



Equivalence testing of mean vector and covariance matrix for multi-populations under a two-step monotone incomplete sample



Shin-ichi Tsukada

School of Education, Meisei University, 2-1-1 Hodokubo, Hino, Tokyo, 191-8506, Japan

ARTICLE INFO

Article history:

Received 2 April 2013

Available online 3 September 2014

AMS 2000 subject classifications:

62H10

62E20

Keywords:

Mean vector

Covariance matrix

Likelihood ratio criterion

Wald-type criterion

Monotone incomplete data

ABSTRACT

This paper investigates the hypothesis testing of a mean vector and covariance matrix for multi-populations in the context of two-step monotone incomplete data drawn from $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a multivariate normal population with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Three null hypotheses are considered, and the likelihood ratio criterion and Wald-type criterion are derived. On the basis of numerical simulations, the test that employs the Wald-type criterion is recommended.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Missing data often appear in practical data analysis. Therefore, rather than conventional statistical analyses, methods that identify missing data should be employed. These methods have been studied by many authors, which include Anderson and Olkin [3], Srivastava [10], Little [8], Kanda and Fujikoshi [7], and Chang and Richards [4,5].

In this paper, three hypothesis tests are considered in the context of two-step monotone incomplete data drawn from $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a $(p + q)$ -dimensional multivariate normal population with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. To begin, suppose that the data are composed of N mutually independent observations consisting of a random sample of n complete observations and $N - n$ additional observations on \mathbf{x} alone. That is,

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{pmatrix}, \begin{pmatrix} \mathbf{x}_2 \\ \mathbf{y}_2 \end{pmatrix}, \dots, \begin{pmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{pmatrix}, \begin{pmatrix} \mathbf{x}_{n+1} \\ * \end{pmatrix}, \dots, \begin{pmatrix} \mathbf{x}_N \\ * \end{pmatrix} \sim N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (1)$$

where \mathbf{x} is a $p \times 1$ vector, \mathbf{y} is a $q \times 1$ vector, and the symbol $*$ denotes the missing data. The data in (1) are usually referred to as two-step monotone incomplete data; these data show the simplest pattern with the missing data. The maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be explicitly expressed and are given by Anderson and Olkin [3]. Properties of the maximum likelihood estimator are discussed by Kanda and Fujikoshi [7] and Chang and Richards [4,5].

Hao and Krishnamoorthy [6] considered a hypothesis test for a k -step monotone incomplete sample where the covariance matrix is equal to a specified matrix, and the mean vector and the covariance matrix are equal to a given vector and matrix, respectively. They derived the likelihood ratio criterion and asymptotic null distribution. Provost [9] considered the mutual

E-mail address: tsukada@ed.meisei-u.ac.jp.

independence of a covariance matrix for a two-step monotone incomplete sample, and derived the likelihood ratio criterion as well. More recently, Chang and Richards [5] derived the likelihood ratio criterion for the hypothesis testing of a mean vector and covariance matrix for a two-step monotone incomplete sample. In this paper, the following hypothesis tests, where the mean vector and covariance matrix are equivalent for multi-populations under a two-step monotone incomplete sample, are considered:

$$\begin{aligned} H_{01} : \Sigma_1 = \cdots = \Sigma_g, & & H_{11} : H_{01} \text{ is not valid,} \\ H_{02} : \mu_1 = \cdots = \mu_g, & \Sigma_1 = \cdots = \Sigma_g, & H_{12} : H_{02} \text{ is not valid,} \\ H_{03} : \mu_1 = \cdots = \mu_g, & & H_{13} : H_{03} \text{ is not valid,} \end{aligned}$$

where μ_i and Σ_i are the population mean vector and population covariance matrix, respectively, of the group i . We derive the likelihood ratio criterion and the Wald-type criterion using the maximum likelihood estimator or the unbiased estimator by Tsukada [11]. These estimators are also used to investigate the size and power of the test.

To begin, preliminary results are described in Section 2. Sections 3–5 discuss hypothesis testing for the null hypotheses H_{01} , H_{02} , and H_{03} , respectively. The necessary criteria for testing the hypotheses are derived in these sections as well. The size and power of the tests that use these criteria are compared by performing numerical simulations. These results are presented in Section 6.

2. Preliminary results

When data are chosen according to (1), the sample means are defined as

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i, \quad \bar{\mathbf{x}}_1 = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i, \quad \bar{\mathbf{x}}_2 = \frac{1}{N-n} \sum_{i=n+1}^N \mathbf{x}_i, \quad \bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i, \quad (2)$$

and the corresponding matrices of sums of squares and products are given by

$$\begin{aligned} \mathbf{W}_{11}^{(1)} &= \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}_1) (\mathbf{x}_i - \bar{\mathbf{x}}_1)', & \mathbf{W}_{12}^{(1)} &= \mathbf{W}_{21}^{(1)'} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}_1) (\mathbf{y}_i - \bar{\mathbf{y}})', \\ \mathbf{W}_{22}^{(1)} &= \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}}) (\mathbf{y}_i - \bar{\mathbf{y}})', & \mathbf{W}^{(2)} &= \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})' - \mathbf{W}_{11}^{(1)}. \end{aligned} \quad (3)$$

The maximum likelihood estimators $\hat{\mu}$ and $\hat{\Sigma}$, obtained by Anderson [1], are as follows:

$$\begin{aligned} \hat{\mu}_1 &= \bar{\mathbf{x}}, & \hat{\mu}_2 &= \bar{\mathbf{y}} - (1 - \tau) \mathbf{W}_{21}^{(1)} \mathbf{W}_{11}^{(1)-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2), \\ \hat{\Sigma}_{11} &= \frac{1}{N} (\mathbf{W}_{11}^{(1)} + \mathbf{W}^{(2)}), & \hat{\Sigma}_{12} &= \hat{\Sigma}_{21}' = \hat{\Sigma}_{11} \mathbf{W}_{11}^{(1)-1} \mathbf{W}_{12}^{(1)}, \\ \hat{\Sigma}_{22} &= \frac{1}{n} \mathbf{W}_{22}^{(1)} + \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12}, \end{aligned} \quad (4)$$

where $\tau = (N - n)/N$. Kanda and Fujikoshi [7] derived the expectation of maximum likelihood estimators $\hat{\mu}$ and $\hat{\Sigma}$ and pointed out that $\hat{\Sigma}$ is not unbiased. Tsukada [11] proposed the following unbiased estimator for Σ ,

$$\begin{aligned} \tilde{\Sigma}_{11} &= \frac{N}{N-1} \hat{\Sigma}_{11}, & \tilde{\Sigma}_{12} &= \tilde{\Sigma}_{21}' = \frac{N}{N-1} \hat{\Sigma}_{12}, \\ \tilde{\Sigma}_{22} &= \frac{N}{N-1} \hat{\Sigma}_{22} - c_0 \hat{\Sigma}_{22-1}, \end{aligned} \quad (5)$$

and showed that the risk of the unbiased estimator is smaller than that of the maximum likelihood estimator with regard to Stein's loss function, where $\hat{\Sigma}_{22-1} = \hat{\Sigma}_{22} - \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12}$ and

$$c_0 = \frac{(N-n)(p+1)(p+2) - n(N-n)}{(N-1)(n-p-2)(n-p-1)}.$$

The asymptotic distribution of the estimators is given by the following theorems.

Theorem 2.1 (Chang and Richards [4]). Suppose $n, N \rightarrow \infty$ with $0 < n/N \leq 1$. For $n > q + 2$, the asymptotic distribution of $\hat{\mu}$ is

$$N_{p+q} \left(\mu, \frac{1}{N} \Sigma + \frac{\tau(n-2)}{n(n-p-2)} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22-1} \end{pmatrix} \right).$$

Theorem 2.2 (Tsukada [11]). Suppose $n, N \rightarrow \infty$ with $0 < n/N \leq 1$. Let $r = (p + q)^2 - pq$ and

$$\dot{\Sigma} = \begin{pmatrix} \dot{\Sigma}_{11} & \dot{\Sigma}_{12} \\ \dot{\Sigma}'_{12} & \dot{\Sigma}_{22} \end{pmatrix}$$

be the maximum likelihood estimator (4) or the unbiased estimator (5) for Σ . The vector

$$\sqrt{N} \left(\text{vec}'(\dot{\Sigma}_{11} - \Sigma_{11}), \text{vec}'(\dot{\Sigma}_{12} - \Sigma_{12}), \text{vec}'(\dot{\Sigma}_{22} - \Sigma_{22}) \right)'$$

is asymptotically distributed as a normal distribution with mean $\mathbf{0}_r$ and covariance matrix

$$\Theta = \begin{pmatrix} \Theta_{11} & \Theta_{12} & \Theta_{13} \\ \Theta'_{12} & \Theta_{22} & \Theta_{23} \\ \Theta'_{13} & \Theta'_{23} & \Theta_{33} \end{pmatrix}, \quad (6)$$

where

$$\begin{aligned} \Theta_{11} &= (\mathbf{I}_{p^2} + \mathbf{K}_{pp})(\Sigma_{11} \otimes \Sigma_{11}), & \Theta_{12} &= (\mathbf{I}_{p^2} + \mathbf{K}_{pp})(\Sigma_{12} \otimes \Sigma_{11}), \\ \Theta_{13} &= (\mathbf{I}_{p^2} + \mathbf{K}_{pp})(\Sigma_{12} \otimes \Sigma_{12}), \\ \Theta_{22} &= \frac{1}{1-\tau}(\Sigma_{22} \otimes \Sigma_{11}) - \frac{\tau}{1-\tau}(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \otimes \Sigma_{11}) + \mathbf{K}_{qp}(\Sigma_{12} \otimes \Sigma_{21}), \\ \Theta_{23} &= \frac{1}{1-\tau}(\Sigma_{22} \otimes \Sigma_{12})(\mathbf{I}_{q^2} + \mathbf{K}_{qq}) - \frac{\tau}{1-\tau}(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \otimes \Sigma_{12})(\mathbf{I}_{q^2} + \mathbf{K}_{qq}), \\ \Theta_{33} &= \frac{1}{1-\tau}(\mathbf{I}_{q^2} + \mathbf{K}_{qq})(\Sigma_{22} \otimes \Sigma_{22}) - \frac{\tau}{1-\tau}(\mathbf{I}_{q^2} + \mathbf{K}_{qq})(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \otimes \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}). \end{aligned}$$

To maximize the likelihood function, the following lemma is considered.

