)\mathbf{J}_n\}$  is an idempotent matrix with rank equal to  $n-1$ . Thus the sum of squares  $SS_A = \mathbf{Y}^T\{\mathbf{I}_n-(1/n)\mathbf{J}_n\}\mathbf{Y}$  has Wishart( $\boldsymbol{\Omega}, n-1, tp$ ) distribution. Here,  $\boldsymbol{\Omega}^{-1/2}$  is the symmetric square root decomposition matrix of  $\boldsymbol{\Omega}^{-1}$  such that  $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Omega}^{-1/2} \cdot \boldsymbol{\Omega}^{-1/2}$ . The distribution of  $\boldsymbol{\Omega}^{-1/2}\mathbf{Y}^T\{\mathbf{I}_n-(1/n)\mathbf{J}_n\}\mathbf{Y}\boldsymbol{\Omega}^{-1/2}$  is hence obtained to be Wishart( $\mathbf{I}_{tp}, n-1, tp$ ) based on a property of Wishart distributions.

Secondly, we show  $(1/n)\boldsymbol{\Omega}^{-1/2}(\mathbf{I}_{tp}-\mathbf{A})\mathbf{Y}^T\mathbf{J}_n\mathbf{Y}(\mathbf{I}_{tp}-\mathbf{A})\boldsymbol{\Omega}^{-1/2}$  has the distribution of Wishart( $\boldsymbol{\Omega}^{-1/2}(\mathbf{I}_{tp}-\mathbf{A})^T\boldsymbol{\Omega}(\mathbf{I}_{tp}-\mathbf{A})\boldsymbol{\Omega}^{-1/2}, 1, tp$ ). The matrix  $(1/n)\mathbf{J}_n$  is an idempotent matrix with rank equal to 1. Thus the statistic  $(1/n)\mathbf{Y}^T\mathbf{J}_n\mathbf{Y}$  has Wishart( $\boldsymbol{\Omega}, 1, tp$ ) distribution. The statistic  $SS_M - SS_A = (1/n)(\mathbf{I}_{tp}-\mathbf{A})\mathbf{Y}^T\mathbf{J}_n\mathbf{Y}(\mathbf{I}_{tp}-\mathbf{A})$  has the Wishart distribution with parameters  $(\mathbf{I}_{tp}-\mathbf{A})\boldsymbol{\Omega}(\mathbf{I}_{tp}-\mathbf{A})$ , 1, and  $tp$ . Hence,  $(1/n)\boldsymbol{\Omega}^{-1/2}(\mathbf{I}_{tp}-\mathbf{A})\mathbf{Y}^T\mathbf{J}_n\mathbf{Y}(\mathbf{I}_{tp}-\mathbf{A})\boldsymbol{\Omega}^{-1/2}$  has the Wishart( $\boldsymbol{\Omega}^{-1/2}(\mathbf{I}_{tp}-\mathbf{A})^T\boldsymbol{\Omega}(\mathbf{I}_{tp}-\mathbf{A})\boldsymbol{\Omega}^{-1/2}, 1, tp$ ) distribution based on a property of Wishart distributions.

Finally,  $\mathbf{Y}^T\{\mathbf{I}_n-(1/n)\mathbf{J}_n\}\mathbf{Y}$  and  $(1/n)(\mathbf{I}_{tp}-\mathbf{A})\mathbf{Y}^T\mathbf{J}_n\mathbf{Y}(\mathbf{I}_{tp}-\mathbf{A})$  are independent to each other because  $\{\mathbf{I}_n-(1/n)\mathbf{J}_n\}\mathbf{J}_n = 0$ .

Now we derive the distribution of the LRT statistics. First,

$$\begin{aligned} M_n &= n \log \left( \frac{|SS_M|}{|SS_A|} \right) \\ &= n \log \left( \frac{|\mathbf{Y}^T\{\mathbf{I}_n-(1/n)\mathbf{J}_n\}\mathbf{Y} + (1/n)(\mathbf{I}_{tp}-\mathbf{A})\mathbf{Y}^T\mathbf{J}_n\mathbf{Y}(\mathbf{I}_{tp}-\mathbf{A})|}{|\mathbf{Y}^T(\mathbf{I}_n-(1/n)\mathbf{J}_n)\mathbf{Y}|} \right) \\ &= n \log \left( \frac{|\boldsymbol{\Omega}^{-1/2}\mathbf{Y}^T\{\mathbf{I}_n-(1/n)\mathbf{J}_n\}\mathbf{Y}\boldsymbol{\Omega}^{-1/2} + (1/n)\boldsymbol{\Omega}^{-1/2}(\mathbf{I}_{tp}-\mathbf{A})\mathbf{Y}^T\mathbf{J}_n\mathbf{Y}(\mathbf{I}_{tp}-\mathbf{A})\boldsymbol{\Omega}^{-1/2}|}{|\boldsymbol{\Omega}^{-1/2}\mathbf{Y}^T\{\mathbf{I}_n-(1/n)\mathbf{J}_n\}\mathbf{Y}\boldsymbol{\Omega}^{-1/2}|} \right), \end{aligned}$$

Using the above characterizations of the sum of squares  $SS_A$  and  $SS_M$ , we have that  $M_n$  is distributed as

$$n \left( \log |\mathbf{C}_n + \mathbf{D}| - \log |\mathbf{C}_n| \right),$$

where  $\mathbf{C}_n$  and  $\mathbf{D}$  are independent Wishart variables,  $\mathbf{C}_n \sim \text{Wishart}(\mathbf{I}_{tp}, n-1, tp)$ , and  $\mathbf{D} \sim \text{Wishart}(\boldsymbol{\Omega}^{-1/2}(\mathbf{I}_{tp}-\mathbf{A})^T\boldsymbol{\Omega}(\mathbf{I}_{tp}-\mathbf{A})\boldsymbol{\Omega}^{-1/2}, 1, tp)$ .

Alternatively,  $SS_A$  can be easily seen to be distributed as  $\boldsymbol{\Omega}^{1/2}\mathbf{C}_n\boldsymbol{\Omega}^{1/2}$ . Hence,  $SS_V =$

$\mathbf{h}(SS_A)$  is distributed as  $\mathbf{h}(\mathbf{\Omega}^{1/2}\mathbf{C}_n\mathbf{\Omega}^{1/2})$ . Therefore,  $\mathbf{V}_n$  is distributed as

$$n \left( \log |\mathbf{\Omega}^{-1/2}\mathbf{h}(\mathbf{\Omega}^{1/2}\mathbf{C}_n\mathbf{\Omega}^{1/2})\mathbf{\Omega}^{-1/2}| - \log |\mathbf{C}_n| \right).$$

Similarly,  $SS_M$  can be easily seen to be distributed as  $\mathbf{\Omega}^{1/2}(\mathbf{C}_n + \mathbf{D})\mathbf{\Omega}^{1/2}$ . Hence,  $SS_W = \mathbf{h}(SS_M)$  is distributed as  $\mathbf{h}(\mathbf{\Omega}^{1/2}(\mathbf{C}_n + \mathbf{D})\mathbf{\Omega}^{1/2}) = \mathbf{h}(\mathbf{\Omega}^{1/2}\mathbf{C}_n\mathbf{\Omega}^{1/2}) + \mathbf{h}(\mathbf{\Omega}^{1/2}\mathbf{D}\mathbf{\Omega}^{1/2})$ . Therefore,  $\mathbf{W}_n$  is distributed as

$$n \left\{ \log |\mathbf{\Omega}^{-1/2} \left( \mathbf{h}(\mathbf{\Omega}^{1/2}\mathbf{C}_n\mathbf{\Omega}^{1/2}) + \mathbf{h}(\mathbf{\Omega}^{1/2}\mathbf{D}\mathbf{\Omega}^{1/2}) \right) \mathbf{\Omega}^{-1/2}| - \log |\mathbf{C}_n| \right\}.$$

## Proof of Theorem 2

We have introduced function  $\mathbf{h}$  to define the statistics  $\mathbf{V}_n$  and  $\mathbf{W}_n$ . Under the null hypothesis (ii) and (iii) in the Introduction,  $(1/n)SS_A$  converges to  $\mathbf{\Omega}$  in probability. Thus, the probability that  $SS_A$  is singular decreases to 0. Because of this, we assume  $SS_A$  is non-singular for simplicity.