Lemma 2.1 (Anderson [2], page 69). Let \mathbf{W} be any $p \times p$ symmetric positive definite matrix and let

$$f(\mathbf{W}) = |\mathbf{W}|^{N/2} \exp \left[-\frac{1}{2} \text{tr} \mathbf{W} \right].$$

Then, $f(\mathbf{W})$ has a maximum value

$$N^{Np/2} \exp \left[-\frac{Np}{2} \right]$$

in the space of all positive definite matrices when $\mathbf{W} = N\mathbf{I}_p$.

Let the distribution of the population Π_k be a $(p + q)$ -variate normal distribution $N_{p+q}(\boldsymbol{\mu}_k, \Sigma_k)$, with mean $\boldsymbol{\mu}_k = (\boldsymbol{\mu}_1^{(k)'}, \boldsymbol{\mu}_2^{(k)'})'$ and covariance matrix

$$\Sigma_k = \begin{pmatrix} \Sigma_{11}^{(k)} & \Sigma_{12}^{(k)} \\ \Sigma_{21}^{(k)} & \Sigma_{22}^{(k)} \end{pmatrix}, \quad (k = 1, \dots, g).$$

Assume that the following observations from the population Π_k , which is of the same form as (1), are available:

$$\begin{pmatrix} \mathbf{x}_1^{(k)} \\ \mathbf{y}_1^{(k)} \end{pmatrix}, \begin{pmatrix} \mathbf{x}_2^{(k)} \\ \mathbf{y}_2^{(k)} \end{pmatrix}, \dots, \begin{pmatrix} \mathbf{x}_{n_k}^{(k)} \\ \mathbf{y}_{n_k}^{(k)} \end{pmatrix}, \begin{pmatrix} \mathbf{x}_{n_k+1}^{(k)} \\ * \end{pmatrix}, \dots, \begin{pmatrix} \mathbf{x}_{N_k}^{(k)} \\ * \end{pmatrix},$$

where all N_k samples are mutually independent. As explained by Yamada et al. [12], it is also assumed that data are randomly missing; it is necessary to assume that their absence is completely random in order to derive the maximum likelihood estimators $\hat{\boldsymbol{\mu}}_k$ and $\hat{\Sigma}_k$.

3. Hypothesis testing for the covariance matrix

In this section, the null hypothesis tested is

$$H_{01} : \Sigma_1 = \dots = \Sigma_g, \quad H_{11} : H_{01} \text{ is not valid.}$$

The likelihood ratio criterion is derived in Section 3.1, and the Wald-type criterion is derived in Section 3.2.

3.1. Likelihood ratio criterion

The likelihood function $L(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_g; \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_g)$ is given by

$$L(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_g; \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_g) = \prod_{k=1}^g c_k \left| \boldsymbol{\Sigma}_{11}^{(k)} \right|^{-N_k/2} \left| \boldsymbol{\Sigma}_{22 \cdot 1}^{(k)} \right|^{-n_k/2} \exp \left[-\frac{1}{2} \phi_1^{(k)} \right] \exp \left[-\frac{1}{2} \phi_2^{(k)} \right],$$

where

$$\begin{aligned} c_k &= (2\pi)^{-(pN_k + qn_k)/2}, \\ \phi_1^{(k)} &= \text{tr} \left\{ \boldsymbol{\Sigma}_{11}^{(k)-1} \left(\mathbf{w}_{11,k}^{(1)} + \mathbf{w}_k^{(2)} \right) \right\} + \text{tr} \left\{ N_k \boldsymbol{\Sigma}_{11}^{(k)-1} \left(\bar{\mathbf{x}}^{(k)} - \boldsymbol{\mu}_1^{(k)} \right) \left(\bar{\mathbf{x}}^{(k)} - \boldsymbol{\mu}_1^{(k)} \right)' \right\}, \\ \phi_2^{(k)} &= \text{tr} \left\{ \boldsymbol{\Sigma}_{22 \cdot 1}^{(k)-1} \left(\mathbf{w}_{22,k}^{(1)} - \mathbf{w}_{21,k}^{(1)} \boldsymbol{\Sigma}_{12}^{(k)*} - \boldsymbol{\Sigma}_{21}^{(k)*} \mathbf{w}_{12,k}^{(1)} + \boldsymbol{\Sigma}_{21}^{(k)*} \mathbf{w}_{11,k}^{(1)} \boldsymbol{\Sigma}_{12}^{(k)*} \right) \right\} \\ &\quad + \text{tr} \left\{ n_k \boldsymbol{\Sigma}_{22 \cdot 1}^{(k)-1} \left[\left(\bar{\mathbf{y}}^{(k)} - \boldsymbol{\mu}_2^{(k)*} - \boldsymbol{\Sigma}_{21}^{(k)*} \bar{\mathbf{x}}_{1,k} \right) \left(\bar{\mathbf{y}}^{(k)} - \boldsymbol{\mu}_2^{(k)*} - \boldsymbol{\Sigma}_{21}^{(k)*} \bar{\mathbf{x}}_{1,k} \right)' \right] \right\}, \\ \bar{\mathbf{x}}_{1,k} &= \frac{1}{n_k} \sum_{j=1}^{n_k} \mathbf{x}_j^{(k)}, \quad \bar{\mathbf{x}}_{2,k} = \frac{1}{N_k - n_k} \sum_{j=n_k+1}^{N_k} \mathbf{x}_j^{(k)}, \\ \bar{\mathbf{x}}^{(k)} &= \frac{1}{N_k} \sum_{j=1}^{N_k} \mathbf{x}_j^{(k)}, \quad \bar{\mathbf{y}}^{(k)} = \frac{1}{n_k} \sum_{j=1}^{n_k} \mathbf{y}_j^{(k)}, \\ \mathbf{w}_{11,k}^{(1)} &= \sum_{j=1}^{n_k} \left(\mathbf{x}_j^{(k)} - \bar{\mathbf{x}}_{1,k} \right) \left(\mathbf{x}_j^{(k)} - \bar{\mathbf{x}}_{1,k} \right)', \quad \mathbf{w}_{12,k}^{(1)} = \sum_{j=1}^{n_k} \left(\mathbf{x}_j^{(k)} - \bar{\mathbf{x}}_{1,k} \right) \left(\mathbf{y}_j^{(k)} - \bar{\mathbf{y}}^{(k)} \right)', \\ \mathbf{w}_{22,k}^{(1)} &= \sum_{j=1}^{n_k} \left(\mathbf{y}_j^{(k)} - \bar{\mathbf{y}}^{(k)} \right) \left(\mathbf{y}_j^{(k)} - \bar{\mathbf{y}}^{(k)} \right)', \\ \mathbf{w}_k^{(2)} &= \sum_{j=n_k+1}^{N_k} \left(\mathbf{x}_j^{(k)} - \bar{\mathbf{x}}_{2,k} \right) \left(\mathbf{x}_j^{(k)} - \bar{\mathbf{x}}_{2,k} \right)' + \frac{n_k(N_k - n_k)}{N_k} \left(\bar{\mathbf{x}}_{1,k} - \bar{\mathbf{x}}_{2,k} \right) \left(\bar{\mathbf{x}}_{1,k} - \bar{\mathbf{x}}_{2,k} \right)', \\ \boldsymbol{\Sigma}_{21}^{(k)*} &= \boldsymbol{\Sigma}_{21}^{(k)} - \boldsymbol{\Sigma}_{11}^{(k)-1}, \quad \boldsymbol{\mu}_2^{(k)*} = \boldsymbol{\mu}_2^{(k)} - \boldsymbol{\Sigma}_{21}^{(k)*} \boldsymbol{\mu}_1^{(k)}, \quad (k = 1, \dots, g). \end{aligned}$$

The statistics with a subscript for the population are defined similarly to those of (3)–(5). The likelihood function is maximized with respect to $\boldsymbol{\mu}_2^{(k)*}$ when

$$\hat{\boldsymbol{\mu}}_2^{(k)*} = \bar{\mathbf{y}}^{(k)} - \hat{\boldsymbol{\Sigma}}_{21}^{(k)*} \bar{\mathbf{x}}_{1,k}, \quad (7)$$

and the maximum likelihood estimator of $\boldsymbol{\mu}_1^{(k)}$ is

$$\hat{\boldsymbol{\mu}}_1^{(k)} = \bar{\mathbf{x}}^{(k)}, \quad (k = 1, \dots, g). \quad (8)$$

Letting $\boldsymbol{\mu}_2^{(k)*} = \hat{\boldsymbol{\mu}}_2^{(k)*}$ and $\boldsymbol{\mu}_1^{(k)} = \hat{\boldsymbol{\mu}}_1^{(k)}$, we have

$$\begin{aligned} L(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_g) &\equiv \text{Max}_{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_g} L(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_g; \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_g) \\ &= \prod_{k=1}^g c_k \left| \boldsymbol{\Sigma}_{11}^{(k)} \right|^{-N_k/2} \left| \boldsymbol{\Sigma}_{22 \cdot 1}^{(k)} \right|^{-n_k/2} \exp \left[-\frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}_{11}^{(k)-1} \left(\mathbf{w}_{11,k}^{(1)} + \mathbf{w}_k^{(2)} \right) \right\} \right] \\ &\quad \times \exp \left[-\frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}_{22 \cdot 1}^{(k)-1} \mathbf{s}_{22}^{(k)*} \right\} \right], \end{aligned} \quad (9)$$

where

$$\mathbf{s}_{22}^{(k)*} = \mathbf{w}_{22,k}^{(1)} - \mathbf{w}_{21,k}^{(1)} \boldsymbol{\Sigma}_{12}^{(k)*} - \boldsymbol{\Sigma}_{21}^{(k)*} \mathbf{w}_{12,k}^{(1)} + \boldsymbol{\Sigma}_{21}^{(k)*} \mathbf{w}_{11,k}^{(1)} \boldsymbol{\Sigma}_{12}^{(k)*}.$$

Using the lemma proved by Anderson [2], the function $L(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_g)$ is maximized with respect to $\boldsymbol{\Sigma}_{11}^{(k)}$ and $\boldsymbol{\Sigma}_{22 \cdot 1}^{(k)}$ as follows:

$$\text{Max}_{\boldsymbol{\Sigma}_{11}^{(k)}, \boldsymbol{\Sigma}_{22 \cdot 1}^{(k)}} L(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_g) = \prod_{k=1}^g c_k N_k^{N_k p/2} n_k^{n_k q/2} \exp \left[-\frac{1}{2} (N_k p + n_k q) \right] \left| \mathbf{w}_{11,k}^{(1)} + \mathbf{w}_k^{(2)} \right|^{-N_k/2} \left| \mathbf{s}_{22}^{(k)*} \right|^{-n_k/2}, \quad (10)$$

when

$$\hat{\Sigma}_{11}^{(k)} = \frac{1}{N_k} \left(\mathbf{W}_{11,k}^{(1)} + \mathbf{W}_k^{(2)} \right) \quad \text{and} \quad \hat{\Sigma}_{22,1}^{(k)} = \frac{1}{n_k} \left\{ \mathbf{W}_{22,k}^{(1)} - \mathbf{W}_{21,k}^{(1)} \left(\mathbf{W}_{11,k}^{(1)} \right)^{-1} \mathbf{W}_{12,k}^{(1)} \right\}.$$

Noting that

$$\begin{aligned} \left| \mathbf{s}_{22}^{(k)*} \right| &= \left| \mathbf{W}_{22,k}^{(1)} - \Sigma_{21}^{(k)*} \mathbf{W}_{12,k}^{(1)} - \mathbf{W}_{21,k}^{(1)} \Sigma_{12}^{(k)*} + \Sigma_{21}^{(k)*} \mathbf{W}_{11,k}^{(1)} \Sigma_{12}^{(k)*} \right| \\ &= \left| \mathbf{W}_{22,k}^{(1)} - \mathbf{W}_{21,k}^{(1)} \left(\mathbf{W}_{11,k}^{(1)} \right)^{-1} \mathbf{W}_{12,k}^{(1)} + \left\{ \Sigma_{21}^{(k)*} - \mathbf{W}_{21,k}^{(1)} \left(\mathbf{W}_{11,k}^{(1)} \right)^{-1} \right\} \mathbf{W}_{11,k}^{(1)} \left\{ \Sigma_{21}^{(k)*} - \mathbf{W}_{21,k}^{(1)} \left(\mathbf{W}_{11,k}^{(1)} \right)^{-1} \right\}' \right|, \end{aligned}$$

the function (10) is maximized with respect to $\Sigma_{21}^{(k)*}$ when

$$\hat{\Sigma}_{21}^{(k)*} = \mathbf{W}_{21,k}^{(1)} \left(\mathbf{W}_{11,k}^{(1)} \right)^{-1}, \quad (k = 1, \dots, g).$$

As a result,

$$\begin{aligned} \text{Max}_{\mu_1, \dots, \mu_g, \Sigma_1, \dots, \Sigma_g} L(\mu_1, \dots, \mu_g; \Sigma_1, \dots, \Sigma_g) &= \prod_{k=1}^g c_k N_k^{N_k p/2} n_k^{n_k q/2} \exp \left[-\frac{1}{2} (N_k p + n_k q) \right] \\ &\quad \times \left| \mathbf{W}_{11,k}^{(1)} + \mathbf{W}_k^{(2)} \right|^{-N_k/2} \left| \mathbf{W}_{22,k}^{(1)} - \mathbf{W}_{21,k}^{(1)} \left(\mathbf{W}_{11,k}^{(1)} \right)^{-1} \mathbf{W}_{12,k}^{(1)} \right|^{-n_k/2}. \end{aligned} \quad (11)$$

The maximum of similar likelihood functions for hypothesis tests involving μ and Σ have also been obtained and can be found explicitly in Chang and Richards [5].

Next, the maximum value of the likelihood function is considered under the null hypothesis. Let $\Sigma_1 = \dots = \Sigma_g \equiv \Sigma$. In this case, the likelihood function is

$$L(\mu_1, \dots, \mu_g; \Sigma) = \prod_{k=1}^g c_k |\Sigma_{11}|^{N_k/2} |\Sigma_{22,1}|^{n_k/2} \exp \left[-\frac{1}{2} \varphi_1^{(k)} \right] \exp \left[-\frac{1}{2} \varphi_2^{(k)} \right], \quad (12)$$

where

$$\begin{aligned} \varphi_1^{(k)} &= \text{tr} \left\{ \Sigma_{11}^{-1} \left(\mathbf{W}_{11,k}^{(1)} + \mathbf{W}_k^{(2)} \right) \right\} + \text{tr} \left\{ N_k \Sigma_{11}^{-1} \left(\bar{\mathbf{x}}^{(k)} - \mu_1^{(k)} \right) \left(\bar{\mathbf{x}}^{(k)} - \mu_1^{(k)} \right)' \right\}, \\ \varphi_2^{(k)} &= \text{tr} \left\{ \Sigma_{22,1}^{-1} \left(\mathbf{W}_{22,k}^{(1)} - \mathbf{W}_{21,k}^{(1)} \Sigma_{12}^* - \Sigma_{21}^* \mathbf{W}_{12,k}^{(1)} + \Sigma_{21}^* \mathbf{W}_{11,k}^{(1)} \Sigma_{12}^* \right) \right\} \\ &\quad + \text{tr} \left\{ n_k \Sigma_{22,1}^{-1} \left[\left(\bar{\mathbf{y}}^{(k)} - \mu_{2,k}^* - \Sigma_{21}^* \bar{\mathbf{x}}_{1,k} \right) \left(\bar{\mathbf{y}}^{(k)} - \mu_{2,k}^* - \Sigma_{21}^* \bar{\mathbf{x}}_{1,k} \right)' \right] \right\}, \\ \Sigma_{21}^* &= \Sigma_{21} - \Sigma_{11}^{-1}, \quad \text{and} \quad \mu_{2,k}^* = \mu_2^{(k)} - \Sigma_{21}^{(k)*} \mu_1^{(k)}. \end{aligned}$$

When $\mu_1^{(k)} = \bar{\mathbf{x}}^{(k)}$ and $\mu_{2,k}^* = \bar{\mathbf{y}}^{(k)} - \Sigma_{21}^* \bar{\mathbf{x}}_{1,k}$, the likelihood function (12) is maximized as follows:

$$\begin{aligned} L(\Sigma) &\equiv \text{Max}_{\mu_1, \dots, \mu_g} L(\mu_1, \dots, \mu_g; \Sigma) \\ &= \prod_{k=1}^g c_k |\Sigma_{11}|^{-N_k/2} |\Sigma_{22,1}|^{-n_k/2} \exp \left[-\frac{1}{2} \text{tr} \left\{ \Sigma_{11}^{-1} \left(\mathbf{W}_{11,k}^{(1)} + \mathbf{W}_k^{(2)} \right) \right\} \right] \\ &\quad \times \exp \left[-\frac{1}{2} \text{tr} \left\{ \Sigma_{22,1}^{-1} \left(\mathbf{W}_{22,k}^{(1)} - \mathbf{W}_{21,k}^{(1)} \Sigma_{12}^* - \Sigma_{21}^* \mathbf{W}_{12,k}^{(1)} + \Sigma_{21}^* \mathbf{W}_{11,k}^{(1)} \Sigma_{12}^* \right) \right\} \right]. \end{aligned}$$

Letting $N = \sum_{k=1}^g N_k$,

$$\begin{aligned} |\Sigma_{11}|^{-N/2} \exp \left[-\frac{1}{2} \text{tr} \left\{ \Sigma_{11}^{-1} \left(\mathbf{W}_{11,k}^{(1)} + \mathbf{W}_k^{(2)} \right) \right\} \right] &= \left| \sum_{k=1}^g \left(\mathbf{W}_{11,k}^{(1)} + \mathbf{W}_k^{(2)} \right) \right|^{-N/2} \left| \Sigma_{11}^{-1} \sum_{k=1}^g \left(\mathbf{W}_{11,k}^{(1)} + \mathbf{W}_k^{(2)} \right) \right|^{N/2} \\ &\quad \times \exp \left[-\frac{1}{2} \text{tr} \left\{ \Sigma_{11}^{-1} \left(\mathbf{W}_{11,k}^{(1)} + \mathbf{W}_k^{(2)} \right) \right\} \right] \end{aligned}$$

attains its maximum when

$$\Sigma_{11}^{-1} \sum_{k=1}^g \left(\mathbf{W}_{11,k}^{(1)} + \mathbf{W}_k^{(2)} \right) = N \mathbf{I}_p.$$