### Proof for $\mathbf{M}_n$

We first compute the asymptotic distribution of  $\mathbf{M}_n$ . We have

$$\begin{aligned} \mathbf{M}_n &= -n \log \left( \frac{|SS_A|}{|SS_M|} \right) \\ &= -n \log \frac{|\mathbf{SS}_A|}{\left| \mathbf{SS}_A + (\mathbf{I}_{tp} - \mathbf{A}) \mathbf{\Omega}^{1/2} \left( \sum_i \mathbf{Z}_i^T / \sqrt{n} \right) \left( \sum_i \mathbf{Z}_i / \sqrt{n} \right) \mathbf{\Omega}^{1/2} (\mathbf{I}_{tp} - \mathbf{A}) \right|} \\ &= -n \log \frac{1}{\left| \mathbf{I}_{tp} + \mathbf{SS}_A^{-1} (\mathbf{I}_{tp} - \mathbf{A}) \mathbf{\Omega}^{1/2} \left( \sum_i \mathbf{Z}_i^T / \sqrt{n} \right) \left( \sum_i \mathbf{Z}_i / \sqrt{n} \right) \mathbf{\Omega}^{1/2} (\mathbf{I}_{tp} - \mathbf{A}) \right|} \\ &= n \log \left| \mathbf{I}_{tp} + \mathbf{SS}_A^{-1} (\mathbf{I}_{tp} - \mathbf{A}) \mathbf{\Omega}^{1/2} \left( \sum_i \mathbf{Z}_i^T / \sqrt{n} \right) \left( \sum_i \mathbf{Z}_i / \sqrt{n} \right) \mathbf{\Omega}^{1/2} (\mathbf{I}_{tp} - \mathbf{A}) \right|. \quad (3) \end{aligned}$$

Note that  $SS_A$  is a consistent estimate of  $n\mathbf{\Omega}$ , and thus, for sufficiently large  $n$ , the equation (3) can be approximated as

$$n \log \left| \mathbf{I}_{tp} + (\mathbf{\Omega}^{-1}/n) (\mathbf{I}_{tp} - \mathbf{A}) \mathbf{\Omega}^{1/2} \left( \sum_i \mathbf{Z}_i^T / \sqrt{n} \right) \left( \sum_i \mathbf{Z}_i / \sqrt{n} \right) \mathbf{\Omega}^{1/2} (\mathbf{I}_{tp} - \mathbf{A}) \right|, \quad (4)$$

which is equivalent to

$$\log \left| 1 + \left( \sum_i \mathbf{Z}_i / \sqrt{n} \right) \boldsymbol{\Omega}^{1/2} (\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega}^{-1} (\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega}^{1/2} \left( \sum_i \mathbf{Z}_i^T / \sqrt{n} \right) \right| \quad (5)$$

by Proposition 1.35 in Eaton (1983). The equation (5) can be approximated by, for sufficiently large  $n$ ,

$$\mathbf{Z} \boldsymbol{\Omega}^{1/2} (\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega}^{-1} (\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega}^{1/2} \mathbf{Z}^T.$$

Therefore,  $\mathbf{M}_n$  has the asymptotic distribution of  $\mathbf{M}_\infty = \mathbf{Z}^T \mathbf{Q} \mathbf{Z}$ , where  $\mathbf{Z}$  is from the  $(tp)$ -variate standard normal distribution and  $\mathbf{Q} = \boldsymbol{\Omega}^{1/2} (\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega}^{-1} (\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega}^{1/2}$ .