That is,

$$\Sigma_{11} = \frac{1}{N} \sum_{k=1}^g \left(\mathbf{w}_{11,k}^{(1)} + \mathbf{w}_k^{(2)} \right),$$

which is equal to

$$\left| \sum_{k=1}^g \left(\mathbf{w}_{11,k}^{(1)} + \mathbf{w}_k^{(2)} \right) \right|^{N/2} N^{pN/2} \exp \left[-\frac{pN}{2} \right].$$

Similarly, letting $n. = \sum_{k=1}^g n_k$ and

$$\mathbf{S}_{22}^* = \sum_{k=1}^g \left(\mathbf{w}_{22,k}^{(1)} - \mathbf{w}_{21,k}^{(1)} \Sigma_{12}^* - \Sigma_{21}^* \mathbf{w}_{12,k}^{(1)} + \Sigma_{21}^* \mathbf{w}_{11,k}^{(1)} \Sigma_{12}^* \right),$$

the maximum of

$$|\Sigma_{22 \cdot 1}|^{-n./2} \exp \left[-\frac{1}{2} \text{tr} \left(\Sigma_{22 \cdot 1}^{-1} \mathbf{S}_{22}^* \right) \right] = |\mathbf{S}_{22}^*|^{-n./2} |\Sigma_{22 \cdot 1}^{-1} \mathbf{S}_{22}^*|^{n./2} \exp \left[-\frac{1}{2} \text{tr} \left\{ \Sigma_{22 \cdot 1}^{-1} \mathbf{S}_{22}^* \right\} \right]$$

is attained when

$$\Sigma_{22 \cdot 1} = \frac{1}{n.} \mathbf{S}_{22}^*,$$

which is equal to

$$|\mathbf{S}_{22}^*|^{n./2} n^{qn./2} \exp \left[-\frac{pn.}{2} \right].$$

Hence, we have

$$\text{Max}_{\Sigma_{11}, \Sigma_{22 \cdot 1}} L(\Sigma) = \left(\prod_{k=1}^g c_k \right) \left| \sum_{k=1}^g \left(\mathbf{w}_{11,k}^{(1)} + \mathbf{w}_k^{(2)} \right) \right|^{N/2} N^{pN/2} \exp \left[-\frac{pN}{2} \right] |\mathbf{S}_{22}^*|^{n./2} n^{qn./2} \exp \left[-\frac{pn.}{2} \right].$$

Let

$$\mathbf{v}_{ij} = \sum_{k=1}^g \mathbf{w}_{ij,k}^{(1)}, \quad (i, j = 1, 2).$$

Since one can rewrite

$$\begin{aligned} |\mathbf{S}_{22}^*| &= |\mathbf{v}_{22} - \mathbf{v}_{21} \Sigma_{12}^* - \Sigma_{21}^* \mathbf{v}_{12} + \Sigma_{21}^* \mathbf{v}_{11} \Sigma_{12}^*| \\ &= \left| \mathbf{v}_{22} - \mathbf{v}_{21} \mathbf{v}_{11}^{-1} \mathbf{v}_{12} + (\Sigma_{21}^* - \mathbf{v}_{21} \mathbf{v}_{11}^{-1}) \mathbf{v}_{11} (\Sigma_{21}^* - \mathbf{v}_{21} \mathbf{v}_{11}^{-1})' \right|, \end{aligned}$$

the function $|\mathbf{S}_{22}^*|$ is maximized when $\Sigma_{21}^* = \mathbf{v}_{21} \mathbf{v}_{11}^{-1}$. Therefore, the maximum value under the null hypothesis is as follows:

$$\text{Max}_{\mu_1, \dots, \mu_g; \Sigma} = \left(\prod_{k=1}^g c_k \right) N^{pN/2} n^{qn./2} \exp \left[-\frac{pN}{2} \right] \exp \left[-\frac{qn.}{2} \right] |\mathbf{v}_{11} + \mathbf{v}^{(2)}|^{-N/2} |\mathbf{v}_{22} - \mathbf{v}_{21} \mathbf{v}_{11}^{-1} \mathbf{v}_{12}|^{-n./2}, \quad (13)$$

where

$$\mathbf{v}^{(2)} = \sum_{k=1}^g \mathbf{w}_k^{(2)}.$$

From (11) and (13), the log-likelihood ratio criterion is

$$\begin{aligned} -2 \log \lambda &= -2 \log \frac{\text{Max}_{H_{01}} L(\mu_1, \dots, \mu_g; \Sigma)}{\text{Max}_{H_{11}} L(\mu_1, \dots, \mu_g; \Sigma_1, \dots, \Sigma_g)} \\ &= N \log |\mathbf{v}_{11} + \mathbf{v}^{(2)}| + n. \log |\mathbf{v}_{22} - \mathbf{v}_{21} \mathbf{v}_{11}^{-1} \mathbf{v}_{12}| - \sum_{k=1}^g N_k \log |\mathbf{w}_{11,k}^{(1)} + \mathbf{w}_k^{(2)}| \\ &\quad - \sum_{k=1}^g n_k \log \left| \mathbf{w}_{22,k}^{(1)} - \mathbf{w}_{21,k}^{(1)} \left(\mathbf{w}_{11,k}^{(1)} \right)^{-1} \mathbf{w}_{12,k}^{(1)} \right| \\ &\quad - pN \log N - qn. \log n. + p \sum_{k=1}^g N_k \log N_k + q \sum_{k=1}^g n_k \log n_k. \end{aligned} \quad (14)$$

Since the number of parameters are $g(p+q) + g(p+q)(p+q+1)/2$ and $g(p+q) + (p+q)(p+q+1)/2$ under the alternative hypothesis and the null hypothesis, respectively, we have the following theorem from an asymptotic property for the log-likelihood ratio criterion.

Theorem 3.1. Suppose $n, N \rightarrow \infty$ with $0 < n/N \leq 1$. Under the null hypothesis H_{01} , the log-likelihood ratio criterion is asymptotically distributed as a chi-squared distribution with f degrees of freedom, where

$$f = \frac{1}{2}(g-1)(p+q)(p+q+1).$$

3.2. Wald-type criterion

In this subsection, the Wald-type criterion is evaluated using the asymptotic covariance matrix of the maximum likelihood estimator and the unbiased estimator for Σ given by Tsukada [11]. From Theorem 2.2, the Wald-type criterion is evaluated for testing the null hypothesis H_{01} . Let $\pi_k = N_k/N$,

$$\begin{aligned} \mathbf{u}^{(k)} &= \left(\text{vec}' \left(\hat{\Sigma}_{11}^{(k)} \right), \text{vec}' \left(\hat{\Sigma}_{12}^{(k)} \right), \text{vec}' \left(\hat{\Sigma}_{22}^{(k)} \right) \right)', \\ \mathbf{v}^{(k)} &= \left(\text{vec}' \left(\tilde{\Sigma}_{11}^{(k)} \right), \text{vec}' \left(\tilde{\Sigma}_{12}^{(k)} \right), \text{vec}' \left(\tilde{\Sigma}_{22}^{(k)} \right) \right)', \\ \mathbf{v}^{(k)} &= \left(\text{vec}' \left(\Sigma_{11}^{(k)} \right), \text{vec}' \left(\Sigma_{12}^{(k)} \right), \text{vec}' \left(\Sigma_{22}^{(k)} \right) \right)', \quad (k = 1, \dots, g), \\ r &= \frac{1}{2}p(p+1) + pq + \frac{1}{2}q(q+1). \end{aligned}$$

Moreover, let \mathbf{Q} be a $(g-1) \times g$ matrix defined by

$$\begin{pmatrix} \sqrt{\pi_2} & -\sqrt{\pi_1} & & & & \\ & \sqrt{\pi_3} & -\sqrt{\pi_2} & & & \\ & & \ddots & \ddots & & \\ & & & \sqrt{\pi_{g-1}} & -\sqrt{\pi_{g-2}} & \\ & \mathbf{0} & & & \sqrt{\pi_g} & -\sqrt{\pi_{g-1}} \end{pmatrix},$$

and let

$$\mathbf{J} = \begin{pmatrix} \mathbf{B}_p' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{pq} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_q' \end{pmatrix},$$

where \mathbf{B}_p is the transition matrix. This matrix is a $p^2 \times p(p+1)/2$ matrix such that

$$\text{vech}(\mathbf{X}) = \mathbf{B}_p' \text{vec}(\mathbf{X}),$$

for a $p \times p$ symmetric matrix \mathbf{X} . The vectors

$$\sqrt{N} \cdot (\mathbf{I}_g \otimes \mathbf{J}) \left(\sqrt{\pi_1} (\mathbf{u}^{(1)} - \mathbf{v}^{(1)})', \dots, \sqrt{\pi_g} (\mathbf{u}^{(g)} - \mathbf{v}^{(g)})' \right)'$$

and

$$\sqrt{N} \cdot (\mathbf{I}_g \otimes \mathbf{J}) \left(\sqrt{\pi_1} (\mathbf{v}^{(1)} - \mathbf{v}^{(1)})', \dots, \sqrt{\pi_g} (\mathbf{v}^{(g)} - \mathbf{v}^{(g)})' \right)'$$

are asymptotically distributed as a normal distribution with mean $\mathbf{0}$ and covariance matrix $(\mathbf{I}_g \otimes \mathbf{J}) \text{diag}(\Theta_1, \Theta_2, \dots, \Theta_g) (\mathbf{I}_g \otimes \mathbf{J})'$, where Θ_k is the asymptotic covariance matrix for the k th group, which is described in (6). Under the null hypothesis, the vectors

$$\sqrt{N} \cdot (\mathbf{Q} \otimes \mathbf{I}_r) (\mathbf{I}_g \otimes \mathbf{J}) \left(\sqrt{\pi_1} \mathbf{u}^{(1)'}, \dots, \sqrt{\pi_g} \mathbf{u}^{(g)'} \right)' \equiv \sqrt{N} \cdot (\mathbf{Q} \otimes \mathbf{I}_r) (\mathbf{I}_g \otimes \mathbf{J}) \mathbf{U}$$

and

$$\sqrt{N} \cdot (\mathbf{Q} \otimes \mathbf{I}_r) (\mathbf{I}_g \otimes \mathbf{J}) \left(\sqrt{\pi_1} \mathbf{v}^{(1)'}, \dots, \sqrt{\pi_g} \mathbf{v}^{(g)'} \right)' \equiv \sqrt{N} \cdot (\mathbf{Q} \otimes \mathbf{I}_r) (\mathbf{I}_g \otimes \mathbf{J}) \mathbf{V}$$

are asymptotically distributed as a normal distribution with mean $\mathbf{0}$ and covariance matrix $(\mathbf{Q} \otimes \mathbf{J}) \text{diag}(\Theta_1, \Theta_2, \dots, \Theta_g) (\mathbf{Q} \otimes \mathbf{J})'$. We have the following theorem by considering the number of parameters similar to the case of the log-likelihood ratio criterion.

Theorem 3.2. Suppose $n, N \rightarrow \infty$ with $0 < n/N \leq 1$. Under the null hypothesis H_{01} , the Wald-type criteria

$$WC_{11} = N \mathbf{U}'(\mathbf{Q} \otimes \mathbf{J})' \left\{ (\mathbf{Q} \otimes \mathbf{J}) \times \text{diag}(\hat{\boldsymbol{\Theta}}_1, \hat{\boldsymbol{\Theta}}_2, \dots, \hat{\boldsymbol{\Theta}}_g)(\mathbf{Q} \otimes \mathbf{J})' \right\}^{-1} (\mathbf{Q} \otimes \mathbf{J}) \mathbf{U} \quad (15)$$

and

$$WC_{12} = N \mathbf{V}'(\mathbf{Q} \otimes \mathbf{J})' \left\{ (\mathbf{Q} \otimes \mathbf{J}) \times \text{diag}(\tilde{\boldsymbol{\Theta}}_1, \tilde{\boldsymbol{\Theta}}_2, \dots, \tilde{\boldsymbol{\Theta}}_g)(\mathbf{Q} \otimes \mathbf{J})' \right\}^{-1} (\mathbf{Q} \otimes \mathbf{J}) \mathbf{V} \quad (16)$$

are asymptotically distributed as a chi-squared distribution with $(g-1)(p+q)(p+q+1)/2$ degrees of freedom, where $\hat{\boldsymbol{\Theta}}_i$ and $\tilde{\boldsymbol{\Theta}}_i$ are the matrices substituted as the maximum likelihood estimators and the unbiased estimators for all parameters, respectively.

4. Hypothesis testing for the mean vector and covariance matrix

The following null hypothesis is considered:

$$H_{02} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_g, \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \dots = \boldsymbol{\Sigma}_g,$$

$H_{12} : H_{02}$ is not valid.

The likelihood ratio criterion and the Wald-type criteria are considered as shown in the previous section.

4.1. Likelihood ratio criterion

Because the maximum value of the likelihood function has already been obtained under the alternative hypothesis, it is now considered under the null hypothesis. Let $\boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_g \equiv \boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2)'$, where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are p - and q -dimensional vectors, respectively, and $\boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_g \equiv \boldsymbol{\Sigma}$. In this case, the likelihood function is

$$L(\boldsymbol{\mu}; \boldsymbol{\Sigma}) = \prod_{k=1}^g c_k |\boldsymbol{\Sigma}_{11}|^{N_k/2} |\boldsymbol{\Sigma}_{22.1}|^{n_k/2} \exp \left[-\frac{1}{2} \varphi_1^{(k)} \right] \exp \left[-\frac{1}{2} \varphi_2^{(k)} \right], \quad (17)$$

where

$$\begin{aligned} \varphi_1^{(k)} &= \text{tr} \left\{ \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{W}_{11,k}^{(1)} + \mathbf{W}_{11,k}^{(2)}) \right\} + \text{tr} \left\{ N_k \boldsymbol{\Sigma}_{11}^{-1} (\bar{\mathbf{x}}^{(k)} - \boldsymbol{\mu}_1) (\bar{\mathbf{x}}^{(k)} - \boldsymbol{\mu}_1)' \right\}, \\ \varphi_2^{(k)} &= \text{tr} \left\{ \boldsymbol{\Sigma}_{22.1}^{-1} (\mathbf{W}_{22,k}^{(1)} - \mathbf{W}_{21,k}^{(1)} \boldsymbol{\Sigma}_{12}^* - \boldsymbol{\Sigma}_{21}^* \mathbf{W}_{12,k}^{(1)} + \boldsymbol{\Sigma}_{21}^* \mathbf{W}_{11,k}^{(1)} \boldsymbol{\Sigma}_{12}^*) \right\} \\ &\quad + \text{tr} \left\{ n_k \boldsymbol{\Sigma}_{22.1}^{-1} [(\bar{\mathbf{y}}^{(k)} - \boldsymbol{\mu}_2^* - \boldsymbol{\Sigma}_{21}^* \bar{\mathbf{x}}_{1,k}) (\bar{\mathbf{y}}^{(k)} - \boldsymbol{\mu}_2^* - \boldsymbol{\Sigma}_{21}^* \bar{\mathbf{x}}_{1,k})'] \right\}, \quad (k = 1, \dots, g), \\ \boldsymbol{\Sigma}_{21}^* &= \boldsymbol{\Sigma}_{21} - \boldsymbol{\Sigma}_{11}^{-1}, \quad \text{and} \quad \boldsymbol{\mu}_2^* = \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}^* \boldsymbol{\mu}_1. \end{aligned}$$

Differentiating $L(\boldsymbol{\mu}; \boldsymbol{\Sigma})$ with respect to $\boldsymbol{\mu}_1$ or $\boldsymbol{\mu}_2$, and setting it to zero yields

$$\hat{\boldsymbol{\mu}}_1 = \frac{1}{N} \sum_{k=1}^g N_k \bar{\mathbf{x}}^{(k)}, \quad \hat{\boldsymbol{\mu}}_2^* = \frac{1}{n} \sum_{k=1}^g n_k \bar{\mathbf{y}}^{(k)} - \boldsymbol{\Sigma}_{21}^* \frac{1}{n} \sum_{k=1}^g n_k \bar{\mathbf{x}}_{1,k}.$$

Considering

$$\begin{aligned} \bar{\mathbf{x}}^{(k)} - \hat{\boldsymbol{\mu}}_1 &= \bar{\mathbf{x}}^{(k)} - \frac{1}{N} \sum_{i=1}^g N_i \bar{\mathbf{x}}^{(i)} \equiv \bar{\mathbf{u}}^{(k)}, \\ \bar{\mathbf{y}}^{(k)} - \hat{\boldsymbol{\mu}}_2^* - \boldsymbol{\Sigma}_{21}^* \bar{\mathbf{x}}_{1,k} &= \bar{\mathbf{y}}^{(k)} - \frac{1}{n} \sum_{i=1}^g n_i \bar{\mathbf{y}}^{(i)} + \boldsymbol{\Sigma}_{21}^* \frac{1}{n} \sum_{i=1}^g n_i \bar{\mathbf{x}}_{1,i} - \boldsymbol{\Sigma}_{21}^* \bar{\mathbf{x}}_{1,k} \\ &= \bar{\mathbf{y}}^{(k)} - \frac{1}{n} \sum_{i=1}^g n_i \bar{\mathbf{y}}^{(i)} - \boldsymbol{\Sigma}_{21}^* \left(\bar{\mathbf{x}}_{1,k} - \frac{1}{n} \sum_{i=1}^g n_i \bar{\mathbf{x}}_{1,i} \right) \\ &\equiv \bar{\mathbf{v}}_k - \boldsymbol{\Sigma}_{21}^* \bar{\mathbf{u}}_{1,k}, \quad (k = 1, \dots, g), \end{aligned}$$

the likelihood function is maximized as follows:

$$\begin{aligned} L(\boldsymbol{\Sigma}) &\equiv \text{Max}_{\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2} L(\boldsymbol{\mu}; \boldsymbol{\Sigma}) \\ &= \prod_{k=1}^g c_k |\boldsymbol{\Sigma}_{11}|^{-N_k/2} |\boldsymbol{\Sigma}_{22.1}|^{-n_k/2} \exp \left[-\frac{1}{2} (\phi_1^{(k)} + \phi_2^{(k)}) \right] \\ &= \left(\prod_{k=1}^g c_k \right) |\boldsymbol{\Sigma}_{11}|^{-N./2} |\boldsymbol{\Sigma}_{22.1}|^{-n./2} \exp \left[-\frac{1}{2} \left(\sum_{k=1}^g \phi_1^{(k)} + \sum_{k=1}^g \phi_2^{(k)} \right) \right], \end{aligned}$$

where

$$\begin{aligned}\phi_1^{(k)} &= \text{tr} \left[\Sigma_{11}^{-1} \left(\mathbf{w}_{11,k}^{(1)} + \mathbf{w}_k^{(2)} \right) \right] + \text{tr} \left[N_k \Sigma_{11}^{-1} \bar{\mathbf{u}}^{(k)} \bar{\mathbf{u}}^{(k)'} \right], \\ \phi_2^{(k)} &= \text{tr} \left[\Sigma_{22,1}^{-1} \left(\mathbf{w}_{22,k}^{(1)} - \mathbf{w}_{21,k}^* \Sigma_{12}^* - \Sigma_{21}^* \mathbf{w}_{12,k}^{(1)} + \Sigma_{21}^* \mathbf{w}_{11,k}^{(1)} \Sigma_{12}^* \right) \right] \\ &\quad + \text{tr} \left[n_k \Sigma_{22,1}^{-1} \left(\bar{\mathbf{v}}_k - \Sigma_{21}^* \bar{\mathbf{u}}_{1,k} \right) \left(\bar{\mathbf{v}}_k - \Sigma_{21}^* \bar{\mathbf{u}}_{1,k} \right)' \right], \quad (k = 1, \dots, g).\end{aligned}$$