Now we show that the matrix  $\mathbf{Q}$  is idempotent, and, thus, the  $\mathbf{M}_\infty$  has the Chi-squared distribution.

**Lemma 1.**

$$\mathbf{R} \equiv (\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega} (\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega}^{-1} = (\mathbf{I}_{tp} - \mathbf{A}). \quad (6)$$

*Proof.* We let

$$\boldsymbol{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1t} \\ \Omega_{21} & \Omega_{22} & \cdots & \Omega_{2t} \\ \vdots & \vdots & \vdots & \vdots \\ \Omega_{t1} & \Omega_{t2} & \cdots & \Omega_{tt} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Omega}^{-1} = \begin{pmatrix} \Omega^{11} & \Omega^{12} & \cdots & \Omega^{1t} \\ \Omega^{21} & \Omega^{22} & \cdots & \Omega^{2t} \\ \vdots & \vdots & \vdots & \vdots \\ \Omega^{t1} & \Omega^{t2} & \cdots & \Omega^{tt} \end{pmatrix}.$$

Since  $\boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} = \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} = \mathbf{I}_{tp}$ , we have

$$\begin{aligned} \sum_k \Omega_{ik} \Omega^{kj} &= \delta_{ij} \mathbf{I}_p \\ \sum_k \Omega^{ik} \Omega_{kj} &= \delta_{ij} \mathbf{I}_p. \end{aligned}$$

Simple algebra shows that

$$\begin{aligned} \mathbf{A} \boldsymbol{\Omega} &= \left( 1/t \right) \mathbf{J}_t \otimes \mathbf{I}_p \boldsymbol{\Omega} \\ &= \left( 1/t \right) \begin{pmatrix} \sum_k \Omega_{k1} & \sum_k \Omega_{k2} & \cdots & \sum_k \Omega_{kt} \\ \sum_k \Omega_{k1} & \sum_k \Omega_{k2} & \cdots & \sum_k \Omega_{kt} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_k \Omega_{k1} & \sum_k \Omega_{k2} & \cdots & \sum_k \Omega_{kt} \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned}\mathbf{A}\Omega^{-1} &= (1/t)\mathbf{J}_t \otimes \mathbf{I}_p \Omega \\ &= (1/t) \begin{pmatrix} \sum_k \Omega^{k1} & \sum_k \Omega^{k2} & \cdots & \sum_k \Omega^{kt} \\ \sum_k \Omega^{k1} & \sum_k \Omega^{k2} & \cdots & \sum_k \Omega^{kt} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_k \Omega^{k1} & \sum_k \Omega^{k2} & \cdots & \sum_k \Omega^{kt} \end{pmatrix}.\end{aligned}$$

Finally,  $\mathbf{A}\Omega\mathbf{A}\Omega^{-1}$  is the product of the above two matrices which is

$$\begin{aligned}& (1/t)^2 \begin{pmatrix} \sum_k \Omega_{k1} & \sum_k \Omega_{k2} & \cdots & \sum_k \Omega_{kt} \\ \sum_k \Omega_{k1} & \sum_k \Omega_{k2} & \cdots & \sum_k \Omega_{kt} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_k \Omega_{k1} & \sum_k \Omega_{k2} & \cdots & \sum_k \Omega_{kt} \end{pmatrix} \begin{pmatrix} \sum_k \Omega^{k1} & \sum_k \Omega^{k2} & \cdots & \sum_k \Omega^{kt} \\ \sum_k \Omega^{k1} & \sum_k \Omega^{k2} & \cdots & \sum_k \Omega^{kt} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_k \Omega^{k1} & \sum_k \Omega^{k2} & \cdots & \sum_k \Omega^{kt} \end{pmatrix} \\ &= (1/t)^2 \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1t} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2t} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{B}_{t1} & \mathbf{B}_{t2} & \cdots & \mathbf{B}_{tt} \end{pmatrix},\end{aligned}\tag{7}$$

where, for every  $i, j = 1, 2, \dots, t$ ,

$$\begin{aligned}\mathbf{B}_{ij} &= \sum_l \Omega^{lj} \left\{ \sum_k \Omega_{k1} + \sum_k \Omega_{k2} + \cdots + \sum_k \Omega_{kt} \right\} \\ &= t \mathbf{I}_p\end{aligned}$$

Thus,  $\mathbf{A}\Omega\mathbf{A}\Omega^{-1} = (1/t)\mathbf{J}_t \otimes \mathbf{I}_p = \mathbf{A}$ , and

$$\begin{aligned}\mathbf{R} &= (\mathbf{I}_{tp} - \mathbf{A})\Omega(\mathbf{I}_{tp} - \mathbf{A})\Omega^{-1} \\ &= \mathbf{I}_{tp} - \mathbf{A} - \mathbf{A} + \mathbf{A}\Omega\mathbf{A}\Omega^{-1} \\ &= \mathbf{I}_{tp} - \mathbf{A} - \mathbf{A} + \mathbf{A} \\ &= \mathbf{I}_{tp} - \mathbf{A}.\end{aligned}$$

□



We then show that  $\mathbf{Q}^2 = \mathbf{Q}$ .

$$\begin{aligned}
\mathbf{Q}^2 &= \mathbf{\Omega}^{1/2}(\mathbf{I}_{tp} - \mathbf{A})\mathbf{\Omega}^{-1}(\mathbf{I}_{tp} - \mathbf{A})\mathbf{\Omega}(\mathbf{I}_{tp} - \mathbf{A})\mathbf{\Omega}^{-1}(\mathbf{I}_{tp} - \mathbf{A})\mathbf{\Omega}^{1/2} \\
&= \mathbf{\Omega}^{1/2}(\mathbf{I}_{tp} - \mathbf{A})\mathbf{\Omega}^{-1}\mathbf{R}(\mathbf{I}_{tp} - \mathbf{A})\mathbf{\Omega}^{1/2} \\
&= \mathbf{\Omega}^{1/2}(\mathbf{I}_{tp} - \mathbf{A})\mathbf{\Omega}^{-1}(\mathbf{I}_{tp} - \mathbf{A})\mathbf{\Omega}^{1/2} = \mathbf{Q},
\end{aligned}$$

because

$$\mathbf{R}(\mathbf{I}_{tp} - \mathbf{A}) = (\mathbf{I}_{tp} - \mathbf{A})(\mathbf{I}_{tp} - \mathbf{A}) = (\mathbf{I}_{tp} - \mathbf{A}).$$

### Proof for $V_n$

We now compute the asymptotic distribution of  $V_n$ .

$$\begin{aligned}
V_n &= -n \left( \log |SS_A| - \log |\mathbf{h}(SS_A)| \right) \\
&= -n \left\{ \log |SS_A| - \log |SS_A + (\mathbf{h}(SS_A) - SS_A)| \right\} \\
&\approx n \log \left| \mathbf{I}_{tp} + (1/n)\mathbf{\Omega}^{-1}(\mathbf{h}(SS_A) - SS_A) \right| \tag{8}
\end{aligned}$$

$$\begin{aligned}
&\approx \text{trace} \left\{ \mathbf{\Omega}^{-1}(\mathbf{h}(SS_A) - SS_A) \right\} \\
&\quad + \frac{1}{2n} \text{trace} \left\{ \mathbf{\Omega}^{-1}(\mathbf{h}(SS_A) - SS_A)\mathbf{\Omega}^{-1}(\mathbf{h}(SS_A) - SS_A) \right\} + o_p(1). \tag{9}
\end{aligned}$$

Here, the approximation (8) is from, for sufficiently large  $n$ ,  $SS_A \approx n\mathbf{\Omega}$ , and (8) is from the results of Martin (1993) that is

$$\log \left| \mathbf{I}_{tp} + A \right| = \sum_{j=1}^{\infty} \text{trace}(A^j) / j.$$