Moreover, setting

$$\begin{aligned}\mathbf{V}^{(2)} &= \sum_{k=1}^g \mathbf{w}_k^{(2)}, \quad \bar{\mathbf{U}} = \sum_{k=1}^g N_k \bar{\mathbf{u}}^{(k)} \bar{\mathbf{u}}^{(k)'}, \quad \mathbf{V}_{ij} = \sum_{k=1}^g \mathbf{w}_{ij,k}^{(1)}, \quad (i, j = 1, 2), \\ \mathbf{D}_1 &= \sum_{k=1}^g n_k \bar{\mathbf{v}}_k \bar{\mathbf{v}}_k', \quad \mathbf{D}_2 = \sum_{k=1}^g n_k \bar{\mathbf{v}}_k \bar{\mathbf{u}}_{1,k}', \quad \mathbf{D}_3 = \sum_{k=1}^g n_k \bar{\mathbf{u}}_{1,k} \bar{\mathbf{u}}_{1,k}',\end{aligned}$$

the maximum of the likelihood function with respect to $\boldsymbol{\mu}$ can be rewritten as

$$\begin{aligned}L(\boldsymbol{\Sigma}) &= \left(\prod_{k=1}^g c_k \right) |\boldsymbol{\Sigma}_{11}|^{-N/2} |\boldsymbol{\Sigma}_{22,1}|^{-n/2} \exp \left[-\frac{1}{2} \text{tr} \left\{ \Sigma_{11}^{-1} (\mathbf{V}_{11} + \mathbf{V}^{(2)} + \bar{\mathbf{U}}) \right\} \right] \\ &\quad \times \exp \left[-\frac{1}{2} \text{tr} \left\{ \Sigma_{22,1}^{-1} (\mathbf{V}_{22} - \mathbf{V}_{21} \Sigma_{12}^* - \Sigma_{21}^* \mathbf{V}_{12} + \Sigma_{21}^* \mathbf{V}_{11} \Sigma_{12}^* + \mathbf{D}_1 - \mathbf{D}_2 \Sigma_{12}^* - \Sigma_{21}^* \mathbf{D}_2' + \Sigma_{21}^* \mathbf{D}_3 \Sigma_{12}^*) \right\} \right].\end{aligned}$$

Applying Lemma 2.1, the maximum of the likelihood function is

$$\begin{aligned}\text{Max}_{\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22,1}} L(\boldsymbol{\Sigma}) &= \left(\prod_{k=1}^g c_k \right) |\mathbf{V}_{11} + \mathbf{V}^{(2)} + \bar{\mathbf{U}}|^{-N/2} |\mathbf{V}_{22} - \mathbf{V}_{21} \Sigma_{12}^* - \Sigma_{21}^* \mathbf{V}_{12} + \Sigma_{21}^* \mathbf{V}_{11} \Sigma_{12}^* \\ &\quad + \mathbf{D}_1 - \mathbf{D}_2 \Sigma_{12}^* - \Sigma_{21}^* \mathbf{D}_2' + \Sigma_{21}^* \mathbf{D}_3 \Sigma_{12}^*|^{-n/2} N^{pN/2} n^{qn/2} \exp \left[-\frac{1}{2} (pN + qn) \right],\end{aligned}$$

when

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{11} &= \frac{1}{N} (\mathbf{V}_{11} + \mathbf{V}^{(2)} + \bar{\mathbf{U}}), \\ \hat{\boldsymbol{\Sigma}}_{22,1} &= \frac{1}{n} (\mathbf{V}_{22} - \mathbf{V}_{21} \Sigma_{12}^* - \Sigma_{21}^* \mathbf{V}_{12} + \Sigma_{21}^* \mathbf{V}_{11} \Sigma_{12}^* + \mathbf{D}_1 - \mathbf{D}_2 \Sigma_{12}^* - \Sigma_{21}^* \mathbf{D}_2' + \Sigma_{21}^* \mathbf{D}_3 \Sigma_{12}^*).\end{aligned}$$

Using the fact that

$$\begin{aligned}& |\mathbf{V}_{22} - \mathbf{V}_{21} \Sigma_{12}^* - \Sigma_{21}^* \mathbf{V}_{12} + \Sigma_{21}^* \mathbf{V}_{11} \Sigma_{12}^* + \mathbf{D}_1 - \mathbf{D}_2 \Sigma_{12}^* - \Sigma_{21}^* \mathbf{D}_2' + \Sigma_{21}^* \mathbf{D}_3 \Sigma_{12}^*|^{-n/2} \\ &= \left| \left\{ \Sigma_{12}^* - (\mathbf{V}_{11} + \mathbf{D}_3)^{-1} (\mathbf{V}_{12} + \mathbf{D}_2') \right\}' (\mathbf{V}_{11} + \mathbf{D}_3) \left\{ \Sigma_{12}^* - (\mathbf{V}_{11} + \mathbf{D}_3)^{-1} (\mathbf{V}_{12} + \mathbf{D}_2') \right\} \right. \\ &\quad \left. + \mathbf{V}_{22} + \mathbf{D}_1 - (\mathbf{V}_{21} + \mathbf{D}_2) (\mathbf{V}_{11} + \mathbf{D}_3)^{-1} (\mathbf{V}_{12} + \mathbf{D}_2') \right|^{-n/2},\end{aligned}$$

the maximum value under the null hypothesis is as follows:

$$\begin{aligned}\text{Max}_{\boldsymbol{\Sigma}} L(\boldsymbol{\Sigma}) &= \left(\prod_{k=1}^g c_k \right) |\mathbf{V}_{11} + \mathbf{V}^{(2)} + \bar{\mathbf{U}}|^{-N/2} |\mathbf{V}_{22} + \mathbf{D}_1 - (\mathbf{V}_{21} + \mathbf{D}_2) (\mathbf{V}_{11} + \mathbf{D}_3)^{-1} (\mathbf{V}_{12} + \mathbf{D}_2')|^{-n/2} \\ &\quad \times N^{pN/2} n^{qn/2} \exp \left[-\frac{1}{2} (pN + qn) \right].\end{aligned}\tag{18}$$

Therefore, the log-likelihood ratio criterion is

$$\begin{aligned}-2 \log \lambda &= N \log |\mathbf{V}_{11} + \mathbf{V}^{(2)} + \bar{\mathbf{U}}| + n \log |\mathbf{V}_{22} + \mathbf{D}_1 - (\mathbf{V}_{21} + \mathbf{D}_2) (\mathbf{V}_{11} + \mathbf{D}_3)^{-1} (\mathbf{V}_{12} + \mathbf{D}_2')| \\ &\quad - \sum_{k=1}^g N_k \log |\mathbf{w}_{11,k}^{(1)} + \mathbf{w}_k^{(2)}| - \sum_{k=1}^g n_k \log \left| \mathbf{w}_{22,k}^{(1)} - \mathbf{w}_{21,k}^{(1)} \left(\mathbf{w}_{11,k}^{(1)} \right)^{-1} \mathbf{w}_{12,k}^{(1)} \right| \\ &\quad - pN \log N - qn \log n + p \sum_{k=1}^g N_k \log N_k + q \sum_{k=1}^g n_k \log n_k.\end{aligned}\tag{19}$$

Since the number of parameters are $g(p+q)(p+q+3)/2$ and $(p+q)(p+q+3)/2$ under the alternative hypothesis and the null hypothesis, respectively, we have the following theorem from an asymptotic property for the log-likelihood ratio criterion.

Theorem 4.1. Suppose $n, N \rightarrow \infty$ with $0 < n/N \leq 1$. Under the null hypothesis H_{02} , the log-likelihood ratio criterion in (19) is asymptotically distributed as a chi-squared distribution with f degrees of freedom, where

$$f = \frac{1}{2}(g-1)(p+q)(p+q+3).$$

4.2. Wald-type criterion

It can be shown that the mean vector $\hat{\boldsymbol{\mu}}$ and the covariance matrices $\hat{\boldsymbol{\Sigma}}$ or $\tilde{\boldsymbol{\Sigma}}$ are asymptotically distributed and independent by the expansion of $\hat{\boldsymbol{\mu}}$ as proposed by Chang and Richards [4]. We denote $\boldsymbol{\Psi}_k$ as the asymptotic covariance matrix for the mean vector $\hat{\boldsymbol{\mu}}_k$ of the k th population. The vector

$$\sqrt{N} \left(\sqrt{\pi_1} (\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)', \dots, \sqrt{\pi_g} (\hat{\boldsymbol{\mu}}_g - \boldsymbol{\mu}_g)' \right)'$$

is asymptotically distributed as a normal distribution with mean $\mathbf{0}$ and covariance matrix $\text{diag}(\boldsymbol{\Psi}_1, \dots, \boldsymbol{\Psi}_g)$. Under the null hypothesis, the vector

$$\sqrt{N} (\mathbf{Q} \otimes \mathbf{I}_{p+q}) \left(\sqrt{\pi_1} \hat{\boldsymbol{\mu}}_1', \dots, \sqrt{\pi_g} \hat{\boldsymbol{\mu}}_g' \right)' \equiv \sqrt{N} (\mathbf{Q} \otimes \mathbf{I}_{p+q}) \mathbf{M}$$

is asymptotically distributed as a normal distribution with mean $\mathbf{0}$ and covariance matrix $(\mathbf{Q} \otimes \mathbf{I}_{p+q}) \text{diag}(\boldsymbol{\Psi}_1, \dots, \boldsymbol{\Psi}_g) (\mathbf{Q} \otimes \mathbf{I}_{p+q})'$, where

$$\boldsymbol{\Psi}_k = \boldsymbol{\Sigma}_k + \frac{N_k \tau_k (n_k - 2)}{n_k (n_k - p - 2)} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22.1}^{(k)} \end{pmatrix}, \quad (k = 1, \dots, g).$$

We have the following theorem by considering the number of parameters similar to the case of the log-likelihood ratio criterion.

Theorem 4.2. Suppose $n, N \rightarrow \infty$ with $0 < n/N \leq 1$. Under the null hypothesis H_{02} , the Wald-type criteria

$$WC_{21} = WC_{31} + WC_{11}, \quad (20)$$

$$WC_{22} = WC_{31} + WC_{12}, \quad (21)$$

$$WC_{23} = WC_{32} + WC_{11}, \quad (22)$$

$$WC_{24} = WC_{32} + WC_{12} \quad (23)$$

are asymptotically distributed as a chi-squared distribution with $(g-1)(p+q)(p+q+3)/2$ degrees of freedom, where

$$WC_{31} = N \mathbf{M}' (\mathbf{Q} \otimes \mathbf{I}_{p+q})' \left((\mathbf{Q} \otimes \mathbf{I}_{p+q}) \text{diag}(\hat{\boldsymbol{\Psi}}_1, \dots, \hat{\boldsymbol{\Psi}}_g) (\mathbf{Q} \otimes \mathbf{I}_{p+q})' \right)^{-1} (\mathbf{Q} \otimes \mathbf{I}_{p+q}) \mathbf{M},$$

$$WC_{32} = N \mathbf{M}' (\mathbf{Q} \otimes \mathbf{I}_{p+q})' \left((\mathbf{Q} \otimes \mathbf{I}_{p+q}) \text{diag}(\tilde{\boldsymbol{\Psi}}_1, \dots, \tilde{\boldsymbol{\Psi}}_g) (\mathbf{Q} \otimes \mathbf{I}_{p+q})' \right)^{-1} (\mathbf{Q} \otimes \mathbf{I}_{p+q}) \mathbf{M},$$

and $\hat{\boldsymbol{\Psi}}_k$ and $\tilde{\boldsymbol{\Psi}}_k$ are the maximum likelihood estimator and unbiased estimator matrices for all parameters, respectively.

5. Hypothesis testing for the mean vector

Consider the following null hypothesis:

$$H_{03} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_g,$$

$$H_{13} : H_{03} \text{ is not valid.}$$

It is difficult to derive a closed-form solution for the maximum likelihood estimator under this particular null hypothesis. Therefore, only the Wald-type criterion is considered. Since the number of parameters are $g(p+q) + g(p+q)(p+q+1)/2$ and $(p+q) + g(p+q)(p+q+1)/2$ under the alternative hypothesis and the null hypothesis, respectively, we have the following theorem from an asymptotic property for the Wald-type criterion.

Theorem 5.1. Suppose $n, N \rightarrow \infty$ with $0 < n/N \leq 1$. Under the null hypothesis H_{03} , Wald-type criteria WC_{31} and WC_{32} are asymptotically distributed as a chi-squared distribution with $(g-1)(p+q)$ degrees of freedom.

Table 1
Balanced case.

τ	0.10	0.25	0.35	0.50
n	100	100	100	100
N	112	134	154	200
n	200	200	200	200
N	223	267	308	400
n	500	500	500	500
N	556	667	770	1000
n	1000	1000	1000	1000
N	1112	1334	1539	2000

Table 2
Unbalanced case.

	Case 1	Case 2	Case 3	Case 4
N_1	556	667	770	1000
N_2	778	934	1077	1400
N_3	1112	1334	1539	2000
τ_1	0.10	0.25	0.35	0.50
τ_2	0.10	0.25	0.35	0.50
τ_3	0.10	0.25	0.35	0.50
	Case 5	Case 6	Case 7	Case 8
N_1	556	556	770	770
N_2	875	875	875	875
N_3	1539	1112	1539	1112
τ_1	0.10	0.10	0.35	0.35
τ_2	0.20	0.20	0.20	0.20
τ_3	0.35	0.10	0.35	0.10

6. Numerical simulation

In this section, the accuracy of the likelihood ratio criterion and Wald-type criteria is investigated by performing numerical simulations. First, let

$$\mu_0 = (0, 0, 0, 0, 0, 0, 0)', \quad \mu_b = \frac{1}{2}(9, 9, 9, 9, 9, 9, 9)',$$

$$\Sigma = \Lambda P \Lambda, \quad \Lambda = \text{diag}(\sigma^6, \sigma^5, \dots, \sigma^2, \sigma, 1), \quad P = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}.$$

We use Σ_0 to denote when $\rho = 0.15$ and $\sigma = \sqrt{2}$, and Σ_b when $\rho = 1.2$ and $\sigma = 3$. The population distribution is a seven-variate normal distribution with mean μ_0 and covariance matrix Σ_0 ; these values are also used as the null hypothesis. Let $p = 4$ and $q = 3$. The sample sizes n and N are listed in Tables 1 and 2, where $n = n_1 = n_2 = n_3$ and $N = N_1 = N_2 = N_3$ in the balanced case, and $n_1 = 500$, $n_2 = 700$, and $n_3 = 1000$ in the unbalanced case. One hundred thousand simulations were performed.

6.1. Hypothesis testing for the covariance matrix

The size and power of the test are investigated using criteria (14)–(16). The critical point is used as the percentile of the chi-squared distribution with $(g - 1)(p + q)(p + q + 1)/2 = 54$ degrees of freedom. For the hypothesis, we set

$$H_{01} : \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_0,$$

$$H_{11} : \Sigma_1 = \Sigma_0, \quad \Sigma_2 = \Sigma_3 = \Sigma_0 + \frac{1.2}{\sqrt{N_3}} \Sigma_b.$$

Table 3 lists the size and power of the test in the balanced case, and Table 4 lists the result in the unbalanced case. As for the size of the test, in the balanced case, it can be shown that the convergence of the test using the likelihood ratio criterion

Table 3

Size and power of test for covariance matrix (balanced case).

<i>n</i>	Size				Power			
	100	200	500	1000	100	200	500	1000
$\tau = 10\%$								
LRC	.0810	.0641	.0555	.0529	.6288	.7326	.8512	.9071
WC ₁₁	.0386	.0434	.0475	.0490	.4860	.7167	.8751	.9273
WC ₁₂	.0385	.0433	.0474	.0489	.4852	.7164	.8751	.9273
$\tau = 25\%$								
LRC	.0801	.0637	.0546	.0512	.5619	.6550	.7725	.8359
WC ₁₁	.0445	.0478	.0486	.0484	.4465	.6441	.8023	.8640
WC ₁₂	.0442	.0475	.0486	.0484	.4446	.6431	.8021	.8639
$\tau = 35\%$								
LRC	.0779	.0626	.0549	.0532	.5179	.5987	.7023	.7682
WC ₁₁	.0496	.0492	.0499	.0510	.4232	.5915	.7345	.8014
WC ₁₂	.0489	.0487	.0497	.0509	.4204	.5901	.7342	.8012
$\tau = 50\%$								
LRC	.0773	.0633	.0549	.0528	.4286	.4880	.5748	.6337
WC ₁₁	.0582	.0547	.0515	.0512	.3646	.4874	.6074	.6696
WC ₁₂	.0569	.0540	.0513	.0510	.3612	.4856	.6065	.6691

Table 4

Size and power of test for covariance matrix (unbalanced case).

	Size				Power			
	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
LRC	.0530	.0534	.0545	.0536	.5906	.5006	.4401	.3437
WC ₁₁	.0544	.0554	.0564	.0565	.7156	.6234	.5537	.4424
WC ₁₂	.0538	.0547	.0558	.0560	.7127	.6200	.5500	.4389
	Case 5	Case 6	Case 7	Case 8	Case 5	Case 6	Case 7	Case 8
LRC	.0550	.0534	.0532	.0550	.4353	.5870	.4419	.5914
WC ₁₁	.0596	.0549	.0552	.0548	.5549	.7115	.5563	.7112
WC ₁₂	.0586	.0545	.0546	.0543	.5510	.7085	.5530	.7085

(LRC) is a little slower than that obtained using the Wald-type criteria. On the other hand, for the unbalanced case, overall convergence is slow; despite this fact, the convergence of the test using the LRC is faster than that obtained using the Wald-type criteria. In both balanced and unbalanced cases, the power of the test using the Wald-type criteria is high under the alternative hypothesis. There is a tendency for the power of the test to be reduced when the sample size n is fixed and the missing ratio is higher. Compared to the balanced case, the power of the test obtained using the Wald-type criteria is greater than that obtained using the LRC in the unbalanced case.

6.2. Hypothesis testing for the mean vector and covariance matrix

In this subsection, the size and power of the test are simulated using criteria (19)–(23). We set

$$\begin{aligned}
 H_{02} : \boldsymbol{\mu}_1 &= \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}_0, \\
 \boldsymbol{\Sigma}_1 &= \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}_0, \\
 H_{12} : \boldsymbol{\mu}_1 &= \boldsymbol{\mu}_0, \quad \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}_0 + \frac{1}{\sqrt{N_3}} \boldsymbol{\mu}_b, \\
 \boldsymbol{\Sigma}_1 &= \boldsymbol{\Sigma}_0, \quad \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}_0 + \frac{1.2}{\sqrt{N_3}} \boldsymbol{\Sigma}_b.
 \end{aligned}$$

In this case, the critical point is used as the percentile of the chi-squared distribution with $(g-1)(p+q)(p+q+3)/2 = 70$ degrees of freedom. Tables 5 and 6 display the size and power of the test in the balanced and unbalanced cases, respectively. In the balanced case, the size of the test follows a tendency similar to that of the hypothesis tests for the covariance matrix. The convergence of the test obtained using the Wald-type criteria is faster than that obtained using the LRC. The Wald-type criteria WC_{21} , WC_{22} , WC_{23} , and WC_{24} are close. In the unbalanced case, the convergence of the test obtained using the Wald-type criteria is slower than that using the LRC. The fact that the power of the test is greater in the balanced case when the sample size n is large does not change the previous subsection. When the sample size n is large and the missing ratio τ is small, the power of the test is increased in the unbalanced case.