As shown in the proof of  $M_n$ ,

$$SS_A = \sum_{i=1}^n \left( \mathbf{\Omega}^{1/2} Z_i Z_i^T \mathbf{\Omega}^{1/2} \right) - \frac{1}{n} \left( \sum_{i=1}^n \mathbf{\Omega}^{1/2} Z_i \right) \left( \sum_{i=1}^n Z_i^T \mathbf{\Omega}^{1/2} \right),$$

where  $Z_i$ s are independent standard (tp)-variate normal random variables. Thus,

$$\mathbf{h}(SS_A) - SS_A \approx \mathbf{h} \left( \sum_{i=1}^n \mathbf{\Omega}^{1/2} Z_i Z_i^T \mathbf{\Omega}^{1/2} \right) - \sum_{i=1}^n \left( \mathbf{\Omega}^{1/2} Z_i Z_i^T \mathbf{\Omega}^{1/2} \right) + O_p(1).$$

Since the first term in (9) is equal to 0,

$$\begin{aligned} V_n \approx \frac{1}{2} \text{trace} \left\{ \mathbf{\Omega}^{-1} \left( \mathbf{h} \left( \sum_{i=1}^n \mathbf{\Omega}^{1/2} Z_i Z_i^T \mathbf{\Omega}^{1/2} \right) - \sum_{i=1}^n \mathbf{\Omega}^{1/2} Z_i Z_i^T \mathbf{\Omega}^{1/2} \right) \right. \\ \left. \mathbf{\Omega}^{-1} \left( \mathbf{h} \left( \sum_{i=1}^n \mathbf{\Omega}^{1/2} Z_i Z_i^T \mathbf{\Omega}^{1/2} \right) - \sum_{i=1}^n \mathbf{\Omega}^{1/2} Z_i Z_i^T \mathbf{\Omega}^{1/2} \right) \right\} + o_p(1). \end{aligned} \quad (10)$$

We use the following lemma without proof.

**Lemma 2.** *Under the null hypothesis, that is  $\Omega_{11} = \Omega_{22} = \dots = \Omega_{tt}$ , for any  $tp \times tp$  matrix  $B$ ,*

$$\mathbf{\Omega}^{-1/2} \mathbf{h}(B) = \mathbf{h}(\mathbf{\Omega}^{-1/2} B) \quad \text{and} \quad \mathbf{h}(B) \mathbf{\Omega}^{-1/2} = \mathbf{h}(B \mathbf{\Omega}^{-1/2}).$$

From Lemma 2, (10) has the distribution with

$$\frac{1}{2} \text{trace} \left\{ \left( \mathbf{h} \left( \sum_{i=1}^n Z_i Z_i^T \right) - \sum_{i=1}^n Z_i Z_i^T \right) \left( \mathbf{h} \left( \sum_{i=1}^n Z_i Z_i^T \right) - \sum_{i=1}^n Z_i Z_i^T \right) \right\}. \quad (11)$$

We now show that (11) has a Chi-squared distribution with the desired d.f. In (11),

$$(1/\sqrt{n}) \left\{ \mathbf{h} \left( \sum_{i=1}^n Z_i Z_i^T \right) - \sum_{i=1}^n Z_i Z_i^T \right\}$$

converges in distribution to

$$\begin{pmatrix} \mathbf{R}_1 & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{R}_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{R}_{t-1} & 0 \\ 0 & 0 & \cdots & 0 & \mathbf{R}_t \end{pmatrix}, \quad (12)$$

where

$$\mathbf{R}_k = \begin{pmatrix} r_{11}^{(k)} & r_{12}^{(k)} & \cdots & r_{1(t-1)}^{(k)} & r_{1t}^{(k)} \\ r_{21}^{(k)} & r_{22}^{(k)} & \cdots & r_{2(t-1)}^{(k)} & r_{2t}^{(k)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{(t-1)1}^{(k)} & r_{(t-1)2}^{(k)} & \cdots & r_{(t-1)(t-1)}^{(k)} & r_{(t-1)t}^{(k)} \\ r_{t1}^{(k)} & r_{t2}^{(k)} & \cdots & r_{t(t-1)}^{(k)} & r_{tt}^{(k)} \end{pmatrix} \quad (13)$$

with  $r_{ij}^{(k)} = r_{ji}^{(k)}$  and  $\sum_k r_{ij}^{(k)} = 0$ . Here, all distinct  $r_{ij}^{(k)}$ s are independent to each other, and  $r_{ii}^{(k)}$ s have the normal distribution with mean 0 and variance 2, and  $r_{ij}^{(k)}$ s have the normal distribution with mean 0 and variance 1. To sum up,

$$\begin{aligned} \mathbf{V}_n &\approx \frac{1}{2} \text{trace} \left\{ \left( \mathbf{h} \left( \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^T \right) - \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^T \right) \left( \mathbf{h} \left( \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^T \right) - \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^T \right) \right\} \\ &\approx \frac{1}{2} \sum_k \sum_{ij} \left\{ (r_{ij}^{(k)})^2 \right\}, \end{aligned}$$

which has the Chi-squared distribution with d.f.  $(t-1)\{p(p+1)/2\}$ .

### Proof for $\mathbf{W}_n$

Finally, we prove the convergence of  $\mathbf{W}_n$ . Note that

$$\begin{aligned} \mathbf{W}_n &= -n \left( \log |\mathbf{h}(SS_M)| - \log |SS_A| \right) \\ &= -n \left\{ \left( \log |\mathbf{h}(SS_A)| - \log |SS_A| \right) + \left( \log |\mathbf{h}(SS_M) - \log |\mathbf{h}(SS_A)| \right) \right\} \\ &= \mathbf{V}_n + n \left( \log |\mathbf{h}(SS_M) - \log |\mathbf{h}(SS_A)| \right) \\ &= \mathbf{V}_n + n \left( \log |\mathbf{h}(SS_M) - \log |\mathbf{h}(SS_A)| \right). \end{aligned} \tag{14}$$

With the arguments similar to the proof of  $\mathbf{M}_n$ , we have

$$\begin{aligned} &n \left( \log |\mathbf{h}(SS_M) - \log |\mathbf{h}(SS_A)| \right) \\ &= n \log \left| \mathbf{I}_{tp} + (1/n) \mathbf{h}(\boldsymbol{\Omega}^{-1}(\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega}^{1/2} \left( \sum_i \mathbf{Z}_i^T / \sqrt{n} \right) \left( \sum_i \mathbf{Z}_i / \sqrt{n} \right) \boldsymbol{\Omega}^{1/2} (\mathbf{I}_{tp} - \mathbf{A})) \right| \\ &\approx \text{trace} \left( \mathbf{h}(\boldsymbol{\Omega}^{-1}(\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega}^{1/2} \left( \sum_i \mathbf{Z}_i^T / \sqrt{n} \right) \left( \sum_i \mathbf{Z}_i / \sqrt{n} \right) \boldsymbol{\Omega}^{1/2} (\mathbf{I}_{tp} - \mathbf{A})) \right) \\ &\approx \text{trace} \left( \boldsymbol{\Omega}^{-1}(\mathbf{I}_{tp} - \mathbf{A}) \boldsymbol{\Omega}^{1/2} \mathbf{Z}^T \mathbf{Z} \boldsymbol{\Omega}^{1/2} (\mathbf{I}_{tp} - \mathbf{A}) \right) \\ &= \mathbf{Z}^T (\mathbf{I}_{tp} - \mathbf{A}) \mathbf{Z}, \end{aligned} \tag{15}$$

which has the Chi-squared distribution with d.f.  $(t-1)p$ . Further, the limiting distribution of  $\mathbf{V}_n$  does not depend on  $\Omega_{ij}$  for  $i, j$ s that is  $i \neq j$ . In evaluating  $SS_A$ , we could assume  $\Omega_{ij} = 0$  without loss of generality, and can get the independence between  $\mathbf{V}_\infty$  and (15). Thus,  $\mathbf{W}_n$  converges in distribution to the sum of two independent Chi-squared distribution  $(t-1)\{p + p(p+1)/2\}$ .