Table 5

Size and power of test for mean vector and covariance matrix (balanced case).

<i>n</i>	Size				Power			
	100	200	500	1000	100	200	500	1000
$\tau = 10\%$								
<i>LRC</i>	.0803	.0643	.0548	.0532	.7497	.8397	.9238	.9573
<i>WC</i> ₂₁	.0492	.0500	.0499	.0505	.7176	.8593	.9440	.9706
<i>WC</i> ₂₂	.0491	.0499	.0498	.0505	.7171	.8591	.9439	.9705
<i>WC</i> ₂₃	.0493	.0500	.0499	.0505	.7126	.8580	.9437	.9705
<i>WC</i> ₂₄	.0491	.0499	.0498	.0505	.7123	.8579	.9436	.9705
$\tau = 25\%$								
<i>LRC</i>	.0795	.0634	.0556	.0524	.6826	.7733	.8642	.9121
<i>WC</i> ₂₁	.0559	.0525	.0515	.0502	.6587	.7971	.8944	.9335
<i>WC</i> ₂₂	.0553	.0523	.0514	.0501	.6574	.7965	.8942	.9334
<i>WC</i> ₂₃	.0559	.0525	.0515	.0502	.6543	.7955	.8939	.9334
<i>WC</i> ₂₄	.0553	.0523	.0514	.0502	.6528	.7951	.8938	.9333
$\tau = 35\%$								
<i>LRC</i>	.0774	.0631	.0557	.0537	.6325	.7119	.8086	.8601
<i>WC</i> ₂₁	.0597	.0538	.0525	.0529	.6157	.7390	.8441	.8880
<i>WC</i> ₂₂	.0590	.0535	.0523	.0528	.6138	.7380	.8437	.8879
<i>WC</i> ₂₃	.0597	.0539	.0525	.0529	.6111	.7370	.8436	.8879
<i>WC</i> ₂₄	.0590	.0535	.0523	.0528	.6091	.7361	.8432	.8878
$\tau = 50\%$								
<i>LRC</i>	.0767	.0614	.0536	.0524	.5354	.6010	.6855	.7433
<i>WC</i> ₂₁	.0673	.0576	.0524	.0519	.5327	.6313	.7259	.7791
<i>WC</i> ₂₂	.0660	.0569	.0522	.0518	.5297	.6299	.7252	.7788
<i>WC</i> ₂₃	.0673	.0576	.0524	.0520	.5287	.6295	.7250	.7786
<i>WC</i> ₂₄	.0660	.0570	.0522	.0518	.5259	.6280	.7245	.7784

Table 6

Size and power of test for mean vector and covariance matrix (unbalanced case).

	Size				Power			
	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
<i>LRC</i>	.0534	.0531	.0542	.0528	.6782	.5903	.5237	.4119
<i>WC</i> ₂₁	.0564	.0563	.0570	.0563	.7950	.7094	.6426	.5178
<i>WC</i> ₂₂	.0559	.0558	.0565	.0558	.7928	.7070	.6395	.5147
<i>WC</i> ₂₃	.0565	.0563	.0570	.0563	.7945	.7088	.6421	.5172
<i>WC</i> ₂₄	.0559	.0558	.0565	.0558	.7923	.7064	.6390	.5141
<i>LRC</i>	.0553	.0537	.0525	.0536	.5126	.6832	.5194	.6898
<i>WC</i> ₂₁	.0593	.0563	.0555	.0542	.6369	.7991	.6390	.8010
<i>WC</i> ₂₂	.0587	.0557	.0549	.0538	.6335	.7970	.6359	.7990
<i>WC</i> ₂₃	.0593	.0563	.0555	.0542	.6362	.7986	.6384	.8005
<i>WC</i> ₂₄	.0587	.0557	.0549	.0538	.6328	.7962	.6353	.7985

6.3. Hypothesis testing for the mean vector

Finally, the size and power of the test obtained using criteria *WC*₃₁ and *WC*₃₂ are investigated. The critical point is used as the percentile of the chi-squared distribution with $(g - 1)(p + q) = 14$ degrees of freedom. The hypothesis used is

$$H_{03} : \mu_1 = \mu_2 = \mu_3 = \mu_0,$$

$$H_{13} : \mu_1 = \mu_0, \mu_2 = \mu_3 = \mu_0 + \frac{1}{\sqrt{N_3}} \mu_b.$$

Tables 7 and 8 detail the size and power of the test in the balanced and unbalanced cases, respectively. Although the sizes of the test obtained using the two Wald-type criteria are approximately equal, the convergence of the test sizes to 5% using criterion *WC*₃₂ is slightly faster. Unlike previous tests, this test is liberal. The power of the test shows a tendency that is similar to that of the previous tests.

7. Conclusions

In this study, hypothesis tests were performed to verify the equivalence between a mean vector and covariance matrix for multi-populations by considering a two-step monotone incomplete sample. The LRC and Wald-type criteria for each

Table 7
Size and power of test for mean vector (balanced case).

<i>n</i>	Size				Power			
	100	200	500	1000	100	200	500	1000
$\tau = 10\%$								
WC_{31}	.0771	.0624	.0552	.0523	.8658	.8532	.8458	.8463
WC_{32}	.0731	.0605	.0545	.0520	.8603	.8500	.8446	.8455
$\tau = 25\%$								
WC_{31}	.0748	.0608	.0542	.0523	.7982	.7779	.7705	.7663
WC_{32}	.0716	.0593	.0536	.0520	.7913	.7743	.7691	.7655
$\tau = 35\%$								
WC_{31}	.0736	.0609	.0548	.0522	.7364	.7180	.7085	.7012
WC_{32}	.0703	.0594	.0543	.0519	.7289	.7140	.7065	.7003
$\tau = 50\%$								
WC_{31}	.0738	.0621	.0554	.0518	.6246	.6044	.5883	.5851
WC_{32}	.0707	.0607	.0551	.0516	.6172	.6003	.5864	.5844

Table 8
Size and power of test for mean vector (unbalanced case).

	Size				Power			
	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
WC_{31}	.0534	.0527	.0535	.0525	.5598	.4791	.4236	.3418
WC_{32}	.0529	.0522	.0531	.0521	.5582	.4774	.4218	.3406
	Case 5	Case 6	Case 7	Case 8	Case 5	Case 6	Case 7	Case 8
WC_{31}	.0536	.0539	.0529	.0528	.4049	.5615	.4236	.5776
WC_{32}	.0531	.0534	.0524	.0525	.4033	.5598	.4218	.5760

hypothesis test were derived, and the size and power of the tests for these criteria were investigated. Considering the hypothesis testing of the covariance matrix, the test employing the LRC was liberal, while the test employing the Wald-type criteria was either conservative or liberal depending on the missing ratio τ . On the basis of simulation results, the test that employed the Wald-type criteria is recommended. As for the hypothesis testing of the mean vector and covariance matrix, all tests were liberal when the missing ratio τ increased. Considering the size and power of the test, the test that employed the Wald-type criteria is recommended. With respect to the hypothesis testing of the covariance matrix, both tests were liberal. Although results obtained using the two Wald-type criteria WC_{31} and WC_{32} were similar, the test that employed WC_{32} is recommended. In future studies, the asymptotic properties of this equivalence test using the likelihood ratio and Wald-type criteria will be investigated.

Acknowledgments

The author would like to sincerely thank the referees for their careful reading and numerous valuable comments that helped to improve the original manuscript.

References

- [1] T.W. Anderson, Maximum likelihood estimators for a multivariate normal distribution when some observations are missing, *J. Amer. Statist. Assoc.* 52 (1957) 200–203.
- [2] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, third ed., Wiley-Interscience, 2003.
- [3] T.W. Anderson, I. Olkin, Maximum likelihood estimation of the parameters of a multivariate normal distribution, *Linear Algebra Appl.* 70 (1985) 147–171.
- [4] W.-Y. Chang, D.St.P. Richards, Finite-sample inference with monotone incomplete multivariate normal data, I, *J. Multivariate Anal.* 100 (2009) 1883–1899.
- [5] W.-Y. Chang, D.St.P. Richards, Finite-sample inference with monotone incomplete multivariate normal data, II, *J. Multivariate Anal.* 101 (2010) 603–620.
- [6] J. Hao, K. Krishnamoorthy, Inferences on a normal covariance matrix and generalized variance with monotone missing data, *J. Multivariate Anal.* 78 (2001) 62–82.
- [7] T. Kanda, Y. Fujikoshi, Some basic properties of the MLE's for a multivariate normal distribution with monotone missing data, *Amer. J. Math. Management Sci.* 18 (1998) 161–190.
- [8] R.J.A. Little, D.B. Rubin, *Statistical Analysis with Missing Data*, second ed., Wiley-Interscience, 2002.
- [9] S.B. Provost, Inference about covariances under missing values, *Environmetrics* 2 (1991) 487–496.
- [10] M.S. Srivastava, Multivariate data with missing observations, *Comm. Statist. Theory Methods* 14 (1985) 775–792.
- [11] S. Tsukada, Unbiased estimators for a covariance matrix under two-step monotone incomplete sample, *Comm. Statist. Theory Methods* 43 (2014) 1613–1629.
- [12] T. Yamada, M.M. Romer, D.St.P. Richards, Kurtosis Tests for Multivariate Normality with Monotone Incomplete Data, Penn State University, 2012, Preprint.