Table 2: Empirical type I error probabilities of the likelihood ratio test  $M_n$  under  $H_0$  for two correlated multivariate normal populations at level  $\alpha = 0.05$ .  $\mu_1 = \mu_2 = 0$ ;  $\sigma_1^2 = 1$ ;  $\rho_{12} = \rho_{21} = 0.3$ ;  $n$  is sample size;  $M_n$ : LRT for equality of mean vectors;  $\chi^2$ : the asymptotic Chi-squared test.

	$\sigma_2^2/\sigma_1^2 = 1$		$\sigma_2^2/\sigma_1^2 = 1.5$		$\sigma_2^2/\sigma_1^2 = 2$	
	$M_n$	$\chi_5^2$	$M_n$	$\chi_5^2$	$M_n$	$\chi_5^2$
<i>Same within-population correlation (<math>\rho_1 = \rho_2 = 0.5</math>)</i>						
$n = 25$	0.02	0.23	0.03	0.22	0.01	0.26
$n = 50$	0.03	0.09	0.02	0.10	0.03	0.14
$n = 100$	0.05	0.08	0.04	0.09	0.05	0.09
$n = 500$	0.05	0.05	0.04	0.06	0.05	0.07
<i>Different within-population correlation (<math>\rho_1 = 0.5, \rho_2 = 0</math>)</i>						
$n = 25$	0.01	0.32	0.03	0.37	0.01	0.36
$n = 50$	0.01	0.19	0.02	0.19	0.02	0.23
$n = 100$	0.03	0.16	0.06	0.18	0.05	0.22
$n = 500$	0.06	0.17	0.05	0.15	0.05	0.18



Table 4: Comparison of the proposed parametric bootstrap procedure and the alternative nonparametric bootstrap procedure, in terms of the empirical rejection probabilities of the likelihood ratio test  $M_n$  for the equality of mean vectors of two correlated multivariate normal populations at level  $\alpha = 0.05$ . Par: parametric bootstrap; Nonpar: nonparametric bootstrap;  $\mu_1 = 0$ ;  $\sigma_1^2 = 1$ ;  $\rho_1 = \rho_2 = 0.5$ ;  $\rho_{12} = \rho_{21} = 0.3$ ;  $n$  is sample size.

		$\mu_2 - \mu_1 = 0$		$\mu_2 - \mu_1 = 0.5$		$\mu_2 - \mu_1 = 1$	
		Par	Nonpar	Par	Nonpar	Par	Nonpar
$n = 25$	$\sigma_2^2/\sigma_1^2 = 1$	0.02	0.00	0.32	0.02	0.96	0.51
	$\sigma_2^2/\sigma_1^2 = 1.5$	0.03	0.00	0.22	0.01	0.84	0.25
	$\sigma_2^2/\sigma_1^2 = 2$	0.01	0.00	0.11	0.01	0.67	0.17
$n = 50$	$\sigma_2^2/\sigma_1^2 = 1$	0.03	0.02	0.90	0.84	1.00	1.00
	$\sigma_2^2/\sigma_1^2 = 1.5$	0.02	0.02	0.73	0.67	1.00	1.00
	$\sigma_2^2/\sigma_1^2 = 2$	0.03	0.01	0.56	0.45	1.00	0.99
$n = 100$	$\sigma_2^2/\sigma_1^2 = 1$	0.05	0.03	1.00	1.00	1.00	1.00
	$\sigma_2^2/\sigma_1^2 = 1.5$	0.04	0.04	0.99	0.98	1.00	1.00
	$\sigma_2^2/\sigma_1^2 = 2$	0.05	0.03	0.92	0.92	1.00	1.